



**HAL**  
open science

# Raffinement adaptatif du processus d'assimilation de données par méthodes de Kalman d'ensemble pour des problèmes non linéaires.

Dimitri Mottet

► **To cite this version:**

Dimitri Mottet. Raffinement adaptatif du processus d'assimilation de données par méthodes de Kalman d'ensemble pour des problèmes non linéaires.. Mathématiques générales [math.GM]. Institut National Polytechnique de Toulouse - INPT, 2021. Français. NNT : 2021INPT0003 . tel-04165257

**HAL Id: tel-04165257**

**<https://theses.hal.science/tel-04165257>**

Submitted on 18 Jul 2023

**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.



Université  
de Toulouse

# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

**Délivré par :**

Institut National Polytechnique de Toulouse (Toulouse INP)

**Discipline ou spécialité :**

Mathématiques Appliquées

---

**Présentée et soutenue par :**

M. DIMITRI MOTTET

le mardi 12 janvier 2021

**Titre :**

Raffinement adaptatif du processus d'assimilation de données par méthodes de Kalman d'ensemble pour des problèmes non linéaires.

---

**Ecole doctorale :**

Mathématiques, Informatique, Télécommunications de Toulouse (MITT)

**Unité de recherche :**

Institut de Recherche en Informatique de Toulouse ( IRIT)

**Directeur(s) de Thèse :**

M. SERGE GRATTON

**Rapporteurs :**

M. ERIC BLAYO, UNIVERSITE GRENOBLE ALPES

Mme LAURE RAYNAUD, CENTRE NATIONAL DE RECHERCHES METEO

**Membre(s) du jury :**

M. OLIVIER THUAL, TOULOUSE INP, Président

M. EHOARN SIMON, TOULOUSE INP, Invité

M. JEAN PHILIPPE ARGAUD, EDF CLAMART, Membre

Mme SELIME GUROL, CERFACS, Membre

M. SERGE GRATTON, TOULOUSE INP, Membre



## *Remerciements*

Le doctorat est pour moi la finalité du parcours scolaire et universitaire, la cerise sur le gâteau, la dernière étape et l'aboutissement d'une vie étudiante. Plusieurs étapes sont à franchir pour y arriver. Pour moi, ce furent les classes préparatoires et l'école d'ingénieur. Ces étapes intermédiaires avaient au même titre que le doctorat elles aussi leur lot d'adversité et d'occasions d'abandonner. Mais rien de cela n'est arrivé. Le doctorat, et a fortiori le parcours étudiant dans son intégralité, sont un travail d'équipe avec soi-même et son entourage.

C'est pourquoi j'aimerais dans un premier temps remercier les personnes qui m'ont entouré professionnellement, à commencer par mon directeur de thèse Serge Gratton. Je le remercie de m'avoir dans un premier temps incité à me plonger davantage dans les mathématiques appliquées en école d'ingénieur et de m'avoir donné l'opportunité de faire cette thèse. Son encadrement, ses conseils, et ses idées ont permis d'aboutir à ce manuscrit et à la présentation qui en suivra. Je remercie ensuite Jean-Philippe Argaud, auprès de qui j'ai passé la majorité de mon temps, pour sa disponibilité, sa sagesse, les discussions scientifiques à n'en plus finir ainsi que tous les savoirs qu'il a pu me transmettre au cours de ces trois ans. J'insisterai particulièrement sur sa bonne humeur inébranlable, qui a su de nombreuses fois me redonner le sourire dans les moments de doute. Je remercie également Ehouarn Simon qui m'a lui aussi accompagné durant mon parcours d'élève-ingénieur, pour la rigueur, l'intuition, et la précision dont il a fait preuve tout au long de cette thèse, qui ont été primordiales pour l'intégralité de ce manuscrit.

Je tiens aussi à remercier à EDF R&D l'ensemble du groupe I23 pour leur amabilité et la facilité avec laquelle j'ai pu m'y intégrer. Je remercie Sibou, avec qui j'ai passé trois ans dans le même bureau pour sa bonne humeur, les discussions intéressantes, et plusieurs moments de rigolade, à qui je souhaite de réussir son post-doc et la carrière académique vers laquelle il semble s'orienter ; Gérald, avec qui les échanges ont toujours été agréables et à sa bienveillance pendant les moments difficiles ; Bruno, pour les appels aux pauses cafés et de nombreuses discussions scientifiques et coups de pouce bienvenus ; Frank, pour sa bonne humeur et plusieurs moments d'échanges qui m'ont fait progresser ; Olivier, pour les discussions cinéma et les encouragements ; Soizic, pour sa gentillesse ; Angélique, pour ses conseils de rédaction et ses références en assimilation de données ; et tout particulièrement Alexei pour son temps et ses conseils professionnels précieux. Je remercie également à I2D Guillaume, Alexandra, Alejandro, Arnaud et Raphaël pour les échanges décontractés et amusants, sans oublier plusieurs bons conseils pour la suite des événements. Pour finir, je remercie à I2C Frédéric, notre "chef de secteur" adoré, et son co-bureau Didier qui devra encore le supporter de longues années (!!), sans qui ces trois années auraient perdues de leur saveur. Une dernière pensée va aux autres (anciens) doctorants et membres de l'IRIT, je remercie Adam, Jonathan, Benoît, Guillaume, Alfredo, Luce, Patrick, Olivier, Joseph, Daniel et tous les autres que j'ai pu oublier.

Le monde professionnel est loin d'être le seul à m'avoir côtoyé durant toutes ces années, il en est même à mes yeux celui qui aura eu le moins d'importance. C'est pourquoi je remercie dans un premier temps mes parents et mes sœurs, sans lesquels je n'aurais pu faire cette thèse. Rien n'aurait été possible sans leur présence, sans les échanges que j'ai

eu avec eux, sans leurs conseils, ni sans leur soutien. Je remercie également mes grands-parents et mon oncle Stéphane pour leur douceur, leur aide, leur écoute, et les escapades provençales.

Parmi les gens qui me sont chers, je tiens avant tout à remercier mon colocataire Enzo, déjà pour son aide lors de mon arrivée en région parisienne, mais surtout pour toutes les soirées pour oublier, les concerts et festivals à courir encore 10 minutes à la fin, la rigolade avec notre humour le plus fin en toutes circonstances, les toits parisiens, son amour de la bonne franquette, les remontées de moral et les conseils dans les périodes difficiles, et surtout les 2 quatre fromages et la calzone ; Raphaël pour la musique, la musique, encore la musique, les soirées, les discussions très tardives ou très matinales c'est selon, parce c'est ça le rock ; Thibaut et Youri, qui depuis la prépa sont là que nous partageons de nombreux moments ensembles, qu'ils soient virtuels ou non ; Ludivine, pour son endurance et son énergie en concert, les discussions potagères, et sa ciboulette d'une robustesse merveilleuse ; Howard, groupe sans lequel beaucoup de moins de tête auraient été secouées, pour leur magnifique cadeau lors de mes 25 ans ; Techno 2000, parce que je n'allais quand-même pas les oublier alors qu'ils m'ont eux permis de le faire ; Anne-Solenn, pour les invitations à Chaville et tout particulièrement à Bayeux, qui fut un endroit merveilleux le temps d'un weekend ; Iris, pour avoir ravivé ma passion éteinte dans beaucoup de domaines, m'avoir ouvert les yeux davantage, m'avoir inspiré dans mes recherches tout en me permettant de les fuir malgré la violente houle, et qui fait partie intégrante de cette thèse ; Olivier et Aurélien, pour m'accompagner depuis Nîmes et surtout pour l'escalade de collines avec un vélo sur le dos dans la garrigue sauvage ; Yacine et Jean-Matthieu, qui m'accompagnent depuis l'N7 et avec qui c'est toujours un plaisir de continuer à parler musique, films, et autres thèmes moins intelligents ; Pierre (et Axelle), dont l'amitié depuis l'enfance ne s'estompera jamais et qui sera toujours là pour m'épauler et m'encourager ; Baptiste, qui lui aussi depuis l'enfance m'accompagne et qui, chaque été, nous permet d'en raviver les souvenirs à la maison familiale ; Sylvain et Antoine, pour m'avoir fait rire depuis l'école d'ingénieur et pour être toujours aussi honteux ; Vincent, pour les discussions à la fois sérieuses et décalées en permanence, dont je suis heureux de pouvoir dire que c'est un très bon ami.

Je pense encore à Julia, Amandine, Kawtar, Carlotta, Jean-Marie, Chloé, Tom, Franklin, Jean-Alain, Arnaud G., Arnaud L., Stéphanie, Camilla, Geoffrey, Gilles, Guillaume, Samba, Hugo, Baron, Potier, Floria, Matthieu S., Matthieu L., Mario, Natalia, Gaylor, Martin(g), Déborah, Tess, Nelly, Océane C., Océane S. (qui cuisine les meilleurs pad thaï), Frédéric M., Maëva, Sophie, Estelle, Morgane, Charlélie, qui ont de près ou de loin participé à cette aventure qu'ont été mes études.

Je remercie également tous ceux que je n'ai pas mentionnés, qui a un moment ou un autre ont côtoyés ma vie et sans qui, au même titre que les autres, ce manuscrit n'aurait jamais abouti.

*A ma famille.*

*A mes amis.*

*A Iris.*



# Table des matières

<b>Remerciements</b>	<b>3</b>
<b>Table des figures</b>	<b>13</b>
<b>1 Introduction</b>	<b>19</b>
<b>I Introduction théorique et état de l'art</b>	<b>23</b>
<b>2 Méthodologie et notations</b>	<b>25</b>
2.1 Bases de l'assimilation de données . . . . .	25
2.1.1 Principe et concepts . . . . .	25
2.1.2 Variables, notations et définitions . . . . .	26
Modèle et vecteur d'état . . . . .	26
Observations . . . . .	27
2.1.3 Les différents types d'erreurs et statistiques d'erreur . . . . .	27
Erreur d'ébauche (ou de prévision) . . . . .	28
Erreur d'analyse . . . . .	28
Erreur d'observation . . . . .	28
Gestion de l'erreur modèle . . . . .	29
Représentation des statistiques d'erreur . . . . .	29
2.1.4 L'estimateur BLUE et ses fondements . . . . .	30
Gain de l'estimation et erreur d'analyse . . . . .	30
Calcul du BLUE . . . . .	31
2.1.5 L'assimilation de données séquentielle . . . . .	32
2.2 Approche variationnelle . . . . .	33
2.3 Résolution du problème d'assimilation de données séquentielle . . . . .	33
2.3.1 Le filtre de Kalman (KALMAN, 1960; JAZWINSKI, 1970) . . . . .	34
2.3.2 Le filtre de Kalman étendu (JAZWINSKI, 1970) . . . . .	35
2.4 Définition de l'EnKF (EVENSEN, 1994) . . . . .	36
2.4.1 Ensemble et statistiques d'erreur . . . . .	36
2.4.2 Algorithme final . . . . .	36
2.4.3 ETKF : une variante déterministe de l'EnKF . . . . .	37
Autres variantes déterministes de l'EnKF . . . . .	39
2.4.4 Assimilation de données asynchrone avec l'AEnKF . . . . .	40
2.4.5 Limitations de l'EnKF . . . . .	41
Déficience du rang de la matrice $\mathbf{P}^a$ et covariances fallacieuses . . . . .	41
Collapse de l'ensemble : une des causes de divergence de l'EnKF . . . . .	42
Autres limitations . . . . .	43
2.5 Systèmes dynamiques raides . . . . .	44



<b>3</b>	<b>État de l'art sur les diagnostics de qualité des observations pour la prévision numérique</b>	<b>45</b>
3.1	Prévisibilité des systèmes dynamiques . . . . .	45
3.1.1	Introduction à la notion . . . . .	45
3.1.2	Formalisation mathématique . . . . .	46
3.2	Inventaire des indicateurs de qualité de prévision existants . . . . .	48
3.2.1	Matrice des sensibilités ou "Hat matrix" . . . . .	48
	Origine et formulation statistique . . . . .	48
	Formulation pour l'assimilation de données variationnelle . . . . .	49
3.2.2	DFS : Degrees of Freedom for Signal . . . . .	50
3.2.3	FSOI : Forecast Sensitivity Observation Impact . . . . .	51
	Caractérisation de l'erreur de prévision . . . . .	51
	FSOI pour les méthodes variationnelles . . . . .	53
	Sensibilité aux perturbations de la matrice R . . . . .	54
	Application récente . . . . .	55
	FSOI pour les méthodes d'ensemble (diagnostics EnFSOI) . . . . .	55
	Récapitulatif . . . . .	56
3.3	Conclusion et orientation des recherches . . . . .	57
<b>4</b>	<b>Description des indicateurs retenus de qualité des observations pour la prévision</b>	<b>59</b>
4.1	Matrice des sensibilités ou "Hat Matrix" pour l'EnKF . . . . .	59
4.2	Essais numériques de vérification de l'indicateur . . . . .	60
4.3	Distance de Wasserstein . . . . .	63
4.3.1	Formulation dans le cas gaussien . . . . .	63
4.3.2	Distance de Wasserstein en assimilation de données : un palliatif à l'erreur modèle . . . . .	64
4.4	Divergence de Kullback-Leibler . . . . .	65
4.4.1	Formulation dans le cas gaussien . . . . .	65
4.4.2	La divergence de Kullback-Leibler en assimilation de données . . . . .	66
<b>II</b>	<b>Contribution théorique et travaux de recherche</b>	<b>69</b>
<b>5</b>	<b>Description et utilisation des indicateurs dans l'EnKF</b>	<b>71</b>
5.1	Introduction et contexte - contrôle de la dynamique de correction par gestion de l'assimilation via diagnostics sur les observations . . . . .	71
5.2	Utilisation des indicateurs à l'arrivée d'une observation - plusieurs cas . . . . .	72
5.2.1	Cas 1 : synchrone . . . . .	73
5.2.2	Cas 2 : asynchrone . . . . .	73
5.3	Indicateurs des sensibilités à l'observation . . . . .	74
5.3.1	Formulation pour l'EnKF . . . . .	74
	Justification de l'hypothèse de diagonalité de R . . . . .	74
	Calcul avec $\mathbf{R} = \sigma^2 \mathbf{I}_p$ . . . . .	75
	Calcul avec $\mathbf{R} = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$ diagonale quelconque . . . . .	77
	Cas ensembliste avec B non inversible . . . . .	78
	Formulation pour utilisation asynchrone dans le cas ensembliste . . . . .	79
5.3.2	Interprétation de l'indicateur SOBS . . . . .	81

	Sélection de l'observation avec effet global maximal sur l'état . . . . .	81
	Sélection de l'observation avec effet maximal sur une dimension de l'état . . . . .	82
	Sélection d'un vecteur d'observation entier pour assimilation . . . . .	82
	Comparaisons du fonctionnement avec la référence "Hat Matrix" . . . . .	83
<b>6</b>	<b>Coût opératoire des indicateurs</b>	<b>85</b>
6.1	Hat matrix . . . . .	85
6.1.1	Comparaison avec la matrice des sensibilités de (CARDINALI, 2013)	86
6.2	Indicateur de sensibilités : $\delta \mathbf{x}'(\epsilon)$ et SOBS . . . . .	87
	Coût pour la formulation (5.17) . . . . .	87
	Coût pour la formulation LETKF (5.28) . . . . .	89
6.3	Distances probabilistes . . . . .	90
6.3.1	Coût de formation des matrices de covariances . . . . .	90
6.3.2	Distance de 2-Wasserstein . . . . .	90
6.3.3	Divergence de Kullback-Leibler . . . . .	90
6.4	Comparaisons entre chaque indicateur . . . . .	91
6.5	Comparaison au coût d'une assimilation . . . . .	92
6.5.1	Coût opératoire de l'assimilation version Evensen 2003 . . . . .	92
6.5.2	Ordre de grandeur pour des applications classiques . . . . .	93
6.5.3	Ordre de grandeur des applications EDF . . . . .	94
<b>III</b>	<b>Application, conclusion et perspectives</b>	<b>97</b>
<b>7</b>	<b>Essais numériques : de la théorie à sa vérification</b>	<b>99</b>
7.1	Introduction : objectifs des essais numériques . . . . .	99
7.2	Indicateurs <i>a posteriori</i> . . . . .	100
	Impact sur la précision de la prévision . . . . .	100
	Comparatif de comportement entre SOBS et la Hat matrix . . . . .	107
	Distances probabilistes en diagnostic de vérification des observations sélectionnées . . . . .	109
7.3	Distances probabilistes pour établissement de stratégies d'assimilation . . . . .	113
7.3.1	Spécificités des distances probabilistes : Wasserstein et Kullback-Leibler . . . . .	113
	Distance de 2-Wasserstein . . . . .	114
	Divergence de Kullback-Leibler . . . . .	114
7.3.2	Seuil de sélection/rejet d'observation . . . . .	114
	Recherche par essais numériques . . . . .	115
7.3.3	Comportement anticipé des deux distances . . . . .	118
	Relation entre erreur de prévision et valeur du seuil . . . . .	121
	Lorenz 63 . . . . .	121
	Oscillateur de Van der Pol . . . . .	123
	Système Xénon . . . . .	123
	Changement du critère de sélection pour Kullback-Leibler . . . . .	128
7.4	Stratégies d'assimilation sur des modèles emblématiques . . . . .	130
7.4.1	Essais avec le système de Lorenz . . . . .	130
	Réduction du nombre d'assimilation . . . . .	130

	Suivi de la dynamique en forte carence d'observation . . . . .	131
7.4.2	Essais avec l'oscillateur de Van der Pol . . . . .	134
	Réduction du nombre d'assimilations sur la fenêtre . . . . .	134
	Estimation du nombre minimal d'observations permettant un suivi des changements de régime . . . . .	135
7.5	Stratégies d'assimilation sur des modèles d'origine industrielle . . . . .	138
7.5.1	Modèle d'évolution de la concentration en Xénon-135 . . . . .	138
7.5.2	Suivi de l'évolution avec pilotage de puissance . . . . .	138
7.5.3	Suivi de l'assimilation en suivi de charge . . . . .	140
<b>8</b>	<b>Conclusion &amp; perspectives</b> . . . . .	<b>143</b>
8.1	Conclusion générale . . . . .	143
8.2	Indicateurs dans un filtre itératif . . . . .	145
8.3	Vers la grande dimension . . . . .	146
8.4	Utilisation asynchrone des indicateurs, adaptation aux lisseurs . . . . .	146
8.5	Utilisation pour l'organisation spatio-temporelle des observations . . . . .	147
8.6	Essais sur des systèmes en configuration industrielle . . . . .	148
<b>A</b>	<b>Annexe</b> . . . . .	<b>151</b>
A.1	Calculs de la définition du BLUE . . . . .	151
A.2	Méthodes variationnelles . . . . .	152
A.2.1	3D-Var . . . . .	152
A.2.2	4D-Var . . . . .	152
A.2.3	4D-Var à contraintes faibles . . . . .	153
A.3	Améliorations de l'EnKF . . . . .	154
A.3.1	Inflation . . . . .	154
	Travaux de Anderson sur l'inflation adaptative . . . . .	154
	EnKF-N (BOCQUET, 2011) . . . . .	156
	EnKS : lisseur de Kalman d'ensemble . . . . .	157
A.3.2	Méthodes itératives : IEnKF et IEnKS . . . . .	159
	IEnKF . . . . .	159
	IEnKF-N . . . . .	160
A.3.3	IEnKF avec erreur modèle additive : IEnKF-Q, et IEnKS . . . . .	161
	IEnKF-Q . . . . .	161
	IEnKS . . . . .	162
A.4	Calculs de l'état de l'art sur les FSOI . . . . .	162
A.5	Etablissement des expressions de Wasserstein et Kullback-Leibler . . . . .	163
A.5.1	Expression de la distance de Wasserstein . . . . .	163
A.5.2	Expression de la divergence de Kullback-Leibler . . . . .	164
A.6	Réduction du nombre d'opérations pour les distances probabilistes . . . . .	164
A.6.1	Mise à jour de l'opérateur H et de l'état analysé selon les observa- tions à supprimer . . . . .	164
A.6.2	Amélioration de la divergence de Kullback-Leibler . . . . .	165
A.6.3	Amélioration de la distance de Wasserstein . . . . .	165
A.6.4	Algorithme de calcul à moindre coût pour les deux distances . . . . .	166
	Calcul de Wasserstein : . . . . .	166
	Calcul de Kullback-Leibler : . . . . .	167
A.6.5	Conclusion . . . . .	167

---

A.7	Configurations des systèmes pour essais numériques . . . . .	167
A.7.1	Lorenz 63 . . . . .	167
	Description du système . . . . .	167
	Configuration des essais numériques . . . . .	169
A.7.2	Oscillateur de Van der Pol . . . . .	171
	Description du système . . . . .	171
	Configuration des essais numériques (configuration 3) . . . . .	172
A.7.3	Système Xénon . . . . .	173
	Rapide introduction sur la fission au sein dans un réacteur nucléaire	173
	Chaîne radioactive et cinétique du Xénon-135 . . . . .	173
	Cinétique du Xénon et pilotage d'un réacteur . . . . .	174
	Arrêt du réacteur (configuration 4) . . . . .	175
	Démarrage du réacteur (configuration 5) . . . . .	176
	Puissance variable (configuration 6) . . . . .	177
	Suivi de charge (configuration 7) . . . . .	179
	Configuration des essais numériques . . . . .	179



# Table des figures

2.1	Adapté depuis (ASCH, BOCQUET et NODET, 2016). Voici trois itérations d'une procédure d'assimilation de données séquentielle comme le filtre de Kalman. Les étoiles représentent les observations et les ronds le vecteur d'état. L'axe des abscisse représente le temps et l'axe des ordonnées les valeurs des vecteurs d'état et d'observation. . . . .	32
4.1	Valeurs diagonales de $\hat{\mathbf{H}}$ , pour 100 vecteurs d'observation, avec intervalle de temps $dt_{obs} = 0.25$ entre chaque vecteur. . . . .	60
4.2	Valeurs diagonales de $\hat{\mathbf{H}}$ , restreintes aux 40 1ers vecteurs d'observation. . . . .	61
4.3	Valeurs diagonales moyennes de $\hat{\mathbf{H}}$ pour 100 expériences. . . . .	62
5.1	Visualisation de $SOBS(\epsilon)$ sur un système Lorenz 63 à 3 dimensions (colonnes) et 3 vecteurs de 3 observations (lignes par blocs de 3) de ce système. . . . .	83
7.1	Tracé de chaque dimension, membres de l'ensemble en gris, et observations de références sur la fenêtre en noir. Aucune suppression d'observation. La courbe en haut à gauche représente l'observation décrite comme négligeable par SOBS en fonction du vecteur d'observation tout au long de la fenêtre. . . . .	101
7.2	Évolution des RMSEs d'analyse (à gauche) sur la fenêtre pour trois essais : celui de référence (natif), celui avec sélection par Hat matrix (Hat), et celui avec sélection par SOBS (sobs). A droite, écarts relatifs en RMSE pour les essais avec indicateurs, comparativement à la RMSE de référence. $dt_{obs} = 0.25$ par essai. . . . .	102
7.3	RMSEs d'analyse (à gauche) sur la fenêtre pour trois essais : celui de référence (natif), celui avec sélection par Hat matrix (Hat), et celui avec sélection par SOBS (sobs). A droite, écarts relatifs en RMSE pour les essais avec indicateurs, comparativement à la RMSE de référence. $dt_{obs} = 0.25$ par essai. . . . .	103
7.4	Simulation de référence pour l'oscillateur de Van der Pol pour la configuration 3 (A.7.2), avec $t_f = 150$ . Toutes les observations sont affichées, soit 15, avec $dt_{obs} = 10$ . L'ensemble est affiché et montre la variété des différents membres. . . . .	104
7.5	Van der Pol. Evolutions des RMSEs d'analyse sur la fenêtre pour trois essais : la référence (natif), la sélection par Hat matrix (Hat), et la sélection par SOBS (sobs). Écarts relatifs à droite. $dt_{obs} = 10$ par essai. . . . .	105
7.6	Même tracé pour 150 assimilations que l'encadré (7.5a), soit $t_f = 1500$ . $dt_{obs} = 10$ . . . . .	105
7.7	Hat matrix pour configuration 1. . . . .	107
7.8	SOBS pour configuration 1. . . . .	107

7.9	Hat matrix pour configuration 2. . . . .	107
7.10	SOBS pour configuration 2. . . . .	107
7.11	Scatter-plots de SOBS versus Hat matrix pour la configuration 2. L'axe des abscisses correspond aux valeurs de la Hat matrix, et l'axe des ordonnées aux valeurs de SOBS ramenées à $[0, 1]$ . . . . .	108
7.12	Hat matrix moyennée pour 100 essais. Configuration 1. . . . .	109
7.13	SOBS moyenné pour 100 essais. Configuration 1. . . . .	109
7.14	Lorenz 63. Essais avec diagnostics a posteriori par distances probabilistes entre les distributions issues de l'analyse de référence et celle de l'analyse avec stratégies d'assimilation par les deux indicateurs. Les abscisses représentent le temps, les ordonnées la valeur du diagnostic $\frac{d(p_1, p_2)}{d(p_1, p_f)}$ . . . . .	110
7.15	Van der Pol. Essais avec diagnostics a posteriori par distances probabilistes entre les distributions issues de l'analyse de référence et celle de l'analyse avec stratégies d'assimilation par les deux indicateurs. Les abscisses représentent le temps, les ordonnées la valeur du diagnostic $\frac{d(p_1, p_2)}{d(p_1, p_f)}$ . . . . .	111
7.16	Evolutions comparées pour Hat, config. 1. . . . .	111
7.17	Evolutions comparées pour SOBS, config. 1. . . . .	111
7.18	Intervalle de valeurs de seuil utiles pour le système Xénon, montrant la décroissance du nombre d'analyses sur la fenêtre. . . . .	115
7.19	Intervalle de valeurs de seuil utiles pour le Lorenz 63, montrant la décroissance du nombre d'assimilation sur la fenêtre (droite), avec évolution des moyennes de chaque RMSE (gauche), en fonction du logarithme naturel de la valeur de seuil (axe des abscisses). . . . .	116
7.20	Intervalle de valeurs de seuil utiles pour le Van der Pol, montrant la décroissance du nombre d'assimilation sur la fenêtre (droite), avec évolution des moyennes de chaque RMSE (gauche), en fonction du logarithme naturel de la valeur de seuil (axe des abscisses). . . . .	116
7.21	Évolution en moyenne des distances probabilistes (ordonnées) en fonction de $\sigma_{obs}^2$ (abscisse) pour le système Xénon en suivi de charge. . . . .	119
7.22	Évolution en moyenne des distances probabilistes (ordonnées) selon le logarithme naturel du seuil choisi (abscisse) pour chaque essai sur le Lorenz 63. La distance de Wasserstein est à gauche, la divergence de Kullback-Leibler à droite. . . . .	120
7.23	Évolution en moyenne des distances probabilistes (ordonnées) selon le logarithme naturel du seuil choisi (abscisse) pour chaque essai sur le Van der Pol. La distance de Wasserstein est à gauche, la divergence de Kullback-Leibler à droite. . . . .	120
7.24	Évolution en moyenne de la distance de Wasserstein (gauche) et de la moyenne de la RMSE de prévision (droite) selon le logarithme naturel du seuil choisi pour chaque essai. (Lorenz 63). . . . .	121
7.25	Évolution en moyenne des composantes de chaque distance probabiliste (ordonnées) en fonction de la variance d'erreur d'observation (abscisse) pour le système Lorenz 63. La distance de Wasserstein est à gauche, la divergence de Kullback-Leibler est à droite. . . . .	122

7.26	Évolution en moyenne de la distance de Wasserstein (gauche) et de la moyenne de la RMSE de prévision (droite) selon le logarithme naturel du seuil choisi pour chaque essai. (Van der Pol)	123
7.27	Évolution en moyenne de la distance de Wasserstein (gauche) et de la moyenne de la RMSE de prévision (droite) selon le logarithme naturel du seuil choisi pour chaque essai. (Système Xénon hors-flux)	124
7.28	Évolution en moyenne de la distance de Wasserstein (gauche) et de la moyenne de la RMSE de prévision (droite) selon le logarithme naturel du seuil choisi pour chaque essai. (Système Xénon sous flux)	124
7.29	Référence pour le système Xénon avec profil de puissance variable (courbe bleue).	125
7.30	Évolution en moyenne de la distance de Wasserstein (gauche) et de la moyenne de la RMSE de prévision (droite) selon le logarithme naturel du seuil choisi pour chaque essai. (Système Xénon avec forçage en puissance variable).	127
7.31	Lorenz 63, Kullback-Leibler. Évolution du nombre d'assimilation sur la fenêtre (gauche) et de la moyenne des RMSEs de prévision sur la fenêtre (droite), en fonction du logarithme naturel du seuil.	128
7.32	Évolution en moyenne des composantes de chaque distance probabiliste en fonction de la variance d'erreur d'observation pour la configuration 3 du Van der Pol. (Wasserstein à gauche, Kullback-Leibler à droite.)	129
7.33	Moyennes des RMSEs d'analyse et de prévision (ordonnées), selon le logarithme naturel du seuil (abscisse).	130
7.34	Nombre d'assimilations pour chaque essai (ordonnées), selon le logarithme naturel du seuil (abscisse).	130
7.35	Coupe selon les dimensions $x_1$ et $x_2$ à plusieurs instants de transition pour le Lorenz 63. Point rouge : moyenne de l'ensemble, point orange : l'observation. Les points bleus constituent l'ensemble.	131
7.36	Lorenz 63. Essai de référence avec toutes les observations. Les observations correspondent aux losanges noirs pour chaque dimension. Le temps est en abscisse, la valeur de chaque dimension en ordonnées.	132
7.37	Lorenz 63. Essai avec 9/40 observations sélectionnées par l'indicateur, 80% de réduction. Temps en abscisse, amplitude de chaque dimension en ordonnées.	133
7.38	Lorenz 63. Essai avec 5/40 observations sélectionnées par l'indicateur, 90% de réduction.	134
7.39	Moyennes des RMSEs d'analyse et de prévision (gauche), nombre d'assimilation pour chaque essai (droite), chacune selon le logarithme naturel du seuil en abscisse.	135
7.40	Perte de la dynamique le Van der Pol. 1 seule observation.	135
7.41	Van der Pol. Essai de référence avec toutes les observations. 375 observations.	136
7.42	Van der Pol. Essai avec 14% des observations de référence assimilées sur la fenêtre.	137
7.43	Van der Pol. Essai avec 7% des observations de référence assimilées sur la fenêtre.	137



7.44	Répartition des observations de référence pour le système Xénon avec forçage en puissance variable. Les observations sont uniformément réparties sur la fenêtre. Utilisation de 518 observations. Iode-135 en haut, Xénon-135 en bas. Temps en secondes en abscisse, concentrations en ordonnées. . . . .	139
7.45	Répartition des observations sélectionnées par Wasserstein pour le système Xénon avec profil de puissance variable. Les observations sont situées autour des zones de changement de régime. Utilisation de 92 observations. Iode-135 en haut, Xénon-135 en bas. Temps en secondes en abscisse, concentrations en ordonnées. . . . .	139
7.46	Répartition des observations de référence pour le système Xénon avec forçage en puissance pour suivi de charge. Les observations sont uniformément réparties sur la fenêtre. Utilisation de 108 observations. Iode-135 en haut, Xénon-135 en bas. Temps en secondes en abscisse, concentrations en ordonnées. . . . .	140
7.47	Répartition des observations sélectionnées par Wasserstein pour le système Xénon avec forçage en puissance pour suivi de charge. Les observations sont concentrées dans les zones de variations importante de la dynamique. Iode-135 en haut, Xénon-135 en bas. Temps en secondes en abscisse, concentrations en ordonnées. . . . .	141
A.1	Adapté depuis (BOCQUET, 2016). Ici, $S=2$ et $L=6$ avec un écart de temps $\Delta t$ en chaque itéré. Seules les observations en noir sont assimilées. . . . .	158
A.2	Attracteur du Lorenz 63 - Vue de face. . . . .	168
A.3	Vue de haut. . . . .	169
A.4	Référence pour les oscillations de la première dimension du Van der Pol . . . . .	171
A.5	Chaîne radioactive du Xénon-135 . . . . .	173
A.6	Evolution des concentrations à l'arrêt du réacteur. Le pic de Xénon représente le réservoir d'Iode-135 qui se "déverse" dans celui du Xénon-135 suite aux désintégrations radioactives. . . . .	176
A.7	Evolution des concentrations à l'arrêt du réacteur. L'établissement des concentrations d'équilibre se fait à la même vitesse pour les deux entités chimiques. . . . .	177
A.8	Référence pour le système Xénon avec forçage en puissance variable. . . . .	178
A.9	Référence pour le système Xénon avec forçage en puissance pour suivi de charge. . . . .	179

# Synthèse

L'objectif de cette thèse est d'adapter le processus d'assimilation de données utilisant l'algorithme de l'EnKF (EVENSEN, 1994; EVENSEN, 2003; EVENSEN, 2009b) pour traiter des systèmes dynamiques industriels de petite à moyenne taille (de  $10^1$  à  $10^5$  dimensions environ) pour EDF, éventuellement raides (HAIRER et WANNER, 1996).

L'assimilation de données est une procédure qui consiste à combiner les informations obtenues par des observations et un modèle d'un système physique pour obtenir une connaissance plus complète de son état que si elle avait été estimée par l'une des sources indépendamment de l'autre. Cette action de combinaison est communément appelée *étape d'analyse* ou *assimilation*. L'EnKF est une méthode appartenant à la classe des algorithmes de filtrage et il permet de se passer du calcul de l'adjoint. Il propage un ensemble d'états possibles du système à partir duquel il estime empiriquement les matrices d'erreurs associées à son estimation de l'état du système, réduisant les coûts de stockage de telles matrices pour des gros systèmes, qui sont aussi nécessaires à l'étape d'analyse. La raideur du système dynamique impose des contraintes sur l'intégration, et peut avoir des effets sur l'erreur modèle et la sensibilité du système aux perturbations.

Dans cette thèse, l'EnKF stochastique est utilisé sur des modèles de petite dimension présentant des caractéristiques représentatives des problèmes industriels, comme la raideur, la périodicité et du chaos. Les systèmes peuvent être fortement observés, à la fois en terme de fréquence d'arrivée des observations qu'en terme de répartition spatiale, comme ils peuvent être carencés en observations. Pour simuler un cas industriel, des observations sur demande pourront être disponibles. Les travaux de recherches portent sur l'estimation de l'importance des observations pour l'analyse à travers la formulation de nouveaux indicateurs, dans le but d'établir des stratégies d'assimilation. Elles consistent à contrôler la dynamique des instants d'assimilation ainsi que le nombre d'observations assimilées à chaque étape d'analyse, sur toute la durée du processus.

Les indicateurs sont classés en deux catégories : ceux *a posteriori*, qui se calculent après l'étape d'analyse, et ceux *a priori*, qui se calculent à l'arrivée d'une observation avant qu'elle soit assimilée. Après reformulation de la Hat matrix (CARDINALI, 2013) pour l'EnKF, un nouvel indicateur *a posteriori* basé sur l'étude des sensibilités de l'analyse aux différentes observations a été formulé, donnant ainsi deux nouveaux indicateurs *a posteriori*. La distance de Wasserstein (FARCHI et al., 2016; LI et al., 2018; TAMANG et al., 2020) et la divergence de Kullback-Leibler symétrisée (GAUTAM et al., 2018) ont été retenues comme indicateurs *a priori*, quantifiant l'éloignement des distributions de probabilités de la prévision et des observations. Le coût numérique de chaque indicateur a été calculé, montrant la faisabilité des calculs sur les algorithmes type ETKF (BISHOP, ETHERTON et MAJUMDAR, 2001; HUNT, KOSTELICH et SZUNYOGH, 2007) qui sont les plus répandus actuellement, et un algorithme de réduction des coûts pour les distances

probabilistes a été formulé. Des stratégies de calcul sont décrites pour diminuer le coût et le nombre de calculs d'indicateurs à faire.

Les expériences de suppression d'observations grâce aux indicateurs a posteriori sur les systèmes choisis montrent que la sélection d'observation jugées peu importantes par ces indicateurs ne détériore pas de façon significative la RMSE d'analyse moyenne sur la fenêtre d'assimilation. Ceci valide leur pertinence pour l'établissement de stratégies d'assimilations. Des différences de fonctionnement ont été mises en lumière entre ces deux indicateurs, laissant penser à une complémentarité des informations qu'ils retournent. Les distances probabilistes sont utilisées a priori. Un seuil de distance entre les distributions de la prévision et des observations est défini par l'utilisateur pour n'assimiler que les vecteurs d'observations dont la distance avec la prévision est inférieure à celui-ci. La valeur de ce seuil est recherchée numériquement. Un contrôle de l'erreur de prévision via le choix du seuil est possible lorsque nous voulons établir une stratégie d'assimilation pour un système donné. Dans le cas des systèmes périodiques, il a été montré qu'il est possible grâce à ces indicateurs de n'assimiler que jusqu'à 10% des observations disponibles sur la fenêtre, en maintenant le suivi des changements de régime du système. Une application sur un système représentatif de problématiques industrielles d'EDF est décrite. Elle montre la capacité des indicateurs à s'adapter à la dynamique d'évolution à la fois du modèle et des observations.

Plusieurs perspectives sont discutées. L'insertion des indicateurs au sein des méthodes de filtrage et de lissage d'ensemble (BOCQUET et SAKOV, 2012; BOCQUET, 2016), des pistes pour la réduction supplémentaire des coûts de calcul relativement à la configuration du problème d'assimilation et le passage à la grande dimension, une utilisation pour des algorithmes d'assimilation asynchrones (SAKOV, EVENSEN et BERTINO, 2010), ainsi que plusieurs systèmes industriels pour de possibles applications.

**Mots-clés :** assimilation de données, EnKF, diagnostics, impact des observations, qualité de prévision

# 1 Introduction

## Contexte général et apports de la thèse

Au fur et à mesure de l'évolution des sciences et de la complexification des objets théoriques étudiés apparaissent les systèmes non-linéaires. Ce type de système est une composante majoritaire de la plupart des formalisations et descriptions mathématiques du monde observable comme par exemple la mécanique des fluides, avec les équations de Navier-Stokes, formulées au XIX<sup>e</sup> siècle pour décrire le mouvement des fluides Newtoniens, regroupant la majeure partie des gaz et des liquides. Ces équations sont sujettes à plusieurs hypothèses de simplification permettant d'obtenir des solutions dans des cas particuliers. Leur résolution dans le cas général est tellement difficile qu'elle constitue l'un des sept problèmes du Millénaire. Cependant, des modèles numériques permettent d'obtenir des solutions approximées satisfaisantes du point de vue précision et représentativité. Dans l'industrie, la complexité des systèmes physiques modélisés et les contraintes de coût en ressources numériques peuvent grandement limiter la qualité de simulation obtenue, menant à un modèle parfois lacunaire.

C'est là qu'intervient l'assimilation de données, qui permet l'amélioration du système de simulation avec des informations provenant directement de l'observation du système physique réel pour palier les lacunes du modèle. Ceci permet d'obtenir le meilleur compromis entre une simulation avec modélisation imparfaite et une prise en compte de la réalité pour la prévision numérique.

Deux approches différentes et complémentaires se distinguent au sein de l'assimilation de données : l'assimilation de données variationnelle issue de la théorie du contrôle optimal qui consiste en la résolution d'un problème de minimisation variationnelle (LE DIMET et TALAGRAND, 1986), et les méthodes de filtrage issues de l'interpolation statistique, dont la méthode fondatrice est le filtre de Kalman (KALMAN, 1960). De cette dernière est née le filtre de Kalman d'ensemble (EVENSEN, 1994 ; EVENSEN, 2003 ; EVENSEN, 2009b), qui a ouvert la voie aux méthodes de filtrage d'ensemble. Les deux approches permettent d'obtenir des résultats similaires (TRAN, 2009), leurs différences reposant sur la façon dont elles sont formulées et implémentées numériquement. Dans le cadre de l'assimilation de données séquentielle, les deux approches permettent de faire du filtrage, utilisation de l'assimilation de données propice à la prévision numérique. Les méthodes de filtrage d'ensemble telles que le Filtre de Kalman d'Ensemble (EnKF) en sont un exemple majeur. C'est cette utilisation pour de la simulation prévisionnelle qui intéresse particulièrement EDF.

La force de l'EnKF réside dans sa capacité à traiter les non-linéarités des opérateurs descriptifs du système d'assimilation de données, tout en conservant la démarche propre au filtre de Kalman, et en se passant du calcul de l'opérateur adjoint propre aux méthodes

variationnelles. Il approxime les deux premiers moments de la distribution de probabilité de l'état du système physique étudié via un ensemble d'échantillons propagés dans le temps. Cela permet de grandement réduire la taille des matrices d'erreurs manipulées dont une dimension va maintenant reposer sur la taille de l'ensemble, tout en prenant implicitement en compte le modèle non-linéaire via l'intégration numérique de l'ensemble grâce à celui-ci. Ceci permet d'obtenir un séduisant compromis entre coût et performance pour une méthode de filtrage destinée à des systèmes de taille moyenne ( $O(10^1)$  ou plus dimensions pour l'état) à grande ( $O(10^5)$  ou plus dimensions pour l'état). Chez EDF, plusieurs systèmes d'intérêt sont de ces ordres de grandeur, tels que l'empoisonnement au Xénon des réacteurs, certaines lois de comportement de matériaux pour les ouvrages de génie civil, ou encore le diffusion de suspensions en eau libre.

L'EnKF a été initialement développé pour répondre à des problématiques de prévision en océanographie, puis adapté à la prévision météorologique. La qualité des prévisions obtenues grâce aux méthodes de filtrage ensemblistes comme variationnelles fait l'objet d'un vaste champ de recherche.

Plusieurs types de diagnostics existent : les diagnostics de cohérence des statistiques d'erreur réalisés dans l'espace des observations, proposés par (DESROZIER et IVANOV, 2001 ; DESROZIER et al., 2005) qui vérifient la bonne propagation des statistiques d'erreur durant l'étape d'analyse pour les méthodes variationnelles ; les diagnostics FSOI (LANGLAND et BAKER, 2004 ; TRÉMOLET, 2008 ; KALNAY et al., 2012) qui cherchent à formuler et estimer l'erreur de prévision faite en fonction de l'assimilation d'observations, qui se dérivent à la fois pour les méthodes variationnelles (TRÉMOLET, 2008) et les méthodes ensemblistes (LIU et KALNAY, 2008 ; KALNAY et al., 2012) ; les DFS pour "Degrees of Freedom for Signal" (LUPU, GAUTHIER et LAROCHE, 2011) et la Hat matrix (CARDINALI, 2004 ; CARDINALI, 2013) qui mesurent la sensibilité de l'analyse aux différents types d'observations assimilées. Tous ces diagnostics sont implémentés et utilisés sur des systèmes opérationnels dans le cadre de la prévision météorologique (CHAPNIK et al., 2006 ; CARDINALI, 2009 ; TODLING, 2013).

Cependant, ces méthodes n'établissent la qualité de la prévision que *a posteriori* de l'analyse, c'est-à-dire qu'il faut que celle-ci soit calculée pour que les diagnostics soient utilisables, et aucun diagnostic *a priori* ne semble exister à ce jour dans la littérature. L'utilisation de diagnostics *a priori* présenterait l'avantage de réduire le nombre d'observations à assimiler sur la fenêtre, et surtout de sélectionner les instants d'assimilation les plus pertinents lorsque l'erreur de prévision aura été estimée comme trop importante.

C'est dans ce contexte que s'inscrivent nos travaux de recherches. Notre objectif est de faire un inventaire des indicateurs existants et de formuler de nouveaux indicateurs de qualité de la prévision, en particulier des indicateurs *a priori*. L'ensemble de ces nouveaux indicateurs permettra d'établir des *stratégies d'assimilation*, qui consistent à piloter à la fois les instants d'assimilation et le nombre d'observations assimilées pendant toute la durée du processus d'assimilation, et leurs performances seront comparées à celles des indicateurs existants.

## Plan du manuscrit

Le manuscrit est organisé en trois parties.

La première partie a pour objectif de présenter la méthodologie et l'état de l'art en prévision numérique par filtrage.

Le chapitre (2) donne le contexte de l'assimilation de données et sa mise en place par les approches variationnelles et ensemblistes. Le chapitre (3) propose un état de l'art des méthodes de prévision numérique, principalement orienté sur les diagnostics existants de qualité de la prévision, tout en donnant un aperçu de l'ensemble des améliorations algorithmiques apportées aux méthodes ensemblistes et variationnelles au cours des dernières années. Le chapitre (4) fait l'inventaire des indicateurs de qualité de la prévision que nous utiliserons dans cette thèse pour établir des stratégies d'assimilation, et décrit leur formulation spécifiquement pour l'EnKF.

La deuxième partie présente les contributions théoriques de nos travaux de recherches effectués.

Le chapitre (5) détaille la construction d'un nouvel indicateur de qualité *a posteriori* basé sur l'étude des sensibilités de l'incrément d'analyse aux observations assimilées, et présente son utilisation. Le chapitre (6) discute du coût numérique de calcul des différents indicateurs, et du surcoût induit lors de leur utilisation au sein de l'EnKF, et propose des pistes pour compenser ce surcoût.

La troisième et dernière partie de ce manuscrit décrit les expériences numériques obtenues en instrumentant l'EnKF avec ces indicateurs.

Le chapitre (7) présente les essais numériques effectués sur plusieurs systèmes présentant des caractéristiques emblématiques des systèmes industriels, ainsi que sur un système industriel simplifié. Il y est notamment montré l'existence d'une relation entre la valeur d'un paramètre nécessaire à l'établissement des stratégies d'assimilation avec indicateurs *a priori*, et l'erreur commise sur la prévision. La conclusion des résultats théoriques et numériques est ensuite faite au sein du chapitre (8), où sont également discutées plusieurs perspectives pour de futurs travaux de recherche.



## Première partie

# Introduction théorique et état de l'art





## 2 Méthodologie et notations

### 2.1 Bases de l'assimilation de données

Dans cette section, il s'agit de fournir une introduction aux enjeux de l'assimilation de données en mathématiques appliquées. Une grande partie de ces informations proviennent de deux thèses de (PONÇOT, 2008) et (DAGET, 2008) ainsi que des références qu'elles mentionnent, notamment le cours de (BOUTTIER et COURTIER, 1999). Le livre de (ASCH, BOCQUET et NODET, 2016) constitue également une bonne référence traitant de ce domaine.

#### 2.1.1 Principe et concepts

L'assimilation de données est un outil scientifique couramment utilisé dans la vie de tous les jours, étant au cœur des prévisions météorologiques. C'est une science qui cherche à combiner de façon optimale les informations que nous pouvons tirer de la théorie disponible sur un système physique que nous appellerons le modèle, ainsi que les informations disponibles en pratique sur ce système physique que nous appellerons les observations, pour fournir la meilleure connaissance possible de l'état de ce système.

Dans le cas de la météorologie, nous disposons d'un modèle mathématique des équations de la physique régissant le système atmosphérique terrestre, qui permet d'en déduire son état présent et éventuellement son futur proche. Cependant, les équations qui le régissent sont extrêmement complexes et il faut appliquer de nombreuses hypothèses simplificatrices et approximations pour les résoudre. Cela donne donc une solution qui contient des erreurs dont nous avons conscience mais que nous ne savons pas nécessairement quantifier. Mais nous avons aussi la possibilité d'observer certaines grandeurs physiques réelles, telles que la pression, la température ou l'humidité pour les comparer avec celles de la solution calculée. L'assimilation de données regroupe alors l'ensemble des méthodes mathématiques qui sont capables de lier les informations provenant à la fois de la simulation ou prévision par le modèle et les observations prises directement sur le système physique réel, et d'obtenir ce que l'on appelle l'état analysé, qui constitue le meilleur compromis possible entre les connaissances apportées par le modèle et celles apportées par les observations, la réalité physique du phénomène.

Pour résoudre mathématiquement le problème qu'est l'*analyse*, c'est-à-dire l'état représentant le meilleur compromis étant données toutes les informations disponibles sur le système, deux grandes écoles existent :

- les méthodes de *filtrage*, issues de l'estimation statistique et conduisant souvent à la résolution d'un système linéaire,

- les méthodes *variationnelles*, issues du contrôle optimal, où l'analyse représente le minimum d'une fonction de coût modélisant les écarts de la prévision par rapport aux observations, il s'agit ici de résoudre un problème d'optimisation.

Sous certaines hypothèses, en particulier de linéarité des opérateurs, les deux approches conduisent aux mêmes résultats.

Les méthodes d'assimilation de données se répartissent parmi deux groupes pour leur utilisation : les méthodes de *prévision* où l'on cherche à connaître le futur, ce qui suppose que toutes les observations proviennent du passé ou présent de l'analyse (comme dans le cas de l'exemple précédent), et les méthodes de *réanalyse/lissage* où l'on cherche à reconstruire le passé, avec des observations provenant aussi du futur de l'analyse, qui permettent de raffiner la connaissance disponible sur les statistiques d'erreur.

Pour formaliser et construire les différents algorithmes d'assimilation, il faut pouvoir représenter l'incertitude sur chaque composante du problème, soit chacune des variables du modèle et chacune des observations disponibles. En mesurant ou en supposant les statistiques d'erreurs, cette incertitude sera définie par une approche probabiliste, grâce à des densités de probabilité, en particulier la distribution Gaussienne. Ces éléments permettent d'écrire l'algorithme d'analyse de manière à ce que l'erreur d'analyse obtenue lors du processus d'assimilation soit minimale selon une norme définie par l'utilisateur. Ceci fait de l'assimilation de données un problème d'optimisation.

## 2.1.2 Variables, notations et définitions

### Modèle et vecteur d'état

Une référence globale en terme de notations et concepts sur l'assimilation de données est disponible chez (IDE et al., 1997), nous nous baserons dessus pour définir nos notations. Bien que le déroulement de la physique modélisée soit de nature continue, nous ne pouvons représenter numériquement les espaces de dimension infinie que par des approximations finies (discrétisation sur un maillage...), c'est pourquoi toutes les variables que nous manipulerons seront en dimension finie. Nous noterons avant tout les dimensions de l'espace des états par  $n \in \mathbb{N}^*$  et de l'espace des observations par  $p \in \mathbb{N}^*$ . Concrètement, il est commode d'avoir en tête que l'espace des états est  $\mathbb{R}^n$  et que l'espace des observations est  $\mathbb{R}^p$ .

Nous pouvons ainsi définir l'opérateur du *modèle*  $\mathbf{x} \mapsto \mathcal{M}(\mathbf{x})$  qui permet de calculer l'état du système physique. Nous y associons le *vecteur d'état*  $\mathbf{x} \in \mathbb{R}^n$  dont les composantes sont les variables d'entrées du modèle. Ces variables peuvent être, selon les applications, des données physiques telles que la pression au sein d'une couche de l'atmosphère en météorologie, des données de fonctionnement comme la position des grappes d'assemblage dans les réacteurs nucléaires ou encore des champs physiques comme la salinité d'une portion d'océan en océanographie.

En général, les variables vectorielles sont discrétisées sur un maillage. Plus la dimension de la simulation (2D, 3D..) est grande, plus celle du vecteur d'état l'est en conséquence ! Plusieurs déclinaisons existent pour le vecteur d'état :

- $\mathbf{x}^t$  : représente *l'état vrai* du système, c'est-à-dire l'état qu'il aurait si nous en avions une connaissance parfaite sans la moindre approximation. Il représente l'état de

référence sur lequel nous effectuons nos mesures et représente un idéal qui nous est inaccessible,

- $\mathbf{x}^b$  ou  $\mathbf{x}^f$  : représente respectivement *l'ébauche* et la *prévision*, qui sont nos connaissances a priori du système obtenues soit par inférence, soit par simulation grâce au modèle,
- $\mathbf{x}^a$  : représente *l'état analysé* du système, celui qui est calculé lors de l'étape d'analyse.

## Observations

L'ensemble des observations recueillies à un temps  $t \in \mathbb{R}$  est contenu dans un vecteur  $\mathbf{y} \in \mathbb{R}^p$ , appelé *vecteur d'observations*. Ces observations peuvent être directement des grandeurs physiques comme le taux d'humidité, la pression, ou bien être indirectement liées à une grandeur physique comme c'est le cas des radiances satellites en météorologie qui correspondent à une transformation du champ de température à différents niveaux de pression.

Les composantes de  $\mathbf{y}$  peuvent donc être de nature totalement différente de celles de  $\mathbf{x}$ , c'est pourquoi les deux vecteurs n'appartiennent en général pas au même espace.

Leur répartition peut varier à la fois en temps et en espace. En météorologie, certaines zones terrestres sont très bien recouvertes de stations comme l'Europe, alors que d'autres sont pauvres en mesures comme les déserts et les océans, du fait qu'il soit bien plus aisé d'instrumenter la terre ferme fortement peuplée plutôt que des zones où les contraintes environnementales sont pénalisantes. Cela mène donc à des hétérogénéités dans la répartition spatiale des différents capteurs.

Pour l'aspect temporel, dans le cas de la neutronique, les mesures du décalage de puissance axiale du cœur sont disponibles environ chaque minute et servent à avoir une idée du flux de neutrons au sein de celui-ci. Elles sont pratiquées sur les bords extérieurs de la cuve et représentent une mesure indirecte du flux neutronique. Les mesures internes de calcul direct de celui-ci ne sont faites que tous les mois. Là encore, la raison de cette disparité est évidente : il n'est pas aisé d'introduire et d'extraire à volonté des instruments de mesure dans un réacteur nucléaire en fonctionnement ! Cela mène à une hétérogénéité en temps sur la précision des informations disponibles, la mesure directe permettant une connaissance très fine du flux à l'instant  $t$ , ainsi que sur la fréquence de disponibilités de deux informations de nature physique différente.

Pour pouvoir effectuer des comparaisons directes entre  $\mathbf{x}$  et  $\mathbf{y}$ , *l'opérateur d'observation*  $\mathbf{y} \mapsto \mathcal{H}(\mathbf{x})$  est introduit. Il sert à mettre en relation l'état dans l'espace des observations et encode le lien implicite entre le vecteur d'état et le vecteur d'observations.

### 2.1.3 Les différents types d'erreurs et statistiques d'erreur

Malgré leurs incertitudes respectives, le modèle et les observations sont les meilleures approches disponibles de l'état vrai du système physique auquel nous n'avons pas accès. Pour pouvoir garantir la confiance dans nos estimations, il est primordial de bien caractériser les erreurs commises lors de la procédure d'assimilation de données.

### Erreur d'ébauche (ou de prévision)

L'erreur d'ébauche notée  $\epsilon^b$  correspond à l'écart entre l'estimation a priori de l'état vrai qu'est l'ébauche et l'état vrai, elle s'écrit :

$$\epsilon^b = \mathbf{x}^b - \mathbf{x}^t. \quad (2.1)$$

Faisant intervenir  $\mathbf{x}^t$ , nous n'y avons pas directement accès.

Sa modélisation sous la forme d'une matrice de covariance d'erreur se note en général  $\mathbf{B}$  et vaut :

$$\mathbf{B} = \mathbb{E} \left[ \left( \epsilon^b - \mathbb{E} [\epsilon^b] \right) \left( \epsilon^b - \mathbb{E} [\epsilon^b] \right)^T \right] \in \mathcal{M}_n(\mathbb{R}). \quad (2.2)$$

$\mathbf{B}$  est symétrique définie positive.

### Erreur d'analyse

De façon similaire, l'erreur d'analyse notée  $\epsilon^a$  correspond à :

$$\epsilon^a = \mathbf{x}^a - \mathbf{x}^t. \quad (2.3)$$

Nous n'y avons également pas accès. Sa matrice de covariance d'erreur associée se note en général  $\mathbf{P}^a$  et vaut :

$$\mathbf{P}^a = \mathbb{E} \left[ \left( \epsilon^a - \mathbb{E} [\epsilon^a] \right) \left( \epsilon^a - \mathbb{E} [\epsilon^a] \right)^T \right] \in \mathcal{M}_n(\mathbb{R}). \quad (2.4)$$

C'est cette erreur que l'on veut minimiser dans le processus d'assimilation.

La trace de  $\mathbf{P}^a$  sert d'estimation de l'erreur de l'état analysé, donc aussi de critère de minimisation. De manière générale, c'est cette grandeur que les méthodes d'assimilation de données cherchent à réduire.

D'autres critères de qualité de l'analyse sont aussi utilisés :

- minimisation de la RMSE (Root Mean Square Error) entre l'état analysé et les observations :

$$RMSE(\mathbf{x}^a, \mathbf{y}) = \frac{1}{2} \sqrt{\frac{1}{p} \sum_{j=1}^p \left( \mathcal{H}(\mathbf{x}_j^a) - \mathbf{y}_j \right)^2} \quad (2.5)$$

appelée RMSE d'analyse,

- maximisation de la fonction de vraisemblance, où la vraisemblance est définie comme la probabilité conditionnelle de l'état vrai  $\mathbf{x}^t$  étant donné les observations  $\mathbf{y}$ .

Il est important de remarquer que dans le cas linéaire gaussien, le critère de maximisation de la vraisemblance et de minimisation de  $Tr(\mathbf{P}^a)$  sont équivalents.

### Erreur d'observation

L'erreur d'observation notée  $\epsilon^{obs}$  correspond à l'écart entre l'état vrai et l'observation :

$$\epsilon^{obs} = \mathbf{y} - \mathcal{H}(\mathbf{x}^t). \quad (2.6)$$

Sa matrice de covariance d'erreur  $\mathbf{R}$  vaut :

$$\mathbf{R} = \mathbb{E} \left[ \left( \epsilon^{obs} - \mathbb{E} [\epsilon^{obs}] \right) \left( \epsilon^{obs} - \mathbb{E} [\epsilon^{obs}] \right)^T \right] \in \mathcal{M}_p(\mathbb{R}). \quad (2.7)$$

Les écarts entre l'observation et l'état vrai peuvent s'expliquer de deux façons :

- erreurs de mesure : la précision des instruments de mesure n'est pas parfaite, et nous devons prendre cela en compte ; ces erreurs sont modélisées de manière stochastique, de moyenne et de variance connues,
- erreurs de représentativité : le lien entre l'état et les observations encodé par  $\mathcal{H}$  est en général complexe, et peut lui-même contenir des modèles en son sein, ce qui rajoute des erreurs plus difficiles à appréhender.

L'erreur de représentativité est une notion vaste et plus complexe que ce qui est décrit ici, c'est encore une question qui fait débat dans la communauté scientifique. L'article de review de (JANJIC et al., 2017) est une bonne source pour approfondir le sujet.

La matrice  $\mathbf{R}$  est symétrique définie positive, pour les mêmes raisons que la matrice  $\mathbf{B}$ . Il est important de remarquer que la qualité de l'analyse repose fortement sur la qualité de l'estimation de ces matrices  $\mathbf{B}$  et  $\mathbf{R}$ .

**Remarque :** il est d'usage de supposer avoir des erreurs d'observations et d'ébauche sans biais et décorréelées, soit

- $\mathbb{E} [\epsilon^b] = 0$ ,
- $\mathbb{E} [\epsilon^{obs}] = 0$ ,
- $\mathbb{E} [(\epsilon^b)(\epsilon^{obs})^T] = 0$ .

Si tel n'est pas le cas, il convient d'estimer le biais déterministe pour que ces hypothèses soient vérifiées, chose faite par les fournisseurs d'observations en météorologie par exemple, parmi d'autres.

### Gestion de l'erreur modèle

L'erreur modèle est une grandeur difficile à quantifier et souvent négligée par simplicité. Ses effets se font tout de même ressentir, et plusieurs méthodes se sont développées pour les prendre en compte au mieux dans le domaine applicatif, à la fois dans le cadre de l'EnKF (HOUTEKAMER et MITCHELL, 2008) et du 4D-Var (TRÉMOLET, 2006a ; TRÉMOLET, 2006b ; TRÉMOLET, 2007). Le biais de prévision constitue une de ses manifestations, comme vu chez (VIDARD, PIACENTINI et LE DIMET, 2004).

Plusieurs travaux théoriques sur sa caractérisation et sa prise en compte ont aussi vu le jour, notamment ceux de (NICOLIS, 2003 ; CARRASSI, VANNITSEN et NICOLIS, 2008 ; HARLIM, 2015) dans un cadre global, et ceux de (SHUTYAEV et al., 2016a ; SHUTYAEV et al., 2016b) pour l'assimilation de données variationnelle.

### Représentation des statistiques d'erreur

Chacune des variables d'erreur décrites précédemment est en pratique inconnue, et peut être considérée comme la réalisation d'un vecteur aléatoire réel. Chaque variable aléatoire réelle à valeurs dans  $\mathbb{R}$  est caractérisée entièrement par sa fonction de répartition  $\mathbb{P}$  pour

$x \in \mathbb{R}$ , qui se définit aisément à partir de sa densité de probabilité  $p$  via la formule

$$\mathbb{P}(x) = \int_{-\infty}^x p(t)dt. \quad (2.8)$$

Avec  $r \in \mathbb{N}$ , les moments  $m_r$  d'ordre  $r$  d'une variable aléatoire permettent de caractériser la dispersion de cette variable et même parfois comme dans le cas gaussien, la densité de probabilité et a fortiori la loi de la variable. Ils sont caractérisés grâce à la densité de probabilité de la variable aléatoire via la formule suivante :

$$m_r = \int_{t \in \mathbb{R}} t^r p(t)dt. \quad (2.9)$$

Dans la pratique, il est très rare d'avoir des informations statistiques au delà du second ordre, qui correspond à la variance. De ce fait, une loi très commode pour modéliser les statistiques d'erreur est la loi gaussienne, qui est entièrement définie de manière unique par ses moments d'ordre 1 et 2, la moyenne et la variance. C'est pour cela que les lois gaussiennes sont majoritairement utilisées en assimilation de données – et de nombreux autres domaines – pour la modélisation des erreurs.

### 2.1.4 L'estimateur BLUE et ses fondements

L'estimateur BLUE est un préalable à l'algorithme du filtre de Kalman que nous allons présenter ici. Les détails des calculs de cette partie se trouvent en annexe (A.1).

#### Gain de l'estimation et erreur d'analyse

Nous supposons qu'il est possible de décomposer l'état analysé  $\mathbf{x}^a$  comme une combinaison linéaire entre l'ébauche  $\mathbf{x}^b$  et les observations  $\mathbf{y}$  :

$$\mathbf{x}^a = \mathbf{L}\mathbf{x}^b + \mathbf{K}\mathbf{y} \quad (2.10)$$

avec  $\mathbf{L} \in \mathcal{M}_n(\mathbb{R})$  et  $\mathbf{K} \in \mathcal{M}_{n,p}(\mathbb{R})$ .

En se remémorant les équations (2.1) et (2.6), nous avons  $\mathbf{y} = \mathcal{H}(\mathbf{x}) + \epsilon^{obs}$ . Nous supposons aussi que l'opérateur d'observation  $\mathcal{H}$  admet pour dérivée en  $\mathbf{x}^t$  la matrice  $\mathbf{H}$ , et que les corrections apportées par l'analyse dépendent linéairement des innovations  $\mathbf{y} - \mathcal{H}(\mathbf{x})$ . Nous pouvons alors calculer l'erreur d'analyse :

$$\epsilon^a = \mathbf{x}^a - \mathbf{x}^t = \mathbf{L}\epsilon^b + \mathbf{K}\epsilon^{obs} + (\mathbf{L} + \mathbf{K}\mathbf{H} - \mathbf{I})\mathbf{x}^t. \quad (2.11)$$

Nous voulons une erreur d'analyse non-biaisée. Vu que les erreurs d'ébauche et d'observations le sont par hypothèses, grâce à l'équation (2.11) cela s'écrit comme suit :

$$\mathbb{E}[\epsilon^a] = \mathbb{E}[\mathbf{x}^a - \mathbf{x}^t] = (\mathbf{L} + \mathbf{K}\mathbf{H} - \mathbf{I})\mathbb{E}[\mathbf{x}^t] = 0. \quad (2.12)$$

Il suffit pour cela que  $\mathbf{L} + \mathbf{K}\mathbf{H} - \mathbf{I} = 0 \iff \mathbf{L} = \mathbf{I} - \mathbf{K}\mathbf{H}$ . Ainsi :

$$\mathbf{x}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{x}^b + \mathbf{K}\mathbf{y} = \mathbf{x}^b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^b). \quad (2.13)$$

Nous voyons que l'analyse s'exprime alors comme une correction de l'ébauche par les innovations, pondérée par la matrice  $\mathbf{K}$  appelée *gain*. L'analyse correspond donc à une régression linéaire, et pour qu'elle minimise l'erreur d'analyse, il suffit de déterminer un gain satisfaisant. Pour cela, nous regardons l'expression de la matrice de covariance de l'erreur d'analyse.

Les erreurs d'observations et d'ébauche sont sans biais et décorréelées, nous déduisons ainsi de l'équation (2.10) que :

$$\epsilon^a = \mathbf{L}\epsilon^b + \mathbf{K}\epsilon^{obs}. \quad (2.14)$$

Par conséquent :

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T, \quad \mathbf{L} = \mathbf{I} - \mathbf{K}\mathbf{H}. \quad (2.15)$$

Nous connaissons désormais l'expression de  $\mathbf{P}^a$ . Un gain satisfaisant  $\mathbf{K}$  serait par exemple tel qu'il minimise l'erreur scalaire de la matrice  $Tr(\mathbf{P}^a)$ , c'est ainsi que nous allons le rechercher et c'est lui qui donnera l'estimateur BLUE.

### Calcul du BLUE

Trouver le gain optimal  $\mathbf{K}^*$  qui minimise  $Tr(\mathbf{P}^a)$  revient à étudier la variation de  $\mathbf{K} \mapsto Tr(\mathbf{P}^a(\mathbf{K}))$  lors d'une variation de  $\mathbf{K}$ .

Nous pouvons écrire que :

$$\begin{aligned} Tr(\mathbf{P}^a(\mathbf{K})) &= Tr[(\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T] \\ &= \langle \mathbf{I} - \mathbf{K}\mathbf{H}, (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} \rangle + \langle \mathbf{K}, \mathbf{R}\mathbf{K} \rangle \\ &= f(\mathbf{K}), \end{aligned}$$

où  $\langle \cdot \rangle$  représente le produit scalaire canonique sur l'espace des matrices.  $\mathbf{K} \mapsto f(\mathbf{K})$  est donc différentiable, et son gradient vaut :

$$\nabla f(\mathbf{K}) = -2\mathbf{B}\mathbf{H}^T + 2\mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T + 2\mathbf{K}\mathbf{R}. \quad (2.16)$$

$\mathbf{K}^*$  est obtenu lorsque  $\nabla f(\mathbf{K}) = 0$ . Cela donne :

$$\mathbf{K}^* = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}, \quad (2.17)$$

et l'estimateur BLUE de l'état analysé vaut :

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}^*(\mathbf{y} - \mathbf{H}\mathbf{x}^b) = \mathbf{x}^b + \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}^b). \quad (2.18)$$

Tout cela suppose que  $\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T$  soit inversible, ce qui est le cas lorsque  $rang(\mathbf{H}) = \max(n, p)$ . Autrement, la pseudo-inverse est utilisée.

**Remarque :** Grâce à la formule de Sherman-Morrison-Woodbury (GOLUB et VAN LOAN, 2013), une expression équivalente de  $\mathbf{K}^*$  peut être calculée :

$$\mathbf{K}^* = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}. \quad (2.19)$$



Selon la configuration du problème d'assimilation, l'expression (2.19) peut devenir plus avantageuse que celle de (2.17). L'inversion matricielle nécessaire à son calcul se fait cette fois dans  $\mathcal{M}_n(\mathbb{R})$  au lieu de  $\mathcal{M}_p(\mathbb{R})$ , ce qui va fortement réduire son coût lorsque  $n \ll p$ . Inversement, lorsque  $n \gg p$ , il faut lui préférer l'expression (2.17). L'expression (2.19) permet aussi de montrer l'équivalence entre le BLUE et l'approche variationnelle, il est possible de prouver que l'analyse BLUE minimise la fonction de coût suivante :

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}). \quad (2.20)$$

### 2.1.5 L'assimilation de données séquentielle

Une procédure d'assimilation de données séquentielle-type se décompose en deux étapes : l'étape de *prévision* et l'étape d'*analyse* ou *assimilation*.

Pour poser le problème, soit un temps  $t_k$  auquel nous avons le résultat d'une précédente prévision (qui peut aussi être une condition initiale pour  $k = 0$ ) noté  $\mathbf{x}_k^b$  pour l'ébauche au temps  $t_k$ , et un vecteur d'observations  $\mathbf{y}_k$  au temps  $t_k$ . A partir de ces deux vecteurs, nous pouvons effectuer une analyse donnant l'état analysé  $\mathbf{x}_k^a$  en appliquant l'estimateur BLUE par exemple. En outre, grâce à un modèle d'évolution temporelle, nous calculons ensuite une nouvelle prévision  $\mathbf{x}_{k+1}^f$  au temps  $t_{k+1}$ , qui deviendra elle-même le nouveau point de départ (l'ébauche) pour l'analyse au temps  $t_{k+1}$ , lorsque nous aurons connaissance de  $\mathbf{y}_{k+1}$ .

Ce processus est illustré par le schéma suivant :

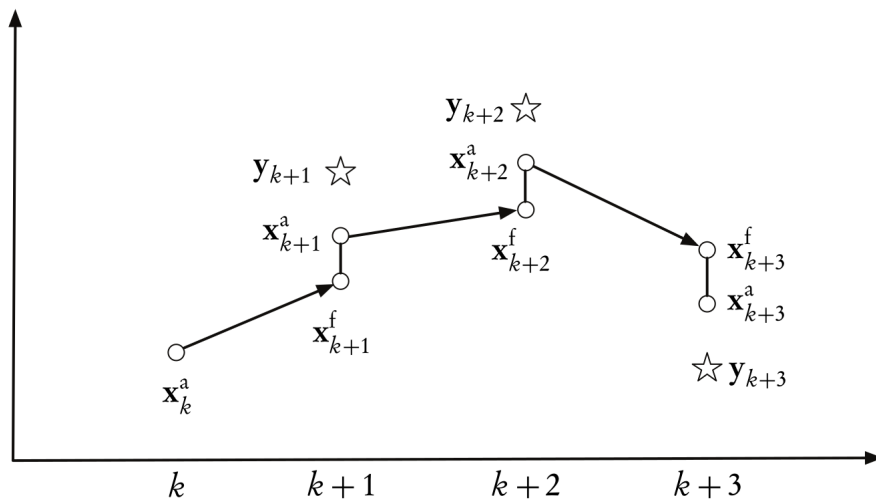


FIGURE 2.1 – Adapté depuis (ASCH, BOCQUET et NODET, 2016). Voici trois itérations d'une procédure d'assimilation de données séquentielle comme le filtre de Kalman. Les étoiles représentent les observations et les ronds le vecteur d'état. L'axe des abscisse représente le temps et l'axe des ordonnées les valeurs des vecteurs d'état et d'observation.

Formellement, en reprenant ce qui a été énoncé depuis le début de cette section, nous pouvons considérer l'état  $\mathbf{x}$  comme un processus stochastique obéissant à l'équation d'évolution suivante :

$$\mathbf{x}_{k+1} = \mathcal{M}_{k+1}(\mathbf{x}_k) + \eta_k, \quad \eta_k \sim \mathcal{N}(0, \mathbf{Q}_k), \quad (2.21)$$

et les observations comme un autre processus stochastique lié à l'état obéissant à :

$$\mathbf{y}_{k+1} = \mathcal{H}_{k+1}(\mathbf{y}_k) + \epsilon_k, \quad \epsilon_k \sim \mathcal{N}(0, \mathbf{R}_k). \quad (2.22)$$

Ici,  $\eta_k$  et  $\epsilon_k$  correspondent respectivement à l'erreur modèle et l'erreur de mesure. De manière générale, elles sont supposées indépendantes.

Le point de vue de la théorie des processus stochastiques permet d'inspecter tout cela sous l'angle des équations différentielles stochastiques, et en suivant la théorie de l'intégration d'Itô, de montrer l'existence et l'unicité des solutions de ce problème de filtrage optimal (ØKSENDAL, 2016).

La prochaine section est consacrée à une brève description des méthodes variationnelles, qui permettent de résoudre ce problème d'assimilation de données séquentielle. Elles permettent également de faire le lien entre chaque séquence  $[t_k, t_{k+1}[$ , c'est-à-dire d'incorporer une dynamique temporelle aux estimateurs statistiques comme le BLUE, à l'instar du filtre de Kalman.

## 2.2 Approche variationnelle

Le point de vue de l'approche variationnelle diffère de celui de l'interpolation statistique car son origine provient du contrôle optimal. Le principe de cette école est de calculer l'analyse comme la solution d'un problème d'optimisation décrit par une fonction de coût  $J$ . De manière générale, cette solution est recherchée par une méthode itérative, ce qui nécessite plusieurs évaluations de la fonction  $J$  et de son gradient  $\nabla J$ .

La méthode 3D-Var est l'équivalent en variationnel de l'estimateur BLUE issue de l'estimation optimale. C'est une méthode statique qui permet de résoudre le problème d'assimilation de données séquentielle, et qui donne un résultat équivalent au BLUE (Cf. sous-section (A.2.1)) dans le cas linéaire. La dimension temporelle est ajoutée par la méthode 4D-Var, par une modification de la fonction de coût.

Nous présentons plus en détail ces méthodes dans l'annexe en section (A.2). Nous allons maintenant regarder la façon de résoudre le problème d'assimilation de données séquentielle grâce aux méthodes héritées de l'estimation optimale.

## 2.3 Résolution du problème d'assimilation de données séquentielle

Cette section décrit les méthodes utilisées pour résoudre le problème d'assimilation de données séquentielle sous le prisme des méthodes de filtrage, méthodes héritées de l'interpolation statistique.

Dans un premier temps sera présentée la méthode fondatrice qu'est le filtre de Kalman, puis son évolution qu'est le filtre de Kalman étendu et enfin la méthode centrale de cette thèse, le filtre de Kalman d'ensemble, qui aura sa propre section lui étant dédiée.

### 2.3.1 Le filtre de Kalman (Kalman, 1960 ; Jazwinski, 1970)

Le filtre de Kalman introduit une dimension temporelle dans l'assimilation de données par interpolation statistique en prenant en compte l'opérateur d'évolution temporelle du modèle. Nous présenterons ici sa version discrète qui est celle qui nous intéresse pour la suite de la thèse.

Soit un modèle de système dynamique défini par l'équation d'évolution suivante :

$$\mathbf{x}_{k+1} = \mathbf{M}_{k+1}\mathbf{x}_k + \eta_{k+1} \quad (2.23)$$

$$\mathbf{P}_{k+1} = \mathbf{M}_{k+1}\mathbf{P}_k\mathbf{M}_{k+1}^T + \mathbf{Q}_{k+1} \quad (2.24)$$

avec  $\mathbf{x}_k \in \mathbb{R}^n$  qui représente l'état du système au temps discret  $t_k > 0$ ,  $\mathbf{P}_k \in \mathcal{M}_n(\mathbb{R})$  la matrice de covariance d'erreur relative à l'état  $\mathbf{x}_k$ , et  $\eta_{k+1} \in \mathbb{R}^n$  un bruit gaussien indépendant de  $\mathbf{x}_k$ .

Les matrices  $\mathbf{M}_k \in \mathcal{M}_{m,n}(\mathbb{R})$  et  $\mathbf{Q}_k \in \mathcal{M}_n(\mathbb{R})$  correspondent respectivement à l'opérateur d'évolution du système et à la matrice de covariance d'erreur de modélisation du système, c'est-à-dire la matrice de covariance des  $\eta_k$ . Ce bruit  $\eta_k$  modélise le fait que pour la plupart des systèmes physiques complexes, la résolution des équations qui les modélisent nécessite d'appliquer des hypothèses simplificatrices ou des approximations qui introduisent une erreur non-quantifiable.

En général et par simplicité, le modèle est considéré comme non-biaisé ce qui signifie que  $\forall k \in \mathbb{N}$ ,  $\mathbb{E}[\eta_k] = 0$ .

Nous disposons en outre de mesures bruitées de ce système dynamique données par :

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k^t + \epsilon_k \in \mathbb{R}^p. \quad (2.25)$$

$\mathbf{H}_k \in \mathcal{M}_{p,n}(\mathbb{R})$  est l'opérateur d'observation,  $\mathbf{x}_k^t \in \mathbb{R}^n$  correspond à l'état-vrai du système, et  $\epsilon_k \in \mathbb{R}^p$  correspond à l'erreur de mesure,  $\epsilon_k \sim \mathcal{N}(0, \mathbf{R}_k)$ , avec  $\mathbf{R}_k \in \mathcal{M}_p(\mathbb{R})$  la matrice de covariance d'erreur de mesure.

Le filtre de Kalman étant une méthode de filtrage séquentielle dynamique, le modèle est intégré dans le temps et, dès que des mesures sont disponibles, celles-ci sont assimilées lors de l'étape d'analyse pour mettre à jour le modèle avant une nouvelle intégration.

Par souci de lisibilité, nous négligeons l'indice temporel  $k$  pour les équations suivantes. Les notations  $\mathbf{x}^f$  et  $\mathbf{P}^f$  représentent la prévision et la matrice de covariance d'erreur de prévisions, décrites à la page suivante par les équations (2.29) et (2.30). L'équation d'analyse (ou d'assimilation) du filtre de Kalman correspond à la construction de :

$$\mathbf{x}_a = \mathbf{x}_f + \mathbf{P}^f\mathbf{H}^T \left( \mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R} \right)^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}_f) \quad (2.26)$$

$$\mathbf{P}_a = \mathbf{P}^f - \mathbf{P}^f\mathbf{H}^T \left( \mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R} \right)^{-1} \mathbf{H}\mathbf{P}^f \quad (2.27)$$

où la matrice

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T \left( \mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R} \right)^{-1} \quad (2.28)$$

représente le **gain de Kalman**. Nous remarquons qu'il s'agit de la même expression que le gain (2.17) pour l'estimateur BLUE.

Cet état analysé nous donne ensuite un nouveau point de départ pour l'intégration du

modèle jusqu'au temps suivant.

En réintroduisant les indices temporels, l'algorithme du filtre de Kalman discret correspond donc à l'alternance entre l'étape de **prévision** pour passer du temps  $t_k$  au temps  $t_{k+1}$  :

$$\mathbf{x}_{k+1}^f = \mathbf{M}_{k+1} \mathbf{x}_k^a \quad (2.29)$$

$$\mathbf{P}_{k+1}^f = \mathbf{M}_{k+1} \mathbf{P}_k^a \mathbf{M}_{k+1}^T + \mathbf{Q}_{k+1} \quad (2.30)$$

et l'étape d'**analyse ou assimilation** au temps  $t_{k+1}$  :

$$\mathbf{x}_{k+1}^a = \mathbf{x}_{k+1}^f + \mathbf{P}_{k+1}^f \mathbf{H}_{k+1}^T \left( \mathbf{H}_{k+1} \mathbf{P}_{k+1}^f \mathbf{H}_{k+1}^T + \mathbf{R} \right)^{-1} (\mathbf{y}_{k+1} - \mathbf{H}_{k+1} \mathbf{x}_{k+1}^f) \quad (2.31)$$

$$\mathbf{P}_{k+1}^a = \mathbf{P}_{k+1}^f - \mathbf{P}_{k+1}^f \mathbf{H}_{k+1}^T \left( \mathbf{H}_{k+1} \mathbf{P}_{k+1}^f \mathbf{H}_{k+1}^T + \mathbf{R} \right)^{-1} \mathbf{H}_{k+1} \mathbf{P}_{k+1}^f \quad (2.32)$$

En assimilation de données séquentielle, la simulation par filtre de Kalman correspond à l'alternance entre ces deux étapes jusqu'à avoir avancé au temps voulu.

Il est important de noter que l'algorithme se formule avec des opérateurs linéaires, il faut donc l'adapter pour l'utiliser avec des systèmes non-linéaires.

### 2.3.2 Le filtre de Kalman étendu (Jazwinski, 1970)

Une modification de l'algorithme du filtre de Kalman permet d'incorporer la prise en compte partielle des non-linéarités dans l'algorithme initial, c'est ce que l'on appelle le "filtre de Kalman étendu" (EKF).

Pour le système dynamique suivant, défini via l'opérateur non linéaire  $\mathbf{x} \in \mathbb{R}^n \mapsto \mathcal{M}(\mathbf{x}) \in \mathbb{R}^n$  :

$$\mathbf{x}_k = \mathcal{M}_k(\mathbf{x}_{k-1}) + \eta_k \quad (2.33)$$

Pour obtenir l'algorithme du filtre de Kalman étendu, il suffit de calculer l'opérateur tangent linéaire (sa jacobienne) de  $\mathcal{M}_k$  au temps voulu (ici  $t_k$ ) que nous noterons  $\mathbf{M}_k$ , et d'utiliser cette approximation dans le calcul de l'évolution de la matrice de covariance d'erreur :

$$\mathbf{P}_k = \mathbf{M}_k \mathbf{P}_{k-1} \mathbf{M}_k^T + \mathbf{Q}_k = \left( \frac{d\mathcal{M}_k}{dt} \right)_{|t=t_k} \mathbf{P}_{k-1} \left( \frac{d\mathcal{M}_k}{dt} \right)_{|t=t_k}^T + \mathbf{Q}_k. \quad (2.34)$$

Dans la majorité des cas, le calcul de l'opérateur tangent linéaire est fastidieux et coûteux en ressource numérique. L'EKF implique donc un surcoût numérique significatif en comparaison du filtre de Kalman classique.

Comme montré dans (EVENSEN, 1992), cette approximation peut s'avérer problématique dans certains cas comme un modèle quasi-géostrophique non linéaire de circulation océanique, faisant apparaître une instabilité non-bornée lors de l'évolution de la matrice  $\mathbf{P}$  selon la formule (2.34), pouvant mener à la divergence du filtre.

Pour permettre le transport des non-linéarités sans avoir à calculer d'opérateur tangent linéaire comme dans l'EKF et s'affranchir des limitations découvertes dans (EVENSEN, 1992), (EVENSEN, 1994) a formulé l'EnKF en s'inspirant des méthodes de Monte-Carlo.

Ceci permet également d'approximer les matrices d'erreurs grâce à un ensemble d'échantillon et de réduire le besoin en stockage pour de grosses matrices. Nous allons décrire son fonctionnement dans la section suivante.

## 2.4 Définition de l'EnKF (Evensen, 1994)

### 2.4.1 Ensemble et statistiques d'erreur

En reprenant les notations précédentes, les matrices de covariance de prévision et d'analyse pour le filtre de Kalman s'écrivent formellement :

$$\mathbf{P}^f = \mathbb{E} \left[ (\mathbf{x}^f - \mathbf{x}^t)(\mathbf{x}^f - \mathbf{x}^t)^T \right] \quad (2.35)$$

$$\mathbf{P}^a = \mathbb{E} \left[ (\mathbf{x}^a - \mathbf{x}^t)(\mathbf{x}^a - \mathbf{x}^t)^T \right]. \quad (2.36)$$

Or nous n'avons pas accès à la connaissance de l'état-vrai. Pour en obtenir une approximation, l'EnKF forme un ensemble d'états possibles  $\mathbf{E}$  et en prend la moyenne qui sera notée  $\bar{\mathbf{x}}^{f,a}$  pour la prévision et l'analyse respectivement. Le  $i^{\text{ème}}$  membre de l'ensemble sera noté  $\mathbf{x}^{i,*}$ , avec  $*$   $\in \{f, a\}$  selon qu'il s'agisse de la prévision ou de l'analyse. Cela permet d'approximer les matrices précédentes par :

$$\mathbf{P}_e^f = \mathbb{E} \left[ (\mathbf{x}^f - \bar{\mathbf{x}}^f)(\mathbf{x}^f - \bar{\mathbf{x}}^f)^T \right] \simeq \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}^{i,f} - \bar{\mathbf{x}}^f)(\mathbf{x}^{i,f} - \bar{\mathbf{x}}^f)^T \quad (2.37)$$

$$\mathbf{P}_e^a = \mathbb{E} \left[ (\mathbf{x}^a - \bar{\mathbf{x}}^a)(\mathbf{x}^a - \bar{\mathbf{x}}^a)^T \right] \simeq \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}^{i,a} - \bar{\mathbf{x}}^a)(\mathbf{x}^{i,a} - \bar{\mathbf{x}}^a)^T, \quad (2.38)$$

où le sous-script  $e$  indique qu'il s'agit d'une approximation ensembliste. Cette écriture permet de calculer directement la matrice d'erreur via l'ensemble, donc de prendre en compte les non-linéarités dans le cas d'un modèle non-linéaire, car les membres seront intégrés dans le temps grâce à celui-ci.

En omettant les notations de prévision/analyse, la matrice  $\mathbf{P}$  peut aussi se factoriser sous une forme utile lors de l'implémentation, via la matrice des anomalies normalisées  $\Delta\mathbf{X}$  définie comme suit :

$$\forall i \in \llbracket 1, N \rrbracket, \Delta\mathbf{X}(:, i) = \frac{\mathbf{x}^i - \bar{\mathbf{x}}}{\sqrt{N-1}} \implies \mathbf{P} = (\Delta\mathbf{X})(\Delta\mathbf{X})^T. \quad (2.39)$$

### 2.4.2 Algorithme final

L'étape de prévision revient donc à calculer, pour un ensemble de  $N$  membres  $\mathbf{E} \in \mathcal{M}_{n,N}(\mathbb{R})$  :

$$\forall i \in \llbracket 1, N \rrbracket, \mathbf{x}_{k+1}^{i,f} = \mathcal{M}_{k+1}(\mathbf{x}_k^{i,a}) + \eta_{k+1} \quad (2.40)$$

$$\bar{\mathbf{x}}_{k+1}^f = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_k^{i,f} \quad (2.41)$$

$$\mathbf{P}_{e,k+1}^f = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_{k+1}^{i,f} - \bar{\mathbf{x}}_{k+1}^f)(\mathbf{x}_{k+1}^{i,f} - \bar{\mathbf{x}}_{k+1}^f)^T. \quad (2.42)$$

Adapter l'étape d'analyse nécessite un peu plus de travail. (BURGERS, VAN LEUWEEN et EVENSEN, 1998) démontre qu'il faut traiter les observations elles aussi comme des variables aléatoires, pour éviter de sous-estimer la variance d'erreur lors de l'assimilation. Dans ce cas, en partant de l'observation  $\mathbf{y} \in \mathbb{R}^p$ , nous construisons un ensemble d'observations  $\mathbf{Y} \in \mathcal{M}_{p,N}(\mathbb{R})$  tel que :

$$\forall i \in \llbracket 1, N \rrbracket, \mathbf{Y}(:, i) = \mathbf{y}_i = \mathbf{y} + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \mathbf{R}). \quad (2.43)$$

Nous définissons ensuite la matrice de covariance d'erreur de l'ensemble d'observation par  $\mathbf{R}_e = \mathbb{E}[\epsilon\epsilon^T]$ , covariance empirique calculée à partir de l'ensemble d'observations perturbées  $\mathbf{Y}$ , qui vérifie  $\lim_{N \rightarrow +\infty} \mathbf{R}_e = \mathbf{R}$ , avec  $\mathbf{R}$  la matrice de covariance d'erreur de mesure qui serait utilisée dans le filtre de Kalman.

Cela nous permet ainsi de définir toutes les matrices nécessaires à l'étape d'analyse, et nous avons alors :

$$\mathbf{K}_e = \mathbf{P}_{e,k+1}^f \mathbf{H}_{k+1}^T \left( \mathbf{H}_{k+1} \mathbf{P}_{e,k+1}^f \mathbf{H}_{k+1}^T + \mathbf{R}_e \right)^{-1} \quad (2.44)$$

$$\bar{\mathbf{x}}_{k+1}^a = \bar{\mathbf{x}}_{k+1}^f + \mathbf{K}_e (\mathbf{y}_i - \mathbf{H}_{k+1} \bar{\mathbf{x}}_{k+1}^f) \quad (2.45)$$

$$\mathbf{P}_{e,k+1}^a = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_{k+1}^{i,a} - \bar{\mathbf{x}}_{k+1}^a) (\mathbf{x}_{k+1}^{i,a} - \bar{\mathbf{x}}_{k+1}^a)^T. \quad (2.46)$$

Le résultat de la deuxième équation est du au fait que l'on cherche, pour chaque membre de l'ensemble, à reproduire l'étape d'assimilation (2.31) du filtre de Kalman via les éléments calculables grâce à l'ensemble, ce qui se traduit par :

$$\forall i \in \llbracket 1, N \rrbracket, \mathbf{x}_{k+1}^{i,a} = \mathbf{x}_{k+1}^{i,f} + \mathbf{P}_{e,k+1}^f \mathbf{H}_{k+1}^T \left( \mathbf{H}_{k+1} \mathbf{P}_{e,k+1}^f \mathbf{H}_{k+1}^T + \mathbf{R}_e \right)^{-1} (\mathbf{y}_i - \mathbf{H}_{k+1} \mathbf{x}_{k+1}^{i,f}). \quad (2.47)$$

Une propriété remarquable de l'EnKF est sa capacité à être parallélisé : la mise à jour à l'analyse comme à la prévision de chaque membre de l'ensemble peut être faite indépendamment de celle des autres, ce qui est intéressant pour son utilisation sur des systèmes en applications réelles.

Il est à noter que cette version de l'algorithme de l'EnKF n'est pas celle originelle de (EVENSEN, 1994), il s'agit de l'EnKF stochastique (du fait de la perturbation des observations) qui prend directement en compte une amélioration de l'algorithme initial.

### 2.4.3 ETKF : une variante déterministe de l'EnKF

La méthode qui consiste à perturber les observations pour permettre à l'EnKF d'obtenir une estimation de variance d'erreur convenable n'est pas sans défaut, la génération des perturbations pour former les  $\mathbf{y}^i$  peut mener à un bruit numérique affectant les performances du filtre (ASCH, BOCQUET et NODET, 2016, p. 172). Une solution pour contourner ce problème est d'utiliser des schémas dits "déterministes" qui ne perturbent pas les observations, comme l'ETKF (Ensemble Transform Kalman Filter) décrit initialement par (BISHOP, ETHERTON et MAJUMDAR, 2001). C'est la variante déterministe de l'EnKF la plus répandue dans la littérature scientifique et les applications opérationnelles. Pour le

décrire, nous nous baserons sur l'article de (HUNT, KOSTELICH et SZUNYOGH, 2007).

Nous supposerons notre opérateur d'évolution modèle  $\mathcal{M}_k$  non-linéaire, et nous factorisons notre matrice de covariance d'erreur d'ébauche  $\mathbf{B}$  par sa matrice d'anomalies d'erreur associée  $\Delta\mathbf{X}^b$  selon l'équation (2.39), soit  $\mathbf{B} = (\Delta\mathbf{X}^b)(\Delta\mathbf{X}^b)^T$ .

Formellement, après calcul de l'étape d'analyse, la moyenne de l'ensemble analysé  $\bar{\mathbf{x}}^a$  est sensée minimiser la fonction de coût suivante :

$$\mathcal{J}(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) + [\mathbf{y} - \mathcal{H}(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y} - \mathcal{H}(\mathbf{x})]. \quad (2.48)$$

Sous sa forme actuelle, la matrice  $\mathbf{B} \in \mathcal{M}_n(\mathbb{R})$  est de rang au plus  $N - 1$  donc non-inversible car en général  $N \ll n, p$ .  $\mathbf{B}$  étant symétrique, son inverse est défini pour sa restriction à l'espace  $S$  engendré par ses colonnes, qui est le même que celui engendré par les colonnes de  $\Delta\mathbf{X}^b$  :  $S$  est donc l'espace engendré par les anomalies d'erreur d'ébauche. Ainsi, lorsque  $\mathbf{x} - \mathbf{x}^b \in S$ , la fonction  $\mathcal{J}$  est bien définie.

Ce changement d'espace dans lequel conduire l'assimilation implique également un changement de variables. L'idée de (HUNT, KOSTELICH et SZUNYOGH, 2007) est de considérer  $\Delta\mathbf{X}^b$  comme une transformation linéaire d'un espace  $N$ -dimensionnel  $\tilde{S}$  vers  $S$ . En prenant  $\mathbf{w} \in \tilde{S}$ , il devient possible de construire  $\mathbf{x} - \mathbf{x}^b = (\Delta\mathbf{X}^b)\mathbf{w} \in S$ , assurant la bonne définition de  $\mathcal{J}$ . Cela est possible en construisant  $\mathbf{w}$  comme un vecteur gaussien suivant une loi  $\mathcal{N}\left(0, \frac{\mathbf{I}_N}{N-1}\right)$ . Dans ce cas,  $\mathbf{x} = \mathbf{x}^b + (\Delta\mathbf{X}^b)\mathbf{w}$  suit une loi  $\mathcal{N}(\mathbf{x}^b, \mathbf{B})$ .

Suivant cette construction, la fonction de coût suivante devient intéressante pour travailler dans  $\tilde{S}$  :

$$\tilde{\mathcal{J}}(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{w}}{N-1} + [\mathbf{y} - \mathcal{H}(\mathbf{x}^b + (\Delta\mathbf{X}^b)\mathbf{w})]^T \mathbf{R}^{-1} [\mathbf{y} - \mathcal{H}(\mathbf{x}^b + (\Delta\mathbf{X}^b)\mathbf{w})]. \quad (2.49)$$

En effet, comme cela si  $\bar{\mathbf{w}}^a$  minimise  $\tilde{\mathcal{J}}$ , alors  $\mathbf{x}^a = \mathbf{x}^b + (\Delta\mathbf{X}^b)\bar{\mathbf{w}}^a$  minimise la fonction  $\mathcal{J}$  originale. En injectant le changement de variable dans la fonction  $\mathcal{J}$  de l'équation (2.48), cela donne une autre expression de  $\tilde{\mathcal{J}}$  :

$$\tilde{\mathcal{J}}(\mathbf{w}) = (N-1)\mathbf{w}^T \left( \mathbf{I}_N - (\Delta\mathbf{X}^b)^T [(\Delta\mathbf{X}^b)(\Delta\mathbf{X}^b)^T]^{-1} (\Delta\mathbf{X}^b) \right) \mathbf{w} + \mathcal{J}(\mathbf{x}^b + (\Delta\mathbf{X}^b)\mathbf{w}). \quad (2.50)$$

L'opérateur  $\mathbf{I}_N - (\Delta\mathbf{X}^b)^T [(\Delta\mathbf{X}^b)(\Delta\mathbf{X}^b)^T]^{-1} (\Delta\mathbf{X}^b)$  correspond au projecteur orthogonal sur  $\ker(\Delta\mathbf{X}^b)$ , assurant que si  $\bar{\mathbf{w}}^a$  minimise  $\tilde{\mathcal{J}}$ , alors le vecteur  $\mathbf{x}^a$  associé minimise  $\mathcal{J}$ . Cela nous permet de travailler essentiellement dans  $\tilde{S}$  en garantissant l'optimalité de la solution  $\mathbf{x}^a$  trouvée.

Le traitement de la non-linéarité de l'opérateur d'observation pour l'analyse se fait ensuite grâce à l'ensemble. Si  $\mathcal{H}$  est suffisamment non-linéaire, plusieurs minima peuvent exister pour  $\tilde{\mathcal{J}}$  ce qui devient problématique. Pour assurer un unique minimum, il suffit de linéariser  $\mathcal{H}$  dans l'espace de l'ensemble pour reformuler la fonction de coût et la rendre quadratique, ce qui revient à évaluer  $\mathcal{H}(\mathbf{x}^b + (\Delta\mathbf{X}^b)\mathbf{w})$  pour un certain  $\mathbf{w} \in \tilde{S}$ .

Une solution pratique pour linéariser est de construire un ensemble d'observations à partir de l'ensemble d'ébauche  $\mathbf{E}^b = (\mathbf{x}^{i,b})_{1 \leq i \leq N}$  dont la moyenne vaut  $\mathbf{x}^b$ , soit  $\forall i \in \llbracket 1, N \rrbracket$ ,  $\mathbf{y}^{i,b} = \mathcal{H}(\mathbf{x}^{i,b})$ .

Notons leur moyenne  $\mathbf{y}^b$ , et construisons de même la matrice d'anomalies d'erreur d'ébauche

$\Delta \mathbf{Y}^b$  pour ces observations :  $\forall i \in \llbracket 1, N \rrbracket$ ,  $\Delta \mathbf{Y}^b(:, i) = \frac{\mathbf{y}^{i,b} - \mathbf{y}^b}{\sqrt{N-1}}$ .

Dans ce cas, nous pouvons faire l'approximation linéaire suivante :

$$\mathcal{H}(\mathbf{x}^b + (\Delta \mathbf{X}^b)\mathbf{w}) \simeq \mathbf{y}^b + \Delta \mathbf{Y}^b \mathbf{w}. \quad (2.51)$$

Grâce à cette approximation, la fonction de coût qui sera minimisée par l'état analysé correspond à l'approximation de  $\tilde{\mathcal{J}}$  suivante :

$$\tilde{\mathcal{J}}'(\mathbf{w}) = (N-1)\mathbf{w}^T \mathbf{w} + [\mathbf{y} - \mathbf{y}^b - (\Delta \mathbf{Y}^b)\mathbf{w}]^T \mathbf{R}^{-1} [\mathbf{y} - \mathbf{y}^b - (\Delta \mathbf{Y}^b)\mathbf{w}]. \quad (2.52)$$

Cette fonction de coût correspond à celle du filtre de Kalman décrite par l'équation (2.20), avec comme ébauche  $\mathbf{w}^b = 0$ , comme matrice de covariance d'erreur d'ébauche  $\frac{\mathbf{I}_N}{N-1}$ , et  $\Delta \mathbf{Y}^b$  qui joue le rôle d'opérateur d'observation. D'après (HUNT, KOSTELICH et SZUNYOGH, 2007), le problème décrit par l'équation précédente a pour solutions :

$$\begin{aligned} \bar{\mathbf{w}}^a &= \tilde{\mathbf{P}}^a (\Delta \mathbf{Y}^b)^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{y}^b) \\ \tilde{\mathbf{P}}^a &= \left[ (N-1)\mathbf{I}_N + (\Delta \mathbf{Y}^b)^T \mathbf{R}^{-1} (\Delta \mathbf{Y}^b)^T \right]^{-1}, \end{aligned}$$

où  $\tilde{\mathbf{P}}^a$  correspond à la matrice de covariance d'erreur associée à  $\bar{\mathbf{w}}^a$ .

Pour obtenir la solution du problème d'analyse complet, il suffit d'appliquer le changement de variable et cela donne :

$$\mathbf{x}^a = \mathbf{x}^b + (\Delta \mathbf{X}^b)\bar{\mathbf{w}}^a \quad (2.53)$$

$$\mathbf{P}^a = (\Delta \mathbf{X}^b)\tilde{\mathbf{P}}^a(\Delta \mathbf{X}^b)^T. \quad (2.54)$$

La construction du nouvel ensemble analysé se fait en formant une matrice  $\Delta \mathbf{X}^a$  qui factorise  $\mathbf{P}^a$  selon la relation (2.39). Il suffit ensuite d'ajouter  $\mathbf{x}^a$  aux colonnes de cette matrice pour obtenir l'ensemble analysé.

Le processus d'application de la localisation pour former le LETKF est également détaillé dans (HUNT, KOSTELICH et SZUNYOGH, 2007, Sec. 2.2.3 et 2.3.4). Y est également décrite une manière efficace d'implémenter l'étape d'analyse.

### Autres variantes déterministes de l'EnKF

Les premières variantes déterministes sont connues sous l'acronyme des EnSRKF pour "Ensemble Square Root Kalman Filters", dont deux formulations sont données chez (BISHOP, ETHERTON et MAJUMDAR, 2001) et (WHITAKER et HAMILL, 2002), pour l'ETKF vu précédemment et l'EnSRF, respectivement.

Une couverture théorique détaillée de ce type de filtre est disponible dans (ASCH, BOCQUET et NODET, 2016, Chap. 6, Sec. 4) et plusieurs applications sont montrées en partie III du même livre. Pour approfondir davantage les algorithmes à racine carrée, le lecteur est renvoyé vers (EVENSEN, 2009a, Chap. 13), ainsi que les références mentionnées dans les éléments de littérature précédemment cités.



### 2.4.4 Assimilation de données asynchrone avec l'AEnKF

L'agencement temporel des étapes d'analyses tel que présenté en sous-section (2.1.5) correspond à de l'assimilation de données synchrone : l'étape d'analyse est calculée après l'arrivée de l'observation, et l'évolution temporelle par le modèle du temps  $t_k$  au temps  $t_{k+1}$  se fait une fois l'étape d'analyse calculée.

Cependant, lorsque le système d'assimilation vérifie certaines conditions, il est possible d'effectuer les étapes d'analyses de façon asynchrone, c'est-à-dire de ne pas calculer l'étape d'analyse au temps d'arrivée de chaque observation. Dans ce cas, le calcul peut se faire en fin de fenêtre d'assimilation par exemple. C'est ce que fait l'AEnKF (Asynchronous EnKF) décrit dans (SAKOV, EVENSEN et BERTINO, 2010).

Présentons quelques notations utiles pour formaliser ce nouvel algorithme. En nous inspirant des notations de l'article, posons :

$$\delta \mathbf{x} = \mathbf{x}^a - \mathbf{x}^f = \Delta \mathbf{X}^f \mathbf{G} \mathbf{s}, \text{ l'incrément d'analyse,} \quad (2.55)$$

$$\delta \mathbf{X} = \Delta \mathbf{X}^a - \Delta \mathbf{X}^f = \Delta \mathbf{X}^f \mathbf{T}, \text{ l'incrément de la matrice d'anomalies d'ensemble} \quad (2.56)$$

pour l'analyse,

$$\mathbf{s} = \mathbf{R}^{-1/2} \frac{(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}})}{N-1}, \text{ le vecteur des innovations normalisé.} \quad (2.57)$$

Voici comment peut se formaliser l'étape d'analyse. Exprimons maintenant la matrice  $\mathbf{S}$ , correspondant aux anomalies normalisées projetées dans l'espace des observations :

$$\mathbf{S} = \frac{\mathbf{R}^{-1/2} \mathbf{H} (\Delta \mathbf{X}^f)}{N-1}. \quad (2.58)$$

Elle nous permet d'explicitier les matrices  $\mathbf{G}$  et  $\mathbf{T}$  vues précédemment. Pour l'EnKF tel que décrit dans la sous-section précédente, celles-ci ont pour expression :

$$\mathbf{G} = (\mathbf{I}_N + \mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T = \mathbf{S}^T (\mathbf{I}_p + \mathbf{S} \mathbf{S}^T)^{-1} \quad (2.59)$$

$$\mathbf{T} = \mathbf{G} (\mathbf{D} - \mathbf{S}). \quad (2.60)$$

Ici, la matrice  $\mathbf{D} = \frac{\mathbf{R}^{-1/2} \tilde{\mathbf{D}}}{N-1}$  correspond à une matrice stockant des perturbations générées selon une distribution gaussienne telle que la moyenne des colonnes de  $\tilde{\mathbf{D}}$  soit nulle, et que  $\frac{\tilde{\mathbf{D}} \tilde{\mathbf{D}}^T}{N-1} \rightarrow \mathbf{R}$ ,  $N \rightarrow +\infty$ .

Les éléments de mise à jour de l'étape d'analyse correspondent ainsi au vecteur  $\mathbf{G} \mathbf{s}$  (eq. 2.55) et à la matrice  $\mathbf{T}$  (eq. 2.56). Nous voyons grâce aux précédentes équations qu'il suffit de connaître le vecteur d'observation  $\mathbf{y}$ , la matrice d'erreur d'observation  $\mathbf{R}$  et l'ensemble projeté dans l'espace des observations  $\mathbf{H} \mathbf{E}^f$  pour les calculer.

Sur une fenêtre d'assimilation  $[t_0, t_1]$ , nous pouvons ainsi exprimer les différentes grandeurs lors d'une assimilation à  $t_0$  :

$$\delta \mathbf{x}_0 = \Delta \mathbf{X}_0 \mathbf{G}_0 \mathbf{s}_0, \quad (2.61)$$

$$\delta \mathbf{X}_0 = \Delta \mathbf{X}_0 \mathbf{T}_0. \quad (2.62)$$

Nous remarquons qu'il s'agit à chaque fois d'une combinaison ou d'une transformation linéaire de la matrice des anomalies de l'ensemble de prévision.

Pour un modèle parfait suffisamment lisse, notons  $\mathbf{M}_{t_0 \rightarrow t_1}$  l'opérateur tangent linéaire du modèle d'évolution, pour aller du temps  $t_0$  au temps  $t_1$ . Pour une norme  $\|\cdot\|$  quelconque, la propagation dans le temps de l'incrément  $\delta \mathbf{x}_0$  vaudra alors :

$$\delta \mathbf{x}_1 = \mathbf{M}_{t_0 \rightarrow t_1} \delta \mathbf{x}_0 + O(\|\delta \mathbf{x}_0\|^2). \quad (2.63)$$

Ainsi, au temps  $t_1$  et dans l'approximation linéaire, les grandeurs suivant sont asymptotiquement équivalentes :

$$\delta \mathbf{x}_1 \sim \mathbf{M}_{t_0 \rightarrow t_1} \delta \mathbf{x}_0 = \mathbf{M}_{t_0 \rightarrow t_1} \Delta \mathbf{X}_0 \mathbf{G}_0 \mathbf{s}_0 \sim \Delta \mathbf{X}_1 \mathbf{G}_0 \mathbf{s}_0, \quad (2.64)$$

$$\delta \mathbf{X}_1 \sim \mathbf{M}_{t_0 \rightarrow t_1} \delta \mathbf{X}_0 = \mathbf{M}_{t_0 \rightarrow t_1} \Delta \mathbf{X}_0 \mathbf{T}_0 \sim \Delta \mathbf{X}_1 \mathbf{T}_0. \quad (2.65)$$

Nous voyons alors avec les équations (2.64) et (2.65) que tant que les évolutions des corrections d'analyse et d'anomalies d'ensemble sont linéaires, les coordonnées de ces mêmes corrections  $\mathbf{G}_0 \mathbf{s}_0$  et  $\mathbf{T}_0$  dans la base engendrée par les colonnes de  $\Delta \mathbf{X}_0$  restent constantes le long de la trajectoire d'évolution du système. Dans ce cas, tant que l'évolution des anomalies d'ensemble est linéaire entre le temps d'arrivée de l'observation et le temps de son assimilation, il est possible d'utiliser les éléments calculés au temps  $t_0$  pour effectuer les mises à jour au temps  $t_1$ .

C'est sur cette idée que s'articule l'AEnKF, il représente une modification générique de l'EnKF traditionnel décrit en sous-section (2.4.2). L'assimilation asynchrone de plusieurs observations est décrite plus tard en sous-section (5.2.2). Un avantage de l'AEnKF pour les systèmes de grande taille est qu'il base ses calculs sur l'ensemble projeté dans l'espace des observations  $\mathbf{H}\mathbf{E}^f \in \mathcal{M}_{p,N}(\mathbb{R})$  plutôt que  $\mathbf{E}^f \in \mathcal{M}_{n,N}(\mathbb{R})$ , ce qui représente un gain de stockage mémoire du fait qu'en général  $n \gg p$ .

Il est important de remarquer que, dans le cas linéaire, les statistiques d'erreurs en fin de fenêtre au temps  $t_1$  sont les mêmes pour l'EnKF traditionnel que celles de l'AEnKF (SAKOV, EVENSEN et BERTINO, 2010, Section 2.2).

D'autres algorithmes asynchrones sont également décrits dans (HUNT et al., 2003) pour l'EnKF, et (HUNT, KOSTELICH et SZUNYOGH, 2007) pour l'ETKF et le LETKF.

### 2.4.5 Limitations de l'EnKF

La principale limitation théorique de l'EnKF vient du fait que la taille de l'ensemble est souvent très limitée en comparaison de celle des modèles physiques sur lesquels il est appliqué. Par exemple, en prévision météorologique, la taille du vecteur d'état est de  $10^7$  alors qu'un ensemble d'environ 100 membres est utilisé, autrement les coûts d'intégration temporelle pour faire évoluer le modèle seraient inacceptables.

Cela conduit à des déficiences de rang des matrices approximées par l'ensemble. Nous allons maintenant en expliquer les conséquences.

#### Déficiences du rang de la matrice $\mathbf{P}^a$ et covariances fallacieuses

L'utilisation d'un ensemble pour approximer les matrices d'erreurs implique lui aussi son lot de défauts. Pour un ensemble  $\mathbf{E} \in \mathcal{M}_{n,N}(\mathbb{R})$ , lorsque  $n \geq N$  comme c'est souvent le

cas en pratique, des erreurs d'échantillonnage vont apparaître lors du calcul de la matrice  $\mathbf{P}$  définie selon les équations (2.42) ou (2.46).

Ces erreurs conduisent à l'existence de covariances fallacieuses entre des variables initialement non-corrélées, du fait de leur nature physique par exemple. Ainsi lors de l'assimilation, ces variables supposées non-corrélées avec l'observation vont tout de même être mises à jour, ce qui peut conduire dans le cas d'un système physique à un non-respect des lois le régissant.

Ces mises à jour fallacieuses peuvent s'estomper avec le temps au fur et à mesure que beaucoup de données sont assimilées, et l'impact sur l'état du système (la moyenne de l'ensemble) peut devenir négligeable. Mais cela conduit aussi à une réduction non voulue de la variance de l'ensemble, pouvant à terme mener à une forte sous-estimation de celle-ci dont la conséquence, si non traitée, est la divergence du filtre. De plus amples explications ainsi que plusieurs exemples se trouvent dans (EVENSEN, 2009a, Chap. 15), avec deux méthodes pour palier le problème des covariances fallacieuses : l'inflation et la localisation.

L'inflation cherche à réduire l'impact des covariances fallacieuses en gonflant artificiellement la variance de l'ensemble au moment de l'analyse, ou bien en multipliant  $\mathbf{P}^a$  par un scalaire  $\alpha > 1$ . Ceci a pour effet d'éviter la divergence du filtre mais ne corrige pas les mises à jour non physiques.

La localisation est une idée inspirée de la physique selon laquelle deux entités fortement éloignées vont être faiblement corrélées. Pour de grands systèmes, ceux-ci sont discrétisés sur une grille de taille finie. Cette localisation peut se faire de deux manières :

- la *localisation de covariance* qui revient à faire un produit de Schur avec une autre matrice dont les coefficients permettent de lisser les covariances fallacieuses dans  $\mathbf{P}$  et d'obtenir une matrice de rang plein,
- la *localisation de domaine*. Elle consiste à mettre à jour chaque point  $x_{i,j}$  de la grille avec le sous-ensemble des observations contenues dans  $\mathbf{y}$  qui sont proches de  $x_{i,j}$ , relativement à un critère de distance généralement dicté par la physique, réduisant ainsi l'effet des observations trop éloignées.

L'objectif de la localisation est de faire en sorte que  $\mathbf{P}^a$  soit de rang plein.

### Collapse de l'ensemble : une des causes de divergence de l'EnKF

Une autre conséquence d'un ensemble de rang déficient est le *collapse*, qui conduit à la divergence du filtre. Ce phénomène est fréquent dans le domaine des filtres à particules (MORZFELD, HODYSS et SNYDER, 2017 et références associées), qui sont des méthodes de Monte-Carlo séquentielles. De même que pour les méthodes précédentes, le filtre à particules va estimer une distribution dans un cadre bayésien sans se placer dans un cadre gaussien comme c'est le cas pour l'EnKF.

En possédant des observations  $\mathbf{y}_k$  et en cherchant à estimer la distribution de l'état  $\mathbf{x}_k$ , l'objectif du filtre à particules est d'estimer la distribution postérieure des variables d'états conditionnellement aux observations, que nous noterons  $p(\mathbf{x}_k | \mathbf{y}_0, \dots, \mathbf{y}_k)$ . En ayant un modèle d'évolution pour l'état tel que  $\mathbf{x}_{k+1} = f(\mathbf{x}_k) + \nu_k$  ou  $\nu_k$  est un bruit connu, la dimension de filtrage est approchée par

$$\int f(\mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_0, \dots, \mathbf{y}_k) d\mathbf{x}_k \simeq \frac{1}{P} \sum_{p=1}^P f(\mathbf{x}_k^{(p)}), \quad (2.66)$$

avec  $\mathbf{x}_k^{(p)}$  la  $p^{\text{ème}}$  particule du filtre à l'instant  $t_k$ .

Lors du déroulement de l'algorithme, chaque particule est pondérée par un poids  $\omega_k$  qui évolue au cours du temps avec comme condition que  $\sum_{k=1}^P \omega_k = 1$ , et qui vont définir la façon dont sera ré-échantillonné l'essaim de particules à chaque prise en compte d'une nouvelle observation. Or, au bout d'un certain long moment sur des systèmes de haute dimension, un nombre très restreint  $m \ll P$  de particules s'accaparent l'essentiel des poids, forçant l'échantillon suivant à une variabilité pathologiquement moindre, menant à la divergence du filtre. C'est le *collapse* ! Il faut alors recréer un nouvel échantillon de particules à partir des dernières statistiques valides pour relancer le filtre.

Le même phénomène se produit sur l'EnKF. L'ensemble  $\mathbf{E}$  doit recouvrir au mieux l'espace des états donc conserver un rang maximal défini par  $\max(n, N)$ . En règle générale,  $N \leq n, p$ ,  $\mathbf{E}$  est donc de rang déficient lors des applications classiques.

Une explication possible de l'apparition du collapse, décrite chez (KEPERT, 2004), repose sur la construction de la matrice de covariance empirique d'erreur d'observation  $\mathbf{R}_e$ . Celle-ci est symétrique positive, donc diagonalisable dans une base orthonormale : il existe  $\mathbf{S} \in \mathcal{M}_N(\mathbb{R})$  et  $\mathbf{D} \in \mathcal{M}_p(\mathbb{R})$  diagonale telles que  $\mathbf{D} = \mathbf{S}\mathbf{R}_e\mathbf{S}^T$ . Dans cette nouvelle base, les observations perturbées s'écrivent  $\tilde{\mathbf{Y}} = \mathbf{S}\mathbf{Y}$ , et l'opérateur d'observation associé  $\tilde{\mathbf{H}} = \mathbf{S}\mathbf{H}$ , avec comme matrice de covariance d'erreur empirique  $\mathbf{D}$ . Du fait que  $\text{rang}(\mathbf{R}_e) \leq N - 1$ ,  $\mathbf{D}$  possède au moins  $p - N + 1$  valeurs diagonales nulles, ce qui revient à dire que  $p - N + 1$  des observations transformées sont parfaites. Or, (LORENC, 2003, Annexe A) démontre que, tant que  $\text{rang}(\mathbf{E}) > 1$ , l'assimilation d'une observation parfaite diminue de 1 le rang de  $\mathbf{E}$ . De ce fait, le rang de l'ensemble analysé sera au plus de  $N - (p - N + 1) = 2N - p + 1$ . Lorsque cette grandeur devient  $\leq 1$ , c'est le *collapse* de l'EnKF.

Le collapse peut être mis en évidence en traçant comparativement l'évolution de la RMSE d'analyse avec l'évolution de l'écart-type de l'ensemble analysé, qui doivent avoir des dynamiques similaires.

Nous pouvons aussi regarder l'évolution des RMSEs d'analyse et de prévision, lorsque celles-ci se rapprochent trop et qu'il n'y a plus de réduction de l'erreur entre la prévision et l'analyse au même temps, cela signifie que les observations ne sont plus prises en compte et qu'il peut y avoir collapse.

### Autres limitations

Plus de détails sont disponibles dans (EVENSEN, 2009a, Chap. 14), où sont présentes plusieurs discussions d'algèbre linéaire sur la pseudo-inversion des matrices de rang déficient ainsi que l'interprétation des équations modifiées de l'analyse qui en découlent. En outre, des discussions sur l'implémentation de l'analyse et le coût des opérations à effectuer y sont aussi présentes avec des exemples numériques.

Une discussion plus ancienne et plus synthétique sur la même thématique est aussi disponible dans (EVENSEN, 2003, Sec. 3-5).

Plusieurs améliorations de l'algorithme de l'EnKF ont vu le jour pour contourner ces limitations, elles sont présentées en détail sous la forme d'un état de l'art en annexe (A.3). Un autre facteur limitant que ces méthodes ne prennent pas en compte directement peut provenir de l'intégration temporelle de l'ensemble. Pour plusieurs systèmes physiques, certaines variables ont des constantes d'évolution temporelle qui diffèrent de plusieurs ordres

de grandeur. Cela constitue ce que l'on appelle un système *raide*.

Ce phénomène de raideur se rencontre chez EDF pour la simulation des cœurs de réacteur, en particulier pour le système modélisant le problème de l'empoisonnement neutronique, expliqué en annexe (sous-section A.7.3), système sur lequel nous effectuerons nos essais numériques au chapitre (7). Nous allons discuter plus en détails du phénomène de raideur dans la section suivante.

## 2.5 Systèmes dynamiques raides

Les équations différentielles raides sont généralement définies comme des équations sur lesquelles les méthodes d'intégration explicites fonctionnent mal. C'est-à-dire qu'elles vont soit mettre un temps très long pour intégrer l'équation en comparaison des méthodes implicites, soit qu'elles vont donner une solution imprécise, voire les deux à la fois (HAIRER et WANNER, 1996; HAIRER, NØRSETT et WANNER, 2000).

Une façon simple de caractériser la raideur d'un système différentiel est de regarder ses échelles caractéristiques de temps. Lorsque l'écart entre celles-ci est grand, le système est dit *raide*.

Exemple de (HALL et WATT, 1976) (initialement publié par Robertson en 1966) :

$$\begin{array}{ll} \text{A:} & y_1' = -0.04y_1 + 10^4 y_2 y_3, & y_1(0) = 1 \\ \text{B:} & y_2' = 0.04y_1 - 10^4 y_2 y_3 + 3 \times 10^7 y_2^2, & y_2(0) = 0 \\ \text{C:} & y_3' = -3 \times 10^7 y_2^2, & y_3(0) = 0 \end{array}$$

Cet exemple décrit 3 réactions chimiques, une lente (constante de temps 0.04), une rapide (constante de temps  $10^4$ ) et une très rapide (constante de temps  $3 \times 10^7$ ).

Cela ne permet cependant de fournir qu'une caractérisation "expérimentale" de la raideur d'un système.

Ils se retrouvent souvent parmi les systèmes industriels et sont donc une difficulté à laquelle il est possible de faire face dans le cadre de l'assimilation de données non-statique. En effet, cela va induire des contraintes sur l'intégration dans le temps du système, qui peuvent se répercuter sur la précision ou le temps de calcul, voire compromettre la stabilité numérique du système de simulation. L'effet de la raideur n'est donc pas à négliger, mais il semble heureusement bien se maîtriser essentiellement via une bonne gestion de l'intégration numérique. (HAIRER et WANNER, 1996) est un bon ouvrage pour approfondir sur le phénomène de raideur, ainsi que les références qu'il mentionne.

# 3 État de l’art sur les diagnostics de qualité des observations pour la prévision numérique

## 3.1 Prévisibilité des systèmes dynamiques

### 3.1.1 Introduction à la notion

D’après (KLEEMAN, 2011), la prévisibilité des systèmes dynamiques a connu ses débuts pendant l’ère moderne avec l’étude du système solaire. La prévision d’événements tels que les éclipses solaires se fait avec une extrême précision du fait de la stabilité du système solaire sur des échelles de temps de l’ordre du siècle. En revanche, lorsque nous passons à des échelles de temps de l’ordre du million d’années, du fait du comportement chaotique intrinsèque au problème à N-corps qui sert de modélisation pour les équations du mouvement des planètes autour du Soleil, des événements tels que des collisions planétaires ou des éjections de planètes hors du système solaire sont possibles : le comportement chaotique du système qui prend le dessus, et la prévision en souffre.

D’autres études ont alors suivi, dont les plus connues sont peut-être la prévision des séismes, la modélisation océanographique et la prévision météorologique, les deux dernières consistant à prévoir les mouvements d’un fluide turbulent. Cependant, comme vu précédemment, la grande complexité de ces systèmes mènent à de nombreuses approximations pour permettre leur simulation, ce qui va de nouveau en faire souffrir la prévision.

Pour mieux prévoir le comportement de tels systèmes, de nombreux travaux ont été réalisés par Edward Lorenz durant le XX<sup>e</sup> siècle. Dans un article fondateur (LORENZ, 1963), il donne les équations d’un système simplifié de l’atmosphère terrestre lui permettant de montrer le caractère chaotique des changements de régimes saisonniers de ce dernier. Ce modèle est à l’heure actuelle encore très utilisé dans la recherche pour vérifier la qualité des nouvelles avancées techniques en terme de résolution/prévision. Dans un autre article (LORENZ, 1969), il discute de la prévisibilité des systèmes météorologiques et décrit trois catégories de systèmes dynamiques dont nous discuterons plus tard. Il essaiera aussi dans (LORENZ, 1965 ; LORENZ, 1982) différentes modélisations plus complexe de l’atmosphère, décrira leur résolution numérique, et discutera de leur capacité de prévisibilité d’un point de vue météorologique et climatologique.

De plus, dans le cas de la météorologie, arriver à obtenir suffisamment d’observations de la surface de la Terre et de l’atmosphère pour connaître au mieux l’état physique courant du système est une tâche ardue. Il en résulte donc souvent une imprécision notable sur la connaissance de la condition initiale servant de base à la prévision.

Parmi plusieurs autres, nous voyons donc que deux déficiences majeures influent sur les capacités de prévision d’un système de simulation. Elles sont pour la première l’*erreur*

modèle dont nous avons déjà discuté en sous-section (2.1.3), et pour la seconde l'erreur de condition initiale, conditions initiales qui sont particulièrement déterminantes pour toutes les équations aux dérivées partielles instationnaires. Chacune va faire levier d'une façon différente sur la qualité des prévisions numériques réalisées. Nous rappelons que pour la prévision météorologique, le modèle est souvent considéré comme *parfait* donc sans erreur modèle. L'incertitude sur les conditions initiales va faire que, pour un système chaotique, deux prévisions issues de deux conditions initiales qui diffèrent légèrement vont, au bout d'un certain laps de temps, différer totalement et ne plus représenter le même climat dans le cas de simulations climatologiques par exemple.

Nous voyons donc qu'il y a de nombreux facteurs qui vont influencer le comportement réel du système dynamique que nous cherchons à modéliser et observer. C'est avec eux que nous allons devoir jouer pour obtenir des prévisions de qualité. Nous appelons maintenant *prévisibilité* la capacité d'un système de simulation incluant de l'assimilation de données à fournir des prévisions numériques vérifiant un critère de qualité que nous allons formaliser par la suite.

### 3.1.2 Formalisation mathématique

Formellement, un système dynamique se définit selon l'équation :

$$\frac{d\mathbf{x}}{dt} = F(\mathbf{x}, t), \quad \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{x}|_{t=0} = \mathbf{x}_0 \quad (3.1)$$

où  $\mathbf{x} = \mathbf{x}(t)$  est l'état du système au temps  $t$ ,  $\mathbf{x}_0$  la condition initiale du système. Faire de la prévision sur ce système dynamique conduit à intégrer numériquement l'équation différentielle qui le régit jusqu'au temps de prévision  $t_f$ .

Nous savons que la modélisation du système physique décrit par l'équation (3.1) est incomplète, qu'il y a une part existante d'erreur modèle que l'on ne peut pas quantifier. De même, il est fort probable qu'il y ait des lacunes sur la connaissance de  $\mathbf{x}_0$ .

De ce fait, toute solution analytique comme numérique de l'équation (3.1) sera différente de l'état-vrai  $\mathbf{x}^t(t)$  du système.

Un système *prédictif* est donc un système qui, sur un intervalle de temps donné (éventuellement infini), maintient l'écart entre la prévision et l'état vrai sous un seuil acceptable, défini a priori empiriquement.

Nous pouvons ainsi donner une définition de l'*horizon de prévisibilité* d'un système :

**Définition 1.** *Horizon de prévisibilité*

L'horizon de prévisibilité d'un système correspond au temps  $T$  à partir duquel l'écart entre la prévision et l'état vrai du système est supérieur à un seuil  $\delta$  défini au préalable :

$$T \text{ horizon de prévisibilité} \iff \sup_{t \in [0, T[} \|\mathbf{x}(t) - \mathbf{x}^t(t)\| \leq \delta$$

**Remarque :** il est important de noter que la norme utilisée ici est quelconque. Nous pouvons donc en choisir une avec des propriétés qui nous intéressent. Par exemple une norme dont le carré est différentiable lorsque l'on fait du calcul des variations, une norme infinie lorsqu'on ne s'intéresse qu'à la plus grande valeur d'un vecteur/matrice etc. Cette remarque sera aussi valable pour toute la suite, sauf explicitement précisé.

De cette définition découle la suivante :

**Définition 2.** *Système prédictif (au sens météorologique)*

Un système est dit prédictif sur  $[t_0, t_f] = I$  lorsqu'il existe un seuil  $\delta > 0$  défini au préalable et indépendant de  $I$ , tel que :

$$\forall t \in I, \|\mathbf{x}(t) - \mathbf{x}^t(t)\| \leq \delta$$

Ceci est valable lorsque  $\mathbf{x}$  est quelconque, aléatoire ou déterministe. Lorsque c'est explicitement une variable aléatoire, l'équation (3.1) décrit l'évolution d'un processus stochastique. Dans ce cas, la définition (déf. 2) définit un type de prévisibilité particulier, que (SOLDATENKO et YUSUPOV, 2017) appelle "prévisibilité de première espèce". Lorsque le système dynamique étudié admet un attracteur, donc que ses trajectoires dans l'espace des phases sont contenues dans un élément de volume fini et que chaque trajectoire ne peut s'éloigner indéfiniment l'une de l'autre, selon (SOLDATENKO et YUSUPOV, 2017), en considérant chaque nouvelle trajectoire issue d'une condition initiale perturbée comme une nouvelle réalisation du processus stochastique  $\mathbf{x}$ , alors les erreurs entre  $\mathbf{x}$  et  $\mathbf{x}^t$  sont bornées. En fixant arbitrairement  $\delta$ , on cherche donc à savoir à court-terme jusqu'à quand notre prévision est acceptable. Nous entendons ici par "court-terme" comme "localement en temps", ou bien "court-terme relativement aux échelles de temps du système étudié". Cette définition s'applique par exemple plus au système météorologique.

La deuxième définition de la prévisibilité provenant de (SOLDATENKO et YUSUPOV, 2017) traite de la prévisibilité au long terme relativement aux échelles de temps du système. Par exemple en climatologie, elle va regarder la température moyenne dans une zone géographique sur 30 ans, plutôt que son évolution sur la globalité de la zone pendant quelques jours comme en météorologie. Elle regarde les moments de la distribution de  $\mathbf{x}$  par rapport à ceux de  $\mathbf{x}^t$  plutôt que ses réalisations. C'est la "prévisibilité de seconde espèce".

**Définition 3.** *Système prédictif (au sens climatologique)*

Un système est dit prédictif sur  $[t_0, t_f] = I$  lorsque tous ses moments d'erreur sont bornés par des constantes positives  $\delta, \delta', \dots$  définies au préalable, indépendantes de  $I$  :

$$\begin{aligned} \forall t \in I, \mathbb{E} \left[ \|\mathbf{x}(t) - \mathbf{x}^t(t)\| \right] &\leq \delta \\ \left| \text{Var}[\mathbf{x}(t)] - \text{Var}[\mathbf{x}^t(t)] \right| &\leq \delta' \\ &\vdots \end{aligned}$$

Cette propriété de prévisibilité d'un système se vérifie sur une série statistique qui regroupe la connaissance de nombreuses analyses passées, par exemple sur plusieurs décennies voire même à l'échelle du siècle pour certaines applications. Lors des mises à jour du système de simulation, des réanalyses sont effectuées pour voir si les modifications apportées ne dégradent pas la prévisibilité du système en regardant si les statistiques d'erreurs conservent leur continuité.

La norme utilisée pour définir la prévisibilité peut se concentrer uniquement sur une propriété ou grandeur physique particulière du système, comme par exemple la salinité en océanographie, visible chez (SAKOV et al., 2012) ou encore (XIE et al., 2019), ou bien la norme d'énergie humide totale (moisture total energy norm) (KIM et KIM, 2019, Eq. (3))



en prévision atmosphérique.

Comme décrit dans (CARRASSI et al., 2018), ces études de réanalyse constituent un diagnostic primordial de prévisibilité dans les domaines de la météorologie, de l'océanographie ou encore de la climatologie.

## 3.2 Inventaire des indicateurs de qualité de prévision existants

Les méthodes présentées précédemment ainsi que leurs améliorations/variantes ont toutes pour objectif d'améliorer la qualité du processus d'assimilation de données et des prévisions ou réanalyses qu'il fournit, en jouant principalement sur la partie "état" ou "erreur modèle" du système.

Une autre voie pour l'amélioration en terme de précision ou de coût numérique des systèmes d'assimilation repose sur une meilleure gestion des observations, et sur la connaissance qu'il est possible de leur extraire. Pour cela, plusieurs indicateurs de qualité de prévision ont été formulés pour estimer la qualité de l'analyse une fois l'assimilation faite, qui cherchent à déterminer l'impact des observations sur l'analyse. Nous allons maintenant en étudier les principaux.

### 3.2.1 Matrice des sensibilités ou "Hat matrix"

#### Origine et formulation statistique

Soit la minimisation des moindres carrés suivante :

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{Ax} - \mathbf{y}\|_{\mathbf{R}^{-1}}^2 \quad (3.2)$$

$\mathbf{y} \in \mathbb{R}^p$  représente les observations,  $\mathbf{A} \in \mathcal{M}_{p,n}(\mathbb{R})$  l'opérateur modèle que nous voulons recalculer par les observations, et  $\mathbf{R} \in \mathcal{M}_p(\mathbb{R})$  la matrice d'incertitude sur les observations. Cette modélisation revient à résoudre un problème statistique de régression linéaire :

$$\mathbf{y} = \mathbf{Ax} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{R}). \quad (3.3)$$

Les hypothèses concernant ce modèle sont les suivantes :

1.  $\epsilon$  est indépendante des erreurs sur  $\mathbf{A}$ ,
2.  $\mathbf{R} = \sigma^2 I_p$  avec  $\sigma \neq 0$ .

Ce problème de minimisation est très sensible aux observations atypiques, hors "normes", comme les valeurs aberrantes. Un diagnostic doit donc être établi pour permettre la détection d'observations dont une faible variation va fortement influencer les caractéristiques du modèle  $\mathbf{A}$ .

Ce problème étant un moindres carrés linéaire, nous avons grâce à nos hypothèses une expression pour nos paramètres estimés :

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}. \quad (3.4)$$

En réinjectant (3.4) dans l'équation de notre modèle (3.3), nous obtenons comme prédiction :

$$\hat{\mathbf{y}} = \mathbf{A}\hat{\mathbf{x}} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}. \quad (3.5)$$

Nous avons donc l'expression de notre matrice de projection, surnommée "hat matrix" parce qu'elle "met un chapeau" sur  $\mathbf{y}$  :

$$\hat{\mathbf{H}} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = (\hat{h}_{ij})_{1 \leq i, j \leq p}. \quad (3.6)$$

Les éléments diagonaux  $\hat{h}_{ii}$  mesurent l'effet ou l'importance de la mesure  $y_i$  avec  $\mathbf{y} = (y_i)_{1 \leq i \leq p}$  dans l'estimation par le modèle de  $\hat{\mathbf{y}}$ . Elle est aussi appelée "matrice d'influence" et notée  $\mathbf{S} = \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{y}}$ .

Son utilisation actuelle dans le cadre de l'assimilation de données se retrouve à l'ECMWF avec les travaux de (CARDINALI, 2004; CARDINALI, 2013).

### Formulation pour l'assimilation de données variationnelle

Dans cette sous-section, nous noterons par commodité  $\mathbf{x}_b$  plutôt que  $\mathbf{x}^b$  et  $\mathbf{x}_a$  plutôt que  $\mathbf{x}^a$ , respectivement pour l'ébauche et l'état analysé.

Dans le cadre variationnel, le problème de minimisation peut être vu comme une régression à la fois pour les observations et l'ébauche, dont les équations sont :

$$\mathbf{y} = \mathbf{H}\Theta + \epsilon_{obs}, \quad (3.7)$$

$$\mathbf{x}_b = \Theta + \epsilon_b. \quad (3.8)$$

$\epsilon_{obs} \in \mathbb{R}^p$  et  $\epsilon_b \in \mathbb{R}^n$  sont supposés de moyenne nulle et de matrices de covariance respectives  $\mathbf{R}$  et  $\mathbf{B}$ .

Cela se résume en un problème de régression augmenté :

$$\mathbf{z} = \begin{bmatrix} \mathbf{y} \\ \mathbf{x}_b \end{bmatrix} = \mathbf{A}\Theta + \epsilon = \begin{bmatrix} \mathbf{H} \\ \mathbf{I}_n \end{bmatrix} \Theta + \begin{bmatrix} \epsilon_{obs} \\ \epsilon_b \end{bmatrix}. \quad (3.9)$$

$\epsilon \in \mathbb{R}^{p+n}$  est de moyenne nulle et a pour matrice de covariance  $\Omega = \begin{pmatrix} \mathbf{R} & 0 \\ 0 & \mathbf{B} \end{pmatrix}$ , qui pondère le système. C'est à partir de ce problème augmenté que s'établit l'expression de la Hat matrix du système. Nous obtenons comme paramètres estimés :

$$\hat{\Theta} = (\mathbf{A}^T\Omega^{-1}\mathbf{A})^{-1}\mathbf{A}^T\Omega^{-1}\mathbf{z}. \quad (3.10)$$

Ce qui nous donne comme solution estimée :

$$\hat{\mathbf{z}} = \mathbf{A}\hat{\Theta} = [\hat{\mathbf{y}}^T, \hat{\mathbf{x}}_b^T]^T = [\mathbf{H}^T\mathbf{x}_a^T, \mathbf{x}_a^T]^T = \mathbf{A}(\mathbf{A}^T\Omega^{-1}\mathbf{A})^{-1}\mathbf{A}^T\Omega^{-1}\mathbf{z}. \quad (3.11)$$

Sa matrice de covariance d'erreur vaut :

$$\mathbf{C} = (\mathbf{A}^T\Omega^{-1}\mathbf{A})^{-1}\mathbf{A}^T\Omega^{-1}. \quad (3.12)$$

La hat matrix s'exprime comme la matrice des sensibilités du système. Elle s'écrit :

$$\hat{\mathbf{H}} = \mathbf{S} = \frac{\partial \hat{\mathbf{z}}}{\partial \mathbf{z}} = \begin{pmatrix} \mathbf{R}^{-1} \mathbf{H} \mathbf{C} \mathbf{H}^T & \mathbf{R}^{-1} \mathbf{H} \mathbf{C} \\ \mathbf{B}^{-1} \mathbf{C} \mathbf{H}^T & \mathbf{B}^{-1} \mathbf{C} \end{pmatrix}. \quad (3.13)$$

La sensibilité qui nous intéresse est celle par rapport aux observations. Elle correspond à la partie de  $\mathbf{S}$  portant spécifiquement sur cette sensibilité là. Nous prendrons dans ce cas :

$$\hat{\mathbf{H}} = \frac{\partial \mathbf{H} \mathbf{x}_a}{\partial \mathbf{y}} = \mathbf{R}^{-1} \mathbf{H} \mathbf{C} \mathbf{H}^T. \quad (3.14)$$

Dans le cas d'un moindres carrés linéaire comme pour l'équation (3.2), les matrices décrites par les équations (3.13) et (3.14) sont confondues car le modèle dépend uniquement des observations. (CARDINALI, 2004, eq. (2.3)-(2.4))

### 3.2.2 DFS : Degrees of Freedom for Signal

Dans le cas de l'estimation optimale, l'analyse est donnée par l'équation (2.13). Les DFS servent à mesurer la sensibilité propre de l'analyse aux différents types d'observation. Ils correspondent à la trace des dérivées partielles de l'image de l'analyse dans l'espace des observations, relativement aux observations (LUPU, GAUTHIER et LAROCHE, 2011, Sec. 2, Eq. 3) :

$$\text{DFS} = \text{Tr} \left\{ \frac{\partial \mathbf{H} \mathbf{x}_a}{\partial \mathbf{y}} \right\} = \text{Tr}(\hat{\mathbf{H}}). \quad (3.15)$$

Nous avons un lien direct entre les DFS et la Hat matrix.

Dans le cas linéaire, les équations (LUPU, GAUTHIER et LAROCHE, 2011, Sec. 2, Eq. 3) et (3.15) impliquent que :

$$\text{DFS} = \text{Tr}(\mathbf{K}^T \mathbf{H}^T) = \text{Tr}(\mathbf{H} \mathbf{K}). \quad (3.16)$$

En pratique, la taille des matrices  $\mathbf{H}$  et  $\mathbf{K}$  empêchent le calcul direct des DFS. Il faut pour cela passer par des méthodes d'approximation.

Trois solutions existent pour contourner cette difficulté : l'estimation de la trace grâce aux vecteurs singuliers principaux de la Hessienne du système d'assimilation (CARDINALI, 2004), des techniques de randomisation se basant sur les diagnostics de cohérence de (DESROZIERS et al., 2005), et la méthode de (CHAPNIK et al., 2006).

L'approche de (DESROZIERS et al., 2005) s'applique à l'analyse du contenu informatif aussi bien de toutes les observations que d'un sous-ensemble de celles-ci dans le cadre des méthodes variationnelles, en ne se basant que sur les innovations par rapport à l'analyse et à l'ébauche. Le point fort de cette technique repose sur le fait qu'elle peut s'adapter à des statistiques d'erreurs mal spécifiées, et qu'en ajoutant une hypothèse de non-corrélation des erreurs d'observation, elle se généralise à tout système d'assimilation de données.

Le calcul de la matrice de sensibilités comme présenté dans (CARDINALI, 2013) constitue une quatrième manière de calculer les DFS d'un système. Cependant, cela revient encore une fois à calculer  $\text{Tr}(\mathbf{H} \mathbf{K})$  une fois l'assimilation faite.

Les DFS constituent donc un autre moyen d'évaluer a posteriori l'information apportée par les observations en suivant une voie légèrement différente des diagnostics FSOI

présentés précédemment. Suite aux travaux de (CARDINALI, 2009; CARDINALI, 2013), ils sont opérationnels à l'ECMWF pour servir à un meilleur paramétrage de leur réseau d'observation.

Il est intéressant de noter qu'à la différence des diagnostics FSOI décrits précédemment, les DFS semblent se généraliser plus facilement aux méthodes non-variationnelles, bien que plusieurs travaux dans ce domaine de recherche existent, que nous allons maintenant regarder.

### 3.2.3 FSOI : Forecast Sensitivity Observation Impact

Les premières expérimentations sur l'impact des observations consistaient à supprimer une ou plusieurs observation  $\mathbf{y}_i$  du vecteur d'observation  $\mathbf{y}$  et à comparer les résultats obtenus après analyse avec et sans ces observations ("expérience de déni de données"). Le problème de cette méthode est qu'elle nécessite de calculer autant d'analyses différentes que l'on veut faire de diagnostics, peu pratique donc pour des systèmes opérationnels.

L'estimation de la sensibilité de la prévision aux corrections/informations apportées par les observations lors de l'analyse est une façon de contrôler la qualité de la prévision au fur et à mesure du processus d'assimilation, tout en permettant de mieux calibrer le système d'observation.

Les diagnostics qui cherchent à formuler et caractériser au mieux l'erreur de prévision sont donc particulièrement intéressants pour étudier la qualité des prévisions faites par un système d'assimilation de données et attester de sa prévisibilité : ce sont les diagnostics FSOI.

Nous rappelons que dans le cadre linéaire, l'analyse à  $t_k$  réduit systématiquement la variance de l'erreur commise à  $t_k$ . Cependant, cette certitude disparaît lorsque nous sommes dans un cadre non-linéaire. Les diagnostics FSOI permettent un suivi de l'évolution de l'erreur de prévision en fonction de l'assimilation d'observation, et permettent de caractériser les observations à fort impact correctif *a posteriori* de l'analyse.

Dans cette partie, nous noterons l'état-vrai  $\mathbf{x}_t$  et pas  $\mathbf{x}^t$  comme nous avons pu le voir plus en amont dans le manuscrit, pour éviter de surcharger la rédaction. Nous introduisons également deux autres notations : celle de la prévision obtenue à partir de l'ébauche  $\mathbf{x}^b$ , notée  $\mathbf{x}_b^f$ , et celle de la prévision obtenue après correction par l'analyse  $\mathbf{x}^a$ , notée  $\mathbf{x}_a^f$ .

#### Caractérisation de l'erreur de prévision

Une façon de caractériser l'impact des observations sur la prévision est de suivre l'évolution d'une estimation de l'erreur de prévision. C'est cette voie qui a été empruntée d'abord par (LANGLAND et BAKER, 2004) puis raffinée par (ERRICO, 2007; GELARO, ZHU et ERRICO, 2007). En reprenant les notations et le fil directeur de (TRÉMOLET, 2007), il s'agit de définir une mesure de l'erreur de prévision :

$$e(\mathbf{x}^f) = (\mathbf{x}^f - \mathbf{x}_t)^T \mathbf{C} (\mathbf{x}^f - \mathbf{x}_t), \quad (3.17)$$

avec  $\mathbf{x}^f$  la prévision,  $\mathbf{x}_t$  l'état-vrai et  $\mathbf{C}$  une matrice de poids symétrique définie-positive. En pratique, l'état-vrai et la prévision ne sont pas directement dans le même espace, le second étant une approximation discrétisée du premier. Pour arriver à faire un lien entre les deux, il suffit de légèrement raffiner la formulation (3.17) en définissant  $\mathbf{x}_t^f$ , le

projeté orthogonal de  $\mathbf{x}_t$  dans l'espace des prévisions, relativement au produit scalaire défini depuis  $\mathbf{C}$ . Dans ce cas, en écrivant la relation d'orthogonalité, et la formulation (3.17) devient :

$$e(\mathbf{x}^f) = (\mathbf{x}^f - \mathbf{x}_t^f)^T \mathbf{C}(\mathbf{x}^f - \mathbf{x}_t^f) + (\mathbf{x}_t^f - \mathbf{x}_t)^T \mathbf{C}(\mathbf{x}_t^f - \mathbf{x}_t) \quad (3.18)$$

qui est une fonction quadratique en  $\mathbf{x}^f$ .

Nous pouvons alors aisément en déduire une approximation au premier ordre des variations de l'erreur (3.18) pour une perturbation  $\delta\mathbf{x}$  :

$$\delta e(\delta\mathbf{x}) = e(\mathbf{x}^f + \delta\mathbf{x}) - e(\mathbf{x}^f) = 2(\delta\mathbf{x})^T \mathbf{C}(\mathbf{x}^f - \mathbf{x}_t^f) + \delta\mathbf{x}^T \mathbf{C}\delta\mathbf{x}. \quad (3.19)$$

Le calcul de cette formule se trouve en annexe (section A.4). Du fait que l'erreur soit quadratique, ce calcul donne une formule exacte de la variation.

De manière générale, les états qui nous intéressent sont la prévision obtenue depuis l'ébauche  $\mathbf{x}_t^f$ , et la prévision corrigée par l'analyse  $\mathbf{x}_t^a$ .

Dans le cadre de l'étude de l'impact des observations, la correction de la prévision résulte de l'assimilation des observations. La réduction de l'erreur de prévision s'exprime alors directement comme une fonction des innovations  $\delta\mathbf{y} = \mathbf{y} - \mathbf{H}\mathbf{x}_t^f$ . (TRÉMOLET, 2007) en donne une approximation au premier ordre :

$$\delta\mathbf{x} = \mathbf{M}\mathbf{K}\delta\mathbf{y}. \quad (3.20)$$

$\mathbf{M}$  correspond à la linéarisation de l'opérateur modèle au temps de l'analyse, et  $\mathbf{K}$  au gain de Kalman du système d'assimilation étudié. Avec cette approximation, nous avons une relation linéaire directe entre  $\delta\mathbf{y}$  et  $\delta\mathbf{x}$  qui permettent d'exprimer les corrections directement à partir des innovations.

Au lieu de travailler dans l'espace des états comme avec la formulation (3.17), (TODLING, 2013) propose de travailler dans l'espace des observations en transformant simplement la formulation (3.17) en :

$$e^y(\mathbf{x}^f) = [\mathcal{H}(\mathbf{x}^f) - \mathbf{y}]^T \mathbf{T}[\mathcal{H}(\mathbf{x}^f) - \mathbf{y}] = \mathbf{d}^T \mathbf{T} \mathbf{d}. \quad (3.21)$$

$\mathbf{T} \in \mathcal{M}_p(\mathbb{R})$  est une matrice de poids symétrique définie-positve, s'interprétant comme une matrice de covariance. Le point de vue "observations moins prévision" ("observation-minus-forecast") de (DESROZIERS et al., 2005) présente certains avantages tels que l'absence de corrélations non-voulues entre l'état de vérification et la prévision, du fait que l'état de vérification soit les observations qui sont indépendantes de la prévision, en comparaison au point de vue "analyse moins forecast". Cependant, les résultats diffèrent de ceux obtenus via l'espace des états, ils ne vont coïncider que lorsque l'analyse est optimale comme dans le cas linéaire, ce qui n'est pas souvent vérifié en pratique. Leur autre intérêt tient dans le fait que la formulation (3.21) fait directement référence à l'impact des observations sur la qualité du cycle d'assimilation de données lorsque l'état de vérification  $\mathbf{x}^f = \mathbf{x}^b$  ou  $\mathbf{x}^a$ .

Cependant, l'auteur mentionne lui-même en conclusion que cette approche n'est ni moins bonne ni meilleure qu'une autre pour estimer l'impact des observations.

### FSOI pour les méthodes variationnelles

L'article de (TRÉMOLET, 2008) propose une application des diagnostics FSOI sur un système d'assimilation de données variationnelle, plus particulièrement les algorithmes 3D- et 4D-Var incrémentaux, décrit dans (COURTIER, THÉPAUT et HOLLINGSWORTH, 1994). Le 4D-Var incrémental est utilisé depuis les années 90 dans la plupart des applications météorologiques.

Les algorithmes 3D- et 4D-Var incrémentaux ont pour objectif la minimisation d'une fonction de la forme :

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{y} - \mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - \mathcal{H}(\mathbf{x})), \quad (3.22)$$

quadratique dans le cas linéaire. (CARDINALI, 2018) montre que l'on peut aussi travailler avec une fonction de coût ne prenant en compte que les observations.

En pratique, cette fonction (3.22) est non-linéaire et difficile à minimiser. La méthode itérative de Gauss-Newton est utilisée car il s'agit d'un problème aux moindres carrés. Elle consiste à approcher itérativement le minimum de  $J$  en calculant plusieurs approximations quadratiques de  $J$  sans calculer sa Hessienne, trop coûteuse, jusqu'à l'obtention d'une solution satisfaisante. Cette méthode donne de bons résultats lorsque le problème est moyennement non-linéaire. Ce procédé constitue ce que l'on appelle la *boucle externe* ou *outer loop*.

Quant à elle, la *boucle interne* ou *inner loop* correspond aux itérations de la procédure de minimisation utilisée pour la recherche du minimum des approximations de  $J$ , au sein de chaque itération de la boucle externe.

La détermination des sensibilités de la prévision aux observations nécessite, pour chaque itération de la boucle externe, la construction des opérateurs adjoints  $\mathbf{M}^T$  et  $\mathbf{K}^T$  respectivement de l'opérateur d'évolution du modèle  $\mathcal{M}$  et du gain de Kalman équivalent  $\mathbf{K} = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1}$ . Pour obtenir la sensibilité complète de la prévision, pour chaque itération de la boucle externe, il reste à calculer l'adjoint de la boucle interne comme décrit dans (TRÉMOLET, 2008, Sec. 4).

Cependant, l'opérateur  $\mathbf{K}$  dépend en général de l'état en sortie de l'itération précédente  $j$  de la boucle externe. Ainsi, pour l'itération suivante  $j + 1$ , la formulation du problème adjoint pour le 4D-Var incrémental nécessite le calcul du tenseur  $\delta \mathbf{K}_j$  qui correspond à la dérivée de l'opérateur  $\mathbf{K}_j$  par rapport à chaque variable de l'état. Ce calcul est très onéreux, et en pratique seule une approximation de  $\delta \mathbf{K}_j$  est calculée par différences finies et pose la question de la précision de cette approximation.

Le 4D-Var incrémental bénéficie des itérations de la boucle externe du fait qu'elles permettent la prise en compte des informations retournées par l'adjoint. Cela est d'autant plus marqué que la dépendance des différents opérateurs à l'état du système est forte. De plus, les informations mises à disposition par le diagnostic FSOI en temps réel permettent une amélioration en temps réel de la façon dont sont traitées les différentes observations, notamment un gain de temps sur la détection d'erreurs à ce niveau (TRÉMOLET, 2007, discussion en section (5.3)). En pratique, il n'est pas possible de calculer directement l'adjoint au second-ordre car son coût est prohibitif.

En revanche, (LE DIMET, NAVON et DAESCU, 2002) montre qu'il est possible d'en extraire

les informations les plus importantes comme par exemple ses valeurs propres ou son conditionnement, à un coût acceptable sans construire la matrice Hessienne  $\in \mathcal{M}_n(\mathbb{R})$ . Pour cela, il suffit de former l'opérateur qui calcule le produit entre la Hessienne et un vecteur quelconque. (LE DIMET, NAVON et DAESCU, 2002, Section. 2) montre que cela revient à résoudre un système linéaire puis à effectuer différentes opérations moins coûteuses qu'un calcul direct de la Hessienne. Ces informations de second-ordre servent ensuite pour le suivi du problème de minimisation, renseignant notamment sur la vitesse de convergence pour un problème sans contraintes, et la détection de minimas locaux pour un problème sous contraintes. Deux méthodes d'approximations de la Hessienne sont aussi discutées : le calcul par éléments finis, et une approximation stochastique de rang  $p \leq n$  donnant de bons résultats en tant que préconditionnement du problème.

### Sensibilité aux perturbations de la matrice $\mathbf{R}$

(DAESCU, 2012) étend l'étude des sensibilités de la prévision aux observations à celle des sensibilités de la prévision aux variations de la matrice de covariance d'erreur d'observation  $\mathbf{R}$ , et donne aussi une expression de la sensibilité de la prévision aux variations de la matrice de covariance d'erreur d'ébauche  $\mathbf{B}$  dans son équation (11) sans approfondir son étude dessus.

En pratique, la matrice  $\mathbf{R}$  est considérée comme diagonale – bien que cela soit en train d'évoluer – avec  $\mathbf{R} = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$  (éventuellement  $\forall i, \sigma_i^2 = \sigma_0^2$ ). Soit  $\mathbf{y}_i \in \mathbb{R}^l$ ,  $l < p$ , un sous-ensemble d'observations du vecteur  $\mathbf{y}$ .

Notons  $\alpha : \begin{cases} \llbracket 1, p \rrbracket \rightarrow \llbracket 1, l \rrbracket \\ k \mapsto \alpha(k) \end{cases}$  la fonction injective qui construit  $\mathbf{y}_i$  à partir des éléments

de  $\mathbf{y}$ , telle que  $\mathbf{y}_i = (y_{\alpha(i)})_{1 \leq i \leq p}$ . Cette opération se met sous la forme d'une matrice de projection  $\mathbf{\Gamma} \in \mathcal{M}_{l,p}(\mathbb{R})$ , telle que  $\mathbf{y}_i = \mathbf{\Gamma}\mathbf{y}$ .

La matrice d'erreur d'observation associée  $\mathbf{R}_i \in \mathcal{M}_l(\mathbb{R})$  vaut  $\mathbf{R}_i = \mathbf{\Gamma}\mathbf{R}\mathbf{\Gamma}^T = \text{diag}(\sigma_{\alpha(i)}^2)$ .

En reprenant l'équation (10) de (DAESCU, 2012) avec nos notations, la variation de l'erreur de prévision (3.21) relativement au sous-ensemble d'observations  $\mathbf{y}_i$  du vecteur d'observation  $\mathbf{y}$  vaut :

$$\frac{\partial e(\mathbf{x}^f)}{\partial \mathbf{R}_i} = \nabla_{\mathbf{y}_i} e(\mathbf{x}^f) [\mathcal{H}_i(\mathbf{x}^f) - \mathbf{y}_i] \mathbf{R}_i^{-1}. \quad (3.23)$$

où  $\mathcal{H}_i : \mathbb{R}^n \rightarrow \mathbb{R}^l$  correspond à l'opérateur d'observation adapté pour le sous-ensemble  $\mathbf{y}_i$ . Cette formule est démontrée dans (DAESCU, 2008).

Avec  $\mathbf{y}_i \in \mathbb{R}^l$  et  $j \in \llbracket 1, l \rrbracket$ , la sensibilité de l'erreur de prévision à chaque variance d'erreur individuelle des observations  $\in \mathbf{y}_i$  vaut :

$$\frac{\partial e(\mathbf{x}^f)}{\partial \mathbf{R}_i(j, j)} = \frac{\partial e(\mathbf{x}^f)}{\partial \mathbf{y}_i(j)} \frac{[\mathcal{H}_i(\mathbf{x}^f) - \mathbf{y}_i]_j}{\mathbf{R}_i(j, j)} \quad (3.24)$$

où  $\mathbf{R}_i(j, j)$  représente la  $j^{\text{ème}}$  valeur diagonale de  $\mathbf{R}_i$ . Cependant, l'erreur  $e(\mathbf{x}^f)$  est implicitement liée à la spécification de la matrice  $\mathbf{R}$  du système d'assimilation :  $e(\mathbf{x}^f) = e(\mathbf{x}^f(\mathbf{R}))$ . Ainsi, une approximation de premier ordre de l'erreur de prévision en fonction d'une perturbation  $\delta \mathbf{R}$  est donnée par :

$$\delta e = e(\mathbf{R} + \delta \mathbf{R}) - e(\mathbf{R}) \simeq \text{Tr} \left( \frac{\partial e}{\partial \mathbf{R}} \delta \mathbf{R}^T \right). \quad (3.25)$$

En redécoupant  $\mathbf{y}$  en sous-ensembles  $\mathbf{y}_i$ , et en adaptant l'équation (3.25) pour ces sous-ensembles, cela permet de déterminer la sensibilité de l'erreur de prévision pour chaque observation contenue dans  $\mathbf{y}$ .

Les auteurs utilisent ensuite cette approche pour améliorer les capacités des systèmes d'assimilation variationnelle à estimer la variation de l'erreur commise lors d'un changement de spécification des matrices d'erreur ainsi que l'ajustement et l'estimation des paramètres des matrices  $\mathbf{B}$  et  $\mathbf{R}$ . Plusieurs essais numériques sont ensuite conduits sur le système de prévision météorologique opérationnel de la NASA, GEOS-5.

### Application récente

Un exemple d'application de cela dans le cadre de la prévision météorologique par des méthodes variationnelles se trouve chez (KIM et KIM, 2019).

Les auteurs y utilisent les diagnostics FSOI pour comparer le 4D-Var classique avec le 4D-Var hybride, qui utilise une estimation par méthode d'ensemble pour construire sa matrice de covariance d'erreur d'ébauche  $\mathbf{B}$ . Nous verrons l'adaptation des FSOI dans le cas des méthodes ensemblistes un peu plus loin en sous-section (3.2.3).

### FSOI pour les méthodes d'ensemble (diagnostics EnFSOI)

Les diagnostics FSOI n'intéressant pas seulement la partie variationnelle de l'assimilation de données, leur adaptation pour les méthodes de filtrage étaient une suite logique. Après l'article de (LANGLAND et BAKER, 2004), (LIU et KALNAY, 2008) a suivi la même voie dans le cadre de l'EnKF appliqué spécifiquement au modèle de Lorenz96. La généralisation à tout type de méthode ensembliste se retrouve dans (KALNAY et al., 2012).

En reprenant les notations de (KALNAY et al., 2012),  $\mathbf{x}_{t|0}^f$  correspond à la prévision au temps  $t$  partant de l'analyse au temps 0 soit  $\mathbf{x}_0^a$ , et  $\mathbf{x}_{t|6}^f$  à la prévision au temps  $t$  depuis l'ébauche antérieure de 6h à  $\mathbf{x}_0^a$ . L'intervalle de temps de 6h est calibré sur celui de la prévision météorologique générale, il peut varier selon le système auquel s'applique le diagnostic.

Définissons  $\mathbf{x}_t^a$  l'analyse au temps  $t$ ,  $\mathbf{e}_{t|0} = \mathbf{x}_{t|0}^f - \mathbf{x}_t^a$  l'erreur entre la prévision ayant connaissance des observations au temps 0 et l'analyse au temps  $t$ , ainsi que  $\mathbf{e}_{t|6} = \mathbf{x}_{t|6}^f - \mathbf{x}_t^a$  l'erreur entre la prévision sans connaissance des observations au temps 0 et l'analyse au temps  $t$ .

D'après (LANGLAND et BAKER, 2004), l'impact des observations au temps 0 se mesure par la fonction de coût suivante :

$$J = \left[ \mathbf{e}_{t|0}^T \mathbf{C} \mathbf{e}_{t|0} - \mathbf{e}_{t|6}^T \mathbf{C} \mathbf{e}_{t|6} \right]. \quad (3.26)$$

En notant  $\mathbf{M}$  l'opérateur tangent linéaire d'évolution du modèle, nous avons  $(\mathbf{x}_{t|0}^f - \mathbf{x}_{t|6}^f) \simeq \mathbf{M}(\mathbf{x}_0^a - \mathbf{x}_{0|6}^b) = \mathbf{M}\delta\mathbf{x}_0$ .

Un peu de calcul et l'utilisation de la formule (3.20) mènent à l'approximation :

$$J \simeq [\mathbf{M}\mathbf{K}\delta\mathbf{y}_0]^T \mathbf{C} (\mathbf{e}_{t|0} + \mathbf{e}_{t|6}). \quad (3.27)$$



L'approche utilisée en 4D-Var est d'approximer directement cette quantité en calculant l'adjoint du modèle, l'adjoint du gain et d'effectuer le produit scalaire final.

Pour l'EnKF, c'est différent. Les anomalies d'analyse au temps 0 sont contenues dans la matrice  $\Delta \mathbf{X}_0^a$  dans l'espace des états, et dans  $\Delta \mathbf{Y}_0^a = \mathbf{H}(\Delta \mathbf{X}_0^a)$  dans l'espace des observations. Pour un ensemble de  $N$  membres, une formulation alternative du gain de Kalman est  $\mathbf{K} = (\Delta \mathbf{X}_0^a)(\Delta \mathbf{X}_0^a)^T \mathbf{H}^T \mathbf{R}^{-1}$ . Nous avons alors  $\mathbf{MK} = \mathbf{M}(\Delta \mathbf{X}_0^a)(\Delta \mathbf{X}_0^a)^T \mathbf{H}^T \mathbf{R}^{-1} \simeq (\Delta \mathbf{X}_{t|0}^f)(\Delta \mathbf{Y}_0^a)^T \mathbf{R}^{-1}$ .

$\Delta \mathbf{X}_{t|0}^f$  représente les anomalies de prévisions normalisées relativement à la moyenne de l'ensemble, comme définies par l'équation (2.39). Ainsi, la fonction de coût se reformule :

$$J \simeq \delta \mathbf{y}_0^T \mathbf{R}^{-1} (\Delta \mathbf{Y}_0^a) (\Delta \mathbf{X}_{t|0}^f)^T \mathbf{C} (\mathbf{e}_{t|0} + \mathbf{e}_{t-6}). \quad (3.28)$$

Il est à noter que  $\Delta \mathbf{X}_{t|0}^f$  peut se calculer directement via le modèle non-linéaire. Ceci assure une validité de l'approximation sur une fenêtre plus longue que celle de l'adjoint en variationnel, et la possibilité d'étude de l'impact des observations sur des prévisions relativement à plus long terme. Pour les prévisions à court-terme au sens de la météorologie, les deux approches donnent des résultats équivalents.

Cependant, l'approche ensembliste se dégrade sur les prévisions au long-terme du fait de l'introduction de la localisation, nécessaire lors des expériences sur des cas réalistes. Une localisation en temps pourrait remédier à cela (KALNAY et al., 2012, Sec. 5).

Un dernier avantage par rapport à l'adjoint réside dans le coût de l'approche ensembliste qui est sensiblement moindre.

Une extension de cette méthode en se basant directement sur les observations plutôt que l'analyse se trouve chez (SOMMER et WEISSMANN, 2016), qui permet aussi de ne pas être gêné par la présence de biais sur l'analyse ou autres erreurs systématiques. Cependant, cette méthode est particulièrement sensible au choix de la matrice  $\mathbf{C}$  correspondant à la norme avec laquelle l'erreur est définie.

(HOTTA et al., 2017) combine les travaux de (DAESCU, 2012) et antérieurs avec ceux de (KALNAY et al., 2012) pour déterminer cette fois la sensibilité de la prévision aux variations sur la matrice  $\mathbf{R}$ , pour permettre un meilleur paramétrage de la matrice et améliorer la qualité des prévisions actuelles. En effet, un paramétrage de  $\mathbf{R}$  guidé par leur méthode se ressent sur la sensibilité de la prévision directement aux observations. Cependant, aucune amélioration statistiquement significative de la qualité des prévisions n'a été observée, des travaux supplémentaires autour de cette méthode sont donc nécessaires.

D'avantage de précisions sur les propriétés des diagnostics FSOI par méthode d'ensemble se trouvent chez (KOTSUKI, KUROSAWA et MIYOSHI, 2019) pour des applications en prévision météorologique.

## Récapitulatif

Les diagnostics FSOI ont été en premier formulés pour les méthodes variationnelles, qui sont historiquement les premières à avoir été utilisées et implémentées dans le cadre de la prévision météorologique. Leur formulation équivalente pour les méthodes d'ensemble a logiquement suivi. Ils ont fait leur preuve et sont maintenant utilisés dans plusieurs systèmes opérationnels.

Dans les deux cas de figure, ce sont des diagnostics faits *a posteriori* de l'étape d'analyse,

l'état analysé  $\mathbf{x}^a$  servant d'état de référence au diagnostic du fait qu'il englobe la totalité de la connaissance disponible à la fois depuis le modèle et les observations. Cependant, d'autres méthodes comme chez (SOMMER et WEISSMANN, 2016) ou encore (CARDINALI, 2018) se servent des observations  $\mathbf{y}$  comme état de référence, qui sont éventuellement disponibles en dehors du temps de l'analyse ou en tout cas ne nécessitent pas le calcul de  $\mathbf{x}^a$  pour être obtenues. Ils peuvent aussi être étendus à l'étude de la sensibilité aux perturbations des caractéristiques d'erreur à la fois du modèle ou des observations comme chez (DAESCU, 2012) et (HOTTA et al., 2017), respectivement.

Cependant, ils sont essentiellement des diagnostics *a posteriori* et ne permettent pas un contrôle de la prévision en amont de l'analyse lorsqu'ils dépendent de  $\mathbf{x}^a$ . Une piste pour obtenir de nouveaux indicateurs pouvant fournir de l'information pour des diagnostics cette fois *a priori* de l'analyse, semble se trouver dans la comparaison directe entre les observations et la prévision comme chez (SOMMER et WEISSMANN, 2016). C'est avec cette idée en tête que nous poursuivons le travail de recherche de cette thèse.

### 3.3 Conclusion et orientation des recherches

Les recherches en prévision numérique par filtrage ont trois axes de développement ayant chacun pour but final d'améliorer la qualité des prévisions obtenues, éventuellement par une réduction de leur coût ou bien par une amélioration de la précision et de la pertinence physique de celles-ci.

Le premier axe est dans l'amélioration des méthodes de filtrage elles-mêmes et a abouti aux méthodes d'ensemble itératives, lisseurs... Ces méthodes s'étendent au delà du pur formalisme ensembliste de l'EnKF en résolvant le problème d'analyse comme un problème de minimisation, à l'instar de l'approche variationnelle. Cela a débouché sur l'IEnKS, méthode itérative la plus récente et la plus prometteuse car elle permettrait des prévisions plus précises et à moindre coûts, du fait de son lissage et de la forte réduction de taille de l'ensemble qu'elle permet.

Le deuxième axe – qui commence à être pris en compte dans les méthodes itératives les plus récentes – est la gestion de l'erreur modèle en assimilation de données. C'est une problématique importante qui commence à être mise en place dans les systèmes opérationnels. Ceci est dû à la complexité intrinsèque de la modélisation de cette erreur difficilement mesurable, contrairement à une variance d'erreur sur un capteur pour les observations par exemple. La théorie sur l'erreur modèle est actuellement riche, et c'est maintenant des développements applicatifs qui sont à l'ordre du jour, notamment sur les systèmes variationnels en prévision météorologique. Ceci permet d'augmenter encore la précision des prévisions faites, avec un certain surcoût.

Le troisième axe est le pendant de l'aspect modèle, il va lui s'intéresser aux observations, et constitue les travaux sur les diagnostics de sensibilité de la prévision aux observations et de l'impact des observations sur l'analyse. Ces diagnostics ont été initialement formulés pour les méthodes variationnelles du fait de l'augmentation massive de données disponibles ces dernières années en prévision atmosphérique. Ils commencent à être formulés à la fois dans le cadre des méthodes ensemblistes et hybrides, ces dernières combinant

l'approche variationnelle à l'approche ensembliste. Ils permettent de discriminer efficacement les observations les plus importantes et sont aujourd'hui utilisés pour faire de la calibration du système d'observation ainsi que de la réanalyse, car trop coûteux encore pour la prévision. Ils sont majoritairement faits *a posteriori* de l'analyse.

Nous sommes intéressés dans nos travaux de recherches par la qualité de nos prévisions faites par un EnKF classique, ainsi que par la conservation voire l'amélioration de la prévisibilité de notre système de manière générale. Cette prévisibilité est une propriété statistique au long terme, elle s'évalue grâce à une série temporelle d'analyses passées de taille conséquente, se déroulant sur l'échelle d'un siècle en météorologie par exemple. Suivre sa qualité implique de régulièrement faire des réanalyses, ce qui n'est pas notre objectif : en effet, nous voulons voir s'il est possible de déterminer *a priori* l'impact d'une observation sur l'analyse et/ou la prévision pour éventuellement ne pas l'assimiler, lorsque nous disposons d'un système fortement observé. L'utilisation des indicateurs *a posteriori* sera plutôt réservée à la calibration du réseau d'observation d'un système, pour justifier la pose d'un nouveau capteur par exemple.

Pour cela nous allons voir comment comparer l'observation et la prévision sans faire l'étape d'analyse, en définissant des indicateurs de qualité *a priori*. Nous définirons aussi des indicateurs *a posteriori* depuis des indicateurs existants pour vérifier les nouveaux que nous proposerons. Ils seront explicités et analysés dans le chapitre suivant.

## 4 Description des indicateurs retenus de qualité des observations pour la prévision

Trois indicateurs sont présentés dans ce chapitre : la Hat matrix dont nous avons déjà discuté, la distance de Wasserstein et la distance de Kullback-Leibler. Tous trois ont déjà servi pour plusieurs applications en assimilation de données. Nous les détaillons également pour étoffer leur description.

### 4.1 Matrice des sensibilités ou "Hat Matrix" pour l'EnKF

Nous avons présenté la Hat matrix dans le cadre d'un moindre carré assez général en sous-section (3.2.1), nous allons maintenant nous concentrer sur son calcul dans le cadre de l'assimilation de données avec l'EnKF.

Dans ce cas, le problème d'assimilation de données peut se formuler au sens des moindres carrés de la façon suivante :

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left\| \begin{bmatrix} \mathbf{B}^{-1/2} \\ \mathbf{R}^{-1/2} \mathbf{H} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{B}^{-1/2} \mathbf{x}_b \\ \mathbf{R}^{-1/2} \mathbf{y} \end{bmatrix} \right\|_2^2 = \min_{\mathbf{x} \in \mathbb{R}^n} \left\| \begin{bmatrix} \mathbf{I}_n \\ \mathbf{H} \end{bmatrix} x - \begin{bmatrix} \mathbf{x}_b \\ \mathbf{y} \end{bmatrix} \right\|_{\mathbf{M}^{-1}}^2 \quad (4.1)$$

en suivant les notations du chapitre 2, avec la matrice de poids  $\mathbf{M} = \begin{pmatrix} \mathbf{B}^{-1} & 0 \\ 0 & \mathbf{R}^{-1} \end{pmatrix}$ .

On cherche maintenant à exprimer la Hat matrix de ce moindre carré. Formons le gain pour le problème complet :

$$\begin{aligned} & \begin{bmatrix} \mathbf{B}^{-1/2} & 0 \\ 0 & \mathbf{R}^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{I}_n \\ \mathbf{H} \end{bmatrix} \left( \begin{bmatrix} \mathbf{I}_n & \mathbf{H}^T \end{bmatrix} \begin{bmatrix} \mathbf{B}^{-1} & 0 \\ 0 & \mathbf{R}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I}_n \\ \mathbf{H} \end{bmatrix} \right)^{-1} \left( \begin{bmatrix} \mathbf{I}_n & \mathbf{H}^T \end{bmatrix} \begin{bmatrix} \mathbf{B}^{-1/2} & 0 \\ 0 & \mathbf{R}^{-1/2} \end{bmatrix} \right) \\ &= \begin{bmatrix} \mathbf{B}^{-1/2} & 0 \\ 0 & \mathbf{R}^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{I}_n \\ \mathbf{H} \end{bmatrix} (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \left( \begin{bmatrix} \mathbf{I}_n & \mathbf{H}^T \end{bmatrix} \begin{bmatrix} \mathbf{B}^{-1/2} & 0 \\ 0 & \mathbf{R}^{-1/2} \end{bmatrix} \right) \\ &= \begin{bmatrix} \mathbf{B}^{-1/2} & 0 \\ 0 & \mathbf{R}^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{C} & \mathbf{C} \mathbf{H}^T \\ \mathbf{H} \mathbf{C} & \mathbf{H} \mathbf{C} \mathbf{H}^T \end{bmatrix} \begin{bmatrix} \mathbf{B}^{-1/2} & 0 \\ 0 & \mathbf{R}^{-1/2} \end{bmatrix}, \quad \text{où } \mathbf{C} = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \end{aligned}$$

La Hat matrix est la matrice correspondant à  $\mathbf{R}^{-1/2}\mathbf{HCH}^T\mathbf{R}^{-1/2}$  qui représente la matrice de covariance de l'estimation par les observations. Nous la noterons  $\hat{\mathbf{H}}$ .

$$\begin{aligned} \hat{\mathbf{H}} &= \mathbf{R}^{-1/2}\mathbf{HCH}^T\mathbf{R}^{-1/2} = \mathbf{R}^{-1/2}\mathbf{H}(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1/2} \\ &= \mathbf{I}_p - \mathbf{R}^{1/2}(\mathbf{R} + \mathbf{HBH}^T)^{-1}\mathbf{R}^{1/2} \quad \text{Sherman-Morrison-Woodbury} \\ \hat{\mathbf{H}} &= \mathbf{I}_p - (\mathbf{I}_p + \mathbf{R}^{-1/2}\mathbf{HBH}^T\mathbf{R}^{-1/2})^{-1} \end{aligned} \quad (4.2)$$

La matrice ainsi obtenue correspond exactement à la transposition de la notion statistique de Hat matrix au domaine de l'assimilation de données, en l'exprimant sous forme d'un problème aux moindres carrés.

Il ne nous reste plus qu'à récupérer la diagonale de  $\hat{\mathbf{H}} = (\hat{h}_{ii})_{i \leq m}$  pour connaître l'impact de chaque observation sur le modèle.

Un calcul similaire est fait dans (CARDINALI, 2013, Sec. (A.1)) pour obtenir la Hat matrix du système complet d'assimilation de données.

## 4.2 Essais numériques de vérification de l'indicateur

Le système utilisé est celui de Lorenz 63 avec le jeu de paramètres classique . Il est totalement observé, avec l'ajout d'une 4<sup>eme</sup> observation supplémentaire a priori peu informative : en notant  $\mathbf{y} = [y_1, y_2, y_3, y_4]^T$  le vecteur d'observation,  $y_4 = 0 + \epsilon_4$ , où  $\epsilon_4$  représente le bruit du à l'erreur sur l'observation.

Il devrait s'en suivre que  $\hat{h}_{4,4} = 0 < \hat{h}_{1,1}, \hat{h}_{2,2}, \hat{h}_{3,3}$ , du fait que l'observation  $y_4$  n'apporte comme information qu'un simple bruit. La configuration utilisée est la configuration 2 décrite en annexe (A.7.1).

$\hat{\mathbf{H}}$  nous renvoie comme valeurs diagonales sur la fenêtre :

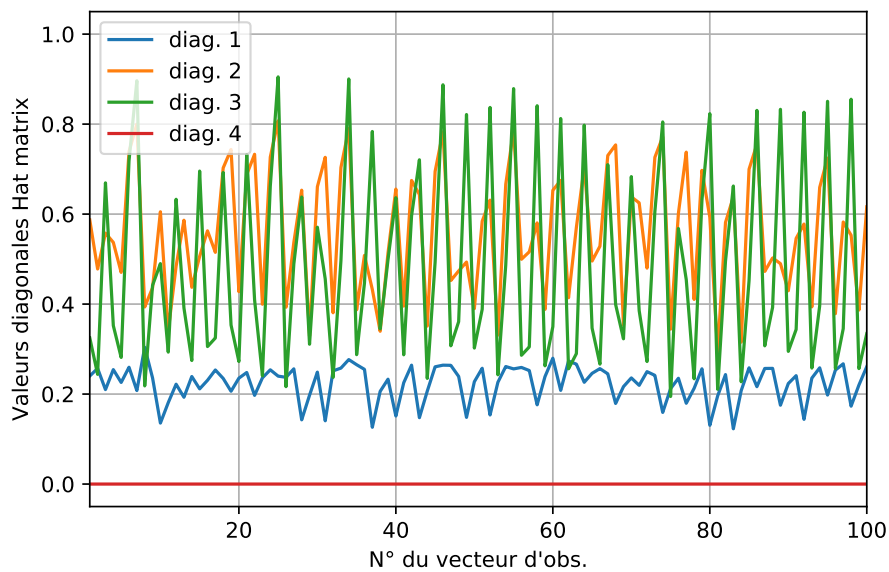


FIGURE 4.1 – Valeurs diagonales de  $\hat{\mathbf{H}}$ , pour 100 vecteurs d'observation, avec intervalle de temps  $dt_{obs} = 0.25$  entre chaque vecteur.

Notre hypothèse que la 4<sup>ème</sup> dimension du vecteur d'observation a peu d'impact sur le résultat obtenu lors de l'assimilation est validée : la Hat matrix détecte bien que  $y_4$  n'apporte aucune information au système d'assimilation de données.

Voyons maintenant les moyennes des RMSEs d'analyse pour des essais successifs d'abord complètement observés puis où l'on supprime l'observation d'importance la plus faible lors de l'assimilation, c'est-à-dire celle qui correspond à la valeur diagonale la plus faible de  $\hat{\mathbf{H}}$ , qui peut varier dans le temps et semble être soit  $y_1$ , soit  $y_4$  d'après (4.1). Voici les valeurs obtenues :

	40 assimilations : RMSE moyenne	1000 assimilations : RMSE moyenne
Sans suppression	0.05801	0.05955
Avec suppression	0.05801	0.05955

Nous voyons que celles-ci sont identiques, du fait que la Hat matrix supprime à chaque la 4<sup>ème</sup> observation lors de l'analyse. Toutes les analyses se déroulent finalement comme si  $\mathbf{y} = [y_1, y_2, y_3]^T$ , ce qui était attendu.

Voici en comparaison le même essai sur le même système sans l'observation  $y_4$ , c'est-à-dire pour la configuration 1 telle que décrite en annexe (A.7.1).

Dans ce cas, les valeurs diagonales de  $\hat{\mathbf{H}}$  sont :

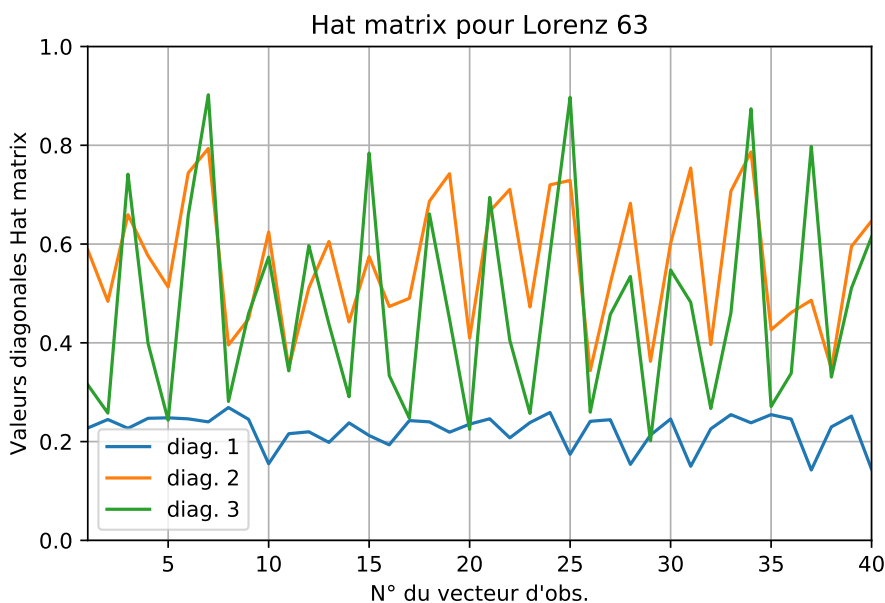


FIGURE 4.2 – Valeurs diagonales de  $\hat{\mathbf{H}}$ , restreintes aux 40 1ers vecteurs d'observation.

Celles-ci varient fortement. Il apparaît en moyenne que les dimensions 2 et 3 soient les importantes et la dimension 1 toujours la moins importante. Nous devons donc nous assurer qu'il s'agit bien d'un comportement systématique et pas d'une configuration particulière due à la graine utilisée pour le générateur pseudo-aléatoire.

En moyennant les valeurs diagonales obtenues sur 100 essais avec une graine du générateur pseudo-aléatoire différente pour chacun, nous obtenons les valeurs d'importance moyennes suivantes :

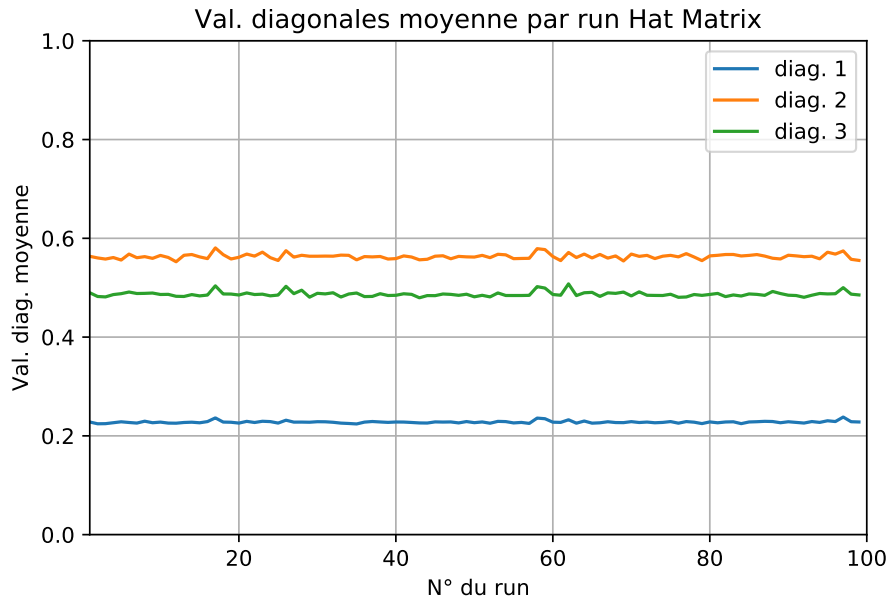


FIGURE 4.3 – Valeurs diagonales moyennes de  $\hat{\mathbf{H}}$  pour 100 expériences.

L'importance accordée à chaque dimension reste en accord avec celle décrite sur la figure (4.2), ce comportement est donc bien systématique.

Nous retrouvons le même comportement pour la 1<sup>ère</sup> dimension du système, qui a encore l'importance la plus faible. En supprimant de nouveau l'observation d'importance la plus faible lors de l'assimilation, nous avons comme moyennes des RMSEs d'analyse :

	40 assimilations : RMSE moyenne	1000 assimilations : RMSE moyenne
Sans suppression	0.06803	0.05955
Avec suppression	0.05896	0.06201

Sur la petite fenêtre de 40 assimilations, nous avons localement une amélioration de la qualité des analyses lorsque nous n'avons pas l'observation artificielle dans le système, et lorsque nous supprimons l'observation de moindre importance correspondant à la 1<sup>ère</sup> dimension du système. Nous retombons sur les valeurs précédentes dans le cas d'une longue fenêtre d'assimilation avec des statistiques qui ont convergé, les perturbations dues à l'observation artificielle semblent donc agir avant tout à échelle locale.

La Hat matrix adaptée pour l'EnKF permet donc bien de discriminer les observations avec un effet moindre sur la mise à jour du vecteur d'état au sein du vecteur  $\mathbf{y}$ . L'impact sur la RMSE au long terme est faible entre un essai où l'on supprime systématiquement l'observation la moins importante lors de l'analyse et celle où l'on assimile systématiquement l'intégralité du contenu de  $\mathbf{y}$ . Cependant des variations locales plus marquées peuvent apparaître.

Ces essais préliminaires nous permettent de voir comment utiliser la Hat matrix dans nos expérimentations futures.

### 4.3 Distance de Wasserstein

La distance de Wasserstein (VILLANI, 2009, chap. 6) tient ses origines de la théorie du transport optimal.

Soit  $(\chi, d)$  un espace complet et séparable. Soit  $p \in [1, +\infty)$ . Pour tout couple de mesures  $(\mu, \nu)$  sur  $\chi$ , soit  $\Gamma(\mu, \nu)$  l'ensemble des couplages de  $\mu$  et  $\nu$ , c'est-à-dire l'ensemble des mesures sur  $\chi \times \chi$  ayant pour densités marginales  $\mu$  et  $\nu$ , respectivement en 1<sup>er</sup> et 2<sup>nd</sup> facteurs.

La distance de Wasserstein d'ordre  $p$  entre  $\mu$  et  $\nu$  est définie comme suit :

$$W_p(\mu, \nu) = \left( \inf_{\pi \in \Gamma(\mu, \nu)} \int_{\chi^2} d(x, y)^p d\pi(x, y) \right)^{1/p} \quad (4.3)$$

Cette métrique permet de comparer deux distributions de probabilités de deux variables  $X$  et  $Y$ , lorsqu'une des variables est issue de l'autre grâce à de petites perturbations non-uniformes (aléatoires comme déterministes).

En l'état actuel, ce n'est pas à strictement parler une distance car elle peut prendre des valeurs infinies, il suffit pour éviter cela de se restreindre à l'espace des mesures de probabilités ayant un *moment d'ordre  $p$  fini*, i.e :

$$\mathbb{P}_p(\chi) = \left\{ \mu \in \mathbb{P}(\chi); \int_{\chi} d(x_0, x)^p \mu(dx) < +\infty \right\}, \quad (4.4)$$

avec  $x_0 \in \chi$  arbitraire, l'espace étant indépendant du point choisi.

#### 4.3.1 Formulation dans le cas gaussien

Soient  $\mu_1 = \mathcal{N}(\mathbf{m}_1, \mathbf{C}_1)$  et  $\mu_2 = \mathcal{N}(\mathbf{m}_2, \mathbf{C}_2)$  deux distributions gaussiennes non-dégénérées sur  $\mathbb{R}^n$ , avec pour moyennes respectives  $\mathbf{m}_1$  et  $\mathbf{m}_2 \in \mathbb{R}^n$  et pour matrices de covariances  $\mathbf{C}_1$  et  $\mathbf{C}_2 \in \mathcal{M}_n(\mathbb{R})$ . Dans ce cas, selon (GIVENS et SHORTT, 1984, section 2, pp. 236-240), la distance de Wasserstein d'ordre 2 (ou 2-Wasserstein) entre  $\mu_1$  et  $\mu_2$  correspond à :

$$W_2(\mu_1, \mu_2)^2 = \|\mathbf{m}_1 - \mathbf{m}_2\|_2^2 + \text{tr} \left[ \mathbf{C}_1 + \mathbf{C}_2 - 2(\mathbf{C}_2^{1/2} \mathbf{C}_1 \mathbf{C}_2^{1/2})^{1/2} \right]. \quad (4.5)$$

Pour de nombreux détails théoriques sur l'origine de cette formule et les propriétés des distances de Wasserstein, le lecteur est invité à se plonger plus amplement dans les références mentionnées précédemment.

Comme remarqué dans (FARCHI et al., 2016, Sec. 2), la norme Euclidienne utilisée ici peut être remplacée par une norme matricielle  $\|\cdot\|_{\mathbf{T}^{-1}}$ , où  $\mathbf{T}$  est une matrice inversible de poids, prenant en compte les écarts d'ordre de grandeur entre les différentes variables du système, ou par exemple  $\mathbf{R}$ , pour prendre en compte les différents niveaux d'erreurs sur chaque observation.

Les systèmes sur lesquels nous effectuerons nos essais ont des variables chacune du même ordre de grandeur, et il en est de même pour les variances d'erreur de chaque observation, nous utiliserons donc directement la formulation (4.5) pour nos calculs de distance.



### 4.3.2 Distance de Wasserstein en assimilation de données : un palliatif à l'erreur modèle

L'utilisation de la distance de Wasserstein se retrouve en assimilation de données dans des domaines d'applications comme l'étude de la propagation atmosphérique de polluants aériens dans (FARCHI et al., 2016), la réduction d'artefacts d'assimilation lors du suivi ou de la prévision de polluants de surface sur des étendues d'eau pour (LI et al., 2018), une meilleure prise en compte de l'erreur modèle et/ou du biais dans (NING et al., 2014; FEYEUX, VIDARD et NODET, 2018) voire encore la correction simultanée des biais de modèle et d'observation pour des approches variationnelles dans (TAMANG et al., 2020).

Dans (FARCHI et al., 2016) la distance de Wasserstein sert pour comparer les distributions spatio-temporelles de sources d'émissions de polluants en palliant les défauts des normes plus "classiques" d'algèbre linéaire dans le sens où elle permet de prendre les champs de diffusion des polluants dans leur globalité sans nécessiter de les échantillonner. Cette échantillonnage peut parfois impliquer un surcoût, mais surtout poser un problème de double-pénalisation de la solution : les informations fournies par des normes sont locales et ponctuelles et lorsque deux termes sources diffèrent spatialement ou temporellement, cette déviation ne peut être estimée avec ces normes classiques, ce qui n'est pas le cas de Wasserstein.

(NING et al., 2014) ont introduit la théorie du transport optimal dans le cadre des méthodes variationnelles et ont montré qu'un biais au niveau de l'ébauche donnait à l'analyse une distribution bimodale, grâce à la reformulation de certains termes de la fonction de coût en distance de Wasserstein plutôt qu'en norme Euclidienne pondérée. (FEYEUX, VIDARD et NODET, 2018) vont plus loin encore en reformulant le problème d'optimisation variationnelle avec comme seule métrique la distance de Wasserstein pour permettre un meilleur comportement de la distribution de la solution lors de l'analyse, sans pour autant se pencher sur le problème du biais.

C'est (TAMANG et al., 2020) qui va se pencher sur ce problème en reprenant les travaux des deux articles précédents. La distance de Wasserstein est ici utilisée pour prendre en compte les biais à la fois de modèle et d'observation dans la solution du 3D-Var via l'ajout d'un terme de régularisation dans la fonction de coût. Dans le cadre applicatif de la météorologie, les observations ayant un biais faible sont les observations *in situ* tels que les radiosondes atmosphériques ou les hygromètres au sol, bien que la précision de la mesure relativement à d'autres comme les images satellites soit moindre. Ce terme de régularisation exprime le coût revenant à transporter la distribution de l'analyse vers la distribution de référence des observations *in situ* qui encapsule la variabilité spatio-temporelle des variables d'états. Ceci permet, après pondération du terme de régularisation, d'empêcher l'apparition de distributions bimodales irréalistes à l'analyse.

Ces applications montrent l'utilisation de la distance de Wasserstein comme moyen de prise en compte et/ou de correctoin des erreurs modèles sous-jacentes à plusieurs configurations en assimilation de données, propriété qui lui vient de sa capacité à mieux rendre compte des différences spatiales entre deux distributions, ou encore entre leurs deux premiers moments en comparaison aux normes classiques. Il est à noter cependant la nécessité d'effectuer des calculs supplémentaires pour formuler cette distance, voire de résoudre un problème d'optimisation plus complexe (et potentiellement non-convexe) lorsqu'elle sert à

formuler une nouvelle fonction de coût. L'importance du compromis entre coût et précision du résultat ressort dans certaines de ces études.

## 4.4 Divergence de Kullback-Leibler

D'après (SOCH et ALLEFELD, 2016), la divergence de Kullback-Leibler (KL) est une métrique non-symétrique entre deux distributions de probabilités  $P$  et  $Q$  définies sur  $\Omega$  et vaut dans le cas discret :

$$KL[P||Q] = \sum_{i \in \Omega} P(i) \log \frac{P(i)}{Q(i)} \quad (4.6)$$

où  $\log$  correspond au logarithme népérien (ou en base 2 si l'on travaille avec des bits).

Dans le cas continu de  $P$  et  $Q$  ayant pour densité de probabilité respectives  $p(x)$  et  $q(x)$  sur un domaine  $X$ , elle vaut :

$$KL[P||Q] = \int_X p(x) \log \frac{p(x)}{q(x)} dx \quad (4.7)$$

Cette divergence mesure la dissimilarité informative entre deux distributions de probabilités. (GAUTAM et al., 2018) propose une méthode simple pour symétriser cette divergence, qui revient simplement à sommer  $KL[P||Q] + KL[Q||P]$  pour donner :

$$KL(P, Q) = KL[P||Q] + KL[Q||P] = \int_X (p(x) + q(x)) \log \frac{p(x)}{q(x)} dx \quad (4.8)$$

dont on vérifie aisément la symétrie.

### 4.4.1 Formulation dans le cas gaussien

D'après (SOCH et ALLEFELD, 2016), pour  $\mathbf{x} \in \mathbb{R}^n$ , soient  $\mu_1 = \mathcal{N}(\mathbf{m}_1, \mathbf{C}_1)_{\mathbf{x}}$  et  $\mu_2 = \mathcal{N}(\mathbf{m}_2, \mathbf{C}_2)_{\mathbf{x}}$ . La divergence de KL s'exprime alors :

$$KL[\mu_1||\mu_2] = \frac{1}{2} \left[ (\mathbf{m}_2 - \mathbf{m}_1)^T \mathbf{C}_2^{-1} (\mathbf{m}_2 - \mathbf{m}_1) + \text{tr}(\mathbf{C}_2^{-1} \mathbf{C}_1) - \log \frac{\det(\mathbf{C}_1)}{\det(\mathbf{C}_2)} - n \right] \quad (4.9)$$

La version symétrisée de cette formule vaut :

$$KL(\mu_1, \mu_2) = \frac{1}{2} [(\mathbf{m}_2 - \mathbf{m}_1)^T (\mathbf{C}_2^{-1} + \mathbf{C}_1^{-1}) (\mathbf{m}_2 - \mathbf{m}_1) + \text{tr}(\mathbf{C}_2^{-1} \mathbf{C}_1 + \mathbf{C}_1^{-1} \mathbf{C}_2) - 2n]. \quad (4.10)$$

Nous utiliserons préférentiellement la formule symétrique dans la suite, car la symétrie permet d'éviter des ambiguïtés à la compréhension selon que nous regardons sous l'angle de  $\mu_1$  ou de  $\mu_2$ , qui correspondraient dans le cas de l'assimilation de données à regarder les choses du point de vue soit de l'état, soit des observations. Cette formule permet de prendre en compte les dissimilarités informatives des deux points de vue et de les fondre ensemble. Un défaut possible est que cela occulte l'information de dissimilarité propre à chaque point de vue, mais un avantage non-négligeable est que cela rapproche

la divergence de Kullback-Leibler d'une distance probabiliste classique, son interprétation et les comparaisons aux autres indicateurs et notamment la distance de Wasserstein en seront simplifiées.

#### 4.4.2 La divergence de Kullback-Leibler en assimilation de données

La divergence de Kullback-Leibler – aussi connue sous le nom d'*entropie relative* – a fait l'objet d'études théoriques sur sa capacité à décrire le contenu informatif des observations et de la prévision par rapport à une statistique de référence en climatologie.

D'après (MAJDA, KLEEMAN et CAI, 2002), c'est une mesure qui décèle tout changement non-trivial sur la moyenne de la prévision, et ce dans un contexte linéaire comme non-linéaire, du fait que l'information qu'elle retourne soit invariante pour tout changement de variable suffisamment lisse. Selon les auteurs, cette propriété la rend particulièrement intéressante pour quantifier le contenu informatif de la prévision. Du fait que nous soyons amenés à traiter les informations de prévision dans un cadre où les non-linéarités peuvent être marquées, ces arguments vont en faveur de notre choix. De plus, (XU, 2007) montre que, dans le cas d'un opérateur d'observation linéaire, il est possible de décrire d'une façon adéquate l'espace des observations, permettant de sélectionner pour l'assimilation les observations les plus pertinentes grâce la divergence de Kullback-Leibler. Le tout avec une perte minimale d'information.

Notre objectif est de comparer le contenu informatif de notre prévision numérique avec celui des observations pour en sélectionner les plus pertinentes, la divergence de Kullback-Leibler s'avère donc être un choix idéal.

Récemment, la divergence de Kullback-Leibler se retrouve plutôt autour des thématiques de sensibilité de capteurs comme dans (GAUTAM et al., 2018) et (HU et al., 2018), où il s'agit, respectivement, d'estimer l'erreur de mesure et la probabilité de fausse alarme d'un capteur, et d'obtenir un consensus entre plusieurs capteurs spatiaux pour déterminer la position correcte d'un objet. Ou bien l'amélioration de méthodes existantes d'assimilation de données comme dans (MANSOURI, NOUNOU et NOUNOU, 2014), pour améliorer la prise en compte des non-linéarités et éviter la divergence du filtre à particules, voire de permettre la prise en compte de contraintes d'inégalités par projection de la distribution de l'ensemble sur l'espace des contraintes dans un EnKF chez (LI et al., 2019).

Pour (MANSOURI, NOUNOU et NOUNOU, 2014), un défaut du filtre à particule avec rééchantillonnage tient dans le fait que l'on est forcé de négliger la dernière observation pour déterminer la distribution d'importance optimale pour affecter un poids à chaque particule, pouvant mener à une divergence du filtre. Leur méthode consiste à minimiser la divergence de Kullback-Leibler entre la distribution a posteriori et une approximation de la vraisemblance du rééchantillonnage prenant en compte la dernière observation. Ceci donne une distribution a posteriori optimale pour réaffecter les poids de chaque particule, et obtenir un meilleur suivi en RMSE de l'état du système en comparaison à d'autres méthodes de filtrage classiques.

Chez (HU et al., 2018), il s'agit d'utiliser la divergence Kullback-Leibler pour la résolution d'un consensus entre densités de probabilités pour un problème d'estimation par filtre de

Kalman, qui revient à calculer le barycentre entre plusieurs densités dans un espace probabilisé. Une résolution approchée du problème d'estimation global est faite localement pour chaque capteur puis chaque capteur communique son résultat à ses voisins. Une fois cela fait, le barycentre des estimations est calculé pour fournir la position estimée de l'objet étudié. L'exemple d'application donnée est celle du suivi GPS d'un objet par plusieurs satellites.

Selon (GAUTAM et al., 2018), la détection de capteurs fautifs se fait en étudiant les séquences des résidus d'erreur d'innovations, d'estimation de l'état, et de covariance. Selon qu'il y ait des erreurs ou non, les erreurs suivront une loi spécifique dépendant des hypothèses et paramètres inhérents au système. La détection de capteurs fautifs revient donc à faire un test statistique différenciant les hypothèses "avec erreur" et "sans erreur". Cela revient à calculer la divergence de Kullback-Leibler entre la distribution des erreurs obtenues via les différents capteurs et la distribution des erreurs décrites par chaque hypothèse du test. Si l'écart est supérieur à un certain seuil, le capteur est vu comme fautif et isolé du reste. Il est à noter que le cas d'application est pour la détection de capteurs fautif au sein de l'instrumentation des pressuriseurs d'un réacteur à eau pressurisée (REP), technologie de réacteur nucléaire très répandue au sein du parc d'EDF. Nous retrouvons dans cet article une utilisation de Kullback-Leibler similaire à Wasserstein dans la correction de biais : il s'agit d'estimer l'écart entre la distribution estimée et une distribution de référence.

Le dernier cas d'application dans (LI et al., 2019) correspond à un EnKF à modèle de mélange gaussien ou "Gaussian-Mixture Model" (GMM) pour l'estimation de la distribution de l'état. Pour prendre en compte des contraintes d'inégalité dans le formalisme du filtre, la divergence de Kullback-Leibler est utilisée pour projeter la solution sans contraintes du GMM-EnKF dans l'espace des contraintes, et ensuite de minimiser cette divergence. Ceci donne en revanche un fonction de coût non-convexe, et nécessite un raffinement du problème d'optimisation en deux sous-problèmes convexes pour sa résolution lors de l'étape de mise à jour du filtre, augmentant le temps de calcul de la solution. Cette méthode permet ainsi de faire du filtrage avec des distributions bimodales et physiquement réalistes du fait qu'elles respectent les contraintes du modèle.

La divergence de Kullback-Leibler semble quant à elle avoir des applications plus autour des observations mais surtout des systèmes de mesures en eux-mêmes, à la fois pour leur amélioration en terme de précision ou d'erreur de détection mais aussi pour permettre de les coordonner lorsqu'il s'agit d'échanger et recouper entre elles les informations de plusieurs capteurs. Une similarité avec la distance de Wasserstein est tout de même à noter, notamment lorsqu'il s'agit de comparer une distribution de référence à une autre.

Maintenant que la présentation globale des indicateurs et des enjeux de la thèse ont été faites, nous allons passer à la deuxième partie, qui présente les travaux de recherches de la thèse, à la fois sur le plan théorique et le plan applicatif.



## Deuxième partie

### Contribution théorique et travaux de recherche



## 5 Description et utilisation des indicateurs dans l'EnKF

### 5.1 Introduction et contexte - contrôle de la dynamique de correction par gestion de l'assimilation via diagnostics sur les observations

Nous cherchons à contrôler la manière dont la prévision du système physique est corrigée en temps et/ou en espace via l'assimilation d'observation, sans chercher à formuler un nouvel algorithme inspiré de l'EnKF comme c'est le cas du RIP-EnKF (KALNAY et YANG, 2010) ou de l'IEEnKF/IEEnKS (SAKOV, OLIVER et BERTINO, 2012; BOCQUET et SAKOV, 2012; BOCQUET et SAKOV, 2014; BOCQUET, 2016) par exemple, où l'on minimise itérativement une fonctionnelle dépendant des observations, permettant entre autres de diminuer la taille de l'ensemble et de tirer une plus grande information de chaque observation. Nous voulons plutôt voir, quel que soit l'algorithme de filtrage/lissage utilisé, si le schéma classique qui consiste à assimiler chaque observation peut être modifié grâce à des indicateurs permettant d'obtenir des informations sur ces mêmes observations sans nécessairement les assimiler toutes.

Dans ce cas, un moyen de gérer la dynamique de correction – c'est à dire de gérer la fréquence à laquelle le système est contrôlé et corrigé par l'assimilation d'observations – est de déterminer la quantité d'information apportée par une observation. Il existe pour cela des indicateurs de qualité de l'observation *a posteriori* – c'est-à-dire après assimilation, comme la matrice des sensibilités de (CARDINALI, 2013) qui a été implémentée sur plusieurs systèmes opérationnels. Nous pouvons aussi mentionner les DFS, GCV et FSOI (CARDINALI, 2013; WHABA et WENDELBERG, 1980; WHABA et al., 1995; KALNAY et al., 2012; TRÉMOLET, 2008), qui ont par ailleurs un lien avec la matrice de sensibilités, ou "Hat matrix" comme nous allons l'appeler par la suite.

Cependant, ces indicateurs permettent d'obtenir une information *a posteriori* sur la qualité informative du système d'observation en comparaison à celle de la prévision, mais pas directement de contrôler au fil du temps les temps d'assimilation des observations que l'on appellera la dynamique d'assimilation. Il nous faut pour cela des indicateurs *a priori* qui ont été décrit dans le précédent chapitre, une façon d'interpréter l'information qu'ils retournent en terme de qualité d'observation, et une façon de les intégrer au sein de l'EnKF pour contrôler la dynamique d'assimilation via des stratégies pour déterminer quand assimiler un vecteur d'observation et/ou pour permettre de sélectionner les observations à assimiler.

A notre connaissance, la littérature discute peu de l'économie d'observations ou du moins de l'économie d'assimilations de vecteurs d'observation. Ceci est en effet contre-intuitif,



surtout lorsque l'on peut lire dans (TODLING, 2013, p. 1485) que sous les bonnes hypothèses, "l'assimilation d'une observation va *toujours* améliorer la qualité de la prévision qui suivra". Il sera donc particulièrement important de bien évaluer les bénéfices et risques de cette démarche.

Cependant, que se passerait-il si l'économie d'assimilations était possible? Ou du moins, s'il était possible de réduire le nombre d'observations assimilées sans dégrader outre-mesure la qualité de la prévision résultant ensuite? Ceci peut être utile lorsque nous avons à faire à un système qui reçoit des observations en nombre conséquent, dont nous voulons évaluer la pertinence, dans un premier temps sur une fenêtre donnée, voire dans un second temps à chaque arrivée d'une observation. Un tel fonctionnement est commode lorsque la qualité ou le temps de lancement d'une prévision à long terme, relativement à l'échelle de temps des fenêtres d'assimilation, dépend de ces mêmes observations et leur temps d'assimilation. En effet, si le calcul de prévision prend un temps de l'ordre de la longueur de la fenêtre, vaut-elle la peine d'être recalculée à chaque fois ou seulement lorsque l'observation apporte un changement marqué lors de son assimilation? Les simulations d'hydraulique chez EDF sont un cas de figure qui pourrait illustrer cette problématique : elles servent à estimer plusieurs périodes de marée dans un estuaire, chaque simulation de prévision prend environ 10h de calcul alors que le système reçoit toutes les 15 min des données provenant de sources diverses dont des cartographies satellites de taille conséquente, assimilées toutes les 12 à 24h. Le passage à une assimilation à l'arrivée des observations est en cours d'étude, il devient alors intéressant de sélectionner les données pertinentes plutôt que de tout assimiler pour réduire la quantité d'information à traiter et éviter un nouveau calcul de prévision coûteux en temps.

Néanmoins, il faut que ce type de diagnostic respecte les contraintes de coûts imposées par la configuration du système et les exigences sur les résultats de l'industriel.

Nous verrons dans un premier temps l'utilisation des indicateurs lorsqu'une observation se présente, puis la définition d'un nouvel indicateur (SOBS) ainsi que des comparaisons de fonctionnement entre les indicateurs retenus (SOBS, distance de Wassertein, distance de Kullback-Leibler) avec celui de référence (Hat matrix).

## 5.2 Utilisation des indicateurs à l'arrivée d'une observation - plusieurs cas

Avant tout, nous prendrons comme indicateur de référence la Hat matrix décrite en section (4.1), du fait qu'elle soit déjà utilisée dans l'opérationnel (CARDINALI, 2013). Son utilisation comme indicateur *a posteriori* sur l'assimilation permettra de regarder l'impact qu'a une dimension du vecteur d'observations sur la qualité de l'assimilation obtenue. Pour détailler cela, plaçons-nous à  $t = t_{obs}$  sur une fenêtre d'assimilation. Une première assimilation sera faite avec  $\mathbf{H}$ , donnant pour résultat  $\mathbf{x}^a$  et  $\mathbf{P}^a$ , et une seconde sera faite en supprimant une observation avec  $\tilde{\mathbf{H}} \in \mathcal{M}_{p-1,n}(\mathbb{R})$ , donnant comme résultat  $\tilde{\mathbf{x}}^a$  et  $\tilde{\mathbf{P}}^a$ . Cela correspond à deux distributions légèrement différentes en fin de fenêtre d'assimilation.

Vouloir supprimer une observation revient à travailler avec des observations provenant d'un sous-espace de  $\mathbb{R}^p$  qui va être  $\mathbb{R}^{p-1}$  dans l'exemple décrit précédemment. Pour cela, nous pourrions aussi travailler directement avec  $\tilde{\mathbf{H}} = [h_1, \dots, h_{i-1}, h_{i+1}, \dots, h_p]^T$  où les  $h_k$  représentent les lignes de  $\mathbf{H}$ . Dans ce cas, cela revient à restreindre l'espace des observations à  $\mathbb{R}^{p-1}$  via une modification de l'opérateur. Des comparaisons sur ces deux méthodes

seront discutées par la suite.

Comme remarqué à la fin de la sous-section (3.2.1), les comparaisons avec la Hat matrix se font exclusivement au modèle des observations sans prendre en compte les effets du modèle de simulation. La quantification de l'impact des observations qui en résulte peut donc être améliorée par l'utilisation d'autres indicateurs qui la prennent en compte, soit pour remplacer le diagnostic de la Hat matrix, soit pour le compléter. L'utilisation de ces indicateurs pourra se faire au moment de l'arrivée d'une observation en la comparant avec la prévision au temps de l'observation, pour voir s'il est nécessaire de l'assimiler ou non.

En accord avec la définition de l'importance d'une observation en sous-section (3.2.1), supposons que la Hat matrix ait décrit la  $i^{eme}$  observation comme négligeable en terme d'impact sur l'assimilation, les distances probabilistes devraient ainsi montrer un faible écart entre  $\mathcal{N}(\mathbf{x}^a, \mathbf{P}^a)$  et  $\mathcal{N}(\tilde{\mathbf{x}}^a, \tilde{\mathbf{P}}^a)$ . De la même manière, SOBS devrait montrer une importance faible pour la  $i^{eme}$  observation relativement aux autres. Nous y reviendrons dans le chapitre sur les essais numériques.

### 5.2.1 Cas 1 : synchrone

Ceci correspond au cas le plus simple. Dans le cas de l'assimilation synchrone, l'assimilation se fait au temps de l'observation  $t = t_{obs}$ . Il suffit alors de calculer les indicateurs avec l'information de l'observation et de la prévision à  $t = t_{obs}$  pour pouvoir contrôler la dynamique d'assimilation. Ce fonctionnement sera celui de la majorité des applications numériques à venir.

### 5.2.2 Cas 2 : asynchrone

Dans le cas de l'assimilation asynchrone, les observations sont assimilées à un temps  $t_a > t_{obs}$ . Ceci se fait via une implémentation différente de l'algorithme de l'EnKF pour donner l'AEnKF pour "Asynchronous Ensemble Kalman Filter" (SAKOV, EVENSEN et BERTINO, 2010).

Pour l'AEnKF, sur une fenêtre d'assimilation  $[t_0, t_f]$  où l'on dispose de  $n_{obs}$  observations arrivant à des temps  $t_0 < t_i < t_f$ , avec  $i \in \llbracket 1, n_{obs} \rrbracket$ , il suffit de sauvegarder les ensembles projetés dans l'espace des observations  $\mathbf{H}_i \mathbf{E}_i$  pour chaque  $t_i$  pour ensuite les accumuler en  $[\mathbf{H}_1 \mathbf{E}_1, \mathbf{H}_2 \mathbf{E}_2, \dots, \mathbf{H}_{n_{obs}} \mathbf{E}_{n_{obs}}]^T$  avec les vecteurs d'observations  $[\mathbf{y}_1, \dots, \mathbf{y}_{n_{obs}}]^T$  et faire une assimilation augmentée à  $t_f$  de tous ces ensembles et observations concaténés à la fois. Dans le cas d'un problème linéaire, ceci donne le même résultat en fin de fenêtre que pour l'algorithme de l'EnKF synchrone (Cf. (SAKOV, EVENSEN et BERTINO, 2010) pour plus d'explications).

L'étape d'analyse étant en aval de tous les temps d'observations dans ce cas, les indicateurs initialement a posteriori peuvent servir à faire du diagnostic *a priori*. En prenant l'exemple de la Hat matrix, avec le formalisme de l'AEnKF de (SAKOV et BERTINO, 2010), former la Hat matrix augmentée en fin de fenêtre d'assimilation à  $t_f$  nécessite de stocker pour les temps d'observations  $t_k$  définis  $\forall k \in \mathbb{N}^*$  par  $t_0 < \dots < t_k < \dots < t_f$  :

- $\mathbf{HE} = [(\mathbf{HE})_1^T, \dots, (\mathbf{HE})_k^T]^T \in \mathcal{M}_{kp, N}(\mathbb{R})$  l'ensemble à différents instants dans l'espace des observations,

- $Y = [\mathbf{Y}_1^T, \dots, \mathbf{Y}_k^T]^T \in \mathcal{M}_{kp,N}(\mathbb{R})$  les ensembles d'observations perturbées,
- $R = \text{diag}[\text{diag}(\mathbf{R}_1), \dots, \text{diag}(\mathbf{R}_k)] \in \mathcal{M}_{kp}(\mathbb{R})$  la matrice d'erreur d'observation augmentée.

Nous avons ensuite besoin de construire les matrices  $\mathbf{HP}_e\mathbf{H}^T = \frac{(\mathbf{H}\Delta\mathbf{X})(\mathbf{H}\Delta\mathbf{X})^T}{N-1} \in \mathcal{M}_{kp}(\mathbb{R})$  et  $\mathbf{P}_e\mathbf{H}^T = \frac{\Delta\mathbf{X}(\mathbf{H}\Delta\mathbf{X})^T}{N-1} \in \mathcal{M}_{n,kp}(\mathbb{R})$  pour pouvoir calculer l'ensemble analysé qui prend en compte toutes ces observations via la formule :

$$\mathbf{E}_a = \mathbf{E}_f + \mathbf{P}_e\mathbf{H}^T(\mathbf{R}^{-1} + \mathbf{HP}_e\mathbf{H}^T)^{-1}(\mathbf{Y} - \mathbf{HE}). \quad (5.1)$$

En se basant sur la formule de l'équation (4.2), la Hat matrix augmentée pour le problème asynchrone est :

$$\hat{\mathbf{H}}_{async} = \mathbf{I}_{kp} - \left( \mathbf{I}_{kp} + \mathbf{R}^{-1/2} \frac{(\mathbf{HE})(\mathbf{HE})^T}{N-1} \mathbf{R}^{-1/2} \right)^{-1} \quad (5.2)$$

dont nous n'aurons besoin de calculer que les éléments diagonaux.

Ceci permet donc de voir quelles observations sont les plus importantes sur l'ensemble de la fenêtre d'un seul coup, et ce *avant* assimilation, contrairement à son fonctionnement synchrone. Son calcul implique de manipuler des matrices de plus grande taille, surcoût pouvant se compenser à la fois avec la réduction du nombre d'observations à assimiler et le fait que nous ne nécessitions que le calcul de ses valeurs diagonales.

Ensuite, grâce aux retours de la Hat matrix (ou autre indicateur) sur l'importance informative de chaque observation, il est possible de choisir laquelle supprimer pour l'assimilation, permettant une gestion de la dynamique d'assimilation en fin de fenêtre uniquement, une fois la prévision faite.

### 5.3 Indicateurs des sensibilités à l'observation

Notre objectif est de déterminer l'importance qu'a une observation sur l'état analysé lors d'une assimilation. Ceci est fait sachant que nous sommes dans un contexte où nous avons beaucoup d'observations sur notre fenêtre, ou bien que la taille du système fait que le processus le plus coûteux lors du filtrage est l'assimilation d'observations. Par ailleurs, il reste toujours intéressant de réduire les coûts d'assimilation lorsque cela est possible, quelle que soit la configuration du système. Nous voulons voir s'il est possible de le faire sans affecter les performances du système, notamment en terme de précision des prévisions.

#### 5.3.1 Formulation pour l'EnKF

##### Justification de l'hypothèse de diagonalité de $\mathbf{R}$

La matrice d'erreur d'observation  $\mathbf{R}$  est considérée comme diagonale définie positive pour simplifier dans un premier temps les calculs : des erreurs d'observation décorréelées et aucune connaissance parfaite d'une observation.

Une seconde justification est que, chez EDF, pour plusieurs systèmes physiques sur lesquels des recherches ont été menées pour caractériser la matrice d'erreur d'observation ont

conduit à des matrices diagonales définies positives : aucune information supplémentaire n'a pu être obtenue, ou du moins aucune caractérisation plus fine de la matrice d'erreur n'a été possible. Même si cela est en cours d'évolution, les matrices diagonales sont encore utilisées pour des applications réelles. Cela permet ainsi une cohérence avec les configurations applicatives.

Cette hypothèse là est donc à la fois cohérente avec les retours d'expérience industriels et simplifie la forme de certains calculs ainsi que leur coût numérique.

L'idée derrière SOBS est de regarder l'effet de la suppression d'une observation sur l'incrément d'analyse, et de quantifier l'importance d'une ou plusieurs observations du vecteur  $\mathbf{y}$  différemment de la Hat matrix. Cela se fait en perturbant la matrice d'erreur d'observation  $\mathbf{R}$  sur sa  $i^{\text{ème}}$  valeur diagonale, pour rendre ensuite la variance d'erreur associée infinie, donc rendre l'observation  $y_i$  négligeable, et étudier l'effet que cela a sur la mise à jour de l'étape d'analyse.

### Calcul avec $\mathbf{R} = \sigma^2 \mathbf{I}_p$

Par simplicité, nous cherchons une première formulation pour SOBS dans le cas où, sans perturbation, la matrice  $\mathbf{R}$  est un multiple de l'identité.

Dans ce cas, la perturbation s'effectue comme suit, en donnant la matrice perturbée  $\mathbf{R}_\epsilon$  :

$$\mathbf{R}_\epsilon = \sigma_0^2 \left( \mathbf{I}_p + \left( \frac{1}{\epsilon} - 1 \right) E_{ii} \right) = \sigma_0^2 \mathbf{D}_i \left( \frac{1}{\epsilon} \right) \quad (5.3)$$

c'est-à-dire que l'erreur d'observation pour la  $i^{\text{ème}}$  observation vaut  $\sigma_0^2/\epsilon$ . L'introduction de ce  $\epsilon > 0$  sera expliquée quelques lignes plus loin.

Ici,  $E_{ii}$  est une matrice élémentaire de la base canonique de  $\mathcal{M}_p(\mathbb{R})$  qui vérifie  $E_{ii}^2 = E_{ii}$  et  $E_{ii}^T = E_{ii}$ , donc :

$$\mathbf{D}_i \left( \frac{1}{\epsilon} \right) = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & & \ddots & & & \\ 0 & \cdots & \cdots & \frac{1}{\epsilon} & 0 & \cdots & 0 \\ \vdots & & & & \ddots & & \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix} \quad (5.4)$$

On remarque de même que  $\forall \epsilon \neq 0, \mathbf{D}_i \left( \frac{1}{\epsilon} \right)^{-1} = \mathbf{D}_i(\epsilon)$ .

$\mathbf{d} = \mathbf{y} - \mathbf{H}\mathbf{x}^b$  représente l'écart aux observations, où  $\mathbf{y} \in \mathbb{R}^p$  est l'observation et  $\mathbf{x}_b \in \mathbb{R}^n$  l'état du système d'assimilation de données avant assimilation.

En faisant varier  $\epsilon$  ou  $1/\epsilon$ , nous pouvons ainsi étudier la sensibilité de l'assimilation à ces observations. Par exemple  $\epsilon \rightarrow 0 \iff 1/\epsilon \rightarrow \infty$ , cela équivaut à donner une erreur infinie sur une mesure donc à la négliger. Cela modélise la suppression de mesures non importantes et permet de simuler cette action.

La variation de l'état  $\mathbf{x}$  par rapport à l'écart aux observations lors de l'assimilation s'écrit :

$$\begin{aligned} \delta \mathbf{x}(\epsilon) &= \left( \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}_\epsilon^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{R}_\epsilon^{-1} \mathbf{d} \\ &= \frac{1}{\sigma_0^2} \left( \mathbf{B}^{-1} + \frac{1}{\sigma_0^2} \mathbf{H}^T \mathbf{D}_i(\epsilon) \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{D}_i(\epsilon) \mathbf{d} \end{aligned} \quad (5.5)$$

En utilisant l'identité  $d(\mathbf{A}^{-1}) = -\mathbf{A}^{-1}(d\mathbf{A})\mathbf{A}^{-1}$ , la dérivée selon  $\epsilon$  de  $\delta\mathbf{x}$  s'écrit :

$$\begin{aligned} \delta\mathbf{x}'(\epsilon) &= \frac{-1}{\sigma_0^4} \left( \mathbf{B}^{-1} + \frac{1}{\sigma_0^2} \mathbf{H}^T \mathbf{D}_i(\epsilon) \mathbf{H} \right)^{-1} (\mathbf{H}^T E_{ii} \mathbf{H}) \left( \mathbf{B}^{-1} + \frac{1}{\sigma_0^2} \mathbf{H}^T \mathbf{D}_i(\epsilon) \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{D}_i(\epsilon) \mathbf{d} \\ &\quad + \frac{1}{\sigma_0^2} \left( \mathbf{B}^{-1} + \frac{1}{\sigma_0^2} \mathbf{H}^T \mathbf{D}_i(\epsilon) \mathbf{H} \right)^{-1} \mathbf{H}^T E_{ii} \mathbf{d} \end{aligned}$$

Avec  $e_i$  le  $i^{\text{eme}}$  vecteur de la base canonique de  $\mathbb{R}^p$ , nous remarquons que  $\mathbf{H}^T E_{ii} = \mathbf{H}^T e_i e_i^T$ , donc  $\mathbf{H}^T E_{ii} \mathbf{d} = h_i d_i$  avec  $\mathbf{H}^T = [h_1, \dots, h_p]$ , d'où :

$$\begin{aligned} \delta\mathbf{x}'(\epsilon) &= \frac{-1}{\sigma_0^4} \left( \mathbf{B}^{-1} + \frac{1}{\sigma_0^2} \mathbf{H}^T \mathbf{D}_i(\epsilon) \mathbf{H} \right)^{-1} (h_i h_i^T) \left( \mathbf{B}^{-1} + \frac{1}{\sigma_0^2} \mathbf{H}^T \mathbf{D}_i(\epsilon) \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{D}_i(\epsilon) \mathbf{d} \\ &\quad + \frac{d_i}{\sigma_0^2} \left( \mathbf{B}^{-1} + \frac{1}{\sigma_0^2} \mathbf{H}^T \mathbf{D}_i(\epsilon) \mathbf{H} \right)^{-1} h_i \end{aligned} \tag{5.6}$$

Nous posons :

$$\mathbf{g}_i(\epsilon) = \left( \mathbf{B}^{-1} + \frac{1}{\sigma_0^2} \mathbf{H}^T \mathbf{D}_i(\epsilon) \mathbf{H} \right)^{-1} h_i \tag{5.7}$$

L'équation (5.6) devient alors :

$$\begin{aligned} (5.6) &= \frac{-1}{\sigma_0^4} \mathbf{g}_i(\epsilon) \mathbf{g}_i(\epsilon)^T [\mathbf{H}^T \mathbf{d} + (\epsilon - 1) d_i h_i] + \frac{d_i}{\sigma_0^2} \mathbf{g}_i(\epsilon) \\ &= \frac{-1}{\sigma_0^4} \mathbf{g}_i(\epsilon)^T \left[ \sum_{j=1}^p d_j h_j + (\epsilon - 1) d_i h_i \right] \mathbf{g}_i(\epsilon) + \frac{d_i}{\sigma_0^2} \mathbf{g}_i(\epsilon) \end{aligned} \tag{5.8}$$

car  $\mathbf{g}_i(\epsilon)^T = e_i^T \mathbf{H} (\mathbf{B}^{-1} + (1/\sigma_0^2) \mathbf{H}^T \mathbf{D}_i(\epsilon) \mathbf{H})^{-1} \in \mathbb{R}^p$  donc  $\mathbf{g}_i(\epsilon)^T [\sum_j d_j h_j + (\epsilon - 1) d_i h_i] \in \mathbb{R}$ .

Ainsi :

$$\begin{aligned} \delta\mathbf{x}'(\epsilon) &= \frac{-1}{\sigma_0^4} \mathbf{g}_i(\epsilon)^T \left( \sum_{j \neq i}^p d_j h_j \right) \mathbf{g}_i(\epsilon) + \frac{d_i}{\sigma_0^2} \left[ 1 - \frac{\epsilon}{\sigma_0^2} \mathbf{g}_i(\epsilon)^T h_i \right] \mathbf{g}_i(\epsilon) \\ &= \frac{e_i^T \mathbf{d}}{\sigma_0^2} A(\epsilon) - \frac{1}{\sigma_0^4} B(\epsilon, d), \quad \text{pour l'implémentation,} \end{aligned} \tag{5.9}$$

Avec :

$$A(\epsilon) = \left( 1 - \frac{\epsilon}{\sigma_0^2} \mathbf{g}_i(\epsilon)^T h_i \right) \mathbf{g}_i(\epsilon), \tag{5.10}$$

$$B(\epsilon, d) = \mathbf{g}_i(\epsilon)^T \left( \sum_{j \neq i}^p d_j h_j \right) \mathbf{g}_i(\epsilon). \tag{5.11}$$

**Calcul avec  $\mathbf{R} = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$  diagonale quelconque**

En notant  $\circ$  le produit de Schur, la matrice perturbée  $\mathbf{R}_\epsilon$  vaut dans ce cas :

$$\mathbf{R}_\epsilon = \text{diag}(\sigma_1^2, \dots, \sigma_p^2) \circ \left[ \mathbf{I}_p + \left( \frac{1}{\epsilon} - 1 \right) E_{ii} \right]. \quad (5.12)$$

Et son inverse vaut :

$$\mathbf{R}_\epsilon^{-1} = \text{diag}(1/\sigma_1^2, \dots, 1/\sigma_p^2) \circ [\mathbf{I}_p + (\epsilon - 1)E_{ii}]. \quad (5.13)$$

Repartons maintenant de  $\delta \mathbf{x}(\epsilon)$  :

$$\begin{aligned} \delta \mathbf{x}(\epsilon) &= \left( \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}_\epsilon^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{R}_\epsilon^{-1} \mathbf{d} \\ &= \left( \mathbf{B}^{-1} + \mathbf{H}^T \left[ \sum_{j=1}^p \frac{E_{jj}}{\sigma_j^2} + \frac{1}{\sigma_i^2} (\epsilon - 1) E_{ii} \right] \mathbf{H} \right)^{-1} \mathbf{H}^T \left[ \sum_{j=1}^p \frac{E_{jj}}{\sigma_j^2} + \frac{1}{\sigma_i^2} (\epsilon - 1) E_{ii} \right] \mathbf{d} \end{aligned} \quad (5.14)$$

En redéveloppant les calculs de la même manière que dans le cas précédent :

$$\begin{aligned} \delta \mathbf{x}'(\epsilon) &= -\frac{1}{\sigma_i^2} \left( \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}_\epsilon^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^T E_{ii} \mathbf{H} \left( \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}_\epsilon^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{R}_\epsilon^{-1} \mathbf{d} \\ &\quad + \left( \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}_\epsilon^{-1} \mathbf{H} \right)^{-1} \frac{d_i h_i}{\sigma_i^2} \\ &= -\frac{\tilde{\mathbf{g}}_i(\epsilon) \tilde{\mathbf{g}}_i(\epsilon)^T}{\sigma_i^2} \mathbf{H}^T \left[ \sum_{j=1}^p \frac{E_{jj}}{\sigma_j^2} + \left( \frac{\epsilon - 1}{\sigma_i^2} \right) E_{ii} \right] \mathbf{d} + \frac{d_i \tilde{\mathbf{g}}_i(\epsilon)}{\sigma_i^2} \end{aligned} \quad (5.15)$$

avec :

$$\tilde{\mathbf{g}}_i(\epsilon) = \left( \mathbf{B}^{-1} + \mathbf{H}^T \left[ \sum_{j=1}^p \frac{E_{jj}}{\sigma_j^2} + \frac{1}{\sigma_i^2} (\epsilon - 1) E_{ii} \right] \mathbf{H} \right)^{-1} h_i \quad (5.16)$$

nous obtenons :

$$\begin{aligned} \delta \mathbf{x}'(\epsilon) &= -\frac{\tilde{\mathbf{g}}_i(\epsilon)^T}{\sigma_i^2} \left[ \sum_{j=1}^p \frac{d_j h_j}{\sigma_j^2} + \left( \frac{\epsilon - 1}{\sigma_i^2} \right) d_i h_i \right] \tilde{\mathbf{g}}_i(\epsilon) + \frac{d_i}{\sigma_i^2} \tilde{\mathbf{g}}_i(\epsilon) \\ &= \frac{e_i^T \mathbf{d}}{\sigma_i^2} \tilde{A}(\epsilon) - \frac{1}{\sigma_i^2} \tilde{B}(\epsilon, d), \quad \text{pour l'implémentation.} \end{aligned} \quad (5.17)$$

Avec :

$$\tilde{A}(\epsilon) = \left( 1 - \frac{\epsilon}{\sigma_i^2} \tilde{\mathbf{g}}_i(\epsilon)^T h_i \right) \tilde{\mathbf{g}}_i(\epsilon), \quad (5.18)$$

$$\tilde{B}(\epsilon, d) = \tilde{\mathbf{g}}_i(\epsilon)^T \left( \sum_{j \neq i}^p \frac{d_j h_j}{\sigma_j^2} \right) \tilde{\mathbf{g}}_i(\epsilon). \quad (5.19)$$

### Cas ensembliste avec $\mathbf{B}$ non inversible

En général pour les applications de l'EnKF,  $n > N$  donc la matrice  $\mathbf{B} \equiv \mathbf{P}^f$  est de rang déficient et singulière. Plaçons nous dans le formalisme du LETKF de (HUNT, KOSTELICH et SZUNYOGH, 2007, Sec. 2.3). Le but de ce filtre est de chercher la solution de l'analyse dans l'espace des anomalies, c'est-à-dire l'espace engendré par la matrice  $\Delta\mathbf{X}$  définie par l'équation (2.39).

D'après l'équation (15) de l'article, l'analyse du LETKF est le minimum de la fonction :

$$J(\mathbf{w}) = \left(\frac{N-1}{2}\right) \mathbf{w}^T \mathbf{w} + \frac{1}{2} [\mathbf{y} - \mathbf{H}(\mathbf{x}^f + \Delta\mathbf{X}^f \mathbf{w})]^T \mathbf{R}^{-1} [\mathbf{y} - \mathbf{H}(\mathbf{x}^f + \Delta\mathbf{X}^f \mathbf{w})]. \quad (5.20)$$

$\Delta\mathbf{X}^f \in \mathcal{M}_N(\mathbb{R})$  est la matrice des anomalies de prévisions, associée à la prévision  $\mathbf{x}^f$  calculée selon l'équation (2.41), qui fait office d'ébauche pour l'analyse. L'état analysé s'exprime alors sous la forme :

$$\mathbf{x}^a = \mathbf{x}^f + \Delta\mathbf{X}^f \mathbf{w}^a. \quad (5.21)$$

Notons  $\mathbf{d} = \mathbf{y} - \mathbf{H}\mathbf{x}^f$ . Le vecteur  $\mathbf{w}^a \in \mathbb{R}^N$  est un vecteur de poids s'appliquant aux colonnes de  $\Delta\mathbf{X}^f$ . Pour déterminer l'expression de  $\mathbf{w}^a$ , regardons la condition nécessaire d'optimalité au premier ordre :

$$\begin{aligned} \nabla J(\mathbf{w}) &= (N-1)\mathbf{w} - (\mathbf{H}\Delta\mathbf{X}^f)^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}^f - \mathbf{H}\Delta\mathbf{X}^f \mathbf{w}) \\ &= [(N-1)\mathbf{I}_N + (\mathbf{H}\Delta\mathbf{X}^f)^T \mathbf{R}^{-1} (\mathbf{H}\Delta\mathbf{X}^f)]^{-1} \mathbf{w} - (\mathbf{H}\Delta\mathbf{X}^f)^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}^f) = 0 \\ \implies \mathbf{w}^a &= [(N-1)\mathbf{I}_N + (\mathbf{H}\Delta\mathbf{X}^f)^T \mathbf{R}^{-1} (\mathbf{H}\Delta\mathbf{X}^f)]^{-1} (\mathbf{H}\Delta\mathbf{X}^f)^T \mathbf{R}^{-1} \mathbf{d}. \end{aligned} \quad (5.22)$$

En reprenant les calculs du cas précédent  $\mathbf{R} = \sigma_0^2 \mathbf{I}_p$ , par analogie, l'incrément d'analyse  $\delta\mathbf{x}$  vaut :

$$\begin{aligned} \delta\mathbf{x}(\epsilon) &= \Delta\mathbf{X}^f \mathbf{w}^a(\epsilon) = \frac{\Delta\mathbf{X}^f}{\sigma_0^2} \left[ (N-1)\mathbf{I}_N + (\mathbf{H}\Delta\mathbf{X}^f)^T \frac{1}{\sigma_0^2} \mathbf{D}_i(\epsilon) (\mathbf{H}\Delta\mathbf{X}^f) \right]^{-1} \\ &\quad \times (\mathbf{H}\Delta\mathbf{X}^f)^T \mathbf{D}_i(\epsilon) \mathbf{d}. \end{aligned} \quad (5.23)$$

Notons  $\Upsilon = \mathbf{H}\Delta\mathbf{X}^f \implies \Upsilon^T = [v_1, \dots, v_p]$  la matrice des anomalies projetée dans l'espace des observations. Dans ce cas,  $\mathbf{w}^a(\epsilon)$  devient :

$$\mathbf{w}^a(\epsilon) = \frac{1}{\sigma_0^2} \left[ (N-1)\mathbf{I}_N + \Upsilon^T \frac{1}{\sigma_0^2} \mathbf{D}_i(\epsilon) \Upsilon \right]^{-1} \Upsilon^T \mathbf{D}_i(\epsilon) \mathbf{d} \quad (5.24)$$

$$\begin{aligned} \implies \mathbf{w}^a &= \frac{-1}{\sigma_0^4} \left[ (N-1)\mathbf{I}_N + \Upsilon^T \frac{1}{\sigma_0^2} \mathbf{D}_i(\epsilon) \Upsilon \right]^{-1} (v_i v_i^T) \left[ (N-1)\mathbf{I}_N + \Upsilon^T \frac{1}{\sigma_0^2} \mathbf{D}_i(\epsilon) \Upsilon \right]^{-1} \\ &\quad \times \Upsilon^T \mathbf{D}_i(\epsilon) \mathbf{d} + \left[ (N-1)\mathbf{I}_N + \Upsilon^T \frac{1}{\sigma_0^2} \mathbf{D}_i(\epsilon) \Upsilon \right]^{-1} \frac{d_i v_i}{\sigma_0^2}. \end{aligned} \quad (5.25)$$

Par analogie, posons de nouveau :

$$\mathbf{g}_i^e(\epsilon) = \left[ (N-1)\mathbf{I}_N + \Upsilon^T \frac{1}{\sigma_0^2} \mathbf{D}_i(\epsilon) \Upsilon \right]^{-1} v_i. \quad (5.26)$$

Dans le formalisme du LETKF, nous obtenons finalement comme expression pour  $\delta \mathbf{x}'(\epsilon)$  :

$$\delta \mathbf{x}'(\epsilon) = \Delta \mathbf{X}^f \left( -\frac{1}{\sigma_0^4} \mathbf{g}_i^e(\epsilon) \mathbf{g}_i^e(\epsilon)^T \left[ \sum_{j=1}^p d_j v_j + (\epsilon - 1) d_i v_i \right] + \frac{d_i}{\sigma_0^2} \mathbf{g}_i^e(\epsilon) \right), \quad (5.27)$$

formule analogue à l'équation (5.8). De même, par analogie avec l'équation (5.17), nous obtenons dans le cas  $\mathbf{R} = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$  :

$$\delta \mathbf{x}'(\epsilon) = \Delta \mathbf{X}^f \left( -\frac{1}{\sigma_i^2} \tilde{\mathbf{g}}_i^e(\epsilon)^T \tilde{\mathbf{g}}_i^e(\epsilon) \left[ \sum_{j=1}^p \frac{d_j v_j}{\sigma_j^2} + \left( \frac{\epsilon - 1}{\sigma_i^2} \right) d_i v_i \right] + \frac{d_i}{\sigma_i^2} \tilde{\mathbf{g}}_i^e(\epsilon) \right). \quad (5.28)$$

### Formulation pour utilisation asynchrone dans le cas ensembliste

Nous nous baserons sur (SAKOV, EVENSEN et BERTINO, 2010). Pour la formulation asynchrone sur la fenêtre  $[t_0, t_f]$ , en notant les temps d'observation  $t_k$  tels que  $\forall k \in \mathbb{N}^*$ ,  $t_0 < \dots < t_k < \dots < t_f$ , avec les mêmes notations qu'en sous-section (5.3.1), nous allons avoir besoin de stocker les matrices permettant de construire  $\mathbf{G}_i^e(\epsilon)$ , une formulation augmentée de  $\mathbf{g}_i^e(\epsilon)$ .

Pour déterminer le vecteur de poids  $\mathbf{w}^a$  adapté, exprimons la fonction de coût du problème d'assimilation de données asynchrone correspondante :

$$\tilde{J}(\mathbf{w}) = \left( \frac{N-1}{2} \right) \mathbf{w}^T \mathbf{w} + \frac{1}{2} \sum_{j=1}^k \left[ \mathbf{y}_j - \mathbf{H}_j(\mathbf{x}_j^f + \Delta \mathbf{X}_j^f \mathbf{w}) \right]^T \mathbf{R}^{-1} \left[ \mathbf{y}_j - \mathbf{H}_j(\mathbf{x}_j^f + \Delta \mathbf{X}_j^f \mathbf{w}) \right]. \quad (5.29)$$

La taille des vecteurs d'observations peut varier au cours du processus d'assimilation. Nous noterons donc  $\forall j \in \llbracket 1, k \rrbracket$ ,  $\mathbf{y}_j \in \mathbb{R}^{p_j}$  avec  $p_j \leq p$ , et  $\tilde{p} = \sum_{j=1}^k p_j$ . Remarquons aussi que  $\mathbf{H}_j \in \mathcal{M}_{p_j, n}(\mathbb{R})$ . Posons alors :

$$\tilde{\Upsilon} = \begin{bmatrix} \mathbf{H}_1 \Delta \mathbf{X}_1^f \\ \vdots \\ \mathbf{H}_k \Delta \mathbf{X}_k^f \end{bmatrix} \in \mathcal{M}_{\tilde{p}, N}(\mathbb{R}), \quad \mathbf{d} = \begin{bmatrix} \mathbf{y}_1 - \mathbf{H}_1 \mathbf{x}_1^f \\ \vdots \\ \mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^f \end{bmatrix} \in \mathbb{R}^{\tilde{p}}. \quad (5.30)$$

Posons également  $\mathbf{R} = \text{diag}(\mathbf{R}_j) \in \mathcal{M}_{\tilde{p}}(\mathbb{R})$  avec  $\forall j$ ,  $\mathbf{R}_j \in \mathcal{M}_{p_j}(\mathbb{R})$  la matrice d'erreur d'observation correspondant au vecteur d'observation  $\mathbf{y}_j$ .



La condition d'optimalité au premier ordre s'exprime alors :

$$\nabla \tilde{J}(\mathbf{w}) = (N-1)\mathbf{w} - \sum_{j=1}^k (\mathbf{H}_j \Delta \mathbf{X}_j^f)^T \mathbf{R}_j^{-1} (\mathbf{y}_j - \mathbf{H}_j \mathbf{x}_j^f - \mathbf{H}_j \Delta \mathbf{X}_j^f \mathbf{w}) = 0 \quad (5.31)$$

$$= (N-1)\mathbf{w} - \tilde{\Upsilon}^T \mathbf{R}^{-1} (\mathbf{d} - \tilde{\Upsilon} \mathbf{w}) \quad (5.32)$$

$$\begin{aligned} \implies \mathbf{w}^a &= \left[ (N-1)\mathbf{I}_N + \sum_{j=1}^k (\mathbf{H}_j \Delta \mathbf{X}_j^f)^T \mathbf{R}_j^{-1} (\mathbf{H}_j \Delta \mathbf{X}_j^f) \right]^{-1} \\ &\times \left( \sum_{j=1}^k (\mathbf{H}_j \Delta \mathbf{X}_j^f)^T \mathbf{R}_j^{-1} (\mathbf{y}_j - \mathbf{H}_j \mathbf{x}_j^f) \right) \end{aligned} \quad (5.33)$$

$$= \left[ (N-1)\mathbf{I}_N + \tilde{\Upsilon}^T \mathbf{R}^{-1} \tilde{\Upsilon} \right]^{-1} \tilde{\Upsilon}^T \mathbf{R}^{-1} \mathbf{d}. \quad (5.34)$$

Nous avons ainsi besoin de stocker chacun des vecteurs d'observation  $\mathbf{y}_i$ , en formant :

$$\mathbf{Y} = [\mathbf{y}_1^T, \dots, \mathbf{y}_k^T]^T \in \mathbb{R}^{kp}. \quad (5.35)$$

$\forall t_k \in [t_0, t_f]$ , nous avons aussi besoin de stocker l'ensemble projeté dans l'espace des observations en formant :

$$\mathbf{HE} = \left[ (\mathbf{H}_1 \mathbf{E}_1)^T, \dots, (\mathbf{H}_k \mathbf{E}_k)^T \right]^T \in \mathcal{M}_{p, kN}(\mathbb{R}). \quad (5.36)$$

A partir de  $\mathbf{HE}$ , il nous reste à calculer  $\overline{\mathbf{HE}}$ , l'ensemble des prévisions projetées dans l'espace des observations soit la "moyenne" de  $\mathbf{HE}$ . Ensuite nous pourrons former  $\mathbf{d}$  comme décrit par l'équation (5.30), étant donné que  $(\overline{\mathbf{HE}})_j = \mathbf{H}_j \mathbf{x}_j^f$ . Nous formerons également  $\tilde{\Upsilon}$  et  $\mathbf{d}$  à partir de  $\mathbf{Y}$  et  $\overline{\mathbf{HE}}$  en nous appuyant sur les équations (5.30) et (2.39).

En nous basant sur les équations (5.4), (5.26), et (5.30), nous pouvons désormais construire :

$$\mathbf{G}_i^e(\epsilon) = \left[ (N-1)\mathbf{I}_N + \frac{1}{\sigma_0^2} \tilde{\Upsilon}^T \mathbf{D}_i(\epsilon) \tilde{\Upsilon} \right]^{-1} \tilde{v}_i. \quad (5.37)$$

Lorsque  $\mathbf{R} = \sigma_0^2 \mathbf{I}_p$ , nous obtenons finalement pour formule de notre indicateur de sensibilité en asynchrone :

$$\delta \mathbf{x}'(\epsilon) = \Delta \mathbf{X}^f \left( -\frac{1}{\sigma_0^4} \mathbf{G}_i^e(\epsilon) \mathbf{G}_i^e(\epsilon)^T \left[ \tilde{\Upsilon}^T \mathbf{d} + (\epsilon - 1) d_i \tilde{v}_i \right] + \frac{d_i}{\sigma_0^2} \mathbf{G}_i^e(\epsilon) \right) \quad (5.38)$$

$$= \Delta \mathbf{X}^f \left( -\frac{1}{\sigma_0^4} \mathbf{G}_i^e(\epsilon) \mathbf{G}_i^e(\epsilon)^T \left[ \sum_{j=1}^k d_j \tilde{v}_j + (\epsilon - 1) d_i \tilde{v}_i \right] + \frac{d_i}{\sigma_0^2} \mathbf{G}_i^e(\epsilon) \right) \quad (5.39)$$

Par analogie avec les calculs précédents, lorsque  $\mathbf{R} = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$ , l'indicateur de sensibilité asynchrone aura pour formule :

$$\delta \mathbf{x}'(\epsilon) = \Delta \mathbf{X}^f \left( -\frac{1}{\sigma_i^2} \tilde{\mathbf{G}}_i^e(\epsilon) \tilde{\mathbf{G}}_i^e(\epsilon)^T \left[ \sum_{j=1}^k \frac{d_j \tilde{v}_j}{\sigma_j^2} + \left( \frac{\epsilon - 1}{\sigma_i^2} \right) d_i \tilde{v}_i \right] + \frac{d_i}{\sigma_i^2} \tilde{\mathbf{G}}_i^e(\epsilon) \right) \quad (5.40)$$

avec cette fois-ci :

$$\tilde{\mathbf{G}}_i^e(\epsilon) = \left[ (N-1)\mathbf{I}_N + \tilde{\mathbf{Y}}^T \left( \sum_{j=1}^p \frac{E_{jj}}{\sigma_j^2} + \left( \frac{\epsilon-1}{\sigma_i^2} \right) E_{ii} \right) \tilde{\mathbf{Y}} \right]^{-1} \tilde{v}_i \quad (5.41)$$

### 5.3.2 Interprétation de l'indicateur SOBS

L'indicateur SOBS retourne une matrice d'influence observation par observation pour chaque état du système, et ce pour chaque vecteur d'observation sélectionné. Il quantifie la variation de l'incrément d'analyse  $\delta\mathbf{x} = \mathbf{x}^a - \mathbf{x}^f$  lors de l'analyse.  $\mathbf{x}^f$  correspond au  $\mathbf{x}^b$  des calculs précédents, il est donc indépendant des nouvelles observations. La variation de  $\delta\mathbf{x}$  ne dépend alors que de  $\mathbf{y}$  à cet instant, ce qui nous donne les égalités :

$$\delta\mathbf{x}'(\epsilon) = \frac{d}{d\epsilon}(\delta\mathbf{x}(\epsilon)) = \frac{d}{d\epsilon}(\mathbf{x}_a(\epsilon) - \mathbf{x}_f) = \frac{d}{d\epsilon}(\mathbf{x}_a(\epsilon)). \quad (5.42)$$

Ici, le vecteur  $\delta\mathbf{x}(\epsilon) \in \mathbb{R}^n$  correspond à celui calculé dans la sous-section précédente (5.3.1) pour un indice d'observation  $i$  fixé, omis des notations par soucis de lisibilité. Nous considérons maintenant la possibilité de calculer  $\delta\mathbf{x}(\epsilon)$  pour  $k$  vecteurs d'observations différents, soit pour un total de  $kp$  observations. Dans ce cas,  $\delta\mathbf{x}(\epsilon)$  devient une matrice  $\in \mathcal{M}_{n, kp}(\mathbb{R})$ .

Ainsi, de façon plus formelle :

$$SOBS(\epsilon) = \frac{d}{d\epsilon}(\delta\mathbf{x}(\epsilon)) = \left( \frac{\partial x_i^a(\epsilon)}{\partial y_j} \right)_{1 \leq i \leq n, 1 \leq j \leq kp} = (s_{ij})_{1 \leq i \leq n, 1 \leq j \leq kp}, \quad k \in \mathbb{N}^* \quad (5.43)$$

avec  $k$  le nombre de vecteurs d'observation avec lesquels le calcul est fait. De manière générale,  $k = 1$ .

Lorsque l'on veut regarder la variation obtenue sur l'état en supprimant une observation, cela revient à faire  $\epsilon \rightarrow 0$  pour que l'erreur sur l'observation  $y_i$  soit infinie.

Il est difficile d'interpréter SOBS sous sa forme matricielle brute. Les variations de  $\delta\mathbf{x}$  relativement à chacune des observations sont signées et nous ne savons pas interpréter leur signe. Une quantité plus facilement interprétable est l'amplitude des variations de  $\delta\mathbf{x}$  relativement à chacune des observations, qui s'exprime grâce à SOBS.

Comme nous ne nous ne pouvons pas interpréter tel quel le signe des dérivées partielles qui composent SOBS et que nous voulons quantifier l'amplitude de la variation, une solution revient à regarder  $|SOBS(\epsilon)|$ . Cela permet de déterminer l'ampleur de l'effet de chaque observation  $y_i$  sur chaque dimension de l'état, et de sélectionner celles que l'on veut conserver ou supprimer via l'amplitude des  $\left| \frac{\partial x_i^a(\epsilon)}{\partial y_j} \right|$ . Nous décrivons par la suite plusieurs utilisation de *SOBS*.

#### Sélection de l'observation avec effet global maximal sur l'état

Pour sélectionner l'observation avec un effet global sur l'état le plus important, il suffit de regarder les normes  $\|\cdot\|_1$  ou  $\|\cdot\|_2$  de chaque ligne de  $|SOBS(\epsilon)|$ . Cela revient simplement à sommer les contributions de chaque observation à la variation de chaque dimension de

l'incrément d'analyse. Par exemple, pour un système avec  $n = p = 3$ , nous avons en première ligne de  $|SOBS(\epsilon)|$  :

$$|SOBS(\epsilon)| = \begin{pmatrix} \left| \frac{\partial x_1}{\partial y_1}(\epsilon) \right| & \left| \frac{\partial x_2}{\partial y_1}(\epsilon) \right| & \left| \frac{\partial x_3}{\partial y_1}(\epsilon) \right| \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \end{pmatrix}.$$

### Sélection de l'observation avec effet maximal sur une dimension de l'état

Ensuite, pour sélectionner l'observation avec le plus d'effet sur la variation d'une dimension particulière de l'état analysé, il suffit de prendre la norme  $\|\cdot\|_\infty$  de chaque colonne de la matrice  $|SOBS(\epsilon)|$  et de voir pour quel indice elle est atteinte. L'indice correspondra à celui de l'observation recherchée. En effet, toujours pour un système avec  $n = p = 3$ , la première colonne de  $|SOBS(\epsilon)|$  vaut :

$$|SOBS(\epsilon)| = \begin{pmatrix} \left| \frac{\partial x_1}{\partial y_1}(\epsilon) \right| & \dots & \dots \\ \left| \frac{\partial x_1}{\partial y_2}(\epsilon) \right| & \dots & \dots \\ \left| \frac{\partial x_1}{\partial y_3}(\epsilon) \right| & \dots & \dots \end{pmatrix}.$$

### Sélection d'un vecteur d'observation entier pour assimilation

Enfin, pour sélectionner un vecteur d'observation avec le plus (ou le moins) d'effet sur la variation de l'état analysé, il suffit, si  $SOBS(\epsilon)$  est calculée sur des observations groupées comme dans le cas de l'exemple illustré en figure (5.1), de découper  $SOBS(\epsilon)$  en sous-matrices de taille  $n \times p$  qui correspondent à l'effet des observations d'un seul vecteur, puis sommer chaque éléments en valeur absolue. Voici ce que nous aurions comme matrice pour un système aux dimensions  $n = p = 3$ , en le calculant selon trois vecteurs d'observation différents, pour chercher lequel éliminer :

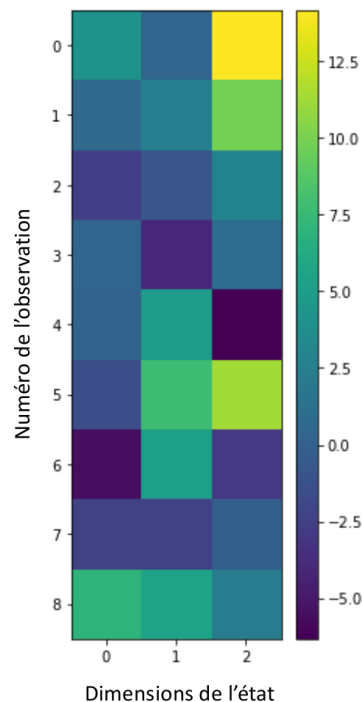


FIGURE 5.1 – Visualisation de  $SOBS(\epsilon)$  sur un système Lorenz 63 à 3 dimensions (colonnes) et 3 vecteurs de 3 observations (lignes par blocs de 3) de ce système.

Ces différents types d'interprétation ont une limite évidente : ils sont fortement sensibles aux valeurs aberrantes. Les valeurs aberrantes sont une sensibilité bien connue de l'EnKF. Cependant, nous recherchons les observations ayant les plus à même de fortement modifier l'analyse, dont font partie les valeurs aberrantes. Cet indicateur est donc adapté pour ce type de diagnostic.

Nous reviendrons au comportement de cet indicateur dans le chapitre sur les essais numériques (Chap. 7).

### Comparaisons du fonctionnement avec la référence "Hat Matrix"

Pour une comparaison plus générale, celle-ci sera faite en se servant de la formule (5.17) pour l'indicateur de sensibilité.

L'indicateur de sensibilité aux observations consiste à calculer la dérivée partielle de l'incrément d'analyse de l'EnKF. Pour connaître l'importance complète de  $\mathbf{y}$ , cela reviendrait à assimiler  $\mathbf{y}$  et à observer la variation obtenue entre  $\mathbf{x}^a$  et  $\mathbf{x}^f$ . C'est donc un indicateur qui fonctionne *a posteriori*, comme la Hat matrix. Il servira préférentiellement à faire du diagnostic de validité d'une dynamique d'assimilation, ou pour la calibration d'un réseau d'observation.

Cependant, au même titre que la Hat matrix, son utilisation dans le cas asynchrone est aussi possible comme mentionné en sous-section (5.3.1) en stockant de nouveau les bonnes variables au temps de chaque observation, comme pour l'AEnKF.

Après avoir décrit l'ensemble des indicateurs qui nous intéressent dans les deux précédents chapitres, nous allons maintenant étudier leur coût numérique pour voir la faisabilité de leur insertion dans l'algorithme de l'EnKF.



## 6 Coût opératoire des indicateurs

Dans tout ce chapitre, sauf mention explicite du contraire, les notations seront :

- $k \in \mathbb{N}$ ,
- $p \in \mathbb{N}$  pour la dimension du vecteur d'observations,
- $n \in \mathbb{N}$  la dimension de l'état,
- $N \ll p, n \in \mathbb{N}$  le nombre de membres de l'ensemble,
- $\mathcal{H} : \mathbb{R}^n \rightarrow \mathbb{R}^p$ ,  $\mathbf{B} \in \mathcal{M}_n(\mathbb{R})$ ,  $\mathbf{R} \in \mathcal{M}_p(\mathbb{R})$ ,
- $\mathbf{H} \in \mathcal{M}_{p,n}(\mathbb{R})$  le linéarisé de  $\mathcal{H}$ ,
- $m_1, m_2 \in \mathbb{R}^k$ ,
- $\mathbf{C}_1, \mathbf{C}_2 \in \mathcal{M}_k(\mathbb{R})$  2 matrices de covariance.

Ce chapitre a pour but de donner une formulation théorique du coût de calcul des différents indicateurs, en supposant que les différentes matrices nécessaires à leur calcul sont déjà construites. C'est pourquoi nous ne comptons pas le coût de formation de  $\mathbf{H}$ .

Les formules sont obtenues sans prendre en considération la structure des différentes matrices et opérateurs  $\mathbf{B}$ ,  $\mathbf{H}$ , ..., sauf parfois le caractère diagonal de  $\mathbf{R}$ . En pratique, et notamment dans le cadre du 4D-Var, beaucoup de ces coûts sont trop importants. Par exemple, la matrice  $\mathbf{B}$  n'est jamais construite (COURTIER, 1997 ; DERBER et BOUTTIER, 1999) voire approchée par une méthode ensembliste (KLEIST et al., 2014 ; KIM et KIM, 2019), plusieurs contournements et simplifications sont effectués pour rendre la méthode opérationnelle (COURTIER, THÉPAUT et HOLLINGSWORTH, 1994).

Ces formules sont données pour guider le processus de réduction des coûts lors de l'implémentation, en identifiant les opérations d'algèbre linéaire à optimiser.

### 6.1 Hat matrix

D'après l'équation (4.2),  $\hat{\mathbf{H}} = \mathbf{I}_p - (\mathbf{I}_p + \mathbf{R}^{-1/2} \mathbf{H} \mathbf{B} \mathbf{H}^T \mathbf{R}^{-1/2})^{-1}$ . Étant donné le fonctionnement de l'EnKF, la seule opération à calculer est l'inversion de matrice.

Par hypothèse,  $\mathbf{R}$  est diagonale définie-positive, calculer la racine carrée de son inverse revient simplement à inverser ses éléments diagonaux puis à prendre leur racine carrée, pour un coût négligeable.

Si l'opérateur d'observation  $\mathcal{H}$  est temporellement constant, son linéarisé  $\mathbf{H} \in \mathcal{M}_{p,n}(\mathbb{R})$  n'a besoin d'être calculé qu'une seule fois en début de fenêtre. Il ne reste donc qu'à former et inverser la matrice entre parenthèses. Comme  $\mathbf{R} = \sigma_0^2 \mathbf{I}_p$ , calculer son inverse ou sa racine carrée revient à multiplier l'identité par un scalaire, pour un coût négligeable.

- 1 produit matriciel :  $\mathbf{R}^{-1/2} \mathbf{H}$ , pour  $pn$  opérations,
- 1 produit matriciel :  $\mathbf{R}^{-1/2} \mathbf{H} \mathbf{B}$ , pour  $2pn^2$  opérations,
- 1 produit matriciel entre une  $p \times n$  et une  $n \times p$ , pour  $2p^2n$  opérations,

- inversion d'une matrice  $p \times p$ , soit  $2p^3$  opérations,

pour un coût total de :

$$\text{Coût}(\hat{\mathbf{H}}) = 2p^3 + 2(p^2n + pn^2) + pn. \quad (6.1)$$

### 6.1.1 Comparaison avec la matrice des sensibilités de (CARDINALI, 2013)

En reprenant les notations de (CARDINALI, 2013, pp. 4-6) pour la nomenclature matricielle, nous avons, après l'assimilation, l'estimation des observations par le modèle donnée par la relation :  $\hat{\mathbf{y}} = \mathbf{H}\mathbf{x}^a = \mathbf{H}\mathbf{K}\mathbf{y} + (\mathbf{I}_p - \mathbf{H}\mathbf{K})\mathbf{H}\mathbf{x}^b$  avec  $\mathbf{K} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}$ . Nous exprimerons cette fois la hat matrix ou matrice des sensibilités comme dans l'article, et nous la noterons  $\mathbf{S}$  pour éviter de confondre sa formule avec celle donnée par l'équation (4.2). Celle-ci s'exprime alors comme suit :

$$\mathbf{S} = \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{y}} = \mathbf{K}^T \mathbf{H}^T = \mathbf{R}^{-1} \mathbf{H} (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \quad (6.2)$$

Grâce à la formule de Sherman-Morrison-Woodbury (GOLUB et VAN LOAN, 2013), l'équation (6.2) se reformule :

$$\begin{aligned} \mathbf{S} &= \mathbf{R}^{-1} \mathbf{H} \left[ \mathbf{B} - \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R}^{-1}) \mathbf{H} \mathbf{B} \right] \mathbf{H}^T \\ &= \mathbf{R}^{-1} \mathbf{H} \mathbf{B} \mathbf{H}^T - \mathbf{R}^{-1} \mathbf{H} \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R}^{-1})^{-1} \mathbf{H} \mathbf{B} \mathbf{H}^T \end{aligned} \quad (6.3)$$

En posant  $\mathbf{V} = \mathbf{H} \mathbf{B} \mathbf{H}^T$ , après des calculs détaillés dans (CARDINALI, 2013, p. 6), la matrice des sensibilités s'écrit :

$$\mathbf{S} = (\mathbf{V} + \mathbf{R})^{-1} \mathbf{V} = (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H} \mathbf{B} \mathbf{H}^T \quad (6.4)$$

Les matrices  $\mathbf{H} \mathbf{B} \mathbf{H}^T$  et  $\mathbf{R}$  étant respectivement symétrique positive et symétrique définie-positive, leur somme est symétrique définie-positive. Nous calculons maintenant le nombre d'opérations nécessaires pour former  $\mathbf{S}$  qui est :

- $\mathbf{H} \mathbf{B} \mathbf{H}^T$  :  $2n^2p + 2p^2n$  opérations pour les 2 produits matriciels,
- $(\mathbf{V} + \mathbf{R})^{-1} \mathbf{V}$  :  $\frac{7}{3}p^3$  opérations pour une factorisation de Cholesky et la résolution de  $p$  systèmes triangulaires de taille  $p$ ,
- $\mathbf{V} + \mathbf{R}$  : coût négligeable.

Le nombre total d'opérations pour former  $\mathbf{S}$  revient ainsi à :

$$\text{Coût}(\mathbf{S}) = \frac{7}{3}p^3 + 2(n^2p + p^2n). \quad (6.5)$$

comparable au coût selon  $p$  de  $\hat{\mathbf{H}}$  pour l'EnKF (6.1).

L'équation (6.2) implique que  $\mathbf{S} = \mathbf{R}^{1/2} \hat{\mathbf{H}} \mathbf{R}^{-1/2}$ , les deux matrices sont semblables<sup>1</sup> donc  $Sp(\mathbf{S}) = Sp(\hat{\mathbf{H}})$  : l'information que l'on peut "extraire" de chacune est la même.

---

1.  $\mathbf{A}$  et  $\mathbf{B}$  semblables  $\iff \exists \mathbf{M} \in GL_n(\mathbb{R}), \mathbf{A} = \mathbf{M}\mathbf{B}\mathbf{M}^{-1}$ .

Nous préférons cependant la formulation de  $\hat{\mathbf{H}}$  qui correspond exactement à la traduction de notre problème d'assimilation de données pour l'EnKF (équation 4.1) de la notion de Hat matrix en statistiques présentée en sous-section (3.2.1).

Cela nous donne un coût global en  $O(p^3)$  pour les deux formules lorsque  $p \gg n$  et  $O(pn^2)$  lorsque  $p \ll n$ . Si  $n \simeq p$ , cela devient  $O(n^3)$ .

## 6.2 Indicateur de sensibilités : $\delta\mathbf{x}'(\epsilon)$ et SOBS

L'indicateur de sensibilités  $\delta\mathbf{x}'(\epsilon)$  et a fortiori SOBS se formulent différemment selon que la matrice  $\mathbf{R}$  est colinéaire à l'identité ou qu'elle soit simplement diagonale positive. Cependant, les deux formules nécessitent le même nombre d'opérations et le même type d'opérations pour être calculées. Le calcul de complexité se fera sur la formule la plus générale des deux,  $\mathbf{R}$  diagonale positive. Nous commencerons par calculer le coût de  $\delta\mathbf{x}'(\epsilon)$ , puis donnerons celui de SOBS, qui revient à calculer  $p$  fois  $\delta\mathbf{x}'(\epsilon)$ , une fois pour chaque observation  $\in \mathbf{y}$ .

### Coût pour la formulation (5.17)

Commençons par détailler le coût de formation de  $\tilde{\mathbf{g}}_i(\epsilon) = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}_\epsilon^{-1} \mathbf{H})^{-1} h_i$ , Cf. l'équation (5.7), qui intervient souvent dans la formulation de  $\delta\mathbf{x}'(\epsilon)$ . La matrice  $\mathbf{B}$  n'est en général pas directement construite car de taille trop importante (elle contiendrait  $10^{18}$  entrées dans le cadre de la météorologie). Pour cela, plusieurs techniques de modélisation sont utilisées, comme une décomposition spécifique en produits de plusieurs opérateurs qui encodent l'effet de  $\mathbf{B}$  sur un vecteur  $\mathbf{x}$  et calculent directement le produit  $\mathbf{B}\mathbf{x}$  dans le cadre du 4D-Var décrite chez (DERBER et BOUTTIER, 1999). Il y a aussi l'approximation à partir de la matrice des anomalies de l'ensemble utilisée en sous-section (5.3.1), comme décrite dans (EVENSEN, 1994; EVENSEN, 2003) qui s'applique dans le cadre du LETKF de (HUNT, KOSTELICH et SZUNYOGH, 2007) mais aussi des méthodes hybrides (KLEIST et al., 2014).

Le coût du calcul de  $\mathbf{B}^{-1}$  pour SOBS doit prendre en compte la formation de  $\mathbf{B}$  via la matrice d'anomalie puis sa pseudo-inversion. Nous utiliserons abusivement la notation  $\mathbf{B}^{-1}$  pour rester cohérent avec les notations de la description de SOBS. La matrice d'anomalie  $\Delta\mathbf{X}$  a été définie par l'équation (2.39). Son coût de formation vient du calcul de la moyenne de  $\mathbf{E} \in \mathcal{M}_{n,N}(\mathbb{R})$  que l'on soustrait ensuite à chacun de ses colonnes, soit  $2nN$  opérations. Ensuite,  $\mathbf{B} = (\Delta\mathbf{X})(\Delta\mathbf{X})^T$ , soit  $2Nn^2$  opérations pour le produit matriciel. Il reste le calcul du pseudo-inverse qui utilisera une SVD tronquée du fait que  $\text{rang}(\mathbf{B}) \leq N - 1 \ll n$  dans le cas général, pour un coût de l'ordre de  $O(Nn^2)$ . Il est à noter que le terme de plus haut degré du coût de la SVD s'applique à la dimension la plus importante de la matrice. Dans le cas où  $n > N$ , le coût sera de  $2Nn^2$  opérations, et si  $n \leq N$ , le coût sera de  $2N^2n$  opérations.

Ainsi, les opérations de construction de  $\tilde{\mathbf{g}}_i(\epsilon)$  coûtent chacune :

- $\mathbf{B}^{-1}$  :  $\begin{cases} n > N, & 2nN(1 + 2n) \\ n \leq N, & 2nN(1 + n + N) \end{cases}$  opérations pour sa formation et le calcul du pseudo-inverse,
- $\mathbf{H}^T \mathbf{R}_\epsilon^{-1} \mathbf{H}$  :  $2pn^2$  opérations pour les produits matriciels,  $\mathbf{R}$  diagonale,



—  $(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}_\epsilon^{-1} \mathbf{H})^{-1} h_i$  :  $2n^2$  opérations pour le produit matrice-vecteur,

pour un coût total de :

$$\text{Coût}(\tilde{\mathbf{g}}_i(\epsilon)) = \begin{cases} n > N, & 2(pn^2 + 2Nn^2 + n^2 + nN), \\ n \leq N, & 2(pn^2 + Nn^2 + n^2 + nN^2 + nN). \end{cases} \quad (6.6)$$

Ensuite, le coût de calcul de  $\tilde{\mathbf{g}}_i(\epsilon)^T \left[ \sum_{j \neq i} \frac{d_j h_j}{\sigma_j^2} \right] \tilde{\mathbf{g}}_i(\epsilon)$  qui constitue le membre de droite de (5.17), avec :

- $p$  sommes de vecteurs  $\in \mathbb{R}^n$  :  $pn$  opérations
- $\tilde{\mathbf{g}}_i(\epsilon)^T \left[ \sum_{j \neq i} \frac{d_j h_j}{\sigma_j^2} \right]$  :  $2n$  opérations pour le produit scalaire,
- $\tilde{\mathbf{g}}_i(\epsilon)^T \left[ \sum_{j \neq i} \frac{d_j h_j}{\sigma_j^2} \right] \tilde{\mathbf{g}}_i(\epsilon)$  :  $n$  opérations pour le produit réel-vecteur final,

soit un total de :

$$\text{Coût} \left( \tilde{\mathbf{g}}_i(\epsilon)^T \left[ \sum_{j \neq i} \frac{d_j h_j}{\sigma_j^2} \right] \tilde{\mathbf{g}}_i(\epsilon) \right) = (p + 3)n. \quad (6.7)$$

Le coût du membre de gauche de (5.17) vaut  $3n$  opérations, il s'agit d'un produit scalaire entre vecteurs de  $\mathbb{R}^p$  et d'un produit réel-vecteur.

Nous ne formons  $\tilde{\mathbf{g}}_i(\epsilon)$  qu'une seule fois. Cela nous fait donc un coût total pour l'indicateur des sensibilités de :

$$\text{Coût}(\delta \mathbf{x}'(\epsilon)) = \begin{cases} n > N, & 2(pn^2 + 2Nn^2 + n^2 + nN) + pn + 6n, \\ n \leq N, & 2(pn^2 + Nn^2 + n^2 + nN^2 + nN) + pn + 6n. \end{cases} \quad (6.8)$$

Nous remarquons un coût dominant en terme de dimension du vecteur d'état, tandis que la dimension du vecteur d'observations semble jouer un rôle moins important dans le coût total de l'indicateur. Lorsque  $n \leq N$ , le coût dominant provient du nombre de membres dans l'ensemble, ce qui n'est en général pas le cas dans les systèmes applicatifs.

Pour obtenir le coût total de SOBS, il faut effectuer  $p$  fois le calcul de  $\delta \mathbf{x}'(\epsilon)$ , ce qui donne dans les deux cas :

$$\text{Coût}(SOBS) = \begin{cases} n > N, & 2(p^2 n^2 + 2pNn^2 + pn^2 + pnN) + p^2 n + 6pn, \\ n \leq N, & 2(p^2 n^2 + pNn^2 + pn^2 + pnN^2 + pnN) + p^2 n + 6pn. \end{cases} \quad (6.9)$$

Asymptotiquement, nous avons un coût en  $O(p^2 n^2)$ . Il reste inchangé que  $n \ll p$  ou  $p \ll n$ , et vaut  $O(n^4)$  lorsque  $n \simeq p$ . Le coût de SOBS est donc de manière générale un ordre de grandeur plus élevé que celui de la Hat matrix, lorsqu'il est formulé dans le cadre d'un EnKF classique comme celui de (EVENSEN, 2003).

**Coût pour la formulation LETKF (5.28)**

Commençons par calculer le coût de  $\mathbf{g}_i^e(\epsilon) = [(N-1)\mathbf{I}_N + \Upsilon^T \mathbf{R}^{-1}(\epsilon) \Upsilon]^{-1} y_i$ , où  $\Upsilon = \mathbf{H} \Delta \mathbf{X}^f = [y_1^T, \dots, y_p^T]^T \in \mathcal{M}_{p,N}(\mathbb{R})$ .

Former  $y_i$  revient à sélectionner une colonne de  $\mathbf{Y}$ , pour un coût négligeable.

- $\mathbf{R}^{-1}$  :  $p$  opérations, inversion d'une matrice diagonale  $> 0$ ,
- $\Upsilon = \mathbf{H} \Delta \mathbf{X}^f$  :  $2pnN$  opérations pour le produit matriciel,
- $Upsilon^T \mathbf{R}^{-1} \Upsilon$  :  $2pN^2 + pN$  opérations pour les produits matriciels,  $\mathbf{R}^{-1}$  diagonale,
- calcul d'une matrice  $\in \mathcal{M}_N(\mathbb{R})$  symétrique définie-positive :  $\frac{1}{3}N^3$  opérations, factorisation de Cholesky,
- résolution de système linéaire finale :  $2N^3$  opérations.

Cela nous fait un total de :

$$\text{Coût}(\mathbf{g}_i^e(\epsilon)) = \frac{7}{3}N^3 + 2pN^2 + 2pnN + pN. \quad (6.10)$$

En nous basant sur la formule (5.28), et en remarquant que :

$$\tilde{\mathbf{g}}_i^e(\epsilon) \tilde{\mathbf{g}}_i^e(\epsilon)^T \left[ \sum_{j=1}^p \frac{d_j v_j}{\sigma_j^2} + \left( \frac{\epsilon-1}{\sigma_i^2} \right) d_i v_i \right] = \tilde{\mathbf{g}}_i^e(\epsilon)^T \left[ \sum_{j=1}^p \frac{d_j v_j}{\sigma_j^2} + \left( \frac{\epsilon-1}{\sigma_i^2} \right) d_i v_i \right] \tilde{\mathbf{g}}_i^e(\epsilon), \quad (6.11)$$

le décompte des opérations est le suivant :

- $\left[ \sum_{j=1}^p \frac{d_j v_j}{\sigma_j^2} + \left( \frac{\epsilon-1}{\sigma_i^2} \right) d_i v_i \right]$  :  $p$  produits réel-vecteur +  $p$  somme de vecteurs  $\in \mathbb{R}^N$ , soit  $2pN$  opérations,
- $\tilde{\mathbf{g}}_i^e(\epsilon)^T \left[ \sum_{j=1}^p \frac{d_j v_j}{\sigma_j^2} + \left( \frac{\epsilon-1}{\sigma_i^2} \right) d_i v_i \right]$  :  $2N$  opérations pour le produit scalaire,
- $\tilde{\mathbf{g}}_i^e(\epsilon)^T \left[ \sum_{j=1}^p \frac{d_j v_j}{\sigma_j^2} + \left( \frac{\epsilon-1}{\sigma_i^2} \right) d_i v_i \right] \tilde{\mathbf{g}}_i^e(\epsilon)$  :  $N$  opérations pour le produit réel-vecteur,
- une somme entre deux vecteurs  $\in \mathbb{R}^N$  :  $N$  opérations,
- un produit matrice vecteur :  $2nN$  opérations.

En prenant en compte le coût de formation de  $\mathbf{g}_i^e(\epsilon)$ , nous obtenons ainsi comme coût de formation :

$$\text{Coût}(\delta\mathbf{x}'(\epsilon)) = \frac{7}{3}N^3 + 2pN^2 + [2(p+n+pn) + 5]N + p. \quad (6.12)$$

Nous devons former  $p$  fois  $\delta\mathbf{x}'(\epsilon)$  pour calculer SOBS, ce qui nous donne comme coût final dans le cas ensembliste :

$$\text{Coût}(SOBS) = \frac{7}{3}pN^3 + 2p^2N^2 + [2(p^2n + pn + p^2) + 5p]N + p^2. \quad (6.13)$$

Nous remarquons cette fois que les termes de plus haut degré portent sur la taille de l'ensemble  $N$ , avec  $N \ll n, p$ , et sur la taille de l'espace des observations qui vérifie

en océanographie et en météorologie  $p \leq n$ . Nous retiendrons un coût asymptotique en  $O(Nnp^2)$ . Cette formule est bien plus avantageuse que celle donnée précédemment, car elle rend l'utilisation de SOBS abordable.

Ces calculs montrent que l'utilisation d'une formulation type LETKF est particulièrement avantageuse non seulement pour le coût en stockage de la matrice  $\mathbf{B}$  ainsi que son éventuelle singularité, pour les différentes opérations effectuées au sein de l'algorithme de filtrage, mais aussi pour l'utilisation des indicateurs *a posteriori*.

## 6.3 Distances probabilistes

### 6.3.1 Coût de formation des matrices de covariances

Il faut cependant y rajouter le coût de formation des matrices  $C_1 = \mathbf{HP}^f\mathbf{H}^T$  et  $C_2 = \mathbf{R} + \mathbf{HP}^f\mathbf{H}^T$ , qui seront utilisés lors des calculs. Il suffit de déterminer le coût de formation de  $\mathbf{HP}^f\mathbf{H}^T$  pour cela.  $\mathbf{HP}^f\mathbf{H}^T = \frac{(\mathbf{H}\Delta\mathbf{X})(\mathbf{H}\Delta\mathbf{X})^T}{N-1}$ , il suffit de calculer  $\mathbf{H}\Delta\mathbf{X}$  et de la multiplier par sa transposée pour avoir le coût de formation de la matrice.

- $\mathbf{H}\Delta\mathbf{X}$  : pour  $2pnN$  opérations,
- $(\mathbf{H}\Delta\mathbf{X})(\mathbf{H}\Delta\mathbf{X})^T$  : pour  $2p^2N$  opérations,

La multiplication par un scalaire et les sommes entre matrices sont négligées, pour un coût total de :

$$\text{Coût}(\mathbf{HP}^f\mathbf{H}^T) = 2(p+n)pN \quad (6.14)$$

qu'il faudra rajouter au coût de calcul de chaque distance probabiliste.

### 6.3.2 Distance de 2-Wasserstein

En se basant sur la formule (4.5), pour  $k \in \{p, n\}$ , son coût se décompose pour chaque éléments en :

- $\|m_1 - m_2\|_2^2$  :  $3k$  opérations, négligeable,
- les matrices étant symétriques, le calcul de leur racine carrée se fait par décomposition de Schur, elle-même calculée par une factorisation QR. Dans le cas des matrices symétriques, cela revient à un coût en  $4k^3/3 + O(k^2)$  opérations pour obtenir la racine carrée. On retiendra donc  $8k^3/3$  opérations pour les deux racines carrées du calcul,
- deux produits matriciels qui ajoutent  $4k^3$  opérations.

Ceci nous donne un coût total de :

$$\text{Coût}(dW) = \frac{20}{3}k^3 + 2(p+n)pN. \quad (6.15)$$

### 6.3.3 Divergence de Kullback-Leibler

En se basant sur la formule (4.9), son coût se décompose de la façon suivante :

- deux constructions de matrices symétriques définies-positives, soit  $\frac{2}{3}k^3$  opérations,
- deux sommes de matrices, soit  $2k^2$  opérations, négligeable
- deux produits matriciels, soit  $4k^3$  opérations,

- deux produits matrice-vecteur, soit  $4k^2$  opérations,
- une somme entre vecteurs, soit  $k$  opérations, négligeable,
- deux calculs de trace, soit  $2k$  opérations, négligeable.

Le tout pour un coût global de :

$$\text{Coût}(dKL) = \frac{14}{3}k^3 + 2(p+n)pN. \quad (6.16)$$

## 6.4 Comparaisons entre chaque indicateur

Dans le cas non-ensembliste, tous ces indicateurs ne peuvent être utilisés tel quel. Il faut pour cela fournir des efforts supplémentaires de réduction de leur coût et/ou en donner un calcul approché.

Etudions maintenant le terme dominant du coût de chaque indicateur en fonction des valeurs de  $n$  et  $p$ . Nous prendrons  $N = O(10^1)$ , comme c'est le cas dans plusieurs applications réelles (parfois même  $N \simeq 10^2$ ). Nous noterons  $SOBS_1$  le coût de SOBS de la formule (6.9) et  $SOBS_2$  celui de la formule (6.13). Voici un tableau récapitulatif des termes dominants du coût de chaque indicateur :

Indicateur	<b>B non-inversible</b>	<b>B inversible</b>
Hat matrix	$p^3, p^2n, pn^2$	$p^3, p^2n, pn^2$
SOBS	$p^2n^2$	$p^2Nn$
Wasserstein	$k^3$	$k^3$
Kullback-Leibler	$k^3$	$k^3$

TABLE 6.1 – Récapitulatif des coûts dominants de chaque indicateur,  $k = n$  ou  $p$ .

Nous avons ainsi les trois ordonnancements suivant :

- $n \simeq p$  :  $SOBS_2 < dW < \hat{\mathbf{H}} < dKL < SOBS_1$ ,
- $n \ll p$  :  $\hat{\mathbf{H}} < dKL < dW < SOBS_2 < SOBS_1$ ,
- $n \gg p$  :  $dW = dKL < SOBS_2 < \hat{\mathbf{H}} < SOBS_1$ .

Dans les trois cas de figure, nous remarquons que  $SOBS_1$  est à chaque fois l'indicateur le plus coûteux. Il sera donc à utiliser préférentiellement sur de petits systèmes, où son calcul implique un faible surcoût.

Lorsque  $n \gg p$ , ce qui est le cas en météorologie, nous remarquons que les distances probabilistes ont le même coût asymptotique et sont en-dessous de la référence qui est la Hat matrix. Elles sont de bons candidats avec lesquels expérimenter pour établir des stratégies d'assimilation, ou plus modestement dans un premier temps, avec lesquels établir de nouveaux diagnostics de qualité de la prévision.

L'intérêt d'utiliser la formulation LETKF, et de la réduction des coûts des matrices stockées apparaît nettement dans le classement de  $SOBS_2$  par rapport à  $SOBS_1$ , en particulier lorsque l'espace des états et l'espace des observations sont de taille équivalente.

Ces rapides comparaisons nous montrent que, pour gagner en capacité d'utilisation des indicateurs et surtout pour permettre des stratégies *a priori* de contrôle synchrone des assimilations, il est intéressant d'utiliser les distances probabilistes car elles sont les seules qui permettent vraiment de formuler de telles stratégies, donc aussi intéressant d'optimiser leur coût avec la réserve que rien jusqu'ici ne laisse penser que ce soit faisable. Dans le cas de l'utilisation *a posteriori* des distances probabilistes, une méthode pour effectuer des mises à jour de rang 1 sur les matrices intervenant dans le calcul des deux distances est décrite en annexe (section A.6).

Pour les stratégies asynchrones, il semble plus intéressant de se pencher sur la Hat matrix qui propose un coût plus faible du fait des faibles constantes multiplicatives devant chaque terme. Cependant, dans le cas *a posteriori*, les matrices d'erreur des distributions à comparer sont pour l'une une mise à jour de rang 1 de l'autre lorsqu'on supprime une observation, ce qui permettrait de répercuter cette mise à jour dans les éléments calculés comme les inverses dans l'expression des distances, pouvant donner un coût réduit intéressant lors du déroulement de l'algorithme d'assimilation.

## 6.5 Comparaison au coût d'une assimilation

### 6.5.1 Coût opératoire de l'assimilation version Evensen 2003

Nous reprenons ici les résultats de (EVENSEN, 2003, pp. 353-354) sur la construction des matrices et du gain de Kalman d'ensemble pour l'assimilation.

En notant  $N \in \mathbb{N}$  la taille de l'ensemble,  $p \in \mathbb{N}$  la taille du vecteur d'observations, et en considérant qu'il n'y pas de réduction de rang par pseudo-inversion par simplicité de calcul, le coût de formation des matrices et des produits matriciels nécessaires à construire le gain sans inversion revient à  $N^2(n + 3p) + pN$  opérations.

L'inversion matricielle se fait au travers d'une modélisation par des matrices de rang  $\leq N$ , et revient à un coût de  $pN$  opérations. Le produit matrice-vecteur donnant la prévision  $\mathbf{x}^f = \frac{1}{N} \mathbf{E} \mathbf{1}^T$ , où  $\mathbf{1} \in \mathbb{R}^N$  est un vecteur rempli de 1, nécessite  $2nN$  opérations. Nous avons donc un coût total de :

$$\text{Coût}(\text{Assim}) = (n + 3p)N^2 + 2(p + n)N \quad (6.17)$$

dans le cas où  $p \gg N$  ce qui est le plus probable dans le cadre industriel. Le coût est donc linéaire en  $n$  et  $p$  qui sont les dimensions les plus importantes du système.

Ceci est plus faible que le coût des autres indicateurs, qui possèdent chacun a minima un terme quadratique en  $n$  ou  $p$ . Il est donc essentiel de s'assurer que les stratégies faisant appel à leur utilisation n'induisent pas un surcoût trop pénalisant en comparaison à la stratégie de base d'assimiler chaque observation. Cela est possible si nous supprimons plusieurs assimilations grâce à un seul calcul d'indicateur.

Il faudra tout de même penser à une autre façon d'utiliser ces indicateurs de diagnostic, par exemple sur des points de contrôle régulier n'amenant pas un gros surcoût, permettant d'assurer la qualité de la prévision en temps réel si besoin. Cela implique des configurations de systèmes industriels préférentielles : celles où l'intégration possède un coût numérique en opération et/ou en temps de calcul supérieur d'un ordre de grandeur au coût d'assimilation de l'observation.

Nous voyons vers quels types de scénarios de test ces informations vont nous guider : des suppressions d'observations par tri selon leur importance sur au moins plusieurs cycles d'assimilation, du calibrage du système d'observation avec sélection de celles qui seront observées sur toute la durée de simulation, un mélange entre ces deux approches, et certainement d'autres possibilités auxquelles nous n'avons pas pensé. Une première idée des configurations des systèmes opérationnels sur lesquels ils se baseront sera donné dans les paragraphes suivant. De plus, le travail de réduction des coûts opératoires des indicateurs va permettre d'étoffer l'éventail disponible de stratégies de contrôle des assimilations.

### 6.5.2 Ordre de grandeur pour des applications classiques

Le domaine des géosciences est un grand utilisateur de l'assimilation de données, notamment pour ce qui est de la météorologie et de l'océanographie. Voici le descriptif de deux configurations opérationnels de tels systèmes.

D'après (DAESCU, 2012), les systèmes opérationnels en météorologie ont pour caractéristiques une dimension du vecteur d'état  $n$  de l'ordre de  $O(10^7 - 10^8)$  ( $O(10^8 - 10^9)$ , respectivement pour les modèles Arpège et Arôme de Météo-France) et des vecteurs d'observation de taille  $p$  de l'ordre de  $O(10^6 - 10^7)$  assimilés à chaque cycle ( $O(10^6)$  pour Arôme,  $O(10^7)$  pour Arpège). Les cycles sont typiquement répétés avec une fréquence de l'ordre de quelques heures (6 à 12h). Dans le cas de Météo France, deux systèmes différents sont utilisés pour des prévisions faites à des résolutions différentes. Le système Arpège pour la prévision globale effectue ces assimilations toutes les 6h qui permettent de réactualiser la prévision à 5 jours. Quant à lui, le système Arôme pour la prévision régionale a un cycle horaire d'assimilation. Il devient alors nécessaire de faire des hypothèses de simplification pour résoudre le problème d'assimilation de données et l'implémenter, comme la linéarisation de l'opérateur d'observation par exemple.

La durée d'une fenêtre d'assimilation se justifie avant tout par le temps nécessaire à la collecte des observations venant de nombreuses sources éparses puis par le temps de calcul à la fois de l'assimilation et de la prévision mise à jour qui en résulte, c'est pour cela que les observations ne sont assimilées à Météo France que toutes les 6h. Les observations arrivées pendant cet intervalle de temps sont "batchées" et assimilées de façon asynchrone en fin de fenêtre.

Dans le cas de l'océanographie, (BERTINO et al., 2012) donne un descriptif détaillé de la complexité du système opérationnel TOPAZ4 servant à la fois à la prévision à court-terme (10 jours) de l'état du nord de l'océan Atlantique et de l'Arctique dans sa totalité, et à la réanalyse sur 10 ans permettant une connaissance de l'état passé de l'océan pour avoir rétrospectivement une idée de son évolution. Au temps de l'écriture de l'article, 10 ans de réanalyse supplémentaires étaient en cours de calcul. La résolution horizontale du modèle est de 15km et chaque maille possède 28 couches hybrides en profondeur, pour une surface à modéliser totalisant environ 70 millions de km carrés. Le vecteur d'état est de taille  $O(10^8)$  et environ un demi-million d'observations sont assimilées lors des cycles hebdomadaires. L'EnKF est la méthode d'assimilation de données utilisées, la prévision se fait en intégrant les 100 membres pour environ 2500h CPU dont les calculs sont heureusement fortement parallélisables. L'analyse quant à elle ne représente que 20h CPU, ici aussi parallélisée, elle est dans ce cas "gratuite". Les 10 ans de réanalyse ont coûté environ

2 millions d'heures CPU, engendrant 10 To (!) de données utiles.

Ceci nous permet de voir la taille classique des systèmes rencontrés à la fois en opérationnel et dans la littérature en assimilation de données. Ces méthodologies pourraient tout aussi bien s'appliquer à plusieurs domaines moins "classiques", notamment chez EDF, c'est d'ailleurs un des enjeux de cette thèse que de donner une vue d'ensemble sur les applications métiers possibles de l'assimilation de données.

### 6.5.3 Ordre de grandeur des applications EDF

Voici le descriptif de quelques domaines d'applications chez EDF qui bénéficieraient de l'assimilation de données. Il s'agit ici de donner des détails techniques sur des applications possibles et de formuler éventuellement quelques scénarios d'utilisation, ainsi que de donner une rapide description du phénomène physique modélisé.

Par exemple, la surveillance de l'état du combustible au sein des cœurs de centrale nucléaire est un élément-clé de sûreté des centrales nucléaires. Le combustible (uranium et/ou plutonium en France) est encapsulé dans des gânes en alliage de zirconium qui forment les crayons combustibles, eux-mêmes regroupés en grappes appelées assemblages combustibles. Ces crayons peuvent supporter des températures de l'ordre de 2000°C avant de rompre, cependant des déformations structurelles peuvent survenir à partir de températures de l'ordre de 800°C, pouvant contribuer à des états accidentels (MARGUET, 2011). Il y a donc un enjeu majeur derrière cela. La simulation à quelques minutes de l'évolution du profil de température au sein des crayons prend environ 1h en temps de calcul, les opérateurs en charge de la surveillance utilisent donc des abaques pour avoir une connaissance de celle-ci à un instant et un état du cœur donné. Une amélioration semble possible à travers l'assimilation de données. Elle serait de permettre la simulation de l'évolution en continu de la température, en implémentant des algorithmes de filtrage permettant la prise en compte des observations indirectes de température au niveau des assemblages combustibles (mesures de puissance en particulier). Cela se ferait avec version simplifiée du modèle existant, cette fois en un temps opérationnel.

Dans le domaine de l'hydraulique, autour de la simulation des rejets d'eau chaude des centrales nucléaires, notamment dans le cas de rejets en bord de mer dans un estuaire, les simulations prennent de quelques dizaines de minutes jusqu'à plusieurs heures de temps de calcul par simulation unitaire pour une période de marée. En effet, de nombreux paramètres physiques comme la houle, la température de l'eau, la quantité et l'évolution des rejets d'eau chaude... sont à prendre en compte et cela devient vite complexe et coûteux. De plus, de nombreuses observations arrivent toutes les 15 minutes sous la forme de retours de marégraphes et de cartographies infrarouges via satellites/drones, à assimiler avec des échéances allant de 12h à 24h. Cela fait donc un nombre conséquent d'observations à prendre en compte. La simulation, conditionnée par l'assimilation, étant déjà coûteuse à calculer, il semble primordial de pouvoir déterminer quelles observations conserver.

Dans ce même cadre de réduction d'observations, l'hydrologie, avec l'étude du comportement d'un bassin hydrologique sous sollicitations de pluie et de température, a cette fois un coût de simulation très faible mais toujours des observations en nombre conséquent : il y a une couverture satellitaire prenant des clichés au sol d'une résolution variant

entre  $10 \times 10\text{m}$  à  $100 \times 100\text{m}$ , pour une surface totalisant  $1000\text{km}^2$ . Cela revient encore à se poser la question de la pertinence ainsi que de la façon de traiter ces observations en préalable à l'assimilation. Nous pouvons aussi mentionner la mécanique des structures où les simulations sont compliquées du fait d'un modèle complexe formé de maillages 3D avec des lois non-linéaires à paramètres variables en espace pour chaque point, menant à des simulations d'une durée variant de l'ordre de la minute, à l'heure voire au jour pour une structure complète et sur la durée de vie simulée. Dans ce cadre, l'assimilation de données pour améliorer la confiance dans la prévision par l'utilisation des covariances et la sélection d'observation qui dénotent des erreurs de mesure ou de réelles ruptures de comportement s'avère intéressante pour diminuer les coûts et/ou la complexité du modèle.

Une remarque importante est à faire sur ce genre de simulation. Le temps de calcul total comprend deux temps séparés, le temps de résolution du modèle qui correspond au calcul effectif de la simulation, et le temps de mise en œuvre de la résolution dans l'environnement de simulation industriel qui correspond au chargement des bibliothèques et du code sur le supercalculateur. Pour certains cas, la part de la résolution est faible, de l'ordre de 10% du temps de calcul complet, le reste provenant de la mise en œuvre. Prenons l'exemple d'une simulation d'ensemble de 100 membres à intégrer de  $t$  à  $t + dt$ , cela reviendra donc à 90% de temps de mise en œuvre et 10% de temps d'intégration/résolution, donc pour  $n$  du même ordre de grandeur que  $p$ , l'écart sera toujours similaire entre assimilation et simulation au niveau du coût numérique, dans ce cas il devient potentiellement plus avantageux d'assimiler moins souvent pour simuler plus longtemps et éviter de devoir être pénalisé par le temps de mise en œuvre. Le décompte d'opérations d'algèbre linéaire n'est donc parfois pas représentatif du temps réel de calcul sur système opérationnel, et cette réflexion doit nous guider dans la formulation de scénarios pour les essais numériques à venir.

Les domaines métiers sont donc très variés, à la fois en terme de phénomène physique décrit que de dimension du modèle et/ou des observations. L'utilisation des méthodes d'assimilation de données ensemblistes munies des indicateurs décrits précédemment intéresse particulièrement EDF. Ce descriptif montre les différentes configurations et ordres de grandeur qu'il faudra traiter dans l'application.

Nous allons maintenant passer aux essais numériques de l'EnKF instrumenté de nos indicateurs sur des systèmes emblématiques dont la configuration et les caractéristiques reprennent celles des systèmes décrits dans cette sous-section, dans le but d'étudier la faisabilité de l'application à de véritables systèmes industriels.





**Troisième partie**

**Application, conclusion et perspectives**



# 7 Essais numériques : de la théorie à sa vérification

## 7.1 Introduction : objectifs des essais numériques

Plusieurs objectifs sont à remplir dans ce chapitre.

Tout d’abord, nous voulons confirmer que les décisions prises grâce aux informations retournées par nos indicateurs nous permettent bien d’établir des stratégies d’assimilation pertinentes. Nous avons pour cela deux types d’indicateurs : les indicateurs *a priori*, dont le calcul se fait avant l’assimilation, et les indicateurs *a posteriori* dont le calcul se fait après l’assimilation.

Dans le cas des indicateurs *a posteriori*, la décision d’assimiler ou non une observation portera sur l’analyse suivante ou les cycles suivants, et la mesure de l’effet de cette prise de décision sur la qualité de l’analyse résultante se fera a minima au temps de la prochaine assimilation. Ceci peut servir à calibrer un système d’observation en sélectionnant seulement les observations les plus pertinentes, ce qui n’implique un surcoût que sur la simulation numérique du système d’assimilation de données opérationnel, surcoût numérique qui se compensera par une configuration plus pertinente des systèmes et instruments de mesure du système physique réel.

Pour les indicateurs *a priori*, la décision sera prise avant l’assimilation : leur objectif consiste surtout à supprimer une étape d’analyse. La mesure de l’effet de cette prise de décision peut se faire directement au temps de l’assimilation actuelle.

Cette mesure de l’effet de chaque décision se fera dans le cadre d’expériences jumelles, où nous avons aisément accès à une simulation de référence lors de laquelle toutes les observations sont assimilées.

Pour illustrer cela, nous utiliserons plusieurs configurations de systèmes dynamiques, qui correspondent chacune à un état spécifique du système, faisant ainsi apparaître par exemple une raideur marquée, une périodicité régulière, un chaos déterministe ou encore des situations réalistes pour certaines applications industrielles.

Nous utiliserons le système de Lorenz 63 (LORENZ, 1963), chaotique déterministe et de petite taille ainsi que représentatif du phénomène de convection cellulaire de l’atmosphère terrestre, l’oscillateur de Van der Pol (TSATSOS, 2006), connu pour sa raideur, sa pseudo-périodicité assez stable ainsi que son côté chaotique, et également le système Xénon (PONÇOT, 2008, Sec. 1.3, REUSS, 2003), qui modélise le phénomène physique d’empoisonnement au Xénon-135 des cœurs de réacteurs nucléaires au cours de leur utilisation. Il possède des régimes de fonctionnement très variés. Une présentation détaillée de ces systèmes est disponible en annexe (section A.7), ainsi que dans les références citées précédemment.

## 7.2 Indicateurs *a posteriori*

SOBS et la Hat matrix sont des indicateurs *a posteriori* qui se calculent une fois l'assimilation faite. Ils permettent de sélectionner des observations au sein du vecteur d'observation  $\mathbf{y}$  n'apportant que peu d'information dans l'estimation de l'état-vrai du système par l'état analysé  $\mathbf{x}_a$ .

Par exemple, cette sélection se fait en triant les valeurs diagonales de la Hat matrix par ordre croissant et en sélectionnant la ou les valeurs les plus faibles, afin de les supprimer du vecteur d'observation  $\mathbf{y}$ . Pour SOBS, celui-ci est calculé pour un seul vecteur d'observation, et nous regardons la valeur des normes de chaque ligne pour les classer par ordre croissant, et effectuer le même type de sélection que décrit pour la Hat matrix. Nous obtenons donc de même un vecteur de valeurs.

L'opérateur d'observation modifié  $\tilde{\mathbf{H}}$  ainsi que  $\tilde{\mathbf{R}}$  sont ensuite mis à jour en récupérant les indices des observations sélectionnées, pour n'assimiler que les observations à conserver, c'est-à-dire celles non contenues dans les listes retournées par SOBS ou la Hat matrix.

### Impact sur la précision de la prévision

Le premier effet de la suppression d'observations grâce à nos indicateurs que nous voulons mesurer est celui sur la RMSE d'analyse. Pour faire la comparaison, notre référence est une simulation où ne sont pas utilisés les indicateurs, donc où sont assimilées toutes les observations à chaque temps d'assimilation. Nous n'espérons pas obtenir une meilleure RMSE pour nos essais avec utilisation de nos indicateurs. Nous voulons voir si les décisions prises grâce à ceux-ci sont pertinentes, c'est-à-dire que les essais avec suppression d'observations ont une RMSE moyennée sur l'ensemble des analyses de la fenêtre proche de celle de la simulation de référence. Nous allons regarder cet effet pour plusieurs configurations de systèmes telles que décrites en introduction et en annexe (section A.7). Dans le cas de la configuration 1, qui correspond au système Lorenz 63, voici le tracé des différentes composantes du système avec les observations vues comme les moins importantes par SOBS :

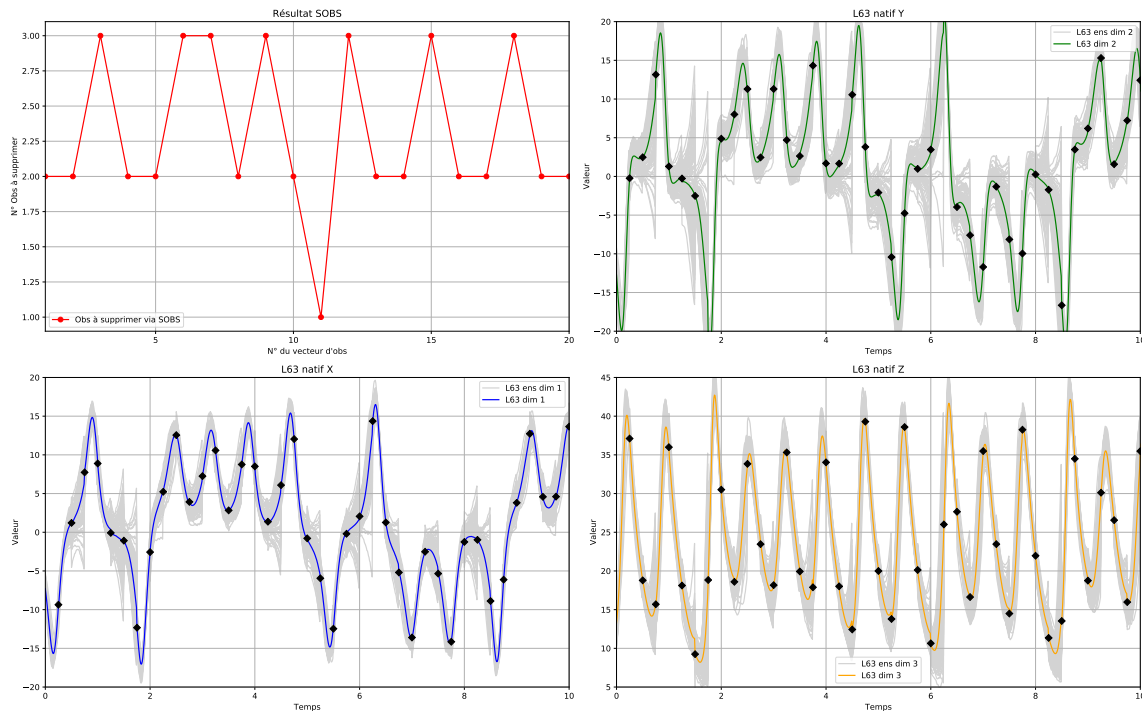
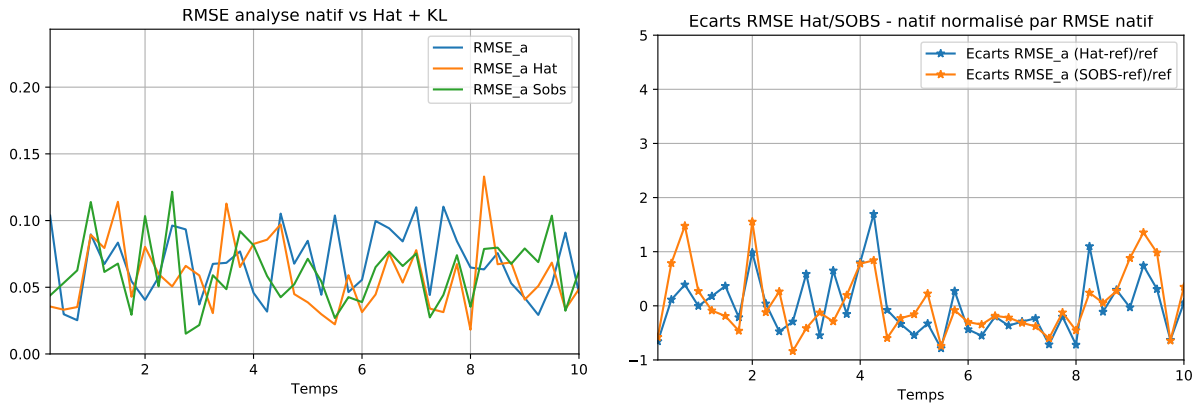


FIGURE 7.1 – Tracé de chaque dimension, membres de l'ensemble en gris, et observations de références sur la fenêtre en noir. Aucune suppression d'observation. La courbe en haut à gauche représente l'observation décrite comme négligeable par SOBS en fonction du vecteur d'observation tout au long de la fenêtre.

Ceci donne une idée de ce qui se déroule sur la fenêtre. Nous remarquons plusieurs transitions qui s'observent sur le tracé de la 1<sup>ère</sup> dimension (courbe bleue), lorsque la courbe traverse l'axe des abscisses. En notant  $x_1$  la 1<sup>ère</sup> dimension, ce phénomène se produit lorsque  $|x_1| \simeq 15$  comme observé dans (EVANS et al., 2004). Ces moments de transition sont la difficulté de cette configuration : ils sont chaotiques et doivent être suivis avec soin lors du processus d'assimilation de données. Ici,  $dt_{obs} = 0.25$ , correspondant à un régime moyennement non-linéaire pour le Lorenz 63. Il n'est pas en carence d'observations.

Voici maintenant l'évolution de la RMSE d'analyse pour chaque analyse sur la fenêtre  $[t_0, t_f]$ , ainsi que l'évolution des écarts relatifs en RMSE pour chaque indicateur, calculés

$$\text{par ecart} = \frac{RMSE_{indic} - RMSE_{ref}}{RMSE_{ref}} :$$



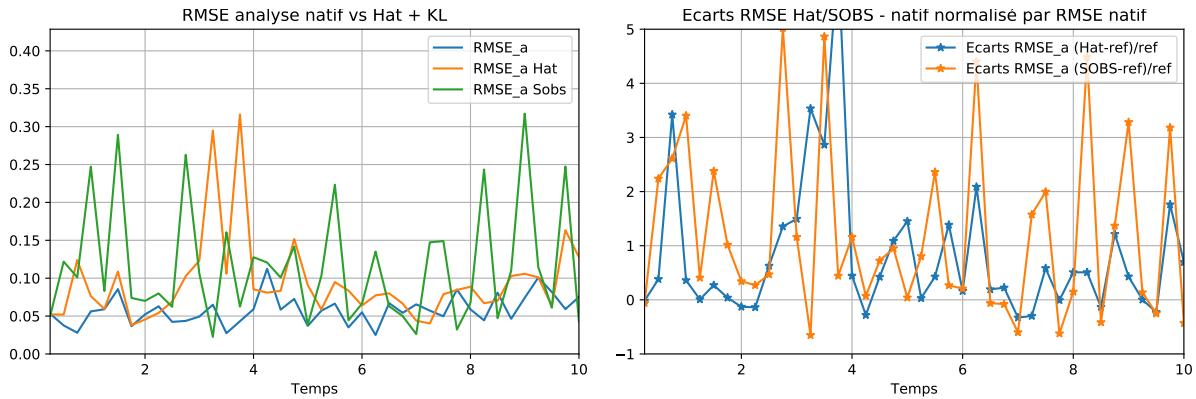
(A) RMSEs d'analyse.

(B) Ecarts relatifs.

FIGURE 7.2 – Évolution des RMSEs d'analyse (à gauche) sur la fenêtre pour trois essais : celui de référence (natif), celui avec sélection par Hat matrix (Hat), et celui avec sélection par SOBS (sobs). A droite, écarts relatifs en RMSE pour les essais avec indicateurs, comparativement à la RMSE de référence.  $dt_{obs} = 0.25$  par essai.

Nous observons parfois des disparités locales, avec de légères augmentations d'erreur en comparaison à la référence pour chacun des indicateurs. Il est intéressant de constater avec les écarts qu'il peut y avoir amélioration de la RMSE d'analyse par rapport à la référence comme c'est le cas entre les temps 6 et 8, cela traduit une dynamique temporairement en faveur de la réduction d'observations. Néanmoins, comme annoncé précédemment, ce n'est ni le but recherché, ni un résultat espéré. Nous apercevons plusieurs fois des moments où les deux indicateurs ont conduit à des décisions différentes, comme au temps  $t = 2.5$  où nous avons dégradation de la RMSE pour SOBS et amélioration de celle-ci pour Hat. Le comportement inverse s'observe au temps  $t = 3.5$ . Globalement, ces deux stratégies conduisent à des résultats similaires.

Dans le cas de la configuration 2, qui correspond au même système de Lorenz 63 avec une observation synthétique supplémentaire comme décrit en section (4.2), nous obtenons des résultats qui diffèrent au niveau de l'évolution des RMSEs d'analyse :



(A) RMSEs d'analyse.

(B) Écarts relatifs.

FIGURE 7.3 – RMSEs d'analyse (à gauche) sur la fenêtre pour trois essais : celui de référence (natif), celui avec sélection par Hat matrix (Hat), et celui avec sélection par SOBS (sobs). A droite, écarts relatifs en RMSE pour les essais avec indicateurs, comparativement à la RMSE de référence.

$$dt_{obs} = 0.25 \text{ par essai.}$$

Cette fois, les deux indicateurs sont au-dessus en terme de RMSE d'analyse sur la majorité de la fenêtre d'assimilation. De plus, nous voyons localement des variations plus importantes par rapport à la RMSE de référence, et principalement dans le sens d'une augmentation. Nous observons une dégradation des informations retournées par les indicateurs, donc de la pertinence des décisions qui en découlent.

Ces effets se constatent sur la moyenne des RMSEs d'analyse sur chaque fenêtre de 40 observations :

	Config. 1 : RMSE moyenne	Config. 2 : RMSE moyenne
Référence	0.06803	0.05801
Avec suppression (Hat matrix)	0.05896	0.09439

où une nette dégradation de la moyenne des RMSEs d'analyse est visible pour la configuration 2 avec observation synthétique, alors que nous observons une légère amélioration de celle-ci pour la configuration 1, le tout relativement aux valeurs de la simulation de référence.

Cependant, ces résultats ne sont pas statistiquement fiables. Pour obtenir un meilleur échantillon statistique ainsi qu'une idée de l'effet à plus long terme, voici les valeurs obtenues avec les mêmes configurations sur une fenêtre comportant 1000 étapes d'analyse :

	Config. 1 : RMSE moyenne	Config. 2 : RMSE moyenne
Référence	0.05955	0.05955
Avec suppression (Hat matrix)	0.06201	0.06201

Les valeurs sont identiques à la 5<sup>e</sup> décimale près. Sur le long terme, les différences entre les deux configurations se lissent. L'écart de RMSE entre la référence et les essais avec indicateurs est marginal dans les deux cas : c'est le résultat que nous attendions.



Il nous reste maintenant à voir dans quelle mesure ceci est vérifié pour d'autres configurations.

La configuration 3 correspond à l'oscillateur de Van der Pol avec en moyenne une à deux observations pour chaque changement de régime. Ceci revient à placer une observation au moment de chaque transition. Le paramètre de raideur du système qui détermine aussi la durée de sa période est faible, la raideur qui en résulte l'est donc aussi, et l'intégration a été adaptée pour. Voici le tracé du filtrage avec l'ensemble pour la simulation de référence. Seule la première dimension du système nous intéresse :

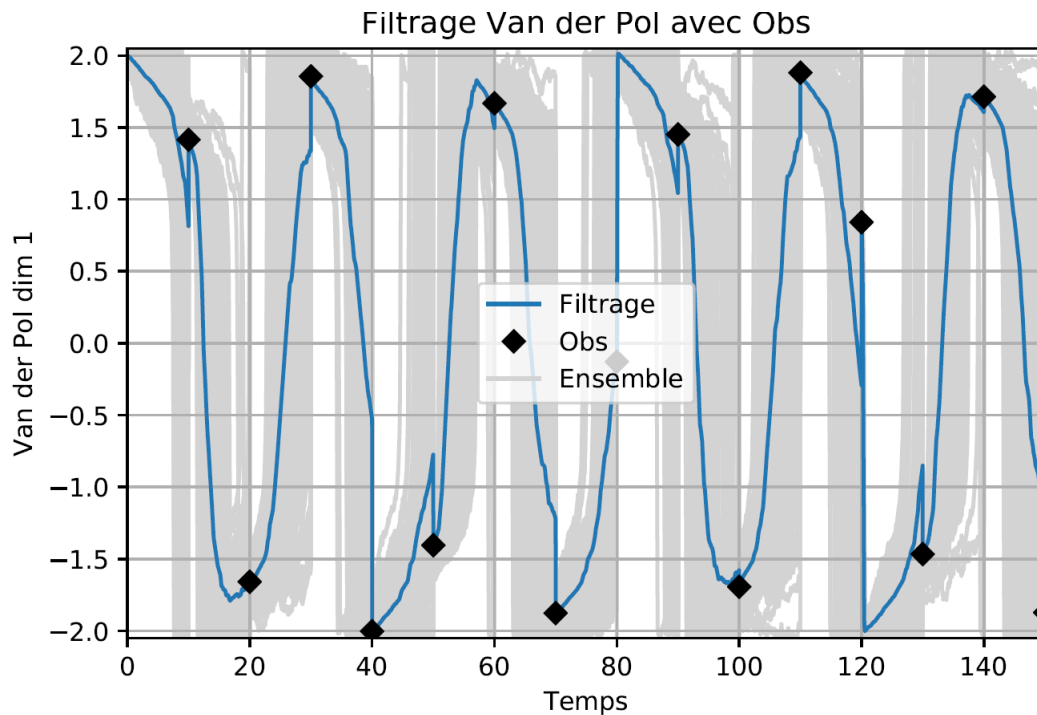
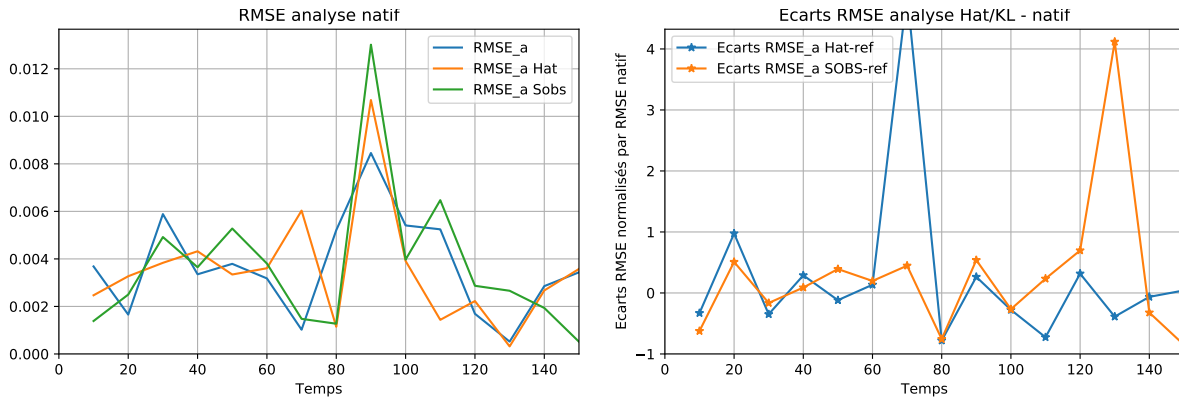


FIGURE 7.4 – Simulation de référence pour l'oscillateur de Van der Pol pour la configuration 3 (A.7.2), avec  $t_f = 150$ . Toutes les observations sont affichées, soit 15, avec  $dt_{obs} = 10$ . L'ensemble est affiché et montre la variété des différents membres.

Il est intéressant de remarquer la façon qu'a par moments l'ensemble d'être décalé au niveau de la transition et la manière dont les observations apportent une correction. Comme précédemment, voici l'évolution de la RMSE d'analyse pour chaque analyse effectuée sur la fenêtre :



(A) RMSEs d'analyse.

(B) Ecart relatifs.

FIGURE 7.5 – Van der Pol. Evolutions des RMSEs d'analyse sur la fenêtre pour trois essais : la référence (natif), la sélection par Hat matrix (Hat), et la sélection par SOBS (sobs). Écart relatifs à droite.  $dt_{obs} = 10$  par essai.

Nous constatons localement un comportement similaire à celui observé pour les deux précédentes configurations : des disparités à propos de l'effet sur la RMSE et moments d'amélioration ou de dégradation par rapport à la référence.

Pour observer l'évolution sur un cycle d'assimilation plus important, voici l'évolution sur une fenêtre correspondant à 150 assimilations qui nous montre que l'évolution de la RMSE est sujette à des variations plus marquées que pour le Lorenz 63. Ceci s'explique par le fait que le système soit moins observé, ce qui conduit inévitablement à des moments de fortes corrections lorsque le filtre rate une transition. Cela est dû à la nature chaotique des pseudo-périodes du Van der Pol. Nous pouvons observer ces corrections sur la figure (7.4).

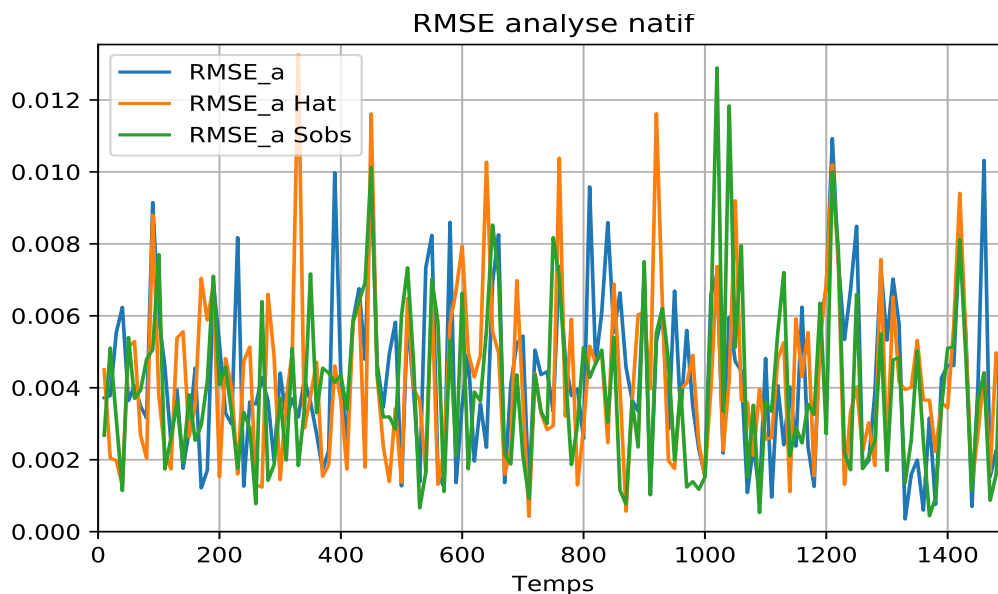


FIGURE 7.6 – Même tracé pour 150 assimilations que l'encadré (7.5a), soit  $t_f = 1500$ .  $dt_{obs} = 10$ .

De même que pour les configurations précédentes, nous n'observons pas d'écart d'ordre de grandeur sur l'évolution des RMSEs d'analyse entre les différents essais. La RMSE a

pour ordre de grandeur  $10^{-3}$ , et les variations se retrouvent à l'ordre de grandeur inférieur,  $10^{-4}$ . Voici les tables de valeurs moyennes de RMSEs :

	15 assim. : RMSE moyenne	150 assim. : RMSE moyenne
Référence	$3.693 \times 10^{-3}$	$4.345 \times 10^{-3}$
Hat matrix	$3.524 \times 10^{-3}$	$4.425 \times 10^{-3}$
SOBS	$3.715 \times 10^{-3}$	$4.354 \times 10^{-3}$

Ici aussi, la dégradation de RMSE est très faible, la pertinence de nos indicateurs est donc confirmée sur ces deux systèmes.

### Comparatif de comportement entre SOBS et la Hat matrix

Les résultats précédents montrent des disparités entre l'évolution en RMSE d'analyse selon que ce soit avec la Hat matrix que nous supprimons les observations ou bien SOBS. Les deux indicateurs ont donc un comportement différent. Cependant, les résultats obtenus avec l'un ou l'autre ne sont pas éloignés de façon significative, nous pouvons donc estimer qu'ils sont équivalents en terme de performances sur la RMSE d'analyse.

Voici les valeurs d'importance accordées aux observations par la Hat matrix et SOBS pour les configurations 1 ( $p = 3$ ) et 2 ( $p = 4$ , observation synthétique), sur une fenêtre correspondant à 100 assimilations :

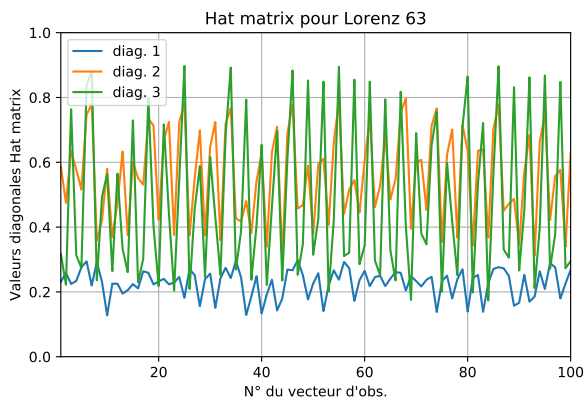


FIGURE 7.7 – Hat matrix pour configuration 1.

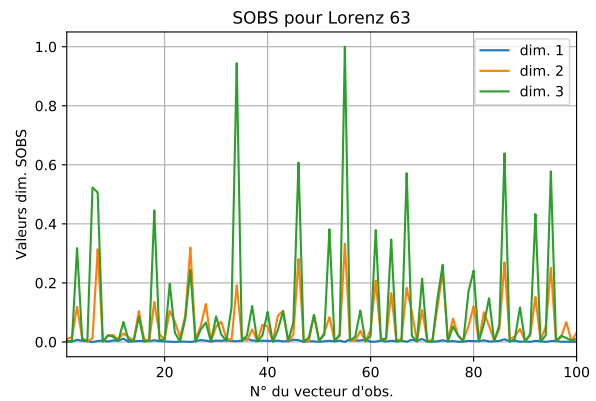


FIGURE 7.8 – SOBS pour configuration 1.

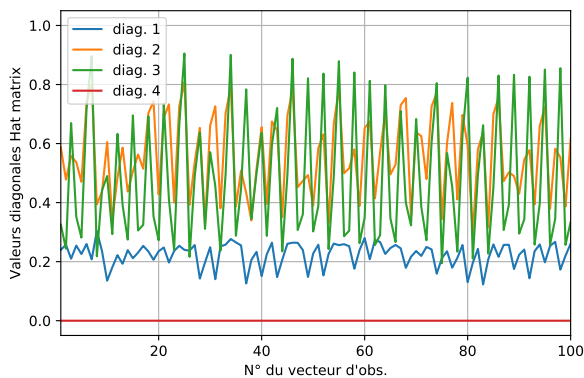


FIGURE 7.9 – Hat matrix pour configuration 2.

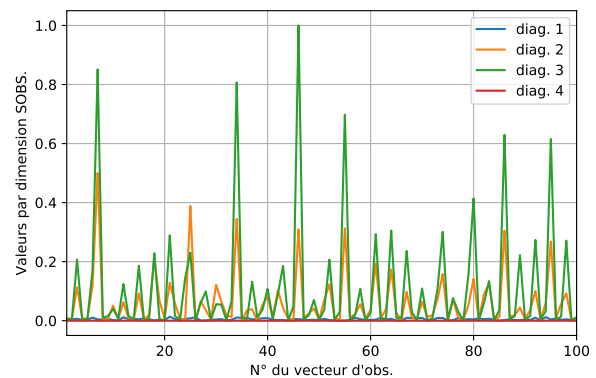


FIGURE 7.10 – SOBS pour configuration 2.

L'axe des abscisses correspond au numéro du vecteur d'observation sur la fenêtre d'assimilation, et l'axe des ordonnées à la valeur de l'indicateur, quantifiant l'importance de chaque observation.

Nous observons des dynamiques différentes pour chaque indicateur, bien qu'ils semblent hiérarchiser les observations de la même façon en terme d'importance.

La figure qui suit montre des nuages de points affichant la relation entre les valeurs prises par la Hat matrix et les valeurs ramenées sur  $[0, 1]$  prises par SOBS pour la configuration 2. Nous voyons effectivement que les valeurs de SOBS croissent avec celles de la Hat

matrix. Un modèle de régression linéaire est affiché. Il n'a pour vocation que d'illustrer la monotonie entre les deux indicateurs, pas celle d'être représentatif de la relation qui les lie. Le fait qu'ils concentrent tous deux leurs valeurs à 0 pour l'observation  $y_4$  (figure en coin inférieur droite) montre que SOBS la détecte aussi comme négligeable. Nous observons le même comportement pour la configuration 1.

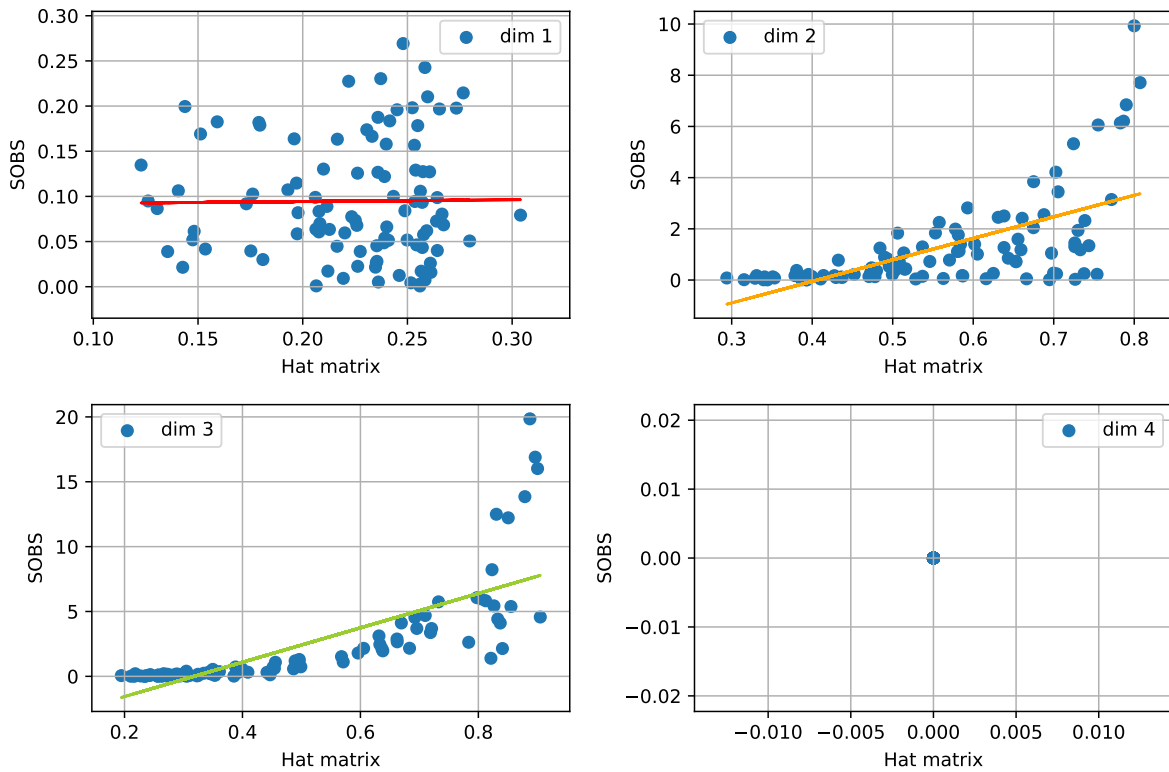


FIGURE 7.11 – Scatter-plots de SOBS versus Hat matrix pour la configuration 2. L'axe des abscisses correspond aux valeurs de la Hat matrix, et l'axe des ordonnées aux valeurs de SOBS ramenées à  $[0, 1]$ .

En reprenant les figures (7.7) et (7.8), nous observons des oscillations marquées chez SOBS qui accorde parfois une importance presque maximale à  $y_3$  lors d'une assimilation puis lui accorde une importance presque nulle lors de la suivante. Le même type d'oscillations s'observe chez la Hat matrix, à la différence que celles-ci sont moins extrêmes et centrées autour d'une valeur moyenne, comme montré sur la figure (7.12). Voici les valeurs moyennes obtenues pour SOBS et la Hat matrix sur la configuration 1 :

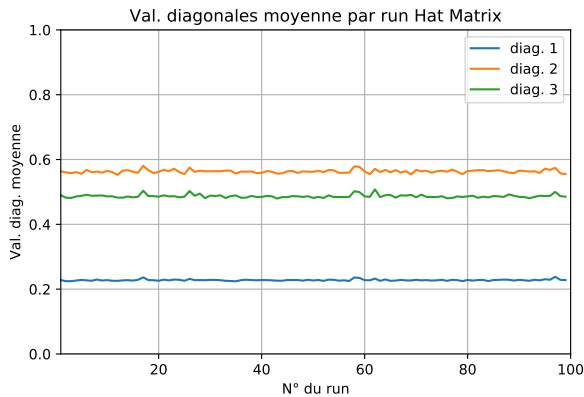


FIGURE 7.12 – Hat matrix moyennée pour 100 essais. Configuration 1.

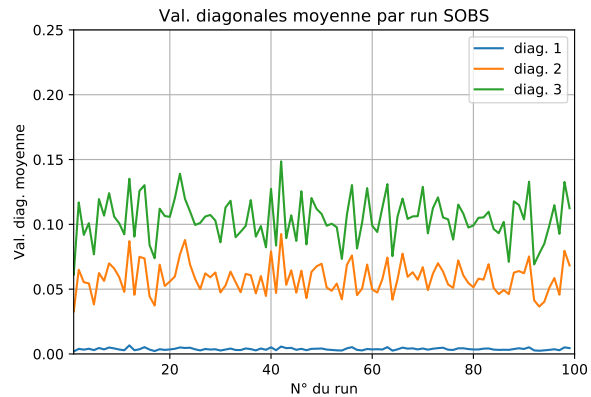


FIGURE 7.13 – SOBS moyenné pour 100 essais. Configuration 1.

En regardant l'évolution des valeurs moyennes la Hat matrix et de SOBS sur 100 simulations du Lorenz 63 en configurations similaires à la configuration 1 sauf pour la graine du générateur aléatoire, nous remarquons que les deux indicateurs ne priorisent pas la même observation (figures (7.12) et (7.13)). C'est la 3<sup>ème</sup> dimension du Lorenz 63 qui est la plus importante pour SOBS, alors que c'est la 2<sup>ème</sup> qui l'est pour la Hat matrix. Ceci explique les différences observées sur la RMSE d'analyse : les deux indicateurs ne sélectionnent parfois pas les mêmes observations au même moment, et auront tendance à hiérarchiser différemment les observations d'un même système.

Nous avons donc ici deux indicateurs qui retournent des informations différentes, tout en menant à un résultat équivalent en terme de RMSE d'analyse lorsque nous décidons de supprimer des observations en accord avec les informations qu'ils retournent. Cette différence peut s'expliquer du fait que la Hat matrix récupère ses informations uniquement des différentes matrices de covariances alors que SOBS s'intéresse directement à la sensibilité de l'état aux observations lors de l'analyse et inclue le résidu  $\mathbf{y} - \mathbf{H}\mathbf{x}$  dans son calcul. Du fait que les résultats obtenus avec chacun des deux soient équivalents, il semble possible de les utiliser de façon complémentaire.

### Distances probabilistes en diagnostic de vérification des observations sélectionnées

Outre l'utilisation a priori, une autre utilisation des distances probabilistes comme diagnostic de vérification des décisions effectuées grâce aux indicateurs Hat matrix et SOBS est possible. Pour cela, nous chercherons à comparer la distance entre les deux distributions de probabilités  $\mathcal{N}(\mathbf{x}^a, \mathbf{P}^a)$  et  $\mathcal{N}(\tilde{\mathbf{x}}^a, \tilde{\mathbf{P}}^a)$ . Elles correspondent respectivement à celle de la référence, et celle où l'on a supprimé une ou plusieurs observations grâce à la Hat matrix ou SOBS. Si nos indicateurs sélectionnent les bonnes observations, alors la distance entre les deux distributions devrait être faible voire quasi-nulle.

Nous ne pouvons pas juger directement de l'ampleur de l'écart entre les deux distributions pour chaque distance et dire s'il est faible sans avoir de jauge à laquelle se référer. Nous définissons comme jauge la distance entre les distributions de la prévision  $p_f = \mathcal{N}(\mathbf{x}^f, \mathbf{P}^f)$  et de l'analyse de référence  $\mathcal{N}(\mathbf{x}^a, \mathbf{P}^a)$ , calculées chacune au temps  $t = t_k$ .

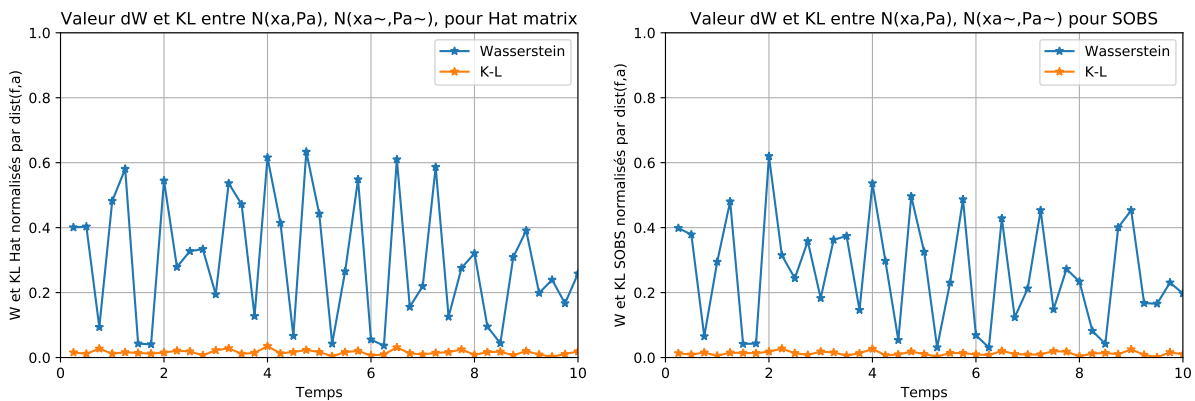
Nous regardons ensuite quel pourcentage de cette jauge vaut la distance entre

$p_1 = \mathcal{N}(\mathbf{x}^a, \mathbf{P}^a)$  et  $p_2 = \mathcal{N}(\tilde{\mathbf{x}}^a, \tilde{\mathbf{P}}^a)$  : plus le pourcentage est faible, plus nos indicateurs sont performants. Ceci nous fournit donc un diagnostic nouveau sur nos propres indicateurs. De plus, couplé au diagnostic de RMSE fait en sous-section (7.2), les distances probabilistes nous servent de "diagnostic de diagnostic" pour obtenir une vérification/quantification supplémentaire de l'effet de la suppression d'observations.

Un autre choix possible pour une jauge est la distance entre la distribution de la prévision projetée dans l'espace des observations et la distribution de l'observation. Ceci serait fait dans un cas où nous avons particulièrement confiance en nos observations.

Discutons notre choix de la distance entre la prévision et l'analyse en tant que jauge. Nous cherchons avec nos simulations à avoir une prévision la plus précise possible, qui doit donc faire en sorte d'avoir des statistiques d'erreurs aussi proches que possible de celles de l'analyse, qui est notre "meilleure" connaissance possible sur l'état du système. L'écart entre ces statistiques peut s'exprimer grâce à une distance probabiliste qui se base sur les moments des deux distributions, comme c'est le cas de Wasserstein et Kullback-Leibler. Nous cherchons aussi à réduire cet écart en nous servant de nos indicateurs pour guider l'analyse, et nous voulons nous assurer que les statistiques d'erreur de notre analyse partielle résultante restent meilleures que celles de notre prévision. C'est pourquoi nous regardons la quantité  $\frac{d(p_1, p_2)}{d(p_1, p_f)}$ , où  $d(p_1, p_f)$  est notre terme de jauge, et voulons nous assurer que celle-ci soit toujours  $< 1$ .

Voici les résultats obtenus avec cette approche pour la configuration 1 du Lorenz 63 et le Van der Pol (configuration 3 sur un temps plus long  $t_f = 1500$ ) :



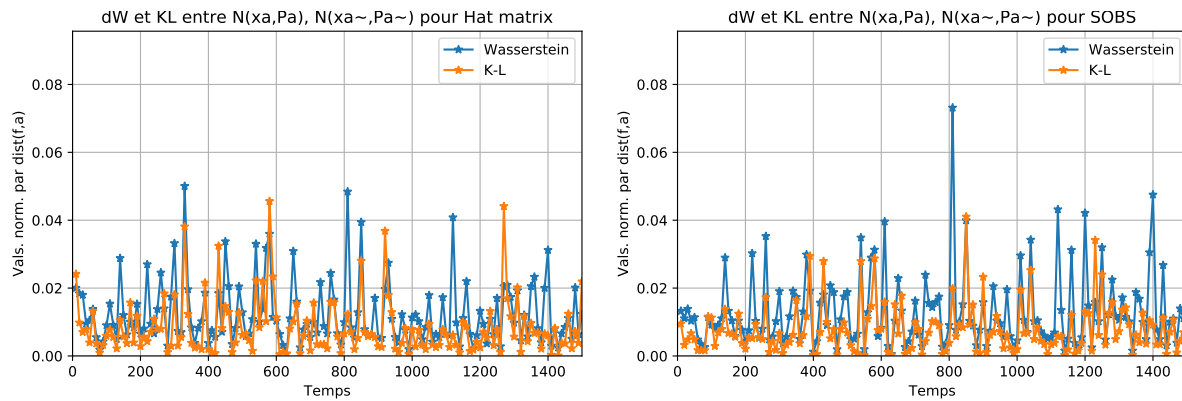
(A) Lorenz 63 avec Hat matrix.

(B) Lorenz 63 avec SOBS.

FIGURE 7.14 – Lorenz 63. Essais avec diagnostics a posteriori par distances probabilistes entre les distributions issues de l'analyse de référence et celle de l'analyse avec stratégies d'assimilation par les deux indicateurs.

Les abscisses représentent le temps, les ordonnées la valeur du diagnostic  $\frac{d(p_1, p_2)}{d(p_1, p_f)}$ .

Nous remarquons des variations plus importantes pour la distance de Wasserstein que pour Kullback-Leibler dans le cas de la configuration 1 du Lorenz 63, et notamment un ordre de grandeur d'écart entre les valeurs de chacune des deux. Ceci n'est pas le cas pour le Van der Pol en configuration 3, où les ordres de grandeur sont les mêmes, la différence



(A) Lorenz 63 avec Hat matrix.

(B) Lorenz 63 avec SOBS.

FIGURE 7.15 – Van der Pol. Essais avec diagnostics a posteriori par distances probabilistes entre les distributions issues de l’analyse de référence et celle de l’analyse avec stratégies d’assimilation par les deux indicateurs. Les abscisses représentent le temps, les ordonnées la valeur du diagnostic

$$\frac{d(p_1, p_2)}{d(p_1, p_f)}$$

la plus notable est celle des pics de Wasserstein qui sont parfois plus prononcés que ceux de Kullback-Leibler.

Une explication possible est la suivante : l’amplitude du Van der Pol est moindre que celle du Lorenz 63 d’où un ordre de grandeur d’écart entre les deux distances pour le Lorenz 63 du fait que le terme en  $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2$  de la distance de Wasserstein sera plus important que pour le Van der Pol. Cela peut aussi expliquer les pics plus prononcés pour la distance de Wasserstein chez le Van der Pol, le terme de résidus de Wasserstein n’étant pas pondéré par les inverses des matrices de covariance comme c’est le cas pour Kullback-Leibler.

En revanche, la comparaison entre les valeurs des deux distances pour la configuration 1 montre qu’elles possèdent la même monotonie. Celle-ci se fait avec un scatter-plot entre les valeurs obtenues par le diagnostic précédent pour la Hat matrix, et celles obtenues par le diagnostic précédent pour SOBS. Un modèle de régression linéaire est affiché pour améliorer l’illustration :

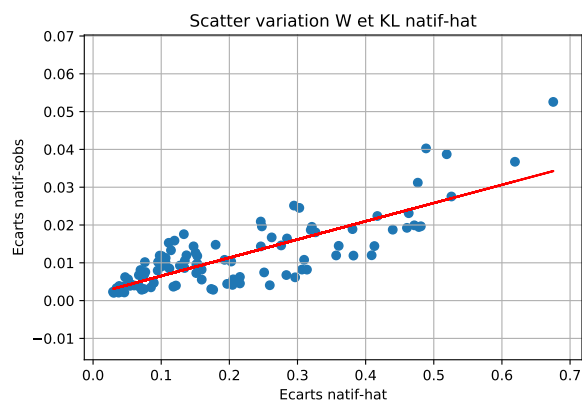


FIGURE 7.16 – Evolutions comparées pour Hat, config. 1.

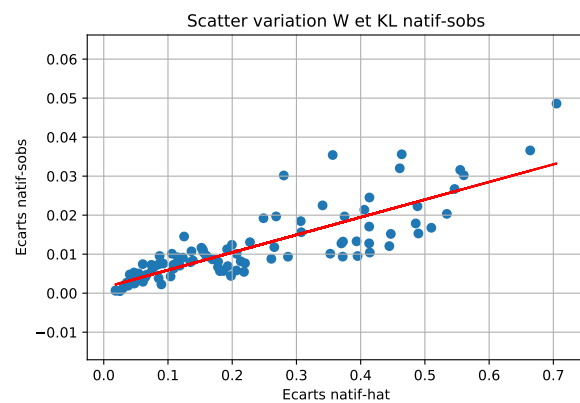


FIGURE 7.17 – Evolutions comparées pour SOBS, config. 1.



Nous récupérons donc des informations équivalentes de la part de l'une ou de l'autre sur la variation des statistiques d'erreurs de chaque distribution. Les résultats obtenus grâce à une des distances sont transposables à l'autre.

## 7.3 Distances probabilistes pour établissement de stratégies d'assimilation

Une stratégie d'assimilation peut définir trois choses :

1. la façon de contrôler les instants auxquels sont assimilées les observations,
2. la façon de contrôler le nombre d'observations assimilées à chaque assimilation,
3. un mélange des deux.

L'utilisation des indicateurs pour les formuler ou du moins les mettre en place peut varier selon les stratégies. Par exemple, le calcul des indicateurs peut se faire soit à chaque arrivée d'observation ou une observation sur deux, etc.

Nous pouvons donc dire qu'une stratégie d'assimilation consiste en une façon particulière d'utiliser les indicateurs pour contrôler la dynamique des instants d'assimilation et/ou le nombre d'observations assimilées, en comparaison à la façon classique qui consiste à assimiler l'intégralité du vecteur d'observations disponible  $\mathbf{y}$  à chaque temps d'assimilation.

Les stratégies d'assimilations explorées en premier sont celles où nous choisissons les moments d'assimilation (cas 1), c'est-à-dire celles où nous décidons ou non d'assimiler une observation lorsqu'elle arrive à l'intérieur de notre fenêtre. Nous entendons par fenêtre d'assimilation un intervalle de temps  $[t_0, t_f]$  sur lequel se déroule le processus d'assimilation, avec plusieurs temps d'assimilations à l'intérieur séparés de  $dt_{obs}$ .

Pour cela, nous nous basons sur l'hypothèse que lorsque la distance entre les distributions de la prévision et de l'observation dépasse un certain seuil fixé par l'utilisateur, l'information contenue dans chacune diffère suffisamment pour qu'il soit intéressant de corriger la prévision avec l'observation en l'assimilant. Cela implique d'avoir une variance d'erreur d'observation  $\sigma_{obs}^2$  faible ou du même ordre de grandeur que celle de prévision  $\sigma_f^2$ . Nous savons qu'en réduisant le nombre d'assimilation sur une fenêtre de longueur fixe, l'effet des non-linéarités va s'amplifier ce qui va mener à une augmentation de la variance de l'ensemble donc à une augmentation de la distance entre la prévision et l'observation au fur et à mesure que le temps sans assimilation augmente lui aussi. Nous cherchons donc une distance probabiliste qui possède ce type de dynamique d'évolution.

L'écart que nous calculons est celui entre les deux distributions  $\mu_1 = \mathcal{N}(\overline{\mathbf{H}\mathbf{x}}, \mathbf{H}\mathbf{P}^f\mathbf{H}^T)$  de la prévision et  $\mu_2 = \mathcal{N}(\mathbf{y}, \mathbf{R} + \mathbf{H}\mathbf{P}^f\mathbf{H}^T)$  de l'observation. Le choix des deux matrices de covariance s'explique du fait qu'elles prennent chacune en compte la dynamique d'évolution des distributions. La matrice de covariance choisie pour les observations fait le lien avec les statistiques d'erreurs qui constituent le gain de Kalman  $\mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{P}^f\mathbf{H}^T)^{-1}$ . De plus, cela conserve le lien entre les dynamiques du modèle et des observations exprimé par  $\mathbf{y} = \mathbf{H}\mathbf{x}^f + \epsilon$ .

### 7.3.1 Spécificités des distances probabilistes : Wasserstein et Kullback-Leibler

Au sens des distributions de probabilités, une distribution gaussienne peut s'interpréter comme une information de confiance que l'on peut avoir en son premier moment (la moyenne), qui se traduit par sa matrice de covariance (second moment). Ces deux distances ont une similitude : elles peuvent être utilisées *a priori* pour déterminer l'écart

informatif de l'observation relativement à la prévision, au sens des distributions de probabilités, mais aussi en diagnostics *a posteriori*, comme décrit en sous-section (7.2). Leur grande différence par rapport aux autres indicateurs est qu'elles portent sur les distributions de probabilités plutôt que directement sur les réalisations de ces distributions comme dans le cas de la Hat matrix ou de l'indicateur de sensibilités.

### Distance de 2-Wasserstein

La distance de 2-Wasserstein (de Wasserstein) compare les deux premiers moments de deux distributions de probabilités au sein d'un même espace probabilisé. En reprenant la formule (4.5) :

$$W_2(\mu_1, \mu_2)^2 = \|\mathbf{m}_1 - \mathbf{m}_2\|_2^2 + \text{tr} \left( \mathbf{C}_1 + \mathbf{C}_2 - 2(\mathbf{C}_2^{1/2} \mathbf{C}_1 \mathbf{C}_2^{1/2})^{1/2} \right) \quad (7.1)$$

Elle permet de déterminer l'éloignement de l'une à l'autre relativement à leurs deux premiers moments, ce qui nous intéresse tout particulièrement dans le cas de l'EnKF, du fait de l'hypothèse de gaussianité des distributions manipulées par la méthode. En effet, dans ce cas une distance de Wasserstein égale à zéro signifie que les deux distributions sont strictement identiques, du fait que les distributions gaussiennes soient définies de manière unique par leurs deux premiers moments.

Dans le cas de l'assimilation de données, en utilisation *a priori*, le calcul sera fait entre les distributions  $\mathcal{N}(\mathbf{m}_1 = \mathbf{H}\mathbf{x}_f, \mathbf{C}_1 = \mathbf{H}\mathbf{P}_f\mathbf{H}^T)$  et  $\mathcal{N}(\mathbf{m}_2 = \mathbf{y}, \mathbf{C}_2 = \mathbf{R} + \mathbf{H}\mathbf{P}_f\mathbf{H}^T)$ , qui correspondent respectivement à celles de l'état transporté dans l'espace des observations et de l'observation courante. Nous choisissons en particulier ces deux distributions gaussiennes car elles ne nécessitent pas d'avoir assimilé l'observation  $\mathbf{y}$  pour être construites, et expriment la dynamique d'évolution de la prévision et de l'observation, respectivement.

### Divergence de Kullback-Leibler

La divergence de Kullback-Leibler ne possède pas toutes les propriétés d'une distance à proprement parler. Par simplicité et pour la clarté de la présentation, elle sera abusivement qualifiée de "distance" durant le chapitre. En reprenant la formule symétrique (4.10) :

$$KL(\mu_1, \mu_2) = \frac{1}{2} [(\mathbf{m}_2 - \mathbf{m}_1)^T (\mathbf{C}_2^{-1} + \mathbf{C}_1^{-1})(\mathbf{m}_2 - \mathbf{m}_1) + \text{tr}(\mathbf{C}_2^{-1} \mathbf{C}_1 + \mathbf{C}_1^{-1} \mathbf{C}_2) - 2n]. \quad (7.2)$$

Elle s'interprète d'une façon différente de la distance de Wasserstein car il s'agit plutôt d'une comparaison de différence en quantité d'information contenue entre deux distributions  $\mu_1$  et  $\mu_2$ .

Dans le cas de l'assimilation de données, les mêmes distributions que Wasserstein seront utilisées pour le calcul.

### 7.3.2 Seuil de sélection/rejet d'observation

La sélection d'observation pose la question du seuil en distance au delà duquel nous estimons qu'il faille assimiler. L'écart entre les distributions de l'observation et de la prévision

va dépendre de beaucoup de paramètres : l'erreur commise sur la prévision, l'erreur de mesure, la finesse de l'intégration temporelle, la qualité de l'approximation linéaire de l'opérateur d'observation, ...

Nous avons décidé de l'estimer grâce à des essais numériques car il n'existe pas d'expression explicite permettant de relier ce seuil et une erreur cible. Le comportement attendu est le suivant : observation d'une monotonie entre la variation de la valeur du seuil et la variation de la moyenne des RMSEs de prévision et/ou la valeur moyenne de la distance probabiliste. Nous nous attendons à ce que moins nous assimilons d'observations sur la fenêtre, plus l'erreur moyenne de prévision augmente, idem pour la distance probabiliste moyenne. La détermination de ce seuil implique un surcoût important, car il faut effectuer plusieurs simulations complètes pour obtenir l'intervalle de valeurs utiles. Il en faudra d'autant plus lorsque nous en désirons une description plus fine. Ce surcoût peut devenir prohibitif en comparaison d'une utilisation des indicateurs *a posteriori*.

En notant notre seuil  $\rho > 0$ , les vecteurs d'observations seront sélectionnés pour l'assimilation lorsque  $d_W(\mu_1, \mu_2) > \rho$  pour la distance de Wasserstein, et  $d_{KL}(\mu_1, \mu_2) < \rho$  pour la divergence de Kullback-Leibler. Le choix des deux critères est expliqué en sous-section (7.3.3).

### Recherche par essais numériques

La recherche numérique de l'intervalle utile de valeurs du seuil s'est faite en le choisissant parmi une liste croissante de valeurs, basées sur les distances moyennes calculées sur une fenêtre d'assimilation sans suppression d'observation.

Ensuite, sur la même fenêtre, nous avons regardé la diminution du nombre d'assimilations effectuées en fonction de la valeur du seuil, pour la distance de Wasserstein. Voici le comportement observé sur plusieurs systèmes :

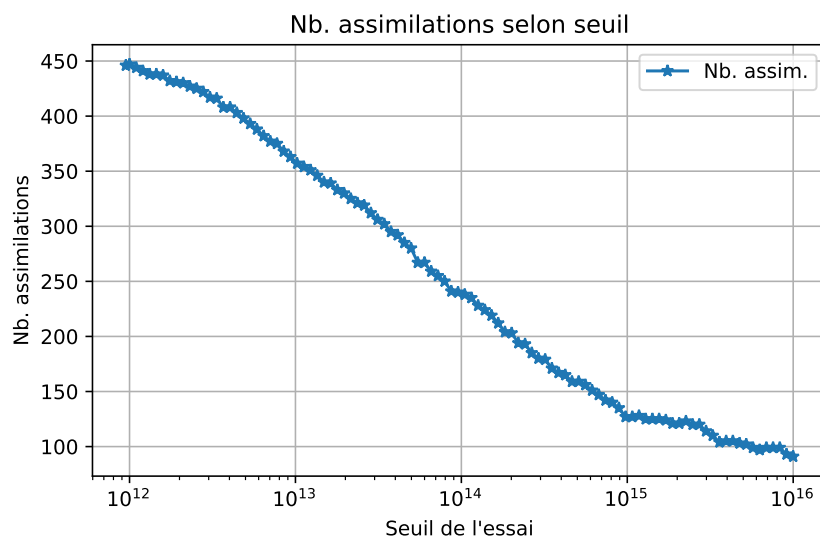


FIGURE 7.18 – Intervalle de valeurs de seuil utiles pour le système Xénon, montrant la décroissance du nombre d'analyses sur la fenêtre.

Nous observons dans les trois cas une diminution progressive du nombre d'observations assimilées à partir de la valeur de seuil initiale. Celle-ci peut être linéairement progressive relativement au logarithme naturel du seuil comme c'est le cas pour le système Xénon (fig.

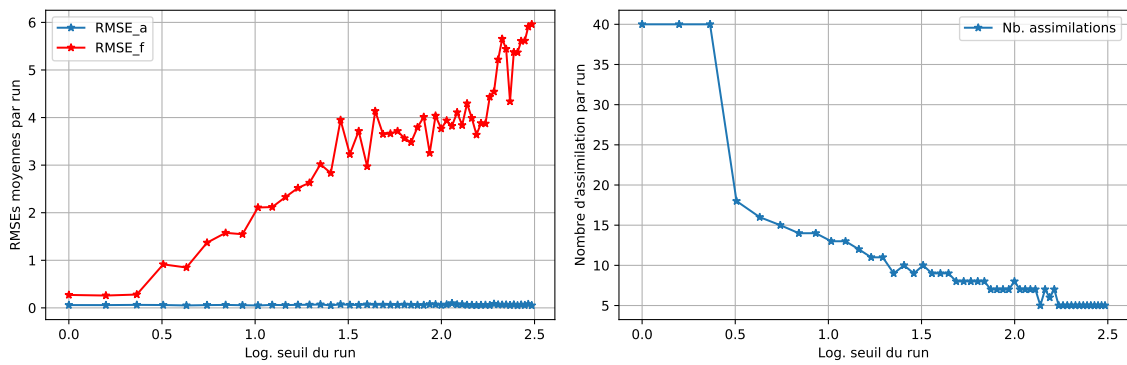


FIGURE 7.19 – Intervalle de valeurs de seuil utiles pour le Lorenz 63, montrant la décroissance du nombre d’assimilation sur la fenêtre (droite), avec évolution des moyennes de chaque RMSE (gauche), en fonction du logarithme naturel de la valeur de seuil (axe des abscisses).

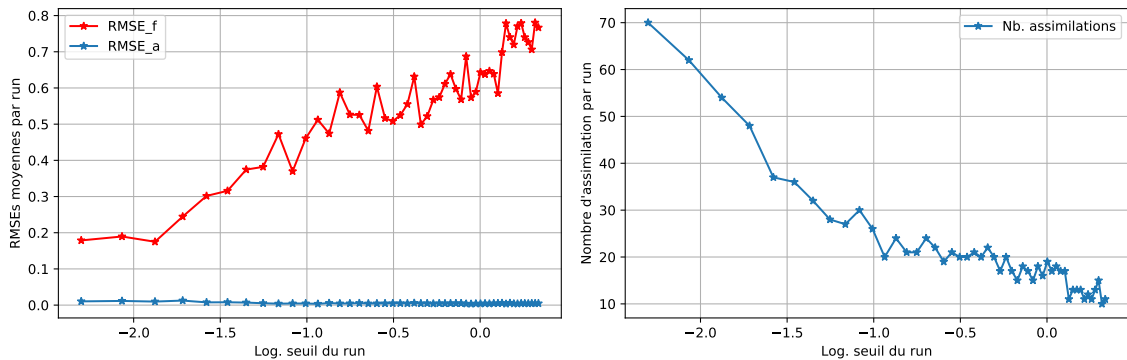


FIGURE 7.20 – Intervalle de valeurs de seuil utiles pour le Van der Pol, montrant la décroissance du nombre d’assimilation sur la fenêtre (droite), avec évolution des moyennes de chaque RMSE (gauche), en fonction du logarithme naturel de la valeur de seuil (axe des abscisses).

7.18) et le Van der Pol (fig. 7.20), ou s’apparenter à un phénomène de rupture comme c’est le cas du Lorenz 63 (fig. 7.19), où le nombre d’assimilation diminue fortement à partir de la valeur pour se stabiliser très vite autour d’un nombre minimal limite d’assimilations pour les valeurs de seuil utilisées.

Chaque système dynamique étant d’amplitude bornée, la distance de Wasserstein entre les distributions l’est aussi, il existe donc une valeur de seuil à partir de laquelle plus aucune observation ne sera assimilée. Nous ne l’avons pas affichée ici car elle n’a aucune utilité pour nos recherches.

Nous remarquons que la moyenne de la RMSE d’analyse reste stable malgré l’augmentation du temps entre chaque analyse au fur et à mesure que la valeur de seuil augmente. Malgré l’erreur plus importante sur la prévision, la correction par les observations reste efficace.

La dégradation de la RMSE de prévision au fur et à mesure de l’augmentation entre les temps d’analyse montre que le type de stratégies d’assimilation formulables grâce aux distances probabilistes semble peu propice à une utilisation pour un suivi avec précision de l’état. En revanche, une utilisation pour un suivi des changements de régimes d’un système reste possible. Cela sera montré aux sous-sections (7.4.1) et (7.4.2).

Nos futurs travaux se baseront sur ces intervalles utiles découverts pour chaque système. Ils seront faits à configuration fixe pour chacun. Les configurations sont détaillées en annexe (section [A.7](#)).

### 7.3.3 Comportement anticipé des deux distances

Nous pouvons exprimer analytiquement les expressions de Wasserstein et Kullback-Leibler dans notre cas, où  $\mathbf{R} = \sigma_{obs}^2 \mathbf{I}_p$ , et où nos matrices  $C_1 = \mathbf{H}\mathbf{P}^f\mathbf{H}^T$  et  $C_2 = \mathbf{R} + \mathbf{H}\mathbf{P}^f\mathbf{H}^T$ . Etant symétriques définies-positives, elles sont diagonalisables dans une base orthonormale et nous avons :

$$C_1 = \mathbf{U}\Sigma_f\mathbf{U}^T, C_2 = \mathbf{U}(\Sigma_f + \sigma_{obs}^2\mathbf{I}_p)\mathbf{U}^T \quad (7.3)$$

où  $\Sigma_f = \text{Diag}(\lambda_i)$ ,  $\lambda_i > 0$ . En reprenant les notations décrites aux sections (4.3) et (4.4), nous obtenons comme expression pour la distance de Wasserstein entre les distributions  $\mu_1$  et  $\mu_2$  :

$$W^2(\mu_1, \mu_2) = \|\mathbf{y} - \mathbf{H}\mathbf{x}^f\|_2^2 + \sum_{i=1}^p \left( \sqrt{\lambda_i + \sigma_{obs}^2} - \sqrt{\lambda_i} \right)^2. \quad (7.4)$$

En notant  $\mathbf{d} = \mathbf{U}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^f) = [d_i]$ , nous avons  $\|\mathbf{y} - \mathbf{H}\mathbf{x}^f\|_2 = \|\mathbf{d}\|_2$ , et dans la même configuration, l'expression équivalente pour la divergence de Kullback-Leibler entre les distributions  $\mu_1$  et  $\mu_2$  est :

$$KL(\mu_1, \mu_2) = \frac{1}{2} \sum_{i=1}^p \left( \frac{d_i^2 - \sigma_{obs}^2}{\lambda_i + \sigma_{obs}^2} + \frac{d_i^2 + \sigma_{obs}^2}{\lambda_i} \right) + p. \quad (7.5)$$

Les calculs détaillés pour établir ces formules sont en annexe (section A.5).

Nous voyons apparaître une similarité de comportement en fonction de l'évolution des valeurs propres de  $C_1$ . Elles représentent les valeurs propres de la matrice  $\mathbf{H}\mathbf{P}^f\mathbf{H}^T$ .

Comme pour  $\sigma_{obs}^2 > 0$ ,  $x \mapsto (\sqrt{x + \sigma_{obs}^2} - \sqrt{x})^2$  est strictement décroissante, la distance de Wasserstein va être d'autant plus importante que  $\lambda_i$  sera petit. Cependant, il est évident que plus  $\sigma_{obs}^2$  augmente, plus il est probable d'assimiler des observations bruitées. De fait, il est important d'adapter le seuil de sélection en fonction de la valeur de  $\sigma_{obs}^2$  pour éviter cela, ou bien d'avoir une valeur de  $\sigma_{obs}^2$  suffisamment petite. C'est pourquoi nos essais se feront à  $\sigma_{obs}^2$  constant et de faible amplitude par rapport aux amplitudes de variation des différents systèmes dynamiques, nous ne serons donc pas confrontés à ce phénomène. La distance de Wasserstein croîtra donc exclusivement de concert avec le résidu.

La divergence de Kullback-Leibler va de même augmenter dans les directions correspondant à de petites valeurs des  $\lambda_i$ , ce qui implique que pour des observations qui tendraient à faire augmenter les  $\lambda_i$  après assimilation, la divergence de Kullback-Leibler va diminuer. L'effet de la pondération du terme de résidu par l'inverse des matrices de covariance ne sera quantifiable qu'après des expériences numériques. Il guidera le choix du critère de sélection des observations avec la divergence.

Par rapport à la variance d'erreur d'observation  $\sigma_{obs}^2$ , le second terme de droite de la distance de Wasserstein va augmenter. Pour la divergence de Kullback-Leibler, plus  $\sigma_{obs}^2$  croîtra, plus le second terme de trace donné par la formule (A.42) va dominer. Nous observerons donc une augmentation des deux distances pour une augmentation de l'erreur d'observation.

Pour un seuil  $\rho > 0$  fixé, nous pouvons en déduire l'évolution de la façon de sélectionner

les observations de chacune des distances : l'augmentation de la valeur de Wasserstein en fonction du bruit va conduire à ce qu'elle assimile plus d'observations car la probabilité que la distance entre l'observation et la prévision soit supérieure au seuil va augmenter. Pour Kullback-Leibler, cela devrait se faire de même. Voici une illustration de l'évolution en moyenne des valeurs des deux termes de chaque distance probabiliste en fonction de la variance d'erreur d'observation de la simulation, pour le système du Xénon avec forçage de puissance pour du suivi de charge :

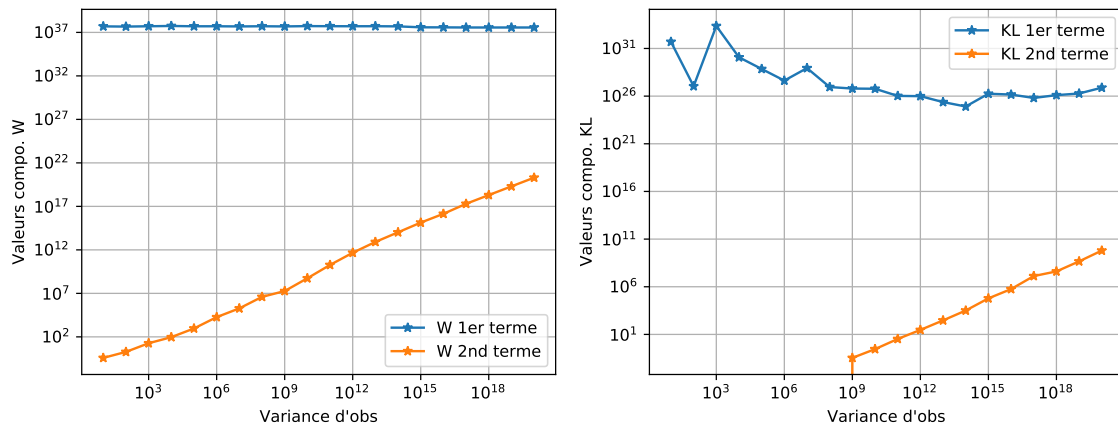


FIGURE 7.21 – Évolution en moyenne des distances probabilistes (ordonnées) en fonction de  $\sigma_{obs}^2$  (abscisse) pour le système Xénon en suivi de charge.

Nous observons comme anticipé pour chacun des seconds termes une augmentation en fonction de  $\sigma_{obs}^2$ . Le premier terme de la distance de Wasserstein étant une norme euclidienne (éventuellement pondérée), il reste constant malgré le bruit. Ce n'est pas le cas pour celui de la divergence de Kullback-Leibler du fait qu'il soit pondéré par les inverses de chaque matrice de covariance (Cf. équation (4.10)). Cependant, celui-ci reste globalement stable à partir de  $\sigma_{obs}^2 = 10^9$  du fait qu'il devienne équivalent à  $\frac{d_i^2}{\lambda_i}$ , indépendant de  $\sigma_{obs}^2$ . Nous voyons que ce sont les deux termes de résidus qui dominent pour les deux distances dans ce cas.

Dans cette configuration pour le système Xénon, il est donc possible d'approximer les deux distances par leurs premiers termes, qui correspondent respectivement à :

$$W_1^2 = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2, \text{ pour Wasserstein,} \quad (7.6)$$

$$KL_1 = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{(C_1^{-1} + C_2^{-1})}, \text{ pour Kullback-Leibler.} \quad (7.7)$$

Cela réduirait le coût numérique de chacune des distances, notamment pour la distance de Wasserstein où les matrices de covariances n'entrent plus en jeu<sup>1</sup>, sans impacter la précision du diagnostic.

Pour illustrer le comportement en fonction de la valeur de seuil choisie, nous effectuons de nouveau un essai qui donne une plage de valeurs utiles de seuil pour réduire le nombre d'assimilations sur une fenêtre de longueur fixe. Le critère de sélection pour la divergence de Kullback-Leibler est pris égal à celui de la distance de Wasserstein, soit  $d_{KL}(\mu_1, \mu_2) > \rho$ .

1. En dehors du cas  $W_1^2 = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\mathbf{T}^{-1}}^2$  mentionné en sous-section (4.3.1)



Voici les comportements observés pour les distances de Wasserstein et de Kullback-Leibler lors de cette expérience :

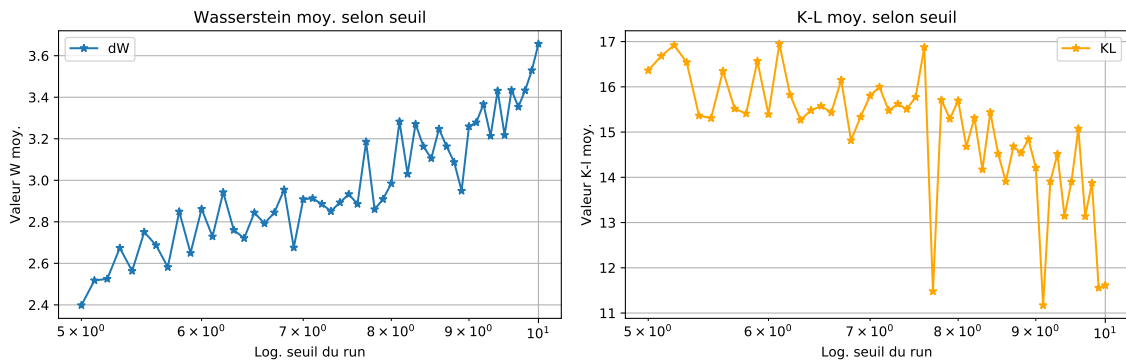


FIGURE 7.22 – Évolution en moyenne des distances probabilistes (ordonnées) selon le logarithme naturel du seuil choisi (abscisse) pour chaque essai sur le Lorenz 63. La distance de Wasserstein est à gauche, la divergence de Kullback-Leibler à droite.

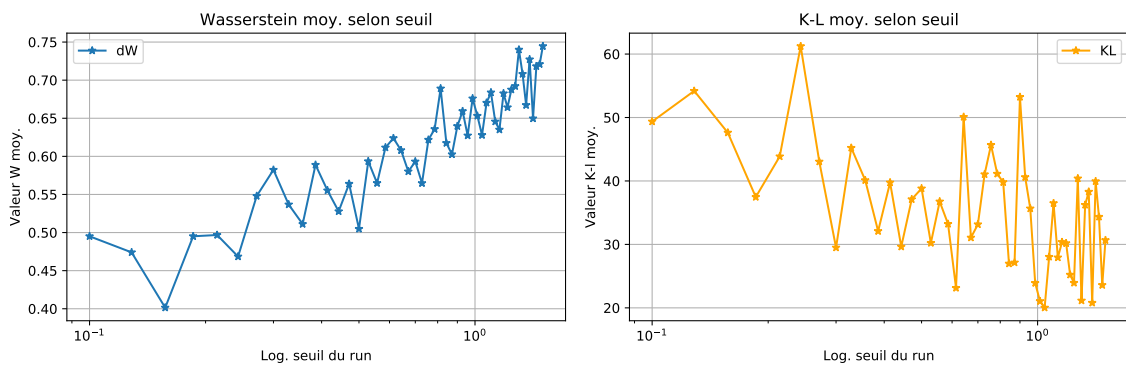


FIGURE 7.23 – Évolution en moyenne des distances probabilistes (ordonnées) selon le logarithme naturel du seuil choisi (abscisse) pour chaque essai sur le Van der Pol. La distance de Wasserstein est à gauche, la divergence de Kullback-Leibler à droite.

Nous y voyons bien apparaître le comportement attendu décrit précédemment pour la distance de Wasserstein, mais pas pour la divergence de Kullback-Leibler. Il semble dans ce cas que l'effet de la pondération du terme de résidu par l'inverse des matrice de covariance ait une influence plus importante. Une possibilité empirique pour obtenir un comportement adapté pour la sélection des étapes d'analyse est d'inverser le critère de sélection pour la divergence de Kullback-Leibler, soit  $d_{KL}(\mu_1, \mu_2) < \rho$ .

Nous allons maintenant examiner la possibilité d'établir une relation entre la valeur du seuil choisie et l'erreur de prévision moyenne commise sur la fenêtre en utilisant nos indicateurs pour établir des stratégies d'assimilation. Ceci sera fait avec la distance de Wasserstein en premier, puis nous montrerons la validité du critère de sélection  $d_{KL}(\mu_1, \mu_2) < \rho$  pour la divergence de Kullback-Leibler.

## Relation entre erreur de prévision et valeur du seuil

### Lorenz 63

Cet essai consiste à faire plusieurs simulations pour la configuration 1 (A.7.1) avec un seuil croissant se situant dans la zone de valeurs utiles qui permettent une réduction du nombre d'assimilations, et à observer l'évolution en moyenne à la fois de la distance de Wasserstein et de la RMSE de prévision, pour voir s'il existe une relation entre les deux. Une régression linéaire a été effectuée dans cette optique.

Nous remarquons qu'une relation linéaire existe entre chaque grandeur : la régression est effectuée entre la moyenne par simulation de la distance de Wasserstein et de la RMSE de prévision par rapport au logarithme naturel de la valeur du seuil.

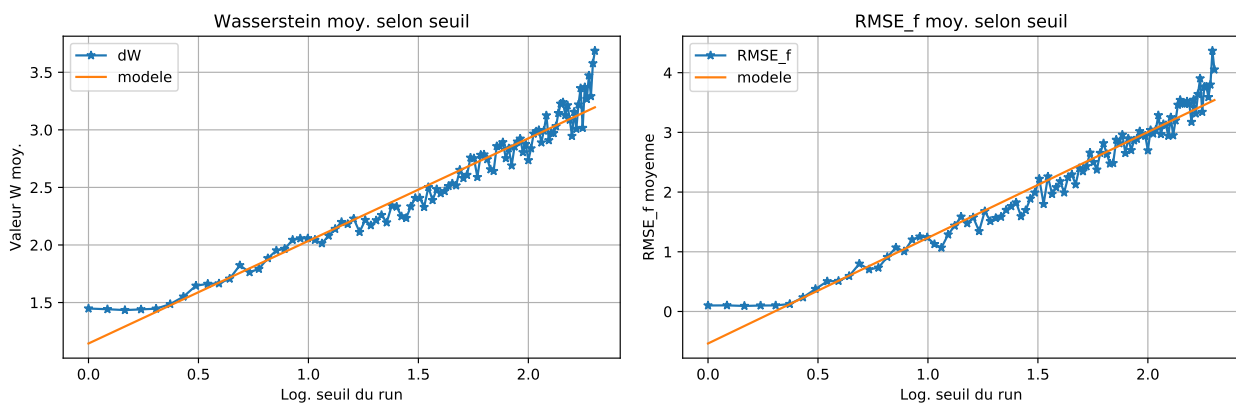


FIGURE 7.24 – Évolution en moyenne de la distance de Wasserstein (gauche) et de la moyenne de la RMSE de prévision (droite) selon le logarithme naturel du seuil choisi pour chaque essai. (Lorenz 63).

Les coefficients de corrélation entre la moyenne de la distance de Wasserstein et la moyenne de la RMSE de prévision sont, respectivement :

- $R_W^2 = 0.9515$ ,
- $R_{RMSE}^2 = 0.9611$ .

L'intérêt que nous portons à ce résultat est qu'il semble signifier que nous pouvons contrôler l'effet du choix de notre seuil sur notre RMSE de prévision, et que lorsque le seuil diminue, la RMSE de prévision diminue. Ce genre de propriété est particulièrement intéressante lorsque nous sommes dans un contexte de pilotage d'un système physique et que nous voulons connaître l'effet de notre manœuvre sur ce même système.

Le cas présent ici diffère de celui du système Xénon mentionné en sous-section (7.3.3), les deux termes de la distance de Wasserstein sont en moyenne d'amplitude équivalente comme illustré ci-après :

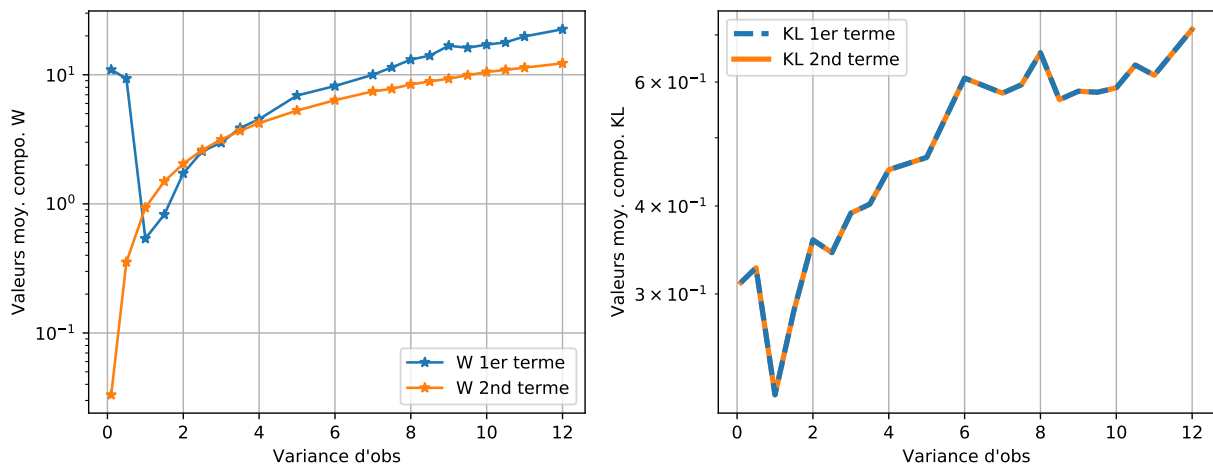


FIGURE 7.25 – Évolution en moyenne des composantes de chaque distance probabiliste (ordonnées) en fonction de la variance d’erreur d’observation (abscisse) pour le système Lorenz 63. La distance de Wasserstein est à gauche, la divergence de Kullback-Leibler est à droite.

Pour la configuration 1 (A.7.1) du Lorenz 63,  $\sigma_{obs}^2 = 2$ , nous voyons sur cette courbe que les deux composantes de la distance de Wasserstein ont une amplitude similaire, aucun ne domine l’autre. La relation décrite sur la figure (7.24) montre donc une relation linéaire qui n’est pas due au fait que ce soit le premier terme de la distance de Wasserstein qui domine.

Nous allons maintenant regarder si ces deux relations apparemment linéaires se vérifient sur d’autres systèmes emblématiques.

### Oscillateur de Van der Pol

Le même type d'essai que précédemment avec le Lorenz 63 est fait pour le Van der Pol. L'erreur modèle additive a pour variance  $5.10^{-4}$  dans l'essai qui va suivre, pour une variable observé d'amplitude en  $O(1)$  :

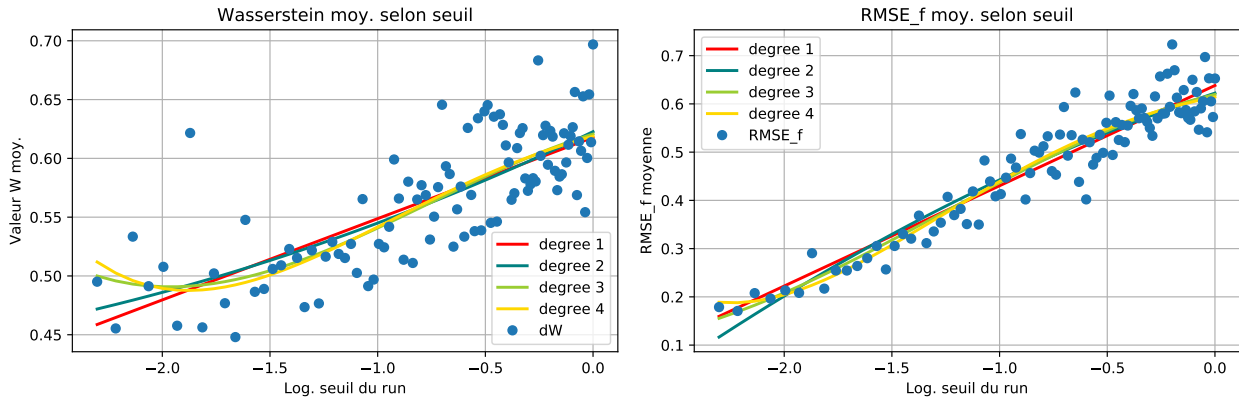


FIGURE 7.26 – Évolution en moyenne de la distance de Wasserstein (gauche) et de la moyenne de la RMSE de prévision (droite) selon le logarithme naturel du seuil choisi pour chaque essai. (Van der Pol)

Les coefficients de corrélation respectifs sont :

- $R_{W}^2 = 0.5881$
- $R_{RMSE}^2 = 0.8978$

Nous observons une dégradation de la relation linéaire entre la distance moyenne de Wasserstein et le logarithme naturel du seuil. Cette relation reste cependant valide entre la moyenne des RMSEs de prévision sur la fenêtre et le logarithme naturel du seuil, ce qui est positif car c'est cette relation qui nous importe véritablement : la RMSE de prévision constitue un critère de qualité pour notre prévision.

Notre choix est de chercher, si elle existe, la relation la plus simple entre le seuil et la RMSE de prévision, et le modèle linéaire est un candidat idéal. Des interpolations polynomiales par des modèles de plus haut degrés ont été effectuées dans l'objectif d'observer s'il n'y avait pas de meilleur candidat en terme de précision et d'adéquation avec la dynamique d'évolution observée entre les deux valeurs. Il est ici difficile de conclure, nous conserverons donc le modèle linéaire qui s'avère satisfaisant.

### Système Xénon

La vérification de la relation linéaire précédente sur deux systèmes emblématiques est très intéressante, cela signifie que dans deux cas particuliers de dynamiques, nous avons une évolution similaire pour deux quantités a priori liées par une relation simple. Le moment est venu de voir si cela se vérifie toujours pour un système industriel simplifié représentatif de la physique au sein d'un cœur de réacteur nucléaire, qui possède une dynamique plus complexe.

Voici donc les résultats obtenus pour une dynamique avec forçage en puissance du système Xénon, dans un cas hors-flux neutronique simulant l'arrêt du réacteur :

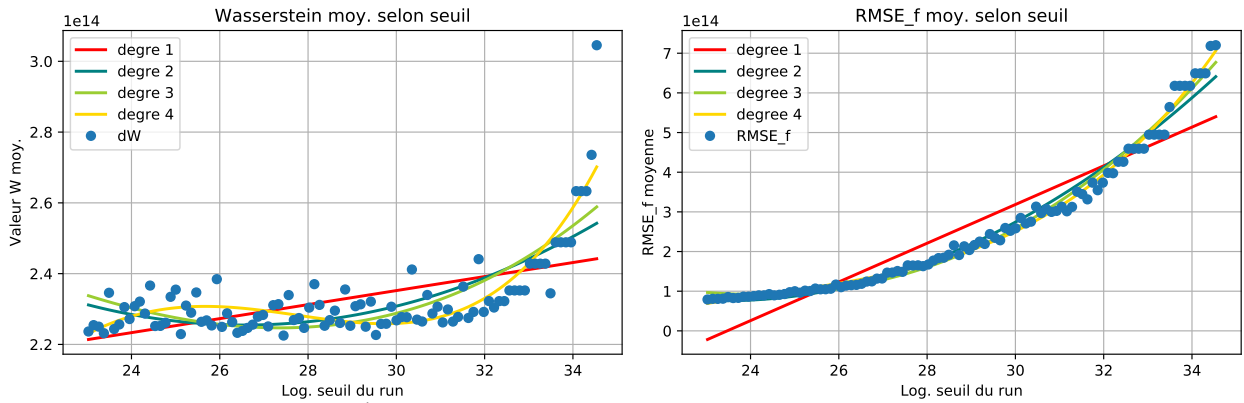


FIGURE 7.27 – Évolution en moyenne de la distance de Wasserstein (gauche) et de la moyenne de la RMSE de prévision (droite) selon le logarithme naturel du seuil choisi pour chaque essai. (Système Xénon hors-flux)

Les coefficients de corrélation respectifs sont :

- $R_W^2 = 0.5608$ ,
- $R_{RMSE}^2 = 0.9811$ .

Voici les résultats sur le même système sous flux neutronique dans le cas d'un démarrage de réacteur :

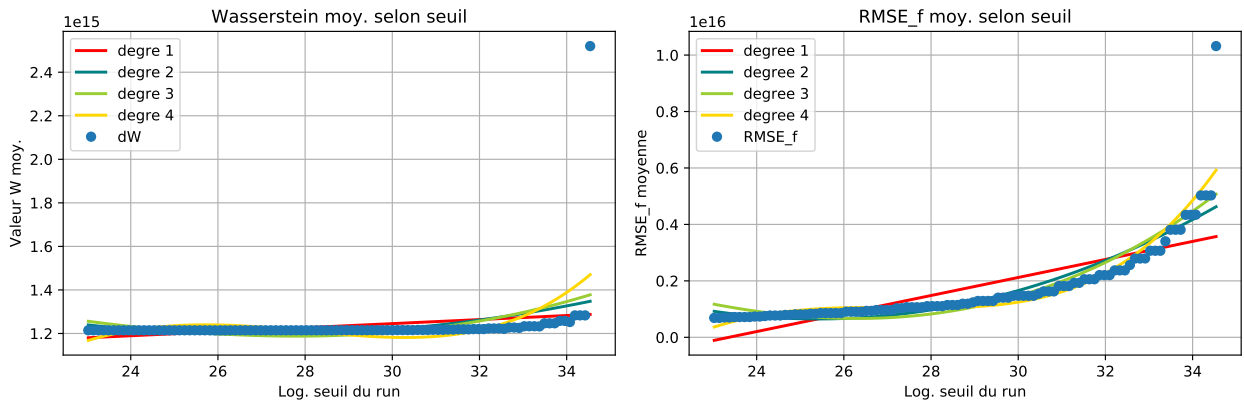


FIGURE 7.28 – Évolution en moyenne de la distance de Wasserstein (gauche) et de la moyenne de la RMSE de prévision (droite) selon le logarithme naturel du seuil choisi pour chaque essai. (Système Xénon sous flux)

Les coefficients de corrélation respectifs sont :

- $R_W^2 = 0.1295$ ,
- $R_{RMSE}^2 = 0.7890$ .

Cette forte baisse en validité des deux coefficients de corrélation est due à la non-linéarité du modèle, qui apparaît nettement de visu. En enlevant les deux valeurs extrêmes pour chaque grandeur, nous obtenons les coefficients suivants :

- $R_W^2 = 0.7106$ ,
- $R_{RMSE}^2 = 0.9334$ .

Il reste possible de faire une approximation linéaire du modèle d'évolution de chacune des deux grandeurs sur de plus petits intervalles de valeurs de seuil, qui reste satisfaisante. En revanche, il est clair que le modèle est non-linéaire, une approximation de plus haut degré sera plus précise.

Une troisième expérience a été faite dans une configuration où le système est soumis à un profil de variations brutales successives de la puissance au sein du réacteur. La puissance module le flux neutronique, donc l'évolution du système Xénon. Voici le programme de puissance associé à l'évolution des deux variables du système :

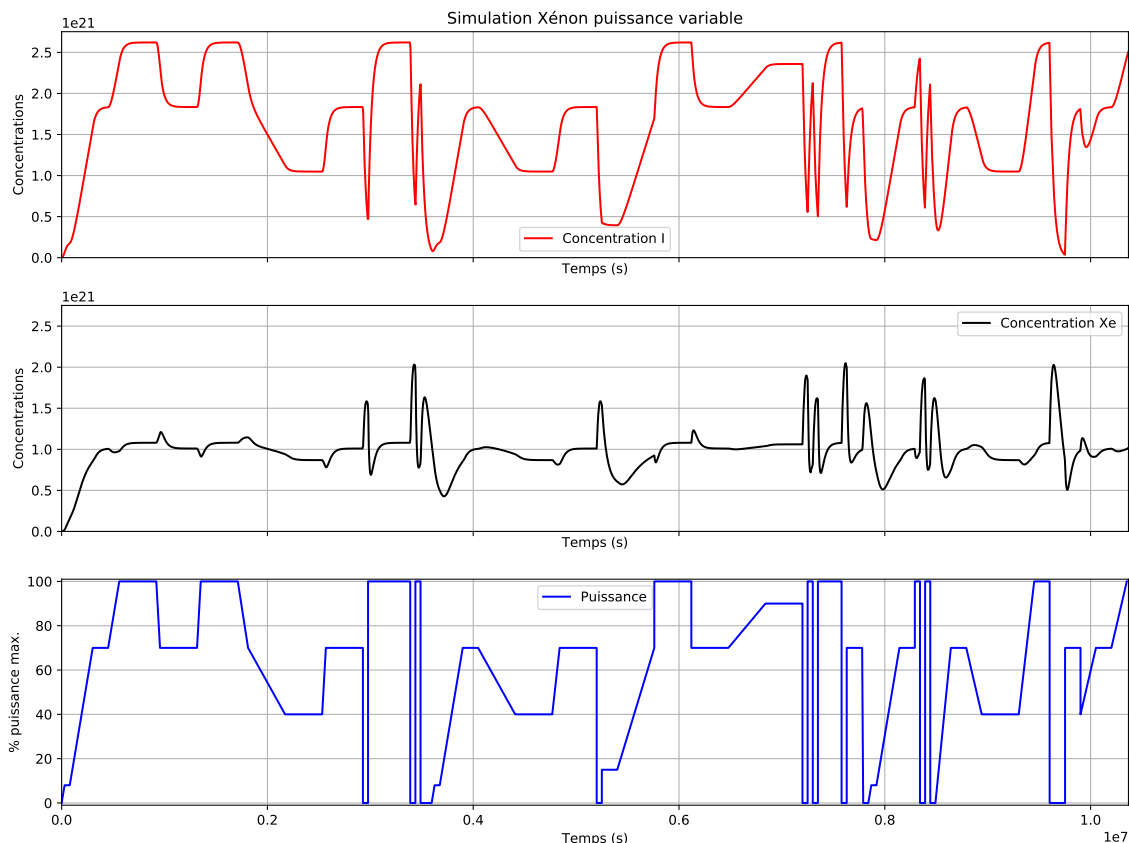


FIGURE 7.29 – Référence pour le système Xénon avec profil de puissance variable (courbe bleue).

Ce programme de puissance n'est pas réaliste par rapport à ce qui est effectué en conditions opérationnelles. Lors du démarrage du réacteur, la montée en puissance est progressive pour garantir sa sécurité. Plusieurs mesures sont effectuées, par exemple aux paliers des 8% et 70% de puissance par rapport à la puissance maximale du réacteur, qui permettent de garantir la sécurité de fonctionnement de la tranche.

De même, hors urgence, l'arrêt du réacteur se fait avec une diminution progressive de la puissance, pour éviter un pic de Xénon-135 comme nous pouvons les observer à chaque fois qu'elle tombe brusquement à zéro sur notre simulation. Un tel pic empêcherait ensuite le réacteur de redémarrer pendant au moins 36h, ce que le pilotage normal évite toujours. Ce programme de forçage artificiel a été conçu pour faire apparaître de nombreux régimes différents sur le système Xénon, notamment des oscillations qui seront difficiles à suivre pour notre système d'assimilation de données. En conditions opérationnelles, le pilote du

réacteur les anticipe et les évite grâce à un pilotage adapté, et va préférer un profil de concentration comme au début de la simulation avant le début des oscillations.

Voici les résultats obtenus en faisant varier la puissance comme décrit précédemment :

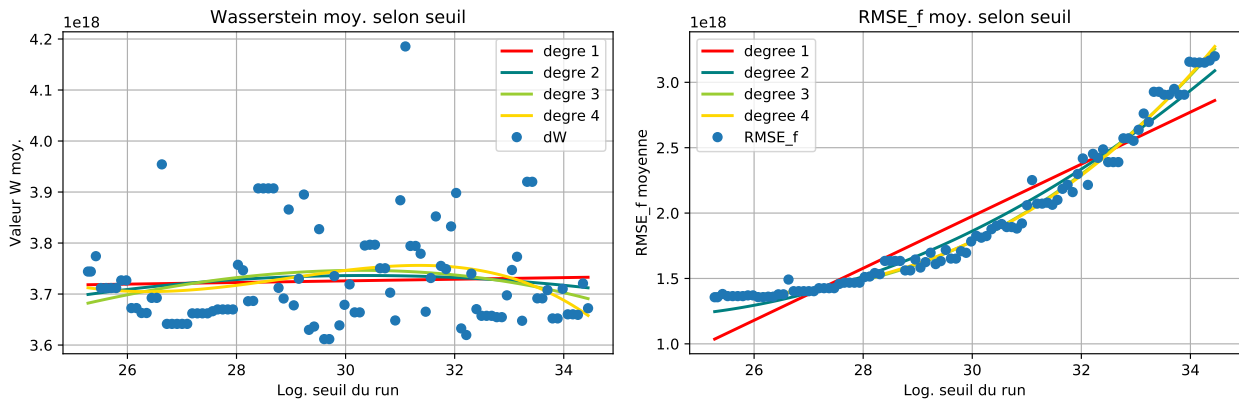


FIGURE 7.30 – Évolution en moyenne de la distance de Wasserstein (gauche) et de la moyenne de la RMSE de prévision (droite) selon le logarithme naturel du seuil choisi pour chaque essai. (Système Xénon avec forçage en puissance variable).

Nous voyons clairement que la monotonie pour Wasserstein n'est pas établie. En revanche, nous retrouvons une bonne approximation linéaire entre le logarithme du seuil et la moyenne des RMSEs de prévision sur la fenêtre, et ce sur un cas où la dynamique du système est complexe, car elle est le résultat d'un forçage extrême du système dynamique. Les valeurs des coefficients de régression sont les suivantes :

- $R_{W}^2 = 0.0285$ ,
- $R_{RMSE}^2 = 0.9742$ .

Elles viennent confirmer nos résultats sur le programme de puissance variable, tout en remarquant encore qu'un modèle de plus haut degré serait plus précis.

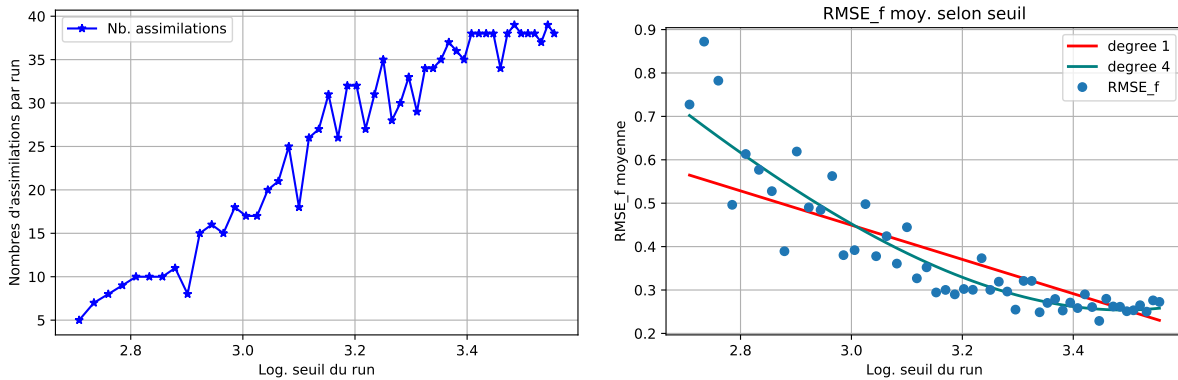
L'approximation linéaire est encore une fois suffisante entre le logarithme du seuil et la moyenne des RMSEs de prévision sur la fenêtre. Cette confirmation de ce comportement sur un système industriel est particulièrement intéressante. En effet, nous avons ici un moyen de régler l'erreur moyenne de prévision en fonction du seuil que nous désirons choisir. Nous pouvons piloter directement le système en ayant une estimation de l'effet sur l'erreur de prévision qu'aura notre pilotage. Nous obtenons également un comportement similaire avec la configuration du système Xénon en suivi de charge (A.7.3), dont le programme de puissance est réaliste.

Nous remarquerons tout de même qu'un modèle quadratique semble bien adapté pour décrire la relation entre seuil et la moyenne des RMSEs de prévision sur la fenêtre, notamment en terme de précision : nous voyons qu'en dehors de la parabole que forment le nuage de points, le modèle linéaire sous-estime la valeur de la moyenne des RMSEs tandis qu'il la surestime une fois à l'intérieur. Cette relation linéaire étant quoiqu'il arrive valable d'après la valeur de  $R_{RMSE}^2 = 0.9742$ , nous nous contenterons de celui-ci étant donné qu'il reste l'estimation la plus simple possible à obtenir entre les deux.



### Changement du critère de sélection pour Kullback-Leibler

Nous allons effectuer de nouveau l'expérience sur le modèle de Lorenz 63 qui a été faite auparavant en sous-section (7.3.3), cette fois-ci en utilisant la divergence de Kullback-Leibler et le critère de sélection qui lui est adapté :  $dKL(\mu_1, \mu_1) < \rho$ . Voici ce que nous observons en terme de nombre d'observations sélectionnées en fonction de la valeur du seuil ainsi qu'en terme d'évolution de la moyenne des RMSEs de prévision sur la fenêtre selon le seuil :



(A) Nombre d'assimilations.

(B) Moyenne des RMSEs de prévision.

FIGURE 7.31 – Lorenz 63, Kullback-Leibler. Évolution du nombre d'assimilation sur la fenêtre (gauche) et de la moyenne des RMSEs de prévision sur la fenêtre (droite), en fonction du logarithme naturel du seuil.

Nous observons bien une croissance du nombre d'assimilations avec la croissance de la valeur du seuil et une diminution de l'erreur de prévision moyenne avec l'augmentation du seuil. La relation entre l'erreur de prévision et la valeur du logarithme naturel du seuil semble cette fois plutôt être doublement quadratique que linéaire. Le coefficient de corrélation linéaire vaut  $R_{RMSE}^2 = 0.7553$ . Ce comportement est similaire à celui obtenu avec la distance de Wasserstein dans les essais précédents. La divergence de Kullback-Leibler s'utilise donc de manière équivalente à la distance de Wasserstein, à la différence près du critère de sélection des observations qui devient l'opposé de celui de cette dernière. La différence principale entre les deux distances semble être dans l'équilibre de la contribution des deux composantes de chacune :

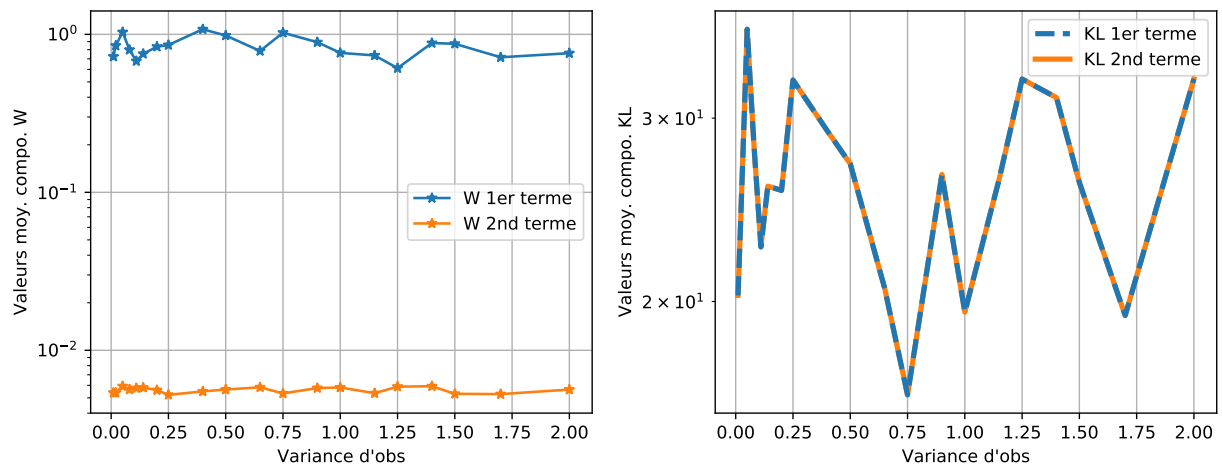


FIGURE 7.32 – Évolution en moyenne des composantes de chaque distance probabiliste en fonction de la variance d'erreur d'observation pour la configuration 3 du Van der Pol. (Wasserstein à gauche, Kullback-Leibler à droite.)

Pour la distance de Wasserstein, celui-ci est variable et le premier terme de résidu peut souvent dominer, pour la divergence de Kullback-Leibler, cela peut varier et un équilibre est trouvé sur le Lorenz 63 (fig. 7.25) ainsi que sur le Van der Pol (fig. 7.32), tandis que c'est le terme de résidu qui domine pour le système Xénon (fig. 7.21).

## 7.4 Stratégies d'assimilation sur des modèles emblématiques

Nous établissons dans un premier temps des stratégies d'assimilation qui consistent à contrôler les instants d'assimilation. D'après nos résultats précédents, ce sont celles qui permettent de réduire le plus le nombre d'observations assimilées. De plus, nous travaillons sur des systèmes de petite dimension ( $\leq 3$ ). Les informations d'importance des observations retournées par nos indicateurs diffèrent des informations de qualité de la prévision. C'est-à-dire que ce que "regardent" les indicateurs n'est pas lié directement à la qualité de la prévision, à la différence des RMSEs d'analyse ou de prévision.

D'après les résultats précédents, il est clair que la suppression d'observation diminue la qualité du suivi de l'état en terme de RMSE : en effet, la prévision étant moins corrigée par les observations, elle s'éloigne automatiquement de l'état-vrai représenté par notre expérience jumelle. Outre que le suivi en précision de l'état, nous sommes intéressés par le suivi des changements de régime du système. Nous allons donc regarder quels effets la suppression d'observation a sur ces deux types de suivi.

### 7.4.1 Essais avec le système de Lorenz

#### Réduction du nombre d'assimilation

Nous nous mettons dans le contexte d'un objectif de réduction du nombre d'assimilations sur la fenêtre complète  $[t_0, t_f]$ . Nous voulons diminuer ce nombre, tout en conservant un suivi convenable de la dynamique du système. Pour cela, nous allons d'abord voir l'évolution des moyennes des RMSEs de prévision et d'analyse sur la fenêtre, en fonction du nombre d'assimilations effectuées. Cela revient à regarder l'évolution de chaque RMSE en fonction du seuil choisi pour chaque simulation. Les voici :

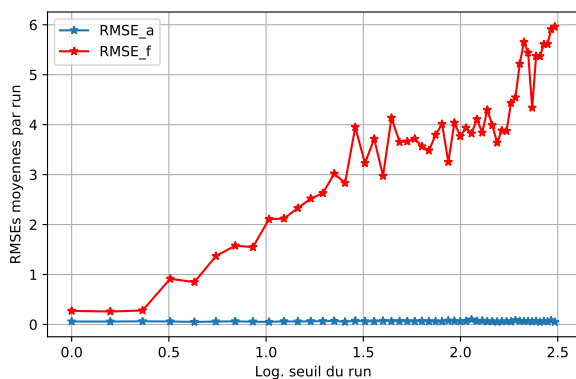


FIGURE 7.33 – Moyennes des RMSEs d'analyse et de prévision (ordonnées), selon le logarithme naturel du seuil (abscisse).

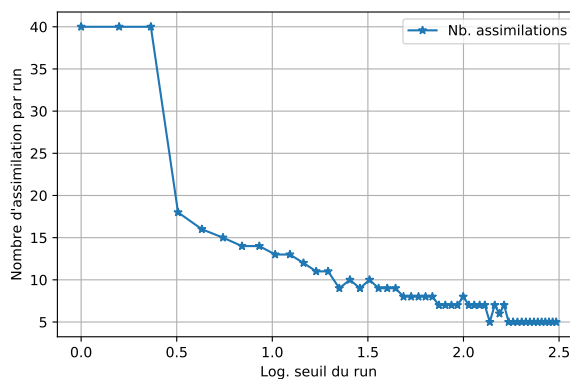


FIGURE 7.34 – Nombre d'assimilations pour chaque essai (ordonnées), selon le logarithme naturel du seuil (abscisse).

Nous voyons une augmentation linéaire de la moyenne de la RMSE de prévision, en accord avec les essais en sous-section (7.3.3). La décroissance du nombre d'assimilation en fonction de la valeur du seuil est affichée à côté : nous passons de 40 assimilations sur la fenêtre avec la valeur initiale, à 5 assimilations avec la valeur finale, soit presque 90%

d'observations assimilées en moins. La RMSE d'analyse reste presque constante et faible par rapport à la RMSE de prévision, l'assimilation apporte donc de bonnes corrections. Néanmoins, la prévision pure reste compliquée en supprimant des instants d'assimilation, seule l'analyse reste suffisamment précise pour être pertinente.

### Suivi de la dynamique en forte carence d'observation

Nous cherchons maintenant à voir s'il est possible de maintenir un suivi des transitions du Lorenz 63 malgré une forte réduction du nombre d'observations assimilées. Ces transitions sont sensibles aux perturbations. La réduction du nombre d'assimilation implique une augmentation de la variance globale de l'ensemble ainsi que de chacune de ses dimensions. Cela peut mener une partie de l'ensemble à effectuer sa transition sur une aile du papillon, et l'autre à effectuer sa transition sur l'autre aile. La moyenne de l'ensemble ne devient plus représentative de l'état et a besoin d'être corrigée par l'assimilation.

Chaque aile se situe dans le plan  $x_1x_2$ , soit du côté des valeurs négatives, soit des valeurs positives. L'amplitude de variation du système sur chaque aile étant la même, cela se traduit par l'effondrement de la moyenne de l'ensemble sur 0 pour les dimensions  $x_1$  et  $x_2$ . C'est un comportement que nous pouvons observer sur ces coupes de l'ensemble en deux dimensions autour de plusieurs instants de transition :

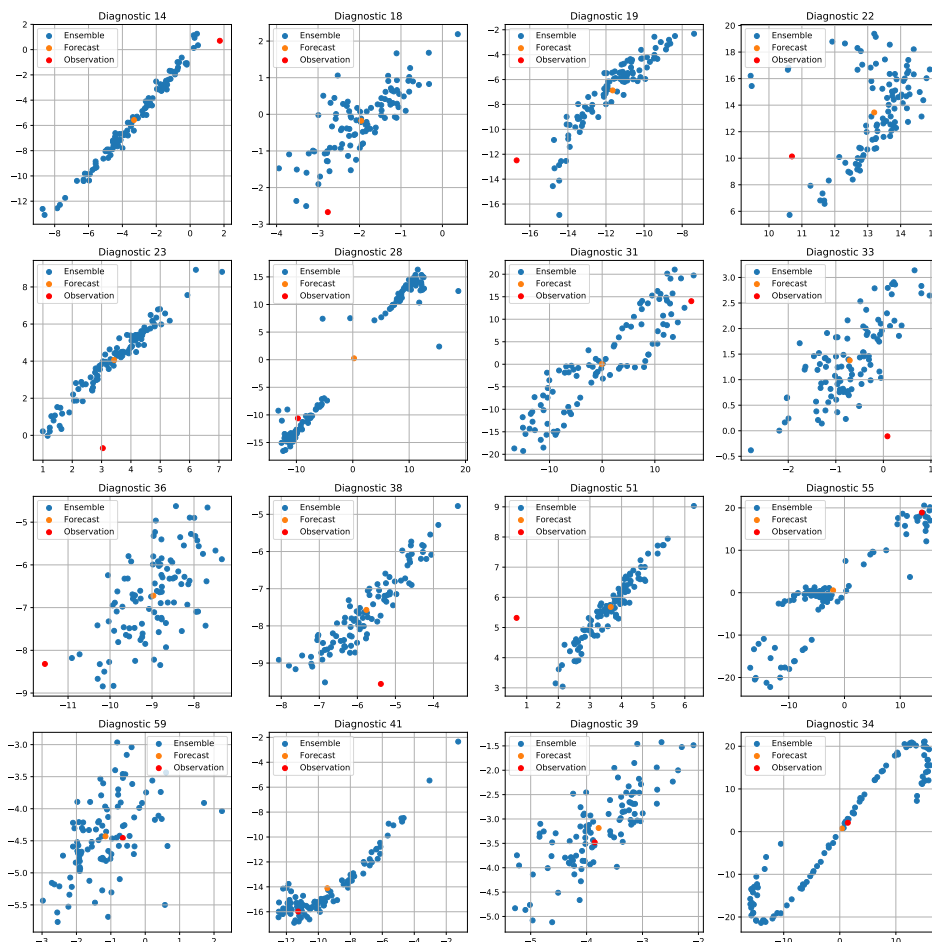


FIGURE 7.35 – Coupe selon les dimensions  $x_1$  et  $x_2$  à plusieurs instants de transition pour le Lorenz 63. Point rouge : moyenne de l'ensemble, point orange : l'observation. Les points bleus constituent l'ensemble.

Nous voulons donc voir jusqu'à quel point nous pouvons réduire le nombre d'observations assimilées sans que ce type de comportement apparaisse. Nous sommes donc particulièrement intéressés par une réduction de l'ordre de 90% du nombre total d'observation sur la fenêtre.

Pour les figures où chaque dimension est représentée séparément des autres, ceci revient à voir l'ensemble qui s'étale dans les deux plans limités par la droite  $x_i = 0$ , avec le vecteur d'état  $\mathbf{x} = [x_i]$ .

Nous allons présenter dans un premier temps l'essai de référence avec assimilation de toutes les prévisions, qui illustre la façon dont sont corrigées les observations pour permettre à l'état estimé par l'EnKF de ne rater aucune transition :

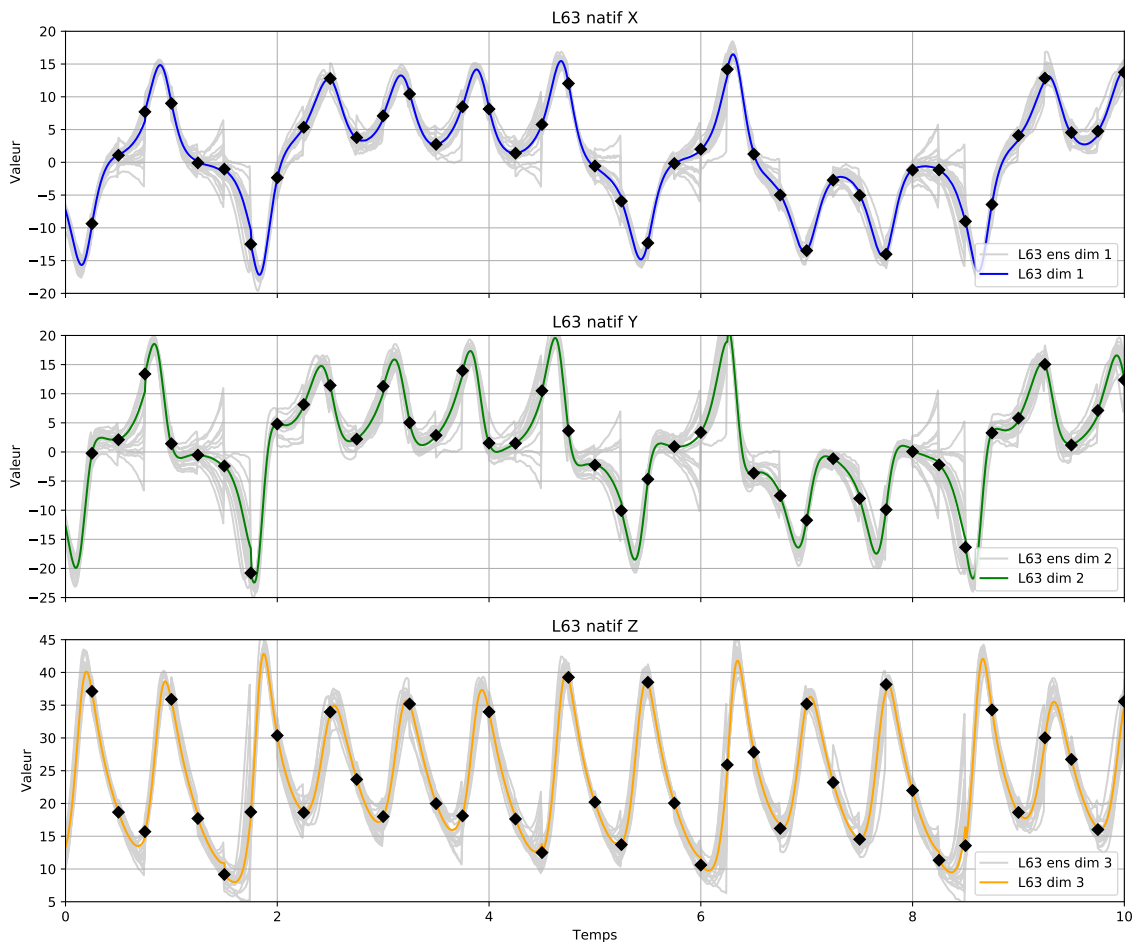


FIGURE 7.36 – Lorenz 63. Essai de référence avec toutes les observations.

Les observations correspondent aux losanges noirs pour chaque dimension.

Le temps est en abscisse, la valeur de chaque dimension en ordonnées.

Nous observons ici qu'avec la fréquence des corrections par l'assimilation, la séparation de l'ensemble en deux sous-ensembles est corrigée suffisamment tôt pour éviter qu'ils transitent chacun sur une aile différente du papillon. Nous observons cependant le début de ce phénomène plusieurs fois le long de la trajectoire, en particulier aux instants de transitions du système. Ils se repèrent en suivant la valeurs de la première dimension lorsque  $x_1 \simeq 15$ .

Nous passons maintenant à des essais avec forte réduction du nombre d'observations où apparaissent des effets dus aux non-linéarités.

Nous observons sur la figure (7.37) l'apparition du phénomène de séparation de l'ensemble à plusieurs endroits (notamment autour des temps  $t = 2$ ,  $t = 6$ ,  $t = 8$ ). C'est à ces endroits précisément que l'utilisation de la distance de Wasserstein pour sélectionner les observation amène à l'assimilation d'une observation. Le seuil fixe le nombre d'observations sur la fenêtre. Pour le cas présent, deux transitions aux temps  $t = 1.75$  et  $t = 6.25$  sont récupérées de justesse par l'assimilation en comparaison à la référence figure (7.36). Le reste des transitions se fait normalement, mais l'amplitude de l'oscillation résultant de la transition est souvent plus faible que celle de la référence, menant à une nouvelle transition trop anticipée si non corrigée comme aux temps  $t = 4.75$  et  $t = 5.25$ . Nous voyons en conséquence que l'indicateur incite à l'assimilation d'une observation aussi à chacun de ces instants.

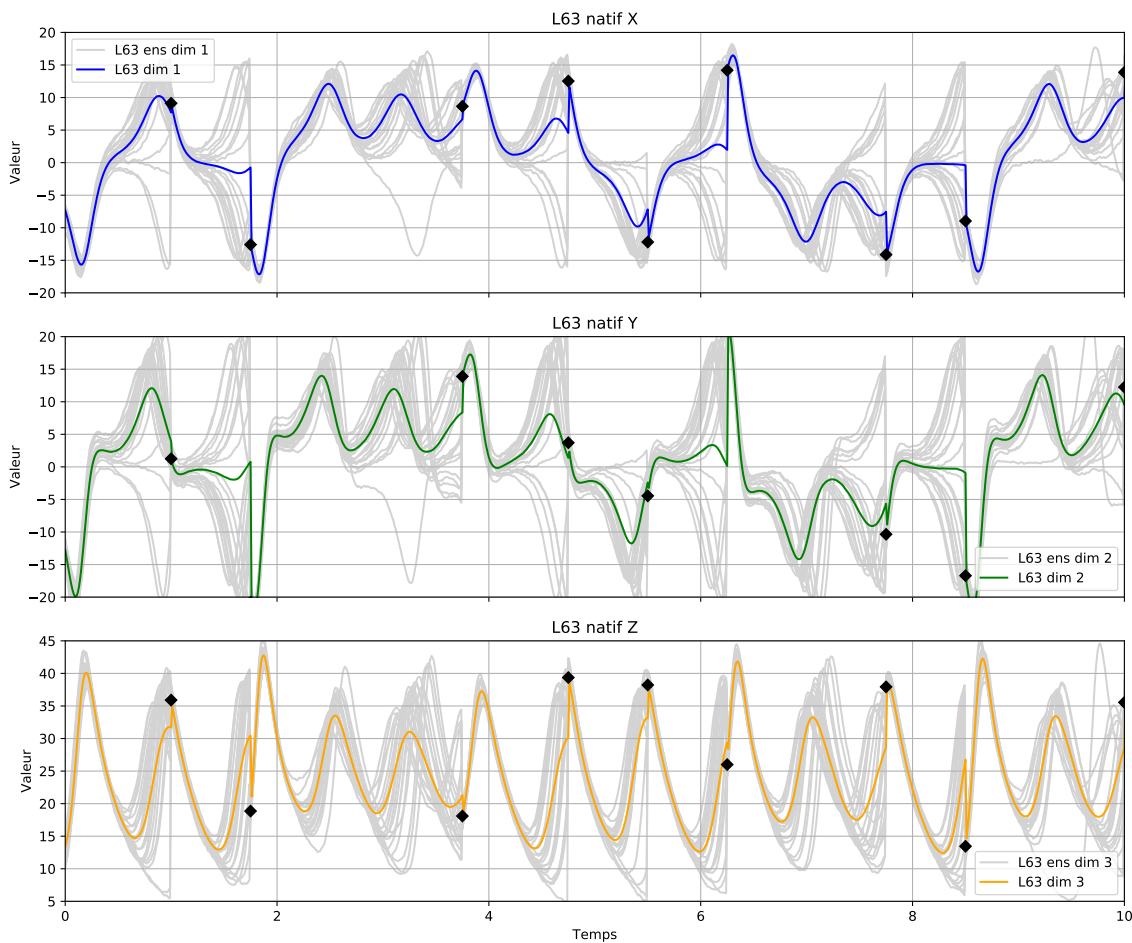


FIGURE 7.37 – Lorenz 63. Essai avec 9/40 observations sélectionnées par l'indicateur, 80% de réduction. Temps en abscisse, amplitude de chaque dimension en ordonnées.

Le même type de comportement s'observe sur la figure (7.38). Les non-linéarités sont plus marquées, cela s'observe notamment sur la dimension  $x_3$  où la variance de l'ensemble sur cette dimension a augmenté en comparaison à la figure (7.37). En dehors des temps  $t = 1.75$ ,  $t = 3.75$  et  $t = 6.25$ , le suivi des transitions reste convenable en ne conservant que 10% du nombre d'observations disponibles sur la fenêtre.

Nous voyons ainsi qu'il est possible de conserver une correction acceptable des suivis de transitions du Lorenz 63 malgré une forte réduction du nombre d'assimilations observées, dans le cadre d'un système fortement observé. La sélection des observations assimilées se faisant sur un nombre important d'observations disponibles sur la fenêtre, cela illustre la capacité de la distance de Wasserstein à prendre en compte l'évolution de la dynamique du système et des observations pour sélectionner les instants d'assimilation les plus opportuns, sur ce type de dynamique.

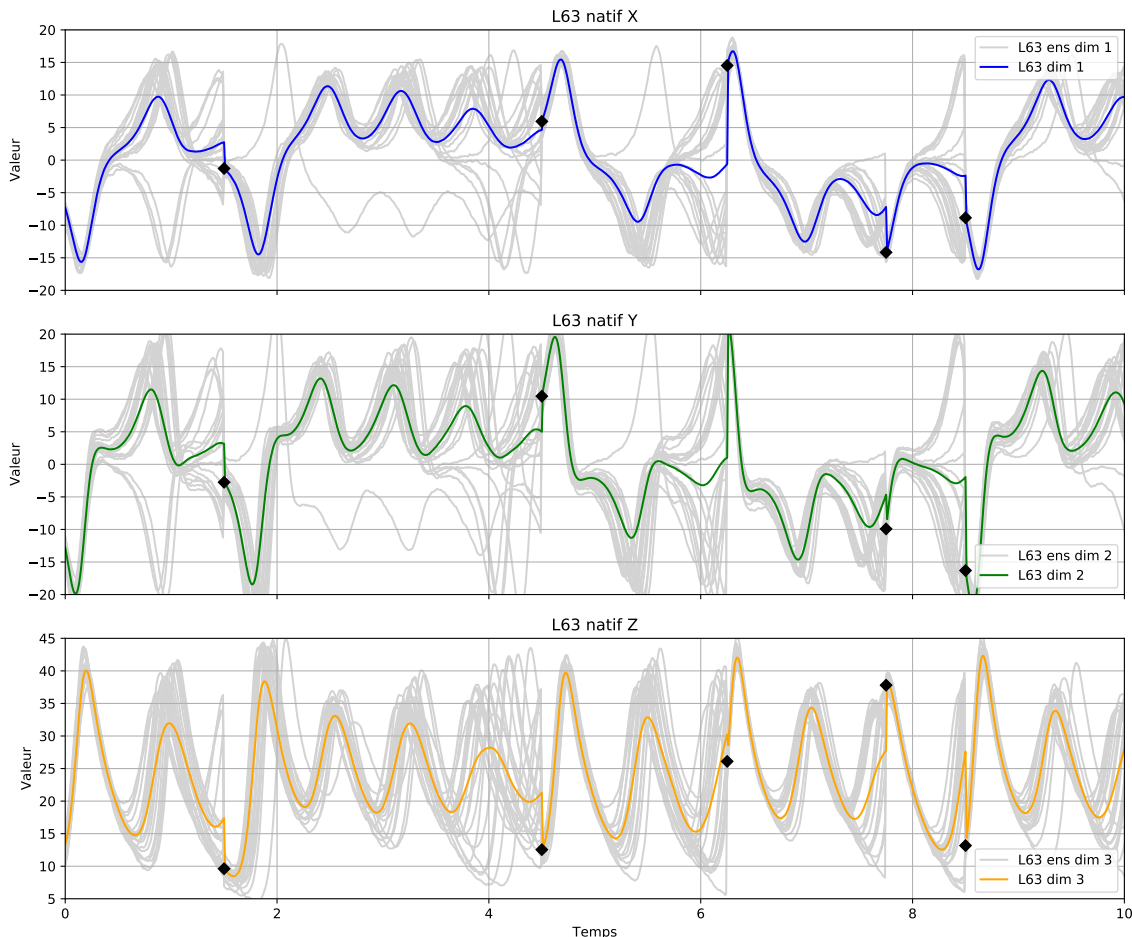


FIGURE 7.38 – Lorenz 63. Essai avec 5/40 observations sélectionnées par l'indicateur, 90% de réduction.

## 7.4.2 Essais avec l'oscillateur de Van der Pol

### Réduction du nombre d'assimilations sur la fenêtre

Comme pour le Lorenz 63, nous nous mettons dans le contexte d'un objectif de réduction du nombre d'assimilations sur la fenêtre complète  $[t_0, t_f]$ , en voulant conserver un suivi acceptable de la dynamique du système, en l'occurrence ici les changements de régime de l'oscillateur de Van der Pol. Du fait de son caractère raide, ce système est plus sensible que le Lorenz 63 aux différents types de perturbations qu'il subit. Il convient donc dans un premier temps de s'assurer que nous observons le même type de comportement qu'avec le Lorenz 63 lorsque nous supprimons des observations sur la fenêtre.

Nous effectuons les mêmes diagnostics que pour le Lorenz 63 en montrant l'évolution des

moyennes des RMSEs de prévision et d'analyse et du nombre d'assimilations, en fonction du seuil choisi :

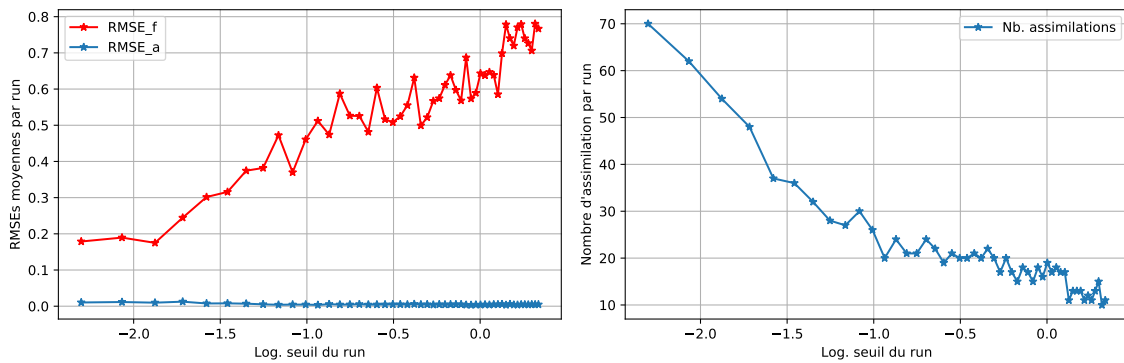


FIGURE 7.39 – Moyennes des RMSEs d'analyse et de prévision (gauche), nombre d'assimilation pour chaque essai (droite), chacune selon le logarithme naturel du seuil en abscisse.

Nous observons de même que la moyenne des RMSEs de prévision sur la fenêtre augmente avec la valeur du seuil, et que la moyenne des RMSEs d'analyse sur la fenêtre reste stable et faible en comparaison de celle de prévision. La différence se situe au niveau de la décroissance du nombre d'assimilations sur la fenêtre en fonction du seuil : celle-ci semble linéaire. Malgré sa sensibilité accrue aux perturbations, nous voyons que nous pouvons effectuer des suppressions d'observations de la même façon qu'avec le Lorenz 63 sur ce système.

### Estimation du nombre minimal d'observations permettant un suivi des changements de régime

Ici, nous voulons réduire au maximum le nombre d'assimilations sur la fenêtre sans rater un seul changement de régime de l'oscillateur. Pour établir le fait que nous capturons bien un changement de régime grâce à la simulation, nous définissons arbitrairement deux seuils en terme d'amplitude des oscillations correspondant aux droites  $y = -0.5$  et  $y = 0.5$ . Si l'oscillation sur une pseudo-période ne sort pas de la zone comprise entre ces deux seuils, nous estimons que le changement de régime n'a pas été capturé. Voici un exemple qui illustre ceci :

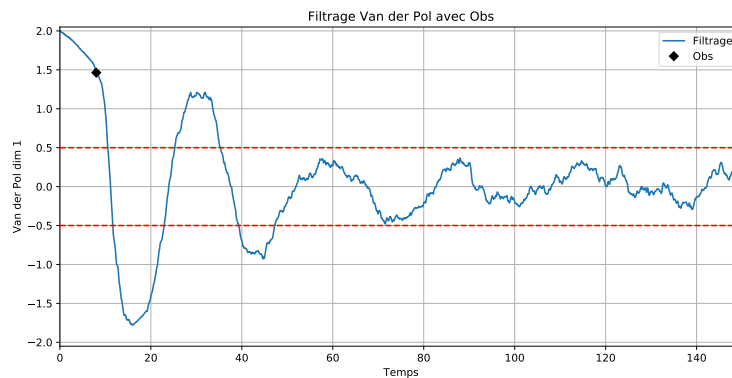


FIGURE 7.40 – Perte de la dynamique le Van der Pol. 1 seule observation.



Nous observons après la fin de la deuxième oscillation au temps  $t = 40$  que le système d'assimilation de données n'arrive plus à suivre les changements de régime de l'oscillateur. Le reste de la trajectoire contenu entre les deux droites en rouge devient difficile à différencier du bruit du aux différentes sources d'erreur dans le système. C'est un exemple extrême où nous sommes clairement en dessous du nombre minimal d'observations  $n_{min}$  que nous recherchons.

Voici la référence avec toutes les observations disponibles sur la fenêtre pour le Van der Pol :

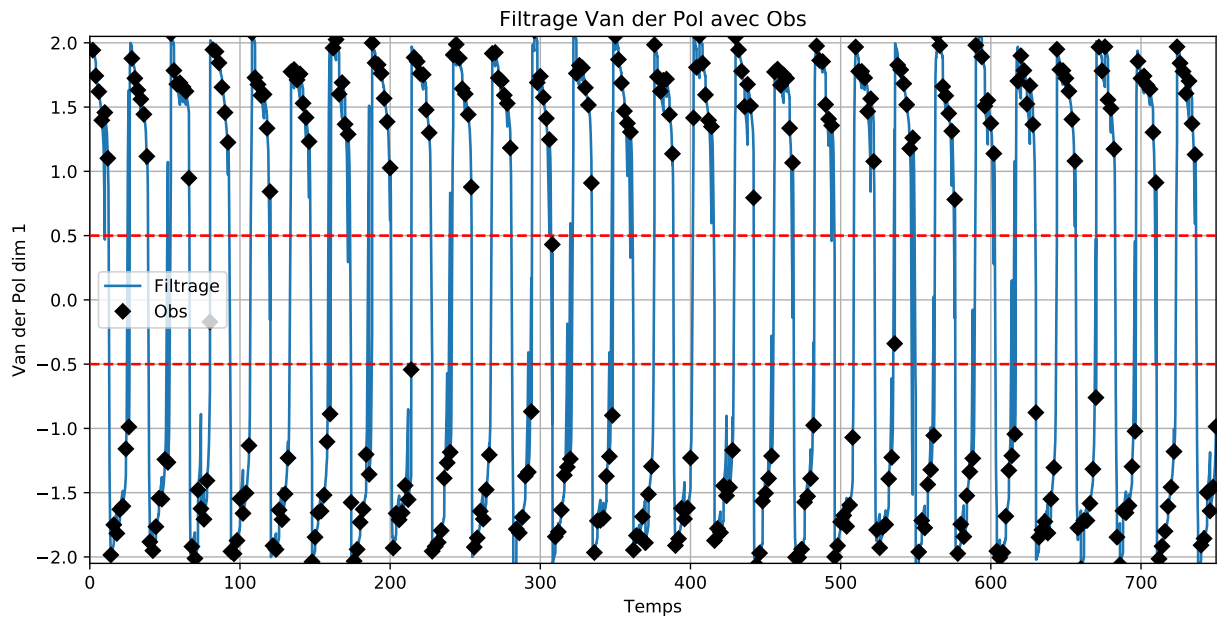


FIGURE 7.41 – Van der Pol. Essai de référence avec toutes les observations.  
375 observations.

Nous voyons que le système est fortement observé et que l'amplitude de chacune de ses oscillations sur la fenêtre dépasse les deux limites marquées en pointillés rouges. Notre objectif en réduisant le nombre d'observation est de conserver l'amplitude des oscillations en dehors de ces limites. Le suivi avec précision de leur amplitude exacte pour chacune des oscillations n'est pas une propriété que nous souhaitons conserver.

Voici les résultats que nous obtenons avec des valeurs de seuil permettant d'assimiler en dessous de 14% (53) des observations disponibles sur la fenêtre complète, puis seulement 7% (25) de ce total d'observations :

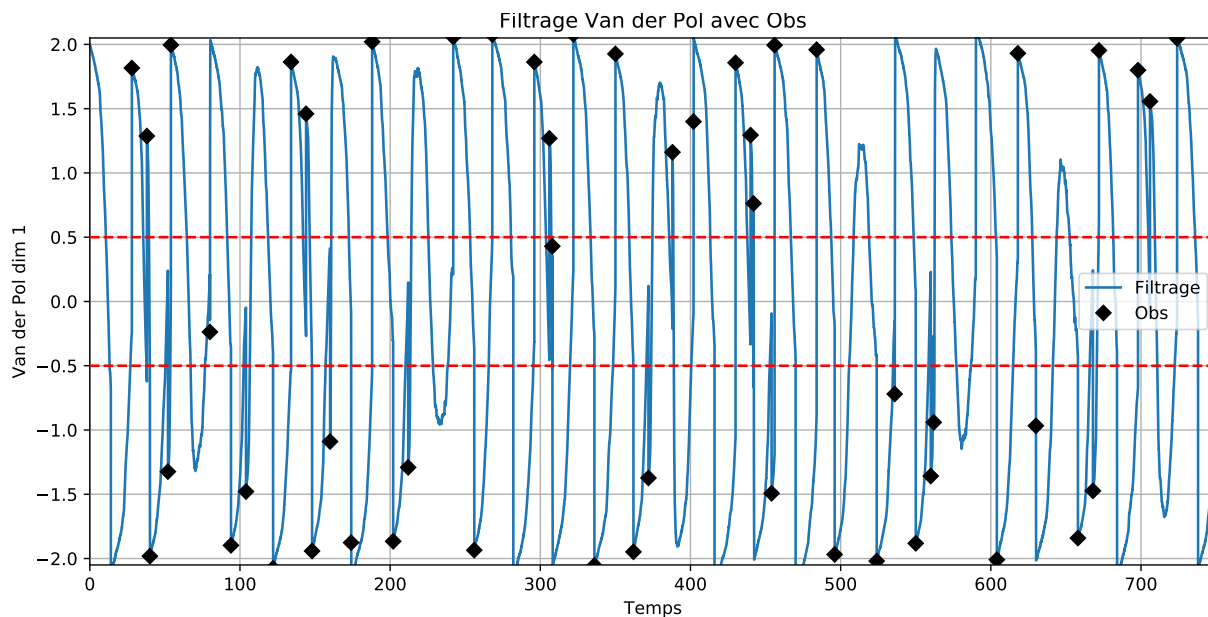


FIGURE 7.42 – Van der Pol. Essai avec 14% des observations de référence assimilées sur la fenêtre.

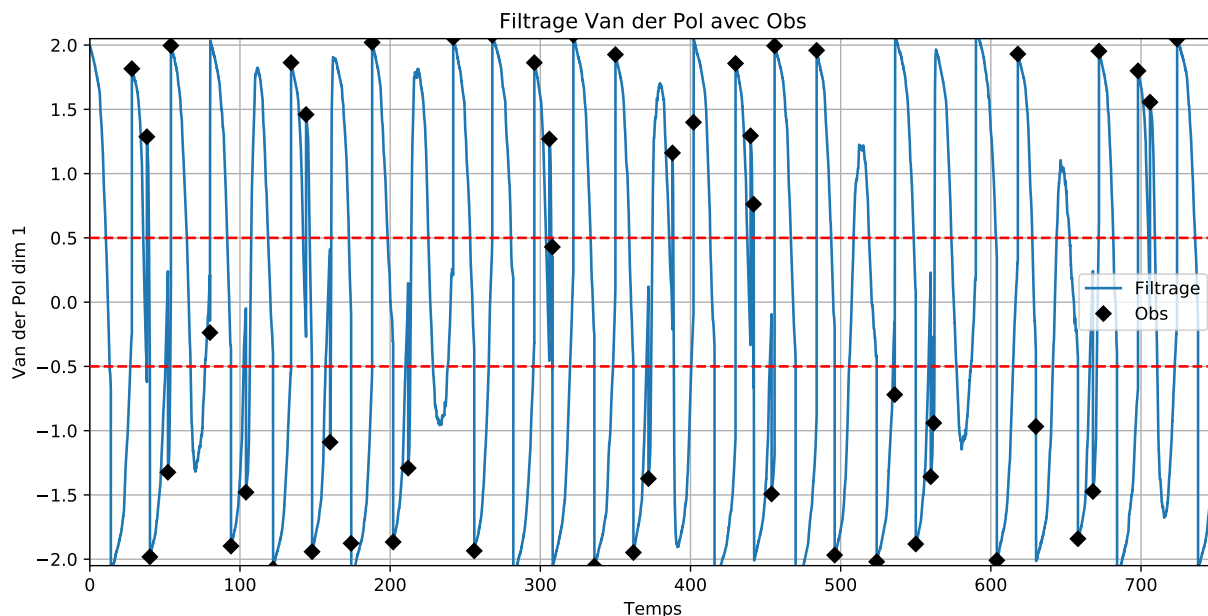


FIGURE 7.43 – Van der Pol. Essai avec 7% des observations de référence assimilées sur la fenêtre.

Nous remarquons qu'il est encore possible de suivre chacune des transitions du système en assimilant seulement 14% des observations disponibles sur la fenêtre, figure (7.42). Nous observons sur la figure (7.43) qu'avec 7% (25) des observations, nous ratons la deuxième transition du système. Cependant, c'est la seule à ne pas être suivie. Ainsi, en conservant toutes les observations sélectionnées lors de cet essai et en en rajoutant une au niveau de la transition ratée, nous aurions un suivi complet des transitions du système. Le nombre minimal d'observations pour le suivi des changements de régime dans cette configuration est donc de 26, soit environ 7% des 375 de référence.

En répétant plusieurs expériences sur une même configuration, nous sommes donc capables d'estimer le nombre minimal d'observation permettant un suivi de l'ensemble des changements de régime pour un système pseudo-périodique chaotique.

## 7.5 Stratégies d'assimilation sur des modèles d'origine industrielle

### 7.5.1 Modèle d'évolution de la concentration en Xénon-135

Le système du Xénon décrit de façon simplifiée l'évolution des concentrations de deux entités chimiques au sein d'un cœur de réacteur nucléaire. Une explication détaillée est disponible en annexe (sous-section [A.7.3](#)).

Lorsque le réacteur est en fonctionnement à puissance constante, un régime stationnaire s'établit et les concentrations des deux entités sont à l'équilibre. Lorsque le réacteur est arrêté brusquement, la production d'Iode-135 est stoppée, et celle-ci va alors se transformer en Xénon-135 par décomposition radioactive. La concentration en Xénon-135, qui est normalement en partie détruit par le flux neutronique lorsque le réacteur fonctionne, va alors fortement augmenter du fait de la décomposition de l'Iode-135. La demi-vie de décomposition de l'Iode-135 est de 6.57h, tandis que celle du Xénon-135 est de 9.14h. Nous assistons alors à ce que l'on appelle le pic de Xénon, où la concentration en Xénon-135 va fortement augmenter et rester au-delà d'une limite qui empêche le redémarrage du réacteur : celui-ci est alors temporairement empoisonné.

Il s'agit d'un système à deux variables piloté par un forçage en puissance, correspondant au couplage des équations d'évolution de la concentration d'Iode-135 et de Xénon-135 dans un réacteur en fonctionnement, légèrement raide.

### 7.5.2 Suivi de l'évolution avec pilotage de puissance

Nous sommes ici dans le cas de la configuration 6 ([A.7.3](#)), décrite par la figure ([7.29](#)). Le programme de puissance défini pour cette configuration sert à faire apparaître de nombreux régimes différents sur le système du Xénon, et à donner une dynamique fortement variée où les non-linéarités des équations sont activées. Cela devrait rendre le système difficile à suivre dans les zones de forte variations de la dynamique. L'indicateur devrait donc sélectionner prioritairement des observations dans ces zones-là pour corriger la simulation. Voici la référence montrant la répartition des observations sur la fenêtre :

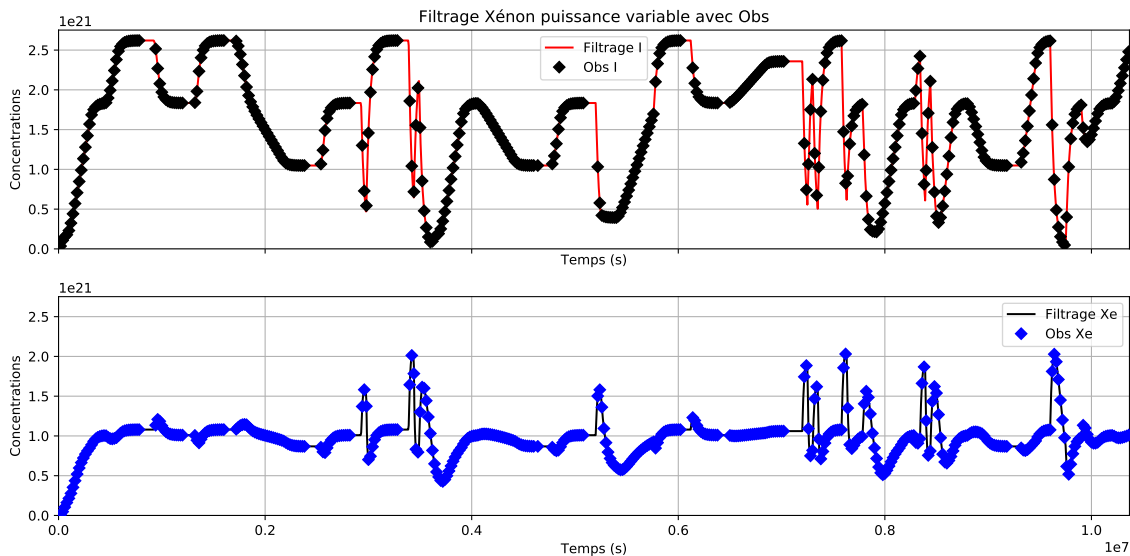


FIGURE 7.44 – Répartition des observations de référence pour le système Xénon avec forçage en puissance variable. Les observations sont uniformément réparties sur la fenêtre. Utilisation de 518 observations. Iode-135 en haut, Xénon-135 en bas. Temps en secondes en abscisse, concentrations en ordonnées.

En réduisant progressivement le nombre d'observations sur la fenêtre en faisant varier le seuil pour l'indicateur, nous ne conservons plus que 92 sur 518, soit environ 18%. Nous obtenons ainsi la figure suivante :

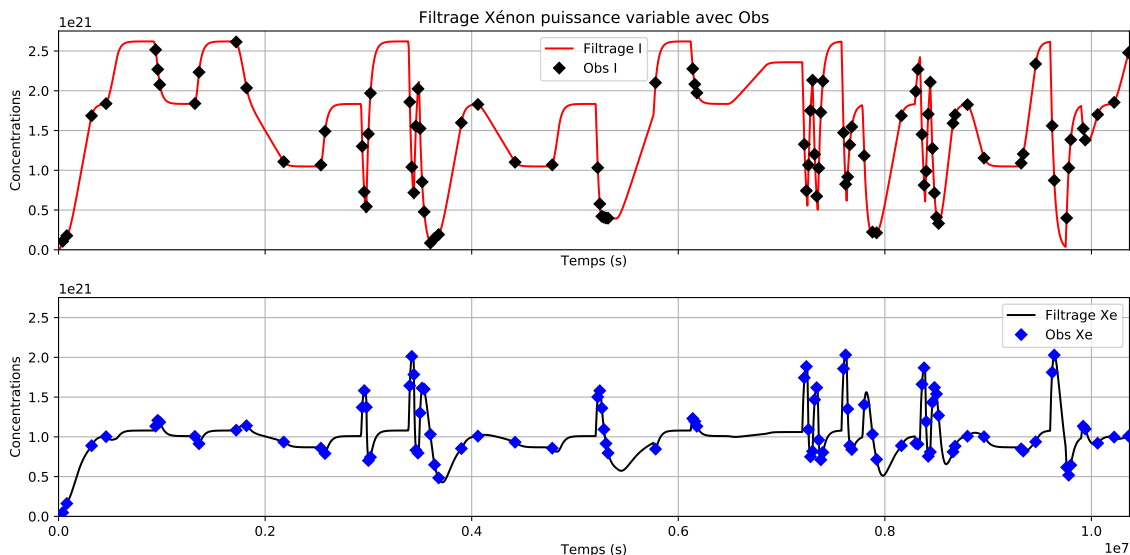


FIGURE 7.45 – Répartition des observations sélectionnées par Wasserstein pour le système Xénon avec profil de puissance variable. Les observations sont situées autour des zones de changement de régime. Utilisation de 92 observations. Iode-135 en haut, Xénon-135 en bas. Temps en secondes en abscisse, concentrations en ordonnées.

Nous n'observons aucun raté dans le suivi des transitions du système et ce malgré une forte réduction du nombre d'observation. La répartition des observations n'est plus uniforme mais reste globalement équilibrée sur la fenêtre du fait que le programme de puissance

ait été conçu pour déstabiliser le système et faire fortement varier sa dynamique. Les observations déjà présentes aux instants de changement de régime du système ont été à chaque fois conservées pour être assimilées.

Cela nous montre que l'indicateur est capable de s'adapter à la dynamique d'évolution du système pour sélectionner les observations situées aux instants de forte variabilité de celui-ci. Il est intéressant de remarquer la forte réduction du nombre d'observations assimilées, d'un facteur 5. A l'instar de l'oscillateur de Van der Pol, l'existence d'un nombre minimal d'observations à assimiler pour maintenir un suivi de chacune des transitions du système est manifeste.

### 7.5.3 Suivi de l'assimilation en suivi de charge

Ces essais sont faits avec la configuration 7 (A.7.3). Nous voyons sur la figure (A.9) trois régimes de variations marquées de la puissance du réacteur, notamment le régime situé entre  $t = 7.5 \times 10^5$  et  $t = 1.25 \times 10^6$ . Ceci va avoir pour effet de faire varier rapidement les concentrations en Iode-135 et en Xénon-135 du système.

Nous voulons observer la façon dont sont réparties les observations sélectionnées par notre indicateur au fur et à mesure que nous réduisons leur nombre sur la fenêtre d'assimilation. Voici la référence qui montre la répartition des observations sur la fenêtre sans les sélectionner :

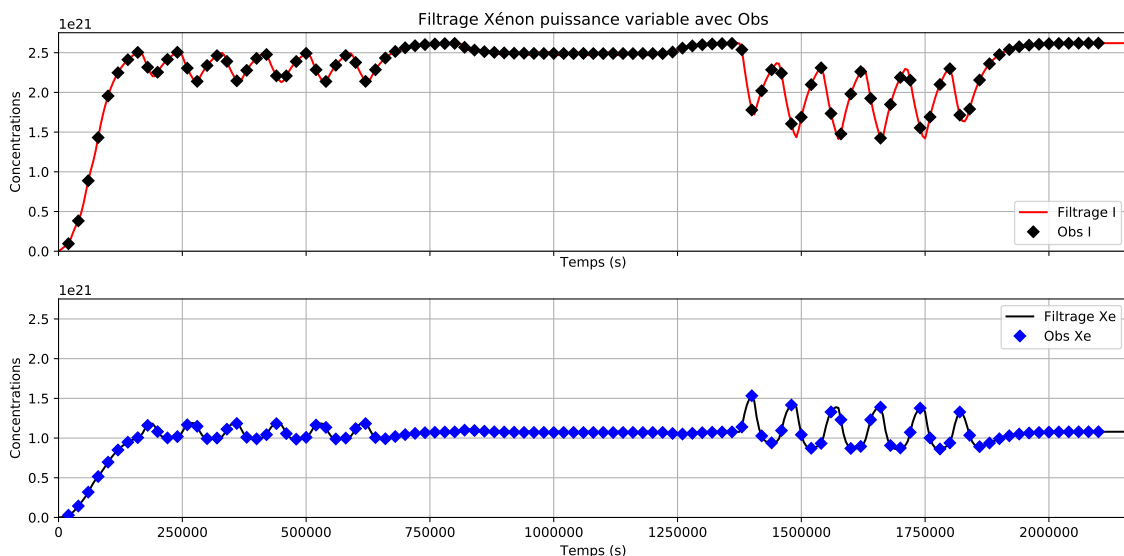


FIGURE 7.46 – Répartition des observations de référence pour le système Xénon avec forçage en puissance pour suivi de charge. Les observations sont uniformément réparties sur la fenêtre. Utilisation de 108 observations. Iode-135 en haut, Xénon-135 en bas. Temps en secondes en abscisse, concentrations en ordonnées.

La réduction progressive du nombre d'observations assimilées nous amènent au résultat suivant, pour 72 observations conservées sur 108, soit  $2/3$  des observations :

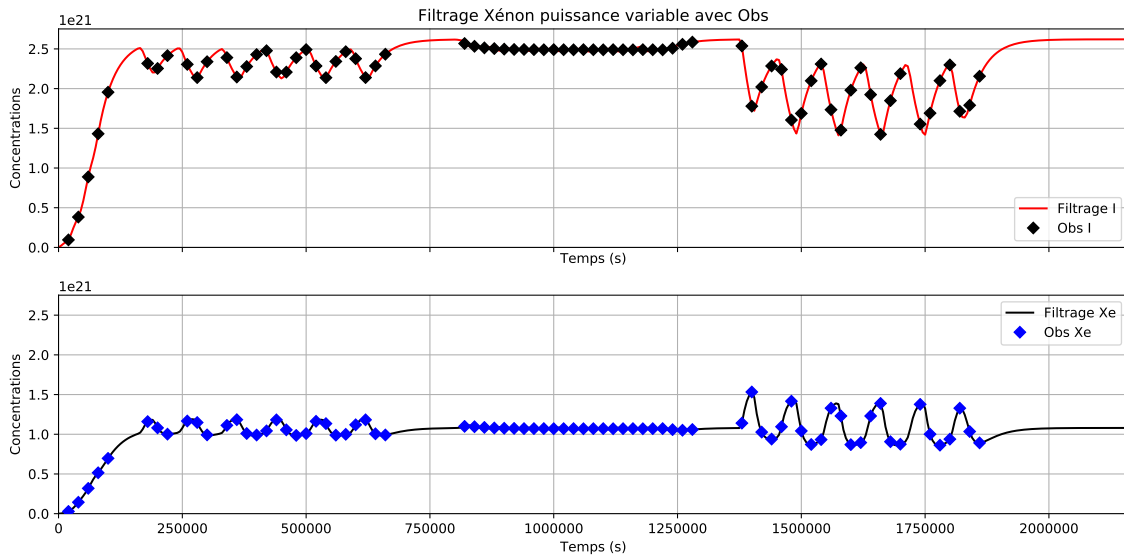


FIGURE 7.47 – Répartition des observations sélectionnées par Wasserstein pour le système Xénon avec forçage en puissance pour suivi de charge. Les observations sont concentrées dans les zones de variations importante de la dynamique. Iode-135 en haut, Xénon-135 en bas. Temps en secondes en abscisse, concentrations en ordonnées.

Nous remarquons que les observations en dehors des zones de forte variabilité de la dynamique ont été éliminées, notamment en fin de fenêtre où le système se stabilise de nouveau, et que les observations sont prioritairement sélectionnées lorsque la puissance du réacteur varie, même dans le cas où les oscillations sont très rapides et à une fréquence inférieure à l'échantillonnage d'observation.

Cela nous confirme que l'indicateur a la capacité de s'adapter à la dynamique d'évolution du système pour la sélection des observations à assimiler. Cette propriété est importante car ce sont à ces instants que le système d'assimilation de données a le besoin d'être le plus précis. Cela ouvre aussi la possibilité de pouvoir calibrer le réseau d'observation pour un système périodique, ce dont nous discuterons dans les perspectives.

Ce résultat constitue le dernier de ceux illustrés dans ce manuscrit. Nous allons maintenant tâcher de conclure sur l'ensemble des recherches discutées dans ce manuscrit, ainsi que de proposer plusieurs perspectives d'ouvertures, dans le prochain chapitre.



## 8 Conclusion & perspectives

### 8.1 Conclusion générale

Trois parties composent ce manuscrit. La première partie traite de la méthodologie et de l'état de l'art actuel en prévision numérique par filtrage. L'objectif de cette partie était de définir le cadre dans lequel se déroulerait les travaux de recherche de la thèse et de mettre en lumière les pistes de recherche à explorer.

Après avoir expliqué le fonctionnement de l'EnKF, nous avons effectué l'état de l'art des méthodes et diagnostics de qualité de la prévision existants actuellement pour la prévision numérique par filtrage. Notre volonté étant de pouvoir établir des stratégies d'assimilation qui servent à piloter à la fois les instants d'assimilation et le nombre d'observations assimilées sur la fenêtre, il en est ressorti deux pistes principales : les indicateurs de qualité *a priori* qui comparent directement l'observation et la prévision sans effectuer l'analyse, et les indicateurs de qualité *a posteriori* qui évaluent l'effet des observations sur l'analyse. Les indicateurs *a priori* étant rares dans la littérature actuelle, nous nous sommes orientés vers deux distances probabilistes pour établir nos stratégies d'assimilation, qui sont la distance de Wasserstein (VILLANI, 2009) et la divergence de Kullback-Leibler (GIVENS et SHORTT, 1984). Ces deux distances ont déjà été utilisées dans le cadre de l'assimilation de données pour des problématiques qui diffèrent de la notre (FARCHI et al., 2016 ; GAUTAM et al., 2018). Cet état de l'art nous a aussi permis d'établir un inventaire des indicateurs de qualité de la prévision existants, qui nous servent de référence tout au long de nos recherches, avec en particulier la Hat matrix (CARDINALI, 2004 ; CARDINALI, 2013).

La deuxième partie porte sur la formulation d'indicateurs de qualité de la prévision et la discussion de leur coût numérique, qui guide leur utilisation en conditions expérimentales. Ces indicateurs de qualité se séparent en deux catégories : les indicateurs *a priori* lorsqu'ils sont calculés avant l'assimilation d'une observation, les indicateurs *a posteriori* lorsqu'ils sont calculés directement après l'étape d'analyse. Nous avons formulé un indicateur de sensibilité aux observations (nommé SOBS) qui étudie la sensibilité de l'incrément d'analyse  $\delta \mathbf{x} = \mathbf{x}^a - \mathbf{x}^f$  aux observations assimilées, qui est un indicateur *a posteriori*. Son approche s'inspire de celle de la Hat matrix (CARDINALI, 2004), il la complète en tirant parti des informations provenant des innovations  $\mathbf{y} - \mathbf{H}\mathbf{x}$ . Il a été formulé pour l'utilisation de l'EnKF classique (EVENSEN, 2003) utilisé dans la thèse, ainsi que pour le LETKF de (HUNT, KOSTELICH et SZUNYOGH, 2007). La seconde formulation est la plus prometteuse du fait qu'elle réduise le coût numérique de l'indicateur, et que le formalisme du LETKF soit à la base des filtres actuellement les plus avancés (BOCQUET et SAKOV, 2012 ; BOCQUET, 2016). Ensuite, nous avons adapté l'utilisation des distances probabilistes pour une utilisation *a priori* et une utilisation *a posteriori*, constituant pour la première un outil utile pour établir des stratégies d'assimilation, pour la seconde un diagnostic supplémentaire pour l'analyse.

Nous avons ensuite déterminé les coûts numériques de chaque indicateur, en prenant en



compte les simplifications disponibles selon la nature des matrices de covariance d'erreur, les dimensions du problème et le type d'algorithme utilisé pour l'EnKF. En comparant ces coûts à celui du calcul de l'analyse de l'EnKF décrit dans (EVENSEN, 2003), nous avons alors précisé le type de stratégies d'assimilation le plus opportun à mettre en œuvre dans nos essais numériques. Le coût de l'assimilation étant un ordre de grandeur plus faible que celui des différents indicateurs, ces derniers sont plus propices à calibrer un réseau d'observation, à la fois en terme de répartition spatiale que temporelle des observations.

La troisième partie correspond à la description des essais numériques guidés par les deux parties précédentes, effectués sur de petits systèmes emblématiques ou représentatif de systèmes industriels réels, et l'interprétation de leurs résultats. Elle illustre le choix de raffinement adaptatif que nous avons choisi pour notre processus d'assimilation de données : l'utilisation d'indicateurs servant à établir des stratégies d'assimilation, qui modulent les instants d'assimilation ainsi que le nombre d'observations assimilées à chaque étape d'analyse.

Ces essais consistent d'abord à étudier l'effet de l'utilisation de chaque indicateur. Pour cela, la qualité de la prévision obtenue en appliquant des stratégies d'assimilation est comparée à celle d'une prévision de référence faite en assimilant toutes les observations disponibles sur la fenêtre, sans modifier les instants d'assimilation. Nous montrons ainsi sur des systèmes dynamiques emblématiques de petite dimension une faible diminution de la précision de l'analyse, parfois malgré la suppression de la moitié des observations disponibles. Nous vérifions ainsi la pertinence de nos indicateurs et confirmons leur utilisation pour l'établissement de stratégies d'assimilation guidées par la partie II. Nous exhibons sur tous nos systèmes d'essais une relation de monotonie entre le paramètre de seuil servant à définir les stratégies d'assimilation *a priori* et la moyenne de l'erreur de prévision obtenue sur la fenêtre d'assimilation, montrant qu'il est possible d'évaluer l'effet direct de l'application d'une stratégie d'assimilation sur l'erreur de prévision commise. Ce résultat est particulièrement intéressant du point de vue opérationnel. Nous discutons également d'une simplification possible de nos indicateurs dans le cas où le terme dominant dans le calcul des distances probabilistes est celui du résidu  $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\mathbf{M}^{-1}}$ , où  $\mathbf{M}$  est symétrique définie-positive : dans ce cas, il suffit d'approximer les distances probabilistes par leur terme de résidu, ce qui réduit considérablement leur coût numérique, tout en permettant d'établir des stratégies d'assimilation convenables comme montré lors des essais.

En outre, lorsqu'il s'agit du critère de qualité de la prévision qui nous intéresse, nous montrons sur plusieurs systèmes la possibilité de maintenir le changement des suivis de régime en réduisant fortement le nombre d'observations assimilées. La réduction du nombre d'analyse peut aller jusqu'à 10% du nombre d'observations disponibles sur la fenêtre d'assimilation, permettant une nette amélioration en terme de coût numérique. Un dernier résultat présente la capacité des distances probabilistes à s'adapter à la dynamique d'évolution du système pour définir des instants d'assimilation en accord avec la variation de l'état du système, en particulier pour le suivi des changements de régime. Cette propriété intéresse EDF pour une possible application sur des systèmes industriels.

Les travaux de recherche théoriques de cette thèse s'inscrivent dans la lignée des nombreux travaux déjà effectués dans le domaine des FSOI (Forecast Sensitivity to Observation Impact), et leur application dans la ligne directrice de ceux effectués ou à développer en recherche appliquée en assimilation de données au sein d'EDF R&D. Ils montrent

l'utilité des indicateurs qui y sont discutés, ainsi que leur aspect général : leur principe de fonctionnement est le même quel que soit le système étudié. Nous allons maintenant discuter de plusieurs perspectives d'approfondissement et de consolidation des résultats obtenus au niveau de la formulation des indicateurs et/ou de leur incorporation au sein de méthodes de filtrages plus avancées que l'EnKF de (EVENSEN, 2003) comme décrit dans cette thèse. D'autres utilisations possibles des différents indicateurs sont aussi envisagées et discutées, notamment sur plusieurs types de systèmes industriels d'intérêt pour EDF.

## 8.2 Indicateurs dans un filtre itératif

Les filtres et lisseurs itératifs (BOCQUET et SAKOV, 2012 ; BOCQUET, 2016) sont actuellement les méthodes ensemblistes les plus avancées, et commencent à être appliqués sur des problématiques industrielles (DEFFORGE et al., 2019). La formulation d'indicateurs pour la sélection des observations les plus importantes lors des itérations de la boucle interne du 4D-Var incrémental a été décrite par (GRATTON et al., 2015). Une réflexion sur l'insertion de nos indicateurs au sein des filtres itératifs est donc proposée.

En reprenant la description de l'IEnKF donnée en annexe (A.3.2), nous voyons que l'étape d'analyse est un processus itératif qui calcule plusieurs états analysés approximatifs intermédiaires  $\mathbf{x}_{j+1}$  en assimilant à chaque fois l'observation  $\mathbf{y}$ , avec  $\mathbf{x}_0 = \mathbf{x}_f$ . Une fois le processus itératif fini, ces approximations donnent l'état analysé final  $\mathbf{x}^a$ . Un point positif pour l'utilisation des indicateurs pour l'IEnKF vient de sa formulation basée sur l'ETKF, qui implique une utilisation des formulations avec un coût plus faible pour certains de nos indicateurs.

Pour les indicateurs *a priori*, lorsque le processus de minimisation est à l'itération  $j$ , il est possible de déterminer si la distribution de  $\mathbf{y}$  est suffisamment proche de celle de  $\mathbf{x}_j$  et d'en faire un critère pour stopper la minimisation. Le coût de calcul de l'indicateur est à comparer avec celui du critère d'arrêt du processus itératif, ainsi que sa pertinence en comparaison d'un critère d'arrêt tel que défini dans (GRATTON, LAWLESS et NICHOLS, 2007, Eq. (11), p. 5), spécifique à l'algorithme de Gauss-Newton adapté pour le 4D-Var. L'utilisation des indicateurs *a priori* avant le processus itératif pour décider d'assimiler  $\mathbf{y}$ , telle que décrite dans les essais du chapitre (7), est toujours possible. La question du coût de l'indicateur reste encore à se poser.

Pour les indicateurs *a posteriori*, la même utilisation que pour les indicateurs *a priori* est possible, avec les mêmes réserves à propos du coût numérique de ceux-ci. Une remarque supplémentaire reste à faire : l'IEnKF est mis en place dans un contexte de fortes non-linéarités où une assimilation classique non-itérative de l'EnKF fournit une analyse de qualité insuffisante. Chaque itération rapproche l'état analyse de l'optimum de la fonction de coût de l'analyse. Dans ce cas, l'importance des observation pourrait varier au cours des itérations, certaines à l'itération  $j$  pourraient se voir substituées à d'autres à l'itération  $j+1$ , et il semble plus opportun de calculer les indicateurs *a posteriori* à la fin du processus itératif.

### 8.3 Vers la grande dimension

Notre utilisation des différents indicateurs présentés dans cette thèse s'est faite sur des systèmes de petite taille uniquement. Le passage à de plus grandes dimensions aura au moins deux conséquences.

D'une part, le coût numérique de chaque indicateur peut devenir prohibitif, Cf. chapitre (6). C'est peut-être même la "bête noire" de ceux-ci, et la grande dimension ( $> O(10^5)$ ) rend nécessaire d'optimiser chaque opération d'algèbre linéaire sous peine de rendre les calculs infaisables. Pour les systèmes intermédiaires de moyenne dimension ( $O(10^1)$  à  $O(10^5)$ ), le besoin de réduction des coûts est moindre, il semble donc plus intéressant d'identifier les étapes les plus coûteuses en fonction de la configuration du système.

Dans le cas de SOBS, nous avons vu qu'il est possible de réduire son coût en passant par un algorithme du type LETKF. En revanche, aucune méthode de réduction du coût n'a été trouvée pour la Hat matrix et seule une réduction marginale des coûts d'utilisation des distances probabilistes a été montrée en annexe (section A.6), cette réduction ne portant pas sur l'expression même des distances mais sur la mise à jour des matrices sur lesquelles elles seront calculées, pour leur utilisation *a posteriori*. Des travaux supplémentaires sur l'optimisation du coût des différents indicateurs sont donc à envisager avant de les mettre en œuvre sur des systèmes de plus grande taille. Ces travaux devraient porter en priorité sur la réduction des opérations matricielles effectuées pour le calcul des distances probabilistes. Pour la Hat matrix, il semble intéressant d'en chercher une formulation équivalente, à moindre coût que l'actuelle.

Hors distances probabilistes, les étapes potentiellement réductibles les plus coûteuses sont les produits matriciels  $\mathbf{H}\mathbf{P}^f\mathbf{H}^T$  et  $\mathbf{H}\mathbf{R}^{-1}\mathbf{H}^T$ , dont les coûts sont ici déjà calculés en prenant en compte la structure de  $\mathbf{P}^f$  et  $\mathbf{R}$ . La réduction du coût des indicateurs Hat matrix et SOBS semble donc reposer essentiellement sur la structures des matrices et/ou opérateurs intervenant dans leur calcul. L'étude des structures des matrices pour chaque cas d'application et l'agencement des opérations en fonction de celles-ci semble être une piste pour optimiser les coûts numériques.

### 8.4 Utilisation asynchrone des indicateurs, adaptation aux lisseurs

Deux formulations asynchrones ont été proposées, une pour la Hat matrix en sous-section (5.2.2) et une pour SOBS en sous-section (5.3.1). Cela étend la variété des stratégies d'assimilations qu'il est possible de formuler.

L'utilisation asynchrone de SOBS reste dans le même esprit que son utilisation synchrone, c'est-à-dire que l'indicateur va déterminer la sensibilité de l'analyse à la somme des innovations  $\sum_{j=1}^k (\mathbf{y}_j - \mathbf{H}_j \mathbf{x}_j^f) \in \mathbb{R}^p$ . Schématiquement, la sélection d'observations portera donc

sur le vecteur d'observation  $\tilde{\mathbf{y}} = \sum_{j=1}^k \mathbf{y}_j$ . Ainsi, sélectionner une observation à supprimer, soit une dimension de  $\tilde{\mathbf{y}}$  à ne pas assimiler, revient à supprimer une dimension de chacun

des  $\tilde{\mathbf{y}}_j$  sur la fenêtre sans prendre en compte la potentielle variabilité temporelle de l'importance informative de chaque observation.

L'utilisation asynchrone de la Hat matrix est formulée différemment, en construisant la Hat matrix du problème augmenté pour les  $k$  observations disponibles sur la fenêtre. Dans ce cas les  $p$  premières valeurs diagonales de  $\hat{\mathbf{H}}$ , correspondent à l'importance des observations contenues dans  $\mathbf{y}_1$ , et ainsi de suite pour le reste de la Hat matrix  $\in \mathcal{M}_{kp}(\mathbb{R})$ . La sélection d'observations porte cette fois sur le vecteur  $\mathbf{Y} = [\mathbf{y}_1^T, \dots, \mathbf{y}_k^T]^T \in \mathbb{R}^{kp}$ , ce qui permet une sélection plus adaptable des observations à supprimer. Néanmoins, cela implique de construire des matrices de taille plus importantes.

Le formalisme des lisseurs non-itératifs (EVENSEN et VAN LEUWEEN, 2000) étant similaire à celui des filtres asynchrones (SAKOV, EVENSEN et BERTINO, 2010), cette formulation est utilisable pour du lissage. Le passage aux lisseurs itératifs peut se faire en s'inspirant des pistes proposées en section (8.2), en ayant toutefois conscience qu'il faille fournir un effort supplémentaire pour adapter ce qui est proposé dans cette section au formalisme des lisseurs itératifs.

Dans un premier temps, des essais supplémentaires sont à effectuer sur un système asynchrone en utilisant ces deux indicateurs. Les performances obtenues seront à comparer avec celles d'une utilisation synchrone. Il serait également intéressant de comparer les deux approches asynchrones entre elles, pour exhiber les avantages et inconvénients de chacune.

Dans un second temps, l'adaptation de ces deux indicateurs pour des lisseurs non-itératifs puis itératifs permettrait de mieux explorer l'ensemble des possibilités qu'ils offrent.

## 8.5 Utilisation pour l'organisation spatio-temporelle des observations

Plusieurs essais ont montré un comportement intéressant sur différents indicateurs : pour la distance de Wassertein aux sous-sections (7.5.2) et (7.5.3), pour la Hat matrix et SOBS en sous-section (7.2). Ces comportements peuvent servir à agencer en temps et en espace les observations à assimiler.

Par exemple, pour un système périodique, la méthodologie serait de sur-observer le système en augmentant fortement la densité spatiale et temporelle des observations sur une première période, puis d'effectuer une première passe d'assimilations en calculant à chaque fois la distance de Wasserstein entre l'observation et la prévision, ainsi que la Hat matrix après chaque analyse. Cela permet de déterminer à la fois les instants d'assimilation opportuns pour l'agencement en temps, ainsi que les observations les plus pertinentes de chaque vecteur d'observation pour l'agencement spatial. Ensuite, de définir un seuil correspondant au pourcentage de vecteurs d'observations que nous voulons retirer du réseau en calculant les différents quantiles de la distribution de valeurs de distances obtenues. Les vecteurs d'observation à assimiler lors des prochaines périodes seront alors ceux contenus dans le quantile que nous choisissons, correspondant au pourcentage voulu. Du fait que l'indicateur basé sur la distance de Wasserstein s'adapte à la dynamique du système, cela

privilégiera les observations situées aux zones où la dynamique du système varie fortement. De même, il est possible de sélectionner les observations les plus importantes à assimiler pour chaque vecteur d'observation grâce aux retours de la Hat matrix et/ou de SOBS, pour réduire le coût de chaque étape d'analyse.

Le surcoût ajouté par la calibration sur la première période sera rapidement lissé au bout de quelques périodes grâce à la réduction du nombre d'étapes d'analyses à calculer sur chacune. Le lissage du surcoût sera d'autant plus rapide que le nombre d'observations supprimées par période sera élevé.

Pour un système non-périodique, du fait qu'il n'y ait aucune garantie de répétition des différents régimes que le système peut décrire, il semble préférable de se contenter de l'utilisation de la Hat matrix et/ou de SOBS. Ils permettront quoiqu'il arrive de déterminer les observations les plus importantes à assimiler.

Pour une utilisation synchrone, les deux utilisations mentionnées précédemment permettent de sélectionner en espace les observations à assimiler. Pour faire apparaître la dimension temporelle dans la Hat matrix, l'utilisation asynchrone semble la plus utile. En effet, dans ce cas il est possible de comparer l'importance globale de chaque vecteur d'observation  $\mathbf{y}_k$  présent sur la fenêtre, au travers de la Hat matrix augmentée définie par l'équation (5.2). Par exemple, cela peut se faire en sommant l'importance de chacune des observations présentes dans les différents vecteurs, et de comparer les sommes obtenues, et de ne pas assimiler le vecteur avec la somme la plus faible. Ce critère n'est donné qu'à titre d'exemple, des travaux supplémentaires pour en définir un qui soit plus fin serait une piste intéressante.

## 8.6 Essais sur des systèmes en configuration industrielle

L'utilisation de ces indicateurs dans le cadre du filtrage sur des systèmes industriels dans lesquels il y a une grande différence entre les intervalles d'observation par rapport au temps de simulation du système intéresse particulièrement EDF. Ces systèmes se retrouvent dans plusieurs domaines variés de la physique.

Un premier exemple provient de l'hydrologie pour l'étude des niveaux de cours d'eau. Un modèle simplifié est souvent utilisé ce qui donne un modèle unidimensionnel voire ponctuel, dont le temps de simulation est de l'ordre de la seconde, tandis que les mesures peuvent arriver avec une fréquence de l'ordre d'une heure, d'un jour, voire au-delà. De plus, ces mesures peuvent correspondre par exemple à un relevé ponctuel du niveau du cours d'eau sur l'ensemble de sa longueur. Les mesures présentent d'importantes hétérogénéités temporelles et spatiales. Le système est dans ce cas très sensible aux observations assimilées. L'utilisation des indicateurs dans ce cas va changer par rapport à ce qui été décrit dans la thèse. Il s'agirait d'en faire des diagnostics complémentaires qui porteront sur l'analyse dans un objectif de vérification du résultat obtenu, de la qualité des mesures assimilées, et éventuellement pour calibrer le réseau de capteurs.

Un deuxième exemple provient de l'hydraulique, dont l'objet d'étude est similaire à l'hydrologie, mais porte cette fois sur une modélisation en trois dimension d'un cours d'eau complet ou d'une portion de celui-ci. Ici, la simulation a une durée de l'ordre de l'heure, alors que les mesures peuvent arriver avec une fréquence de l'ordre de la seconde. La répartition spatiale des mesures est très hétérogènes, car celles-ci sont concentrées autour des ouvrages de génie civil comme les barrages, les centrales nucléaires... L'hétérogénéité porte aussi sur la nature physique des différentes mesures qui peuvent être homogènes à des débits ou des températures par exemple. Dans ce cas, il est intéressant de pouvoir sélectionner les mesures les plus pertinentes par rapport à la simulation calculée pour au moins deux raisons :

- la réduction du nombre d'observations à assimiler, qui semble ici particulièrement important,
- la réduction du nombre de capteurs nécessaire à installer sur les différents types de structures pour ne conserver que les plus utiles.

Les systèmes de mécanique des structures possèdent des propriétés similaires. Par exemple, les études de fatigue nécessitent un nombre répété de simulations d'un ordre de grandeur de  $10^4$ , avec des observations ayant une fréquence annuelle dans le cas de l'étude des vieillissements du béton, dont l'étude se fait sur une durée de 50 ans. Ces systèmes font également de bons candidats à l'utilisation des indicateurs.

Un troisième type de systèmes se situant à l'intermédiaire entre les deux précédents sont les simulations neutroniques pour l'intérieur du cœur du réacteur. Les simulations ont une durée de l'ordre de la dizaine de minutes et les observations sont régulières, tout en étant encore une fois hétérogènes en nature et en espace. Au même titre que le modèle de l'empoisonnement au Xénon-135 sur lequel ont été faits plusieurs essais dans cette thèse, des études plus poussées sur ces autres modèles de neutroniques constituent une perspective intéressante pour EDF.



# A Annexe

## A.1 Calculs de la définition du BLUE

Calcul de  $\epsilon^a$  :

$$\begin{aligned}
 \epsilon^a &= \mathbf{x}^a - \mathbf{x}^t = \mathbf{L}(\mathbf{x}^b + \mathbf{x}^t - \mathbf{x}^t) + \mathbf{K}(\mathcal{H}(\mathbf{x}^t) + \epsilon^{obs}) - \mathbf{x}^t \\
 &= \mathbf{L}(\mathbf{x}^b + \mathbf{x}^t - \mathbf{x}^t) + \mathbf{K}(\mathbf{H}\mathbf{x}^t + \epsilon^{obs}) - \mathbf{x}^t \\
 &= \mathbf{L}\epsilon^b + \mathbf{K}\epsilon^{obs} + (\mathbf{L} + \mathbf{K}\mathbf{H} - \mathbf{I})\mathbf{x}^t.
 \end{aligned} \tag{A.1}$$

Calcul du biais de l'erreur d'analyse :

$$\begin{aligned}
 \mathbb{E}[\epsilon^a] &= \mathbb{E}[\mathbf{x}^a - \mathbf{x}^t] = 0 \\
 &= \mathbf{L}\mathbb{E}[\epsilon^b] + \mathbf{K}\mathbb{E}[\epsilon^{obs}] + (\mathbf{L} + \mathbf{K}\mathbf{H} - \mathbf{I})\mathbb{E}[\mathbf{x}^t] \\
 &= (\mathbf{L} + \mathbf{K}\mathbf{H} - \mathbf{I})\mathbb{E}[\mathbf{x}^t].
 \end{aligned} \tag{A.2}$$

Calcul de  $\mathbf{P}^a$  :

$$\begin{aligned}
 \mathbf{P}^a &= \mathbb{E}[(\mathbf{L}\epsilon^b + \mathbf{K}\epsilon^{obs})(\mathbf{L}\epsilon^b + \mathbf{K}\epsilon^{obs})^T] \\
 &= \mathbb{E}[\mathbf{L}(\epsilon^b)(\epsilon^b)^T\mathbf{L}^T] + \mathbb{E}[\mathbf{K}(\epsilon^{obs})(\epsilon^{obs})^T\mathbf{K}^T] \\
 &= \mathbf{L}\mathbf{B}\mathbf{L}^T + \mathbf{K}\mathbf{R}\mathbf{K}^T \\
 &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T, \quad \mathbf{L} = \mathbf{I} - \mathbf{K}\mathbf{H}.
 \end{aligned} \tag{A.3}$$

Calcul de la dérivée :

$$\nabla f(\mathbf{K}) = \nabla \langle \mathbf{I} - \mathbf{K}\mathbf{H}, \mathbf{B}(\mathbf{I} - \mathbf{K}\mathbf{H}) \rangle + \nabla \langle \mathbf{K}, \mathbf{R}\mathbf{K} \rangle \tag{A.4}$$

$$= -2\mathbf{B}\mathbf{H}^T + 2\mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T + 2\mathbf{K}\mathbf{R}. \tag{A.5}$$

Calcul du gain optimal  $\mathbf{K}^*$  :

$$\begin{aligned}
 \nabla f(\mathbf{K}^*) &= 0 \\
 \implies \mathbf{K}^*(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T) &= \mathbf{B}\mathbf{H}^T \\
 \implies \mathbf{K}^* &= \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}
 \end{aligned}$$



## A.2 Méthodes variationnelles

### A.2.1 3D-Var

La méthode 3D-Var est une méthode statique. Pour l'appliquer, il n'y a besoin que de l'ébauche et des observations au temps courant  $t_k$ . Sa fonction de coût associée est

$$J_{3D-Var}(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{y} - \mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - \mathcal{H}(\mathbf{x})). \quad (\text{A.6})$$

Il s'agit d'une somme pondérée entre les écarts à l'ébauche et les écarts aux observations. La pondération est faite grâce aux matrices  $\mathbf{B}$  et  $\mathbf{R}$  décrites précédemment : elles permettent d'accorder plus ou moins de poids à l'un des deux écarts en fonction de la confiance que nous avons en l'ébauche ou les observations.

Cette méthode consiste donc à chercher le compromis entre ébauche et observations qui permet à l'état de se rapprocher au mieux des données au temps courant, tout en tenant compte des incertitudes sur ces dernières.

Nous remarquons ici que l'opérateur  $\mathcal{H}$  peut être non-linéaire, à la différence du filtre de Kalman. De plus, il est possible d'ajouter sous forme de contraintes sur  $J$  des vérifications de lois physiques comme une loi de conservation par exemple.

Elle a notamment été longtemps utilisée en météorologie avant d'être supplantée par son homologue non-séquentiel, la méthode 4D-Var, qui sera décrite juste après. Cependant, il existe plusieurs difficultés pour l'application de cette méthode. Dans un premier temps, minimiser la fonction de coût nécessite le calcul de son gradient :

$$\nabla J_{3D-Var}(\mathbf{x}) = \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y} - \mathcal{H}(\mathbf{x})). \quad (\text{A.7})$$

$\mathbf{H}^T \in \mathcal{M}_{p,n}(\mathbb{R})$  représente l'adjoint de l'opérateur  $\mathbf{H}$ , linéarisé de  $\mathcal{H}$  lorsque ce dernier est non-linéaire. Lorsque  $\mathcal{H}$  (de facto  $\mathbf{H}$ ) est un modèle de système physique, il peut devenir laborieux de calculer le gradient, à la fois pour des problèmes de coût numérique que pour des problèmes de développement.

La modélisation des matrices  $\mathbf{B}$  et  $\mathbf{R}$  est aussi souvent laborieuse, plusieurs travaux de recherches sont en cours à EDF sur cette thématique.

Dernier point : si  $\mathcal{H}$  est non-linéaire, la convexité du problème est mise à mal et il est possible de rencontrer des minima locaux lors du processus de minimisation.

$\mathbf{B}$  et  $\mathbf{R}$  étant symétriques définies positives, lorsque  $\mathcal{H} \equiv \mathbf{H}$  est linéaire, le problème est convexe et la méthode 3D-Var correspond à l'estimateur BLUE discuté auparavant en sous-section (2.1.4). Nous obtenons ainsi la matrice de covariance d'erreur d'analyse par la formule :

$$\mathbf{P}^a = \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}. \quad (\text{A.8})$$

### A.2.2 4D-Var

Cette méthode a été introduite pour la première fois dans (LE DIMET et TALAGRAN, 1986). Une bonne phrase introductive pour la méthode 4D-Var vient de la thèse de (PONÇOT, 2008) : "la méthode 4D-Var est à la méthode 3D-Var ce que le filtre de Kalman est à l'estimation optimale : elle introduit une dimension temporelle dans l'assimilation

des données". Cela signifie qu'on ne cherche plus à recalculer seulement un état à un temps donné mais toute une trajectoire sur une fenêtre de temps donnée, appelée *fenêtre d'assimilation*, en assimilant toutes les observations présentes à l'intérieur de celle-ci. Ce n'est donc plus une méthode séquentielle mais une méthode temporelle : l'ensemble des observations disponibles sur la fenêtre d'assimilation sont prises en compte lors de la correction, un état peut donc être mis à jour avec des observations qui lui sont futures. Le 4D-Var est donc un *lisneur*.

En notant  $n_{obs}$  le nombre d'observations disponibles sur la fenêtre, la fonction de coût du 4D-Var s'écrit :

$$J_{4D-Var}(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2} \sum_{i=0}^{n_{obs}-1} (\mathbf{y}_i + \mathcal{H}_i(\mathcal{M}_{0 \rightarrow i}(\mathbf{x})))^T \mathbf{R}_i^{-1}(\mathbf{y}_i + \mathcal{H}_i(\mathcal{M}_{0 \rightarrow i}(\mathbf{x}))) \quad (\text{A.9})$$

où les opérateurs  $\mathcal{M}_{0 \rightarrow i}$  font évoluer l'état du temps  $t_0$  au temps  $t_i$  sur la fenêtre. Il s'agit d'une somme d'écart aux observations à différents temps et à l'ébauche, et lorsque  $n_{obs} = 1$ , nous retombons bien sur  $J_{3D-Var}$ , car  $\mathcal{M}_{0 \rightarrow 0} = \mathbf{Id}$ .

Les matrices de covariance d'erreur et l'opérateur d'observation peuvent dépendre du temps, c'est pourquoi ils sont indexés par  $i$ .

Comme pour le 3D-Var, la minimisation se fait en général via une méthode itérative et nous devons donc calculer le gradient de la fonction de coût dont la formule est :

$$\nabla J_{4D-Var}(\mathbf{x}) = \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \sum_{i=0}^{n_{obs}-1} (\mathbf{H}_{|\mathbf{x}} \mathbf{M}_{0 \rightarrow i|\mathbf{x}})^T \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathcal{H}(\mathcal{M}_{0 \rightarrow i}(\mathbf{x}))). \quad (\text{A.10})$$

$\mathbf{M}_{0 \rightarrow i|\mathbf{x}}$  correspond au linéaire tangent de l'opérateur d'évolution du modèle calculé en  $\mathbf{x}$ . Le calcul de ce gradient nécessite de calculer deux opérateurs adjoints. Si cela n'a pas été pris en compte au moment de la conception du code permettant de l'implémenter, la méthode devient difficile à utiliser : la modification des gros codes industriels qui peuvent l'utiliser demande une quantité de travail conséquente.

### A.2.3 4D-Var à contraintes faibles

Dans le domaine des sciences atmosphériques, le 4D-Var habituellement mis en place considère le modèle comme parfait, c'est ce qui lui vaut le nom de "4D-Var à contraintes fortes". Lorsque nous relâchons cette hypothèse, nous prenons alors nécessairement en compte une certaine quantité d'erreur dans le modèle, au même titre que pour les autres éléments du processus d'assimilation. C'est ce qui donne le 4D-Var à contraintes faibles. Lorsque le modèle évolue selon  $\mathbf{x}_{k+1} = \mathcal{M}_{k+1}(\mathbf{x}_k) + \eta_{k+1}$ , où  $\eta_{k+1}$  est un terme d'erreur, il suffit d'ajouter un terme supplémentaire prenant la prenant en compte à la fonction de coût (équation A.9) pour obtenir la formulation avec le terme de forçage de l'erreur modèle en variable de contrôle :

$$J^Q(\eta) = \frac{1}{2} \sum_{i=1}^{n_{obs}} \eta_i^T \mathbf{Q}_i^{-1} \eta_i. \quad (\text{A.11})$$

Ou bien en prenant directement en compte les écarts entre l'état  $\mathbf{x}_i$  par rapport à l'estimation obtenue par le modèle  $\mathcal{M}_i(\mathbf{x}_{i-1})$ , pour obtenir la formulation avec l'état en variable de contrôle :

$$J^Q(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^{n_{obs}} (\mathbf{x}_i - \mathcal{M}_i(\mathbf{x}_{i-1}))^T \mathbf{Q}_i^{-1} (\mathbf{x}_i - \mathcal{M}_i(\mathbf{x}_{i-1})). \quad (\text{A.12})$$

Ces formulations sont issues de (TRÉMOLET, 2006a; TRÉMOLET, 2006b). Leur prise en compte nécessite ensuite une légère modification de l'algorithme de minimisation. Un gradient supplémentaire par rapport à  $\eta$  doit être calculé pour obtenir les conditions d'optimalité au premier ordre.

Le type d'erreur modèle corrigé est en général une erreur de biais dans le modèle (VIDARD, PIACENTINI et LE DIMET, 2004; TRÉMOLET, 2010; LALOYAUX et BONAVIDA, 2020). Cette erreur est systématique dans le sens où elle apparaît à chaque fois que le modèle est appliqué, et peut être globale ou bien spatialement localisée à certaines régions de l'atmosphère comme en météorologie, ainsi qu'indépendante des différents cycles d'assimilation. Corriger des erreurs plus complexes nécessite une paramétrisation de la matrice  $\mathbf{Q}$  encore en travaux, augmente la difficulté du problème de minimisation. De plus, le savoir sur la manière de prendre en compte des erreurs stochastiques est encore lacunaire.

D'après (LALOYAUX et BONAVIDA, 2020) et (TRÉMOLET, 2010), les implémentations actuelles à l'ECMWF du 4D-Var à contraintes faibles n'apportaient pas d'améliorations significatives et l'assimilation était même restreinte à une certaine zone de l'atmosphère car une dégradation trop importante de l'analyse était autrement observée.

Cependant les recherches continuent, et les travaux récents de Laloyaux pour mettre à jour l'implémentation du 4D-Var à contraintes faibles apportent une réduction de 50% du biais de température sur les zones à forte convection. De plus, des travaux d'hybridation du 4D-Var avec des méthodes ensemblistes pour l'estimation des matrices d'erreurs d'ébauche et de modèle semblent prometteuses dans ce domaine.

En outre, les résultats des méthodes d'ensemble notamment l'IEEnKS qui est aussi un lisseur viendront alimenter la recherche sur l'erreur modèle et seront transposables au 4D-Var à contraintes faibles.

## A.3 Améliorations de l'EnKF

### A.3.1 Inflation

Le mécanisme d'inflation est avant tout utile au bon fonctionnement de l'EnKF, mais dans le cas de l'inflation multiplicative par exemple, le choix du coefficient  $\alpha > 1$  idéal à chaque analyse nécessite paramétrage laborieux qui dépend du cas d'application. De plus, ce paramétrage est aussi nécessaire lorsque l'inflation se fait directement sur les membres de l'ensemble.

Plusieurs travaux de recherche ont été effectués pour faire en sorte d'éviter le processus de paramétrage et de trouver un facteur optimal à moindre effort.

#### Travaux de Anderson sur l'inflation adaptative

Ces travaux semblent avoir débuté avec (ANDERSON, 2007), dans le cadre de l'inflation de covariance dans l'espace des états.

Cette méthode d'inflation consiste à transformer linéairement chaque membre de l'ensemble par un facteur  $\lambda > 1$ . Pour  $\mathbf{E} \in \mathcal{M}_{n,N}(\mathbb{R})$ ,  $i \in \llbracket 1, n \rrbracket$  l'indice des composantes de chaque membre, et  $j \in \llbracket 1, N \rrbracket$  l'indice des membres de  $\mathbf{E}$ , il s'agit d'effectuer l'opération :

$$\mathbf{x}^j(i) = \sqrt{\lambda} (\mathbf{x}^j(i) - \bar{\mathbf{x}}(i)) + \bar{\mathbf{x}}(i), \quad (\text{A.13})$$

où  $\bar{\mathbf{x}}(i)$  représente la  $i^{\text{eme}}$  composante de la moyenne de l'ensemble. Cette opération se fait juste avant l'assimilation.

L'idée est que l'inflation de covariance peut être vue comme un modèle simple de correction de la sous-estimation de variance de l'EnKF. Le facteur d'inflation  $\lambda$  correspond donc à un vecteur d'état unidimensionnel de ce modèle de correction de variance d'erreur, que l'on peut estimer grâce au formalisme bayésien et aux observations.

L'inflation adaptative de covariance revient ainsi à augmenter la dimension du problème d'assimilation avec le facteur d'inflation, soit passer d'un état  $\in \mathbb{R}^n$  à un état  $\in \mathbb{R}^{n+1}$ , ce qui permet dynamiquement d'en obtenir une valeur optimale, à un surcoût faible en comparaison du coût d'une assimilation classique, et pour des performances équivalentes à celles obtenues avec un  $\lambda$  bien paramétré. Cependant, cela nécessite de choisir une bonne modélisation initiale de la distribution de  $\lambda$  pour être efficace.

Une possibilité d'amélioration mentionnée par l'auteur est de déterminer quels membres et même quelles dimensions des membres de l'ensemble nécessitent de l'inflation, car dans la pratique, les variables d'états ne suivent pas la même dynamique et peuvent avoir des propriétés bien différentes les unes des autres. C'est ce qui a constitué la suite des travaux d'Anderson dans (ANDERSON, 2009).

L'inflation s'effectue cette fois en temps et en espace, en prenant  $\lambda = [\lambda_1, \dots, \lambda_N] \in \mathbb{R}^N$  où chaque  $\lambda_i$  est associé à un membre de l'ensemble  $\mathbf{x}^i$ . Le principe est ensuite le même : il suffit d'augmenter le problème d'assimilation de données avec  $\lambda$ . L'utilisation de cet algorithme montre son intérêt lors de son application au "National Center for Atmospheric Research Community Atmospheric Model" nord-américain, où un ordre de grandeur d'écart se retrouve sur les facteurs d'inflation pour des zones géographiquement éloignées. Cela permet une amélioration de la précision des simulations obtenues.

Cependant, certaines limitations sont encore présentes. L'inflation multiplicative comme décrite ici corrige essentiellement les erreurs d'échantillonnage de l'ensemble, et il n'est pas encore bien clair si elle doit être effectuée juste avant l'analyse ou bien juste après pour permettre des prévisions avec une variance d'erreur de meilleure qualité.

Déterminer le moment le plus opportun où effectuer l'inflation est l'objet de travaux de recherches, la correction d'erreurs au-delà de celles d'échantillonnage trouvera peut-être sa solution dans l'inflation additive qui pallie mieux les déficiences de modélisation mais qui nécessite des efforts de mise en place plus conséquent pour son application sur l'ensemble, à la différence de l'inflation multiplicative (formule A.13). Une hybridation des deux méthodes pourrait fournir des algorithmes adaptés à tout type de système.

Il est à noter qu'une bonne détermination de  $\lambda$  permet une meilleure connaissance des statistiques d'erreur modèle associée au système et pourrait servir à l'amélioration de la modélisation du système physique étudié (ANDERSON, 2007, Sec. 8).

### EnKF-N (Bocquet, 2011)

Comme vu précédemment, les erreurs d'échantillonnage sont un défaut intrinsèque aux méthodes d'ensemble. L'objectif de (BOCQUET, 2011) est de formuler une variante de l'EnKF plus robuste par rapport à cela, en prenant une voie différente de celle empruntée par Anderson.

L'analyse de l'EnKF est faite en supposant la distribution de l'ensemble gaussienne, donc déterminée entièrement par sa moyenne et sa matrice de covariance. L'objectif de l'EnKF-N est de prendre en compte au mieux la nature discrète et non nécessairement gaussienne de l'ensemble en tant que distribution a priori lors de la mise à jour.

A l'étape d'analyse, l'a priori choisi  $p(\mathbf{x}|\bar{\mathbf{x}}, \mathbf{P})$  est celui du vecteur d'état  $\mathbf{x}$ , conditionné sur la moyenne empirique de l'ensemble  $\bar{\mathbf{x}}$  et sa matrice de covariance empirique associée  $\mathbf{P}$ , et supposé gaussien.

Du fait que  $\bar{\mathbf{x}}$  et  $\mathbf{P}$  soient échantillonnées, l'hypothèse gaussienne sur  $p(\mathbf{x}|\bar{\mathbf{x}}, \mathbf{P})$  conduit à une perte d'information. L'auteur cherche donc plutôt à estimer  $p(\mathbf{x}|\mathbf{x}^1, \dots, \mathbf{x}^N)$ , qui est l'a priori sur  $\mathbf{x}$  conditionné par les membres de l'ensemble. Chaque membre de l'ensemble est supposé être tiré depuis une distribution gaussienne d'ébauche  $\mathcal{N}(\mathbf{x}^b, \mathbf{B})$ . Ainsi, en intégrant sur l'ensemble des possibles vecteurs  $\mathbf{x}^b$  et matrices définies-positives  $\mathbf{B}$ , nous avons la distribution de l'état conditionnellement à l'ensemble qui vaut :

$$p(\mathbf{x}|\mathbf{x}^1, \dots, \mathbf{x}^N) = \int p(\mathbf{x}|\mathbf{x}^b, \mathbf{B})p(\mathbf{x}^b, \mathbf{B}|\mathbf{x}^1, \dots, \mathbf{x}^N)d\mathbf{x}^b d\mathbf{B}. \quad (\text{A.14})$$

$p(\mathbf{x}|\mathbf{x}^b, \mathbf{B})$  correspond à la densité de probabilité de l'état conditionné par l'ensemble tiré selon  $\mathcal{N}(\mathbf{x}^b, \mathbf{B})$ , et  $p(\mathbf{x}^b, \mathbf{B}|\mathbf{x}^1, \dots, \mathbf{x}^N)$  à la vraisemblance de l'ébauche relativement à l'ensemble.

En appliquant la règle de Bayes à la vraisemblance, nous obtenons alors :

$$p(\mathbf{x}|\mathbf{x}^1, \dots, \mathbf{x}^N) = \frac{1}{p(\mathbf{x}^1, \dots, \mathbf{x}^N)} \int p(\mathbf{x}|\mathbf{x}^b, \mathbf{B})p(\mathbf{x}^1, \dots, \mathbf{x}^N|\mathbf{x}^b, \mathbf{B})p(\mathbf{x}^b, \mathbf{B})d\mathbf{x}^b d\mathbf{B}. \quad (\text{A.15})$$

$p(\mathbf{x}^b, \mathbf{B})$  est l'a priori sur l'ébauche. Cette forme est plus utile car les probabilités conditionnelles aux gaussiennes s'explicitent facilement. Il ne manque plus qu'à choisir un a priori pour pouvoir tout calculer, dont le choix est discuté en détails dans (BOCQUET, 2011, Sec. 2.2).

En voyant alors l'analyse sous l'angle d'un problème d'optimisation, cette mise en perspective permet une expression différente de la contribution de l'ébauche  $J^b$  à la fonction de coût du 3D-Var équivalente :

$$J^a(\mathbf{x}) = J^{obs}(\mathbf{x}) + J^b(\mathbf{x}) = (\text{A.6}), \quad (\text{A.16})$$

avec  $J^{obs}(\mathbf{x}) = \frac{1}{2}(\mathbf{y} - \mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - \mathcal{H}(\mathbf{x}))$ .

Un exemple de terme d'ébauche issu de l'article est donné par :

$$\mathcal{J}^b(\mathbf{x}) = \frac{N}{2} \ln \left| \frac{N}{N+1}(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T + (N-1)\mathbf{P} \right|. \quad (\text{A.17})$$

Cette modification rend la fonction  $J^a$  non-convexe. Calculer l'état analysé nécessitera donc d'approcher le minimum de (A.16) grâce à un algorithme de minimisation, à la différence de l'analyse de l'EnKF classique qui est explicite. Cette différence peut introduire un surcoût : algorithme à mettre en place, le processus de minimisation peut rencontrer des minimas locaux, ... De plus, la précision et les temps de calcul vont dépendre du choix du point de départ de l'algorithme.

Ce nouvel a priori pour l'analyse définit une nouvelle classe de filtres permettant de prendre en compte l'incertitude sur les statistiques empiriques obtenues via l'ensemble, présente du fait de l'échantillonnage.

En comparaison à l'ETKF (BISHOP, ETHERTON et MAJUMDAR, 2001), l'ETKF-N donne des résultats similaires sur des applications classiques, et même meilleurs lorsqu'un certain délai est franchi entre les analyses sur le Lorenz 63 (LORENZ, 1963), et sans nécessiter d'inflation. En revanche, les résultats sont mitigés lorsque l'on passe à une dimension plus élevée avec le Lorenz 96 (LORENZ, 2002), où une localisation ainsi qu'une inflation sont parfois nécessaires pour éviter la divergence du filtre.

Dans un contexte applicatif où l'erreur modèle et les non-linéarités mènent à des dérives par rapport au modèle Gaussien pour les statistiques, cet EnKF-N réduisant fortement le besoin d'inflation pourrait permettre une quantification de l'erreur modèle ou de l'impact des non-linéarités lorsque justement le besoin d'inflation se fait ressentir.

Cet article est avant tout une *preuve de concept* et permet d'explorer une nouvelle voie pour le développement de méthodes pour prendre en compte des effets souvent négligés par commodité dans les algorithmes ensemblistes classiques, notamment l'erreur modèle. Il est intéressant de remarquer que les travaux sur l'inflation mènent – pour deux approches différentes – à de nouvelles idées sur la caractérisation des statistiques d'erreur modèle des systèmes étudiés. Une amélioration de ce concept a ensuite été proposée dans (BOCQUET, RAANES et HANNART, 2015), qui permet un nouveau choix de distribution a priori  $p(\mathbf{x}^b|\mathbf{B})$  plus informatif, résultant en de meilleures statistiques d'erreur.

### EnKS : lisseur de Kalman d'ensemble

Les lisseurs sont des méthodes temporelles, comme le 4D-Var. Ils vont rétro-propager les informations des observations assimilées à chaque point de la trajectoire, qui aura connaissance à la fois du passé, présent et futur de son propre temps, ce qui aura finalement pour effet de lisser la trajectoire et d'améliorer les statistiques d'erreurs tout au long de celle-ci.

Ceci peut se formaliser à l'aide du point de vue bayésien et de l'évolution de la densité de probabilité conditionnelle de l'état. Avec  $\mathbf{x}_k$  l'état à  $t_k$  et  $\mathbf{y}_k$  le vecteur d'observation associé, notons  $p(\mathbf{x}_k|\mathbf{y}_1, \dots, \mathbf{y}_k)$  la probabilité conditionnelle de l'état selon les observations. Dans ce cas, sur la fenêtre  $[t_0, \dots, t_k]$  le fonctionnement d'un filtre avec ébauche  $\mathbf{x}^b$  à  $t_0$  et avec prévision  $\mathbf{x}_1$  à  $t_1$  peut se voir séquentiellement comme suit :

$$p(\mathbf{x}_k|\mathbf{y}_1, \dots, \mathbf{y}_k) \propto p(\mathbf{x}_k|\mathbf{y}_1, \dots, \mathbf{y}_{k-1})p(\mathbf{y}_k|\mathbf{x}_k). \quad (\text{A.18})$$

L'information est propagée dans le temps à  $t_k$  conditionnellement à toutes les observations passées  $\mathbf{y}_1, \dots, \mathbf{y}_k$ .

De manière similaire, en supposant que l'évolution du modèle soit un processus de Markov,

un lisseur complet aura comme évolution :

$$p(\mathbf{x}^b, \dots, \mathbf{x}_k | \mathbf{y}_1, \dots, \mathbf{y}_k) \propto p(\mathbf{x}^b, \dots, \mathbf{x}_{k-1} | \mathbf{y}_1, \dots, \mathbf{y}_{k-1}) p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{y}_k | \mathbf{x}_k). \quad (\text{A.19})$$

Ici l'information est propagée  $\forall t_i \in [t_0, \dots, t_k]$ , chaque point de la trajectoire sera donc modifié. Cependant, l'estimation finale du filtre coïncidera avec celle du lisseur, une seule observation étant disponible à  $t_k$ .

En pratique, il n'est pas possible de lisser la totalité de la trajectoire pour des raisons de coût. C'est pourquoi les "lagged filters" ont été mis en place. Ils ne rétropropagent les nouvelles observations que sur les  $L \in \mathbb{N}^*$  itérés les plus récents. Par exemple, un IEnKF est un lisseur avec un lag de  $L = 1$  : seul le dernier itéré  $\mathbf{x}_k$  va avoir connaissance de la dernière observation.

Cela donne donc une évolution de la distribution légèrement modifiée, à savoir :

$$p(\mathbf{x}_{k-L}, \dots, \mathbf{x}_k | \mathbf{y}_{k-L}, \dots, \mathbf{y}_k) \propto p(\mathbf{x}_{k-L-1}, \dots, \mathbf{x}_{k-1} | \mathbf{y}_{k-L-1}, \dots, \mathbf{y}_{k-1}) \times p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{y}_k | \mathbf{x}_k). \quad (\text{A.20})$$

Les filtres à lag peuvent aussi disposer d'un décalage (shift)  $S \in \mathbb{N}^*$  qui correspond au nombre d'observations assimilées et rétropropagées à chaque nouvelle analyse. C'est le nombre de pas de temps séparant chaque fenêtre d'assimilation. Lorsque  $S < L$ , les fenêtres d'assimilation se chevauchent.

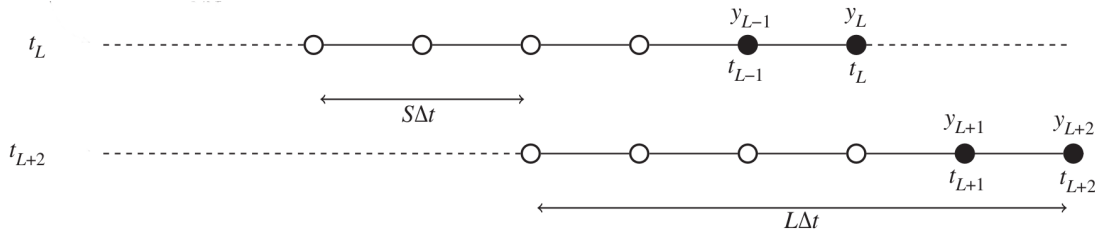


FIGURE A.1 – Adapté depuis (BOCQUET, 2016). Ici,  $S=2$  et  $L=6$  avec un écart de temps  $\Delta t$  en chaque itéré. Seules les observations en noir sont assimilées.

L'avantage principal de la formulation de Kalman d'ensemble pour le lisseur est que cela permet d'éviter l'intégration en temps inverse du modèle pour la rétropropagation des informations assimilées. Cela passe par l'ensemble qui échantillonne les statistiques d'erreurs, et permet de plus de définir la matrice de covariance d'ébauche à partir des données statistiques *a priori*.

Ce même échantillonnage permet de réduire la taille des matrices de covariance d'erreur à stocker, donc le coût de l'analyse. Cependant, les mêmes limitations que l'EnKF dues à la nature ensembliste de la méthode se retrouvent chez l'EnKS, avec de nouvelles difficultés associées à la dimension temporelle.

Pour de la prévision pure, il n'y a pas plus d'intérêt à utiliser un lisseur qu'un filtre. Le lisseur sert avant tout à caractériser la distribution complète de l'état sur toute la fenêtre  $p(\mathbf{x}^b, \dots, \mathbf{x}_k | \mathbf{y}_1, \dots, \mathbf{y}_k)$ , qui permet de raffiner la connaissance des paramètres du système. Ils sont préférentiellement utilisés pour faire de la réanalyse ou de l'estimation de paramètres, aussi connue sous le nom de "recalage de paramètres".

### A.3.2 Méthodes itératives : IEnKF et IEnKS

Le régime non-linéaire de certains systèmes physiques impacte fortement la qualité de l'analyse, du fait que c'est le linéarisé  $\mathbf{H}$  de l'opérateur d'observation  $\mathcal{H}$  qui est utilisé pour former le gain de Kalman de formule (2.28). Ceci encapsule une relation de sensibilité linéaire aux innovations  $\mathbf{y} - \mathcal{H}(\mathbf{x})$ , qui n'est plus valable pour ces systèmes et conduit à des analyses fortement sous-optimales. Une autre source de non-linéarités provient aussi de l'opérateur d'évolution  $\mathcal{M}$ .

Les méthodes itératives sont une solution face à cela. La première méthode itérative est décrite et testée dans (JAZWINSKI, 1970, pp. 349-360), il s'agit de la version itérative du filtre de Kalman étendu : l'IEKF. Cette méthode nécessite cependant toujours une formulation exacte de la matrice de covariance du système, devenant rapidement un facteur limitant pour un système de grande taille. Du fait de la factorisation de la matrice de covariance grâce à l'ensemble, le filtre de Kalman d'ensemble itératif ou "Iterative Ensemble Kalman Filter" (IEnKF) n'a pas ce problème et constitue le prolongement direct de l'IEKF pour les problèmes de grande taille. Plusieurs travaux de mise en œuvre de cette méthode se trouvent chez (GU et OLIVER, 2007; GAOMING et REYNOLDS, 2007). Pour la décrire, nous intéresserons principalement à ceux de (SAKOV, OLIVER et BERTINO, 2012) pour l'IEnKF puis à ceux de (BOCQUET et SAKOV, 2014; BOCQUET, 2016) pour l'IEnKS.

#### IEnKF

L'IEnKF est une méthode d'assimilation de données séquentielle ensembliste qui, comme l'EnKF-N, voit l'étape d'analyse comme un problème d'optimisation.

La fonction de coût à minimiser est celle du 3D-Var (A.6). En reprenant l'équation du gradient du 3D-Var (A.7), la condition d'optimalité du premier ordre s'écrit :

$$\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) - \nabla_{\mathbf{x}}[\mathcal{H} \circ \mathcal{M}(\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{y} - \mathcal{H} \circ \mathcal{M}(\mathbf{x})] = 0. \quad (\text{A.21})$$

Ici,  $\mathcal{M}$  représente l'opérateur d'évolution du modèle pour transporter l'état au temps des observations où se fera l'analyse, et  $\nabla_{\mathbf{x}}$  la Jacobienne par rapport à la variable d'état  $\mathbf{x}$ . Notons  $\mathbf{x}_i$  l'approximation pour l'état analysé  $\mathbf{x}^a$  à la  $i^{\text{eme}}$  itération du processus de minimisation. Pour  $\mathcal{H}$  et  $\mathcal{M}$  suffisamment lisses, et en supposant que les itérés soient suffisamment proches à un certain moment, nous pouvons écrire :

$$\mathcal{M}(\mathbf{x}_{i+1}) = \mathcal{M}(\mathbf{x}_i) + \mathbf{M}_i(\mathbf{x}_{i+1} - \mathbf{x}_i) + O(\|\mathbf{x}_{i+1} - \mathbf{x}_i\|^2) \quad (\text{A.22})$$

$$\nabla_{\mathbf{x}}[\mathcal{H} \circ \mathcal{M}(\mathbf{x}_{i+1})] = \mathbf{H}_i \mathbf{M}_i + O(\|\mathbf{x}_{i+1} - \mathbf{x}_i\|) \quad (\text{A.23})$$

avec  $\mathbf{M}_i = \nabla_{\mathbf{x}} \mathcal{M}(\mathbf{x}_i)$  et  $\mathbf{H}_i = \nabla_{\mathbf{x}} \mathcal{H}(\mathbf{x}_i)$ .

En suivant les calculs de (SAKOV, OLIVER et BERTINO, 2012, Sec. 2, Eq. (4-14)), ces approximations permettent de construire successivement plusieurs approximations de l'analyse avec :

$$\mathbf{B}_{i+1} = [\mathbf{B}^{-1} + (\mathbf{H}_i \mathbf{M}_i)^T \mathbf{R}^{-1} \mathbf{H}_i \mathbf{M}_i]^{-1} \quad (\text{A.24})$$

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \mathbf{K}_i[\mathbf{y} - \mathcal{H} \circ \mathcal{M}(\mathbf{x}_i)] + \mathbf{B}_{i+1}(\mathbf{B})^{-1}(\mathbf{x}^b - \mathbf{x}_i) \quad (\text{A.25})$$

$$= \mathbf{x}^b + \mathbf{K}_i[\mathbf{y} - \mathcal{H} \circ \mathcal{M}(\mathbf{x}_i) + \mathbf{H}_i \mathbf{M}_i(\mathbf{x}_i - \mathbf{x}^b)] \quad (\text{A.26})$$



où les  $\mathbf{B}_i$  et  $\mathbf{x}_i$  représentent respectivement les approximations de la matrice de covariance d'erreur de l'état et de l'analyse, ainsi que :

$$\mathbf{K}_i = \mathbf{B}_{i+1}(\mathbf{H}_i\mathbf{M}_i)^T\mathbf{R}^{-1} \quad (\text{A.27})$$

$$= \mathbf{B}(\mathbf{H}_i\mathbf{M}_i)^T[(\mathbf{H}_i\mathbf{M}_i)\mathbf{B}(\mathbf{H}_i\mathbf{M}_i)^T + \mathbf{R}^{-1}]^{-1} \quad (\text{A.28})$$

qui représente les approximations successives du gain de Kalman associé à chaque itération.

L'adaptation de ces équations au formalisme ensembliste via des matrices d'anomalies est explicité dans les équations (15) à (24) de (SAKOV, OLIVER et BERTINO, 2012).

Le formalisme utilisé ici pour le processus de minimisation reprend la procédure de Newton. Il est cependant possible d'adapter la méthode itérative utilisée en redéfinissant les équations du procédé de minimisation.

Trois régimes de non-linéarité fortes sont décrits et testés sur différents systèmes dans l'article :

1. observations rares mais précises et propagation non-linéaires des anomalies de l'ensemble au début du processus itératif,
2. observations très rares mais très précises où le système devient potentiellement instable,
3. observations fréquentes mais peu précises ne permettant pas une propagation linéaire des anomalies durant la minimisation.

L'IEnKF est adapté aux systèmes vérifiant la 1<sup>ère</sup> définition, mais ceux-ci nécessitent une forme quasi-optimale du problème d'assimilation : une erreur modèle faible voire négligeable, un ensemble de taille suffisante pour les approximations, et un réseau d'observation suffisamment fourni en temps comme en espace.

L'IEnKF surpasse largement l'EnKF à la fois en terme de RMSE, de nécessité moindre d'utiliser l'inflation, et de concordance entre l'évolution de l'écart-type de l'ensemble et de la RMSE. Ceci est d'autant plus flagrant pour des systèmes à grande dimension, lorsque  $N \leq n, p$  et que l'ensemble est de rang déficient.

De manière générale, la prévision obtenue dans un régime fortement non-linéaire reste quoiqu'il arrive nettement plus précise avec l'IEnKF. Malgré le surcoût dû à la minimisation itérative, la prévisibilité du système est a minima maintenue et au mieux améliorée du fait qu'il devient possible – du moins en théorie – de s'attaquer maintenant à des problèmes fortement non-linéaires. Ceci implique donc la possibilité de revoir à la baisse le nombre d'hypothèses simplificatrices faites sur des systèmes actuels, comme par exemple en prévision atmosphérique, et ouvre une nouvelle voie de recherche pour l'assimilation de données dans ce domaine. De plus, un gain est possible sur le nombre de membres de l'ensemble pouvant contrebalancer le surcoût du processus itératif et même permettre dans certains cas de l'éviter.

## IEnKF-N

Dans (BOCQUET et SAKOV, 2012), les auteurs fusionnent leurs deux méthodes en une pour éviter le laborieux paramétrage des paramètres d'inflation lors de l'utilisation de l'IEnKF. L'idée à la base du cumul de ces deux méthodes est que l'EnKF-N va raffiner

l'information apporté par l'ébauche et la connaissance *a priori*, là où l'IEEnKF va raffiner l'information apportée par les innovations.

La formulation de l'IEEnKF-N consiste à détailler plus encore le problème d'optimisation décrit par l'équation (A.16), en le regardant du point de vue Lagrangien pour obtenir le minimum par un élégant procédé faisant appel au problème dual et résolvant ce dernier, après démonstration de son équivalence au problème primal. Il s'agit ensuite de reformuler la fonction de coût d'analyse de l'IEEnKF en y introduisant celle obtenue grâce à l'EnKF-N. Celle-ci n'est pas quadratique mais possède un minimum global (BOCQUET et SAKOV, 2012, Eq. (37)) et sera donc minimisé par un algorithme de Levenberg-Marquardt, plus élaboré mais aussi plus coûteux qu'une méthode de Newton comme utilisée initialement chez (SAKOV, OLIVER et BERTINO, 2012).

Cette méthode est aussi efficace qu'un IEEnKF avec paramétrage d'inflation effectué et plusieurs aspects restent encore à étudier. Elle a permis de poser les bases pour la formulation d'un filtre de Kalman d'ensemble efficace, notamment de l'IEEnKS que nous allons décrire dans la sous-section suivante.

### A.3.3 IEEnKF avec erreur modèle additive : IEEnKF-Q, et IEEnKS

#### IEEnKF-Q

L'IEEnKF classique initialement formulé par (BOCQUET et SAKOV, 2012) faisait l'hypothèse d'un modèle parfait. Le prolongement des travaux de recherche sur cette méthode s'est logiquement orienté vers la prise en compte de l'erreur modèle au sein de la méthode pour améliorer ses performances de simulation.

D'une manière similaire au 4D-Var à contraintes faible, c'est l'erreur additive comme dans les équations du filtre de Kalman (2.29 et 2.30) qui est utilisée, notamment pour sa simplicité. Simplicité surtout apparente car il faut arriver à bien caractériser la matrice  $\mathbf{Q}$ . Le principe repose sur l'ajout à la fonction de coût (A.6) d'un terme d'écart entre la prévision au temps de l'analyse  $\mathbf{x}_{k+1}^f$  (intégrée depuis  $\mathbf{x}_k^a$ ) et l'intégration du modèle depuis la prévision au temps précédent  $\mathcal{M}(\mathbf{x}_k^f)$  :

$$J^m(\mathbf{x}_{k+1}^f, \mathbf{x}_k^f) = \frac{1}{2} \left( \mathbf{x}_{k+1}^f - \mathcal{M}(\mathbf{x}_k^f) \right)^T \mathbf{Q}^{-1} \left( \mathbf{x}_{k+1}^f - \mathcal{M}(\mathbf{x}_k^f) \right), \quad (\text{A.29})$$

et à dériver le problème de minimisation modifié qui en résulte, de la même manière que présenté en sous-section (A.3.2). Cela donne l'IEEnKF-Q.

En présence d'erreur modèle, l'IEEnKF-Q permet d'augmenter les performances par rapports à l'IEEnKF classique et à l'EnKF, notamment en régime fortement non-linéaire où la différence devient importante quand l'amplitude de l'erreur modèle croît. Il bénéficie aussi au mieux d'une réduction de la taille de l'ensemble en étant la méthode qui fournit à chaque fois la RMSE la plus faible en comparaison aux autres, notamment lorsque la localisation telle que décrite dans (BOCQUET, 2016) est mise en place. Nous retrouvons donc ici encore une fois cette propriété pour les méthodes itératives.

Dans le cadre de faibles non-linéarités, ses performances sont légèrement meilleures que les autres méthodes mais peut-être pas assez pour être significatives.

Cette méthode permet donc d’explorer d’avantage l’ensemble des possibilités qu’offre le formalisme de l’IEnKF, et fournit une piste théorique sur la prise en compte de l’erreur modèle. Cependant, l’erreur modèle additive est avant tout un outil théorique. Elle sert de méthode empirique de régularisation du problème de minimisation de l’analyse pour le 4D-Var lorsque celui-ci est appliqué à de longues fenêtres d’assimilation, mais reste peu représentative des erreurs rencontrées en applications réelles.

Quoiqu’il arrive, la prise en compte de l’erreur modèle permet ici une amélioration à la fois du processus d’assimilation menant à la prévision et de la précision de la prévision. Il serait intéressant de voir l’effet que cela a avec une réanalyse sur les nouvelles statistiques d’erreurs obtenues.

## IEnKS

L’IEnKS (BOCQUET et SAKOV, 2014) est l’évolution depuis l’IEnKF et l’IEnKF-N, qui sont des lisseurs avec un retard (lag) de 1 vers un lisseur avec un retard  $L > 1$ . C’est une méthode d’ensemble variationnelle, ne nécessitant pas d’opérateur adjoint pour la propagation en sens temporel inverse des mises à jours à chaque assimilation, ni les opérateurs tangents linéaires des modèles d’évolution et d’observations, car ceux-ci ainsi que les éléments nécessaires à l’optimisation variationnelle sont calculés grâce à l’ensemble. La localisation a été adaptée à cette méthode (BOCQUET, 2016) et une application sur un code industriel d’EDF a suivi dans (DEFFORGE et al., 2019).

L’IEnKS est globalement meilleur en filtrage que l’EnKF pour des régimes moyennement linéaires, amélioration d’autant plus marquée que  $L$  est grand. Il en est de même en comparaison à l’EnKS, à partir d’une certaine valeur de  $L$ .

Toutes performances égales par ailleurs, l’IEnKS nécessite un ensemble trois fois moins grand que l’IEnKF dès lors que les non-linéarités sont suffisamment fortes pour que l’hypothèse gaussienne soit invalidée. Suite aux essais faits pour (DEFFORGE et al., 2019), il a été remarqué que l’IEnKS est particulièrement intéressant lorsqu’on l’utilise sur un système avec lequel les opérateurs tangents linéaires ne sont pas à recalculer à chaque fois, car sa réduction drastique du nombre de membres nécessaire vient le rendre, à performances égales avec les autres méthodes, très compétitif en terme de coût.

D’autres travaux sur un lisseur itératif construit pour de l’hydrologie sont disponibles chez (JU et al., 2018). Il s’agit d’un lisseur itératif formulé différemment de l’IEnKS qui cherche à estimer le lien entre les paramètres du modèle et ses sorties grâce à un processus Gaussien adaptatif.

## A.4 Calculs de l’état de l’art sur les FSOI

Calcul de la formule (3.19) :

Nous rappelons que  $\mathbf{C}$  est une matrice symétrique positive, éventuellement diagonale.  $\langle . \rangle$  représente le produit scalaire canonique  $\mathbf{x}^T \mathbf{y}$ .

$$\delta e(\delta \mathbf{x}) = e(\mathbf{x}^f + \delta \mathbf{x} - \mathbf{x}_t^f) - e(\mathbf{x}^f - \mathbf{x}_t^f) \quad (\text{A.30})$$

$$= (\mathbf{x}^f + \delta \mathbf{x} - \mathbf{x}_t^f)^T \mathbf{C}(\mathbf{x}^f + \delta \mathbf{x} - \mathbf{x}_t^f) - (\mathbf{x}^f - \mathbf{x}_t^f)^T \mathbf{C}(\mathbf{x}^f - \mathbf{x}_t^f) \quad (\text{A.31})$$

$$= \delta \mathbf{x}^T \mathbf{C} \delta \mathbf{x} + \delta \mathbf{x}^T \mathbf{C}(\mathbf{x}^f - \mathbf{x}_t^f) + (\mathbf{x}^f - \mathbf{x}_t^f)^T \mathbf{C} \delta \mathbf{x} \quad (\text{A.32})$$

$$= \langle \delta \mathbf{x}, \mathbf{C}(\mathbf{x}^f - \mathbf{x}_t^f) \rangle + \langle \mathbf{x}^f - \mathbf{x}_t^f, \mathbf{C} \delta \mathbf{x} \rangle + \langle \delta \mathbf{x}^T, \mathbf{C} \delta \mathbf{x} \rangle \quad (\text{A.33})$$

$$= \langle \delta \mathbf{x}, \mathbf{C}(\mathbf{x}^f - \mathbf{x}_t^f) \rangle + \langle \mathbf{C}(\mathbf{x}^f - \mathbf{x}_t^f), \delta \mathbf{x} \rangle + \langle \delta \mathbf{x}^T, \mathbf{C} \delta \mathbf{x} \rangle, \quad \mathbf{C} \text{ symétrique} \quad (\text{A.34})$$

$$= 2 \langle \delta \mathbf{x}, \mathbf{C}(\mathbf{x}^f - \mathbf{x}_t^f) \rangle + \langle \delta \mathbf{x}^T, \mathbf{C} \delta \mathbf{x} \rangle, \quad \text{symétrie produit scalaire} \quad (\text{A.35})$$

$$= 2(\delta \mathbf{x})^T \mathbf{C}(\mathbf{x}^f - \mathbf{x}_t^f) + \delta \mathbf{x}^T \mathbf{C} \delta \mathbf{x}. \quad (\text{A.36})$$

## A.5 Etablissement des expressions de Wasserstein et Kullback-Leibler

### A.5.1 Expression de la distance de Wasserstein

Le terme à exprimer est  $\text{Tr} \left( C_1 + C_2 - 2(C_2^{1/2} C_1 C_2^{1/2})^{1/2} \right)$ .

Nous avons  $C_1 = \mathbf{H} \mathbf{P}^f \mathbf{H}^T$  et  $C_2 = \mathbf{R} + \mathbf{H} \mathbf{P}^f \mathbf{H}^T$ . Elles sont symétriques définies positives, donc  $\exists \mathbf{U}$  orthogonale,  $\Sigma_f = \text{Diag}(\lambda_i)$  telles que :

$$\begin{aligned} C_1 &= \mathbf{U} \Sigma_f \mathbf{U}^T, \\ C_2 &= \mathbf{U} (\Sigma_f + \sigma_{obs}^2 \mathbf{I}_p) \mathbf{U}^T \end{aligned}$$

Ainsi :

$$\text{Tr}(C_1 + C_2) = \sum_{i=1}^p (2\lambda_i + \sigma_{obs}^2) = p\sigma_{obs}^2 + 2 \sum_{i=1}^p \lambda_i. \quad (\text{A.37})$$

Par propriétés fondamentales de l'opérateur trace, nous avons :

$$\begin{aligned} \text{Tr}((C_2^{1/2} C_1 C_2^{1/2})^{1/2}) &= \text{Tr} \left[ \left( (\Sigma_f + \sigma_{obs}^2 \mathbf{I}_p)^{1/2} \Sigma_f (\Sigma_f + \sigma_{obs}^2 \mathbf{I}_p)^{1/2} \right)^{1/2} \right] \\ &= \text{Tr} \left[ (\Sigma_f + \sigma_{obs}^2 \mathbf{I}_p)^{1/4} (\Sigma_f)^{1/2} (\Sigma_f + \sigma_{obs}^2 \mathbf{I}_p)^{1/4} \right] \quad (\text{A.38}) \end{aligned}$$

$$= \text{Tr} \left[ (\Sigma_f (\Sigma_f + \sigma_{obs}^2 \mathbf{I}_p))^{1/2} \right], \quad \text{car matrices diagonales}$$

$$= \sum_{i=1}^p \sqrt{\lambda_i (\lambda_i + \sigma_{obs}^2)} \quad (\text{A.39})$$

En sommant les termes (A.37) et (A.39), et en remarquant que  $2\lambda_i + \sigma_{obs}^2 - 2\sqrt{\lambda_i (\lambda_i + \sigma_{obs}^2)} = (\sqrt{\lambda_i + \sigma_{obs}^2} - \sqrt{\lambda_i})^2$ , nous obtenons :

$$W^2(\mu_1, \mu_2) = \|\mathbf{y} - \mathbf{H} \mathbf{x}^f\|_2^2 + \sum_{i=1}^p \left( \sqrt{\lambda_i + \sigma_{obs}^2} - \sqrt{\lambda_i} \right)^2. \quad (\text{A.40})$$

### A.5.2 Expression de la divergence de Kullback-Leibler

Avec les mêmes notations que précédemment, et en notant  $\mathbf{d} = \mathbf{U}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^f) = [d_i]$ , nous avons :

$$2KL(\mu_1, \mu_2) = \mathbf{d}^T \left( \Sigma_f^{-1} + (\Sigma_f + \sigma_{obs}^2 \mathbf{I}_p)^{-1} \right) \mathbf{d} \\ + Tr \left[ (\Sigma_f + \sigma_{obs}^2 \mathbf{I}_p)^{-1} \Sigma_f + \Sigma_f^{-1} (\Sigma_f + \sigma_{obs}^2 \mathbf{I}_p) \right].$$

Or :

$$(\Sigma_f + \sigma_{obs}^2 \mathbf{I}_p)^{-1} \Sigma_f = (\Sigma_f + \sigma_{obs}^2 \mathbf{I}_p)^{-1} (\Sigma_f + \sigma_{obs}^2 \mathbf{I}_p) - \sigma_{obs}^2 (\Sigma_f + \sigma_{obs}^2 \mathbf{I}_p)^{-1} \\ = \mathbf{I}_p - \sigma_{obs}^2 (\Sigma_f + \sigma_{obs}^2 \mathbf{I}_p)^{-1}.$$

Ainsi :

$$Tr \left[ (\Sigma_f + \sigma_{obs}^2 \mathbf{I}_p)^{-1} \Sigma_f + \Sigma_f^{-1} (\Sigma_f + \sigma_{obs}^2 \mathbf{I}_p) \right] = \sum_{i=1}^p \left( 1 - \frac{\sigma_{obs}^2}{\lambda_i + \sigma_{obs}^2} + 1 + \frac{\sigma_{obs}^2}{\lambda_i} \right).$$

En réinjectant dans l'expression de Kullback-Leibler :

$$KL(\mu_1, \mu_2) = \frac{1}{2} \sum_{i=1}^p d_i^2 \left( \frac{1}{\lambda_i} + \frac{1}{\lambda_i + \sigma_{obs}^2} \right) \quad (\text{A.41})$$

$$+ \frac{1}{2} \sum_{i=1}^p \left( \frac{-\sigma_{obs}^2}{\lambda_i + \sigma_{obs}^2} + \frac{\sigma_{obs}^2}{\lambda_i} \right) + p \quad (\text{A.42})$$

$$KL(\mu_1, \mu_2) = \frac{1}{2} \sum_{i=1}^p \left( \frac{d_i^2 - \sigma_{obs}^2}{\lambda_i + \sigma_{obs}^2} + \frac{d_i^2 + \sigma_{obs}^2}{\lambda_i} \right) + p.$$

## A.6 Réduction du nombre d'opérations pour les distances probabilistes

Par commodité de rédaction, nous noterons dans cette section  $\tilde{\mathbf{x}}_a$  et  $\tilde{\mathbf{P}}_a$  plutôt que  $\tilde{\mathbf{x}}^a$  et  $\tilde{\mathbf{P}}^a$ , respectivement pour l'état analysé avec mise à jour de rang et sa matrice de covariance d'erreur associée.

### A.6.1 Mise à jour de l'opérateur $\mathbf{H}$ et de l'état analysé selon les observations à supprimer

Comme vu en introduction, la suppression d'une observation revient à mettre à jour l'opérateur d'observation. La méthode d'annulation d'une ligne revient à modifier la formulation du problème aux moindres carrés de l'équation (4.1) d'une façon qui ne se transcrit pas comme une modification de rang  $k \geq 1$ . Ceci imposerait donc de recalculer la solution du problème d'assimilation à chaque étape plutôt que pouvoir la mettre à jour directement via une modification de rang 1, comme il est possible de le faire (BJÖRK, 1996, chap. 2, p. 137). La méthode permettant de modifier  $\mathbf{H}$  sera donc la suppression de la ligne correspondante sur l'opérateur. En reprenant l'équation (4.1) :

$\min_{\mathbf{x} \in \mathbb{R}^n} \left\| \begin{bmatrix} \mathbf{B}^{-1/2} \\ \mathbf{R}^{-1/2} \mathbf{H} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{B}^{-1/2} \mathbf{x}^b \\ \mathbf{R}^{-1/2} \mathbf{y} \end{bmatrix} \right\|_2$ , cela revient d'après (BJÖRK, 1996) à mettre à jour

à la fois la décomposition QR de la matrice  $\mathbf{A} = \begin{bmatrix} \mathbf{B}^{-1/2} \\ \mathbf{R}^{-1/2} \mathbf{H} \end{bmatrix} = \mathbf{Q} \mathbf{R}_A$  et la dimension de la matrice  $\mathbf{R}^{-1/2}$ . En notant  $\tilde{\mathbf{Q}} \tilde{\mathbf{R}}_A$  la décomposition QR mise à jour en supprimant la ligne de  $\mathbf{H}$  correspondant à l'observation négligée et  $\tilde{\mathbf{b}}$  le second membre avec réduction de la taille de  $\mathbf{R}^{-1/2}$ , il ne reste qu'à résoudre le système linéaire simplifié  $\tilde{\mathbf{R}}_A \mathbf{x} = \tilde{\mathbf{Q}}^T \tilde{\mathbf{b}}$ , où l'observation aura une dimension réduite, qui va donner  $\tilde{\mathbf{x}}_a$ . Nous pouvons ainsi séquentiellement mettre à jour l'état analysé  $\mathbf{x}_a$  selon les observations que l'on désire supprimer. En suivant cette idée, la matrice de covariance d'erreur d'analyse modifiée sera donnée par  $\tilde{\mathbf{P}}_a = (\tilde{\mathbf{R}}_A^T \tilde{\mathbf{R}}_A)^{-1} = (\tilde{\mathbf{R}}_A)^{-1} (\tilde{\mathbf{R}}_A)^{-T}$ . Ceci revient à résoudre deux systèmes triangulaires et à faire un produit entre deux matrices triangulaires, ce qui est moins coûteux que de recalculer directement l'inverse de la matrice  $\tilde{\mathbf{P}}_a$ .

### A.6.2 Amélioration de la divergence de Kullback-Leibler

L'utilisation de mises à jour de rang 1 pour les inverses en passant par des factorisations QR semble le plus prometteur :  $\mathbf{A} = \mathbf{Q} \mathbf{R} \implies \mathbf{A}^{-1} = \mathbf{R}^{-1} \mathbf{Q}^T$ , une mise à jour de la factorisation QR de  $\mathbf{A}$  permet ensuite facilement de reporter la mise à jour sur son inverse. Nous pouvons donc réutiliser l'algorithme de (BJÖRK, 1996, p. 137) pour faire cela. Il est similaire à l'algorithme décrit dans (DENNIS et SCHNABEL, 1996, pp. 55-58) qui s'effectue pour un coût en  $O(n^2)$ .

Lorsque nous voulons supprimer plusieurs observations, disons  $k \in \mathbb{N}^*$ , il suffit d'appliquer séquentiellement des mises à jour de rang 1 selon le même procédé, pour un coût en  $O(kn^2)$ . En général,  $k \ll n$ , on conserve donc bien une diminution d'un ordre de grandeur sur le coût.

L'inversion consiste ensuite en un produit entre deux matrices triangulaires pour un coût en  $O(n^2)$ . Le coût total de la mise à jour de l'inverse revient à du  $O((k+1)n^2)$ , un ordre de grandeur inférieur aux  $O\left(\frac{4n^3}{3}\right)$  du calcul sans mise à jour.

### A.6.3 Amélioration de la distance de Wasserstein

L'élément le plus coûteux dans le cas de Wasserstein est le calcul de racines carrées de matrice,  $(C_2^{1/2} C_1 C_2^{1/2})^{1/2}$  semble devoir être recalculée à chaque fois. Cependant, modifier  $\mathbf{H}$  revient à modifier le problème aux moindres carrés comme décrit précédemment,  $\tilde{\mathbf{R}}$  étant recalculée, la mise à jour de  $\mathbf{P}_a$  sera simple.

Le calcul d'une racine carrée de matrice se fait à partir d'une décomposition de Schur de la matrice, i.e  $\mathbf{A} = \mathbf{Q} \mathbf{U} \mathbf{Q}^T$ , avec  $\mathbf{U}$  triangulaire supérieure et  $Sp(\mathbf{U}) = Sp(\mathbf{A})$ , dans ce cas  $\mathbf{A}^{1/2} = \mathbf{Q} \mathbf{U}^{1/2} \mathbf{Q}^T$ . Elle existe bien car les matrices manipulées sont symétriques définies-positives, donc diagonalisables. Cette décomposition de Schur est calculée grâce à l'algorithme QR (FRANCIS, 1961), détaillé ci-après sous sa forme la plus simple :

1. Initialiser  $\mathbf{A}_1 = \mathbf{A}$ ,  $\mathbf{Q} = \mathbf{I}$
2. Pour  $k \geq 1$  jusqu'à convergence faire :
  - (a) Calculer la factorisation QR de  $\mathbf{A}_k$  :  $\mathbf{A}_k = \mathbf{Q}_k \mathbf{R}_k$

- (b) Définir  $\mathbf{A}_{k+1} = \mathbf{R}_k \mathbf{Q}_k$
  - (c) Accumuler  $\mathbf{Q} = \mathbf{Q}_k \mathbf{Q}$
  - (d) Revenir à l'étape (a) avec  $k = k + 1$
3. Avec  $k_\infty$  l'indice de convergence<sup>1</sup>, construire  $\mathbf{A} = \mathbf{Q} \mathbf{A}_{k_\infty} \mathbf{Q}^T = \mathbf{Q} \mathbf{U} \mathbf{Q}^T$

Cela nous donne, avec  $\mathbf{A}_k = \mathbf{Q}_k \mathbf{R}_k \iff \mathbf{R}_k = \mathbf{Q}_k^T \mathbf{A}_k$ , que  $\mathbf{A}_{k+1} = \mathbf{R}_k \mathbf{Q}_k = \mathbf{Q}_k^T \mathbf{A}_k \mathbf{Q}_k = \mathbf{Q}_k^T \mathbf{R}_{k-1} \mathbf{Q}_{k-1} \mathbf{Q}_k = \mathbf{Q}_k^T \mathbf{Q}_{k-1}^T \mathbf{A}_{k-1} \mathbf{Q}_{k-1} \mathbf{Q}_k = \dots = (\mathbf{Q}_k^T \dots \mathbf{Q}_1^T) \mathbf{A}_1 (\mathbf{Q}_1 \dots \mathbf{Q}_k) = \mathbf{Q}^T \mathbf{A} \mathbf{Q}$  avec  $\mathbf{Q} = \mathbf{Q}_1 \dots \mathbf{Q}_k$ .

Ainsi, il semble compliqué de pouvoir mettre à jour toute cette décomposition QR avec une mise à jour de rang 1 pour l'ensemble du processus de construction de la décomposition de Schur. En effet, en supposant que nous mettons à jour la matrice initiale  $\mathbf{A}$ , nous allons avoir  $\mathbf{A} = \mathbf{A}_1 = \mathbf{Q}_1 \mathbf{R}_1 = \tilde{\mathbf{Q}}_1 \tilde{\mathbf{R}}_1$ , ce qui coûte  $O(p, n)$ , selon les dimensions de la matrice. Nous formerons donc  $\mathbf{A}_2 = \tilde{\mathbf{R}}_1 \tilde{\mathbf{Q}}_1 = \mathbf{Q}_2 \mathbf{R}_2$  pour la suite, en étant toujours obligé de refaire la décomposition QR de cette matrice, même en ayant eu les mises à jour sur  $\mathbf{A}$ . Nous observons clairement qu'il en sera de même pour les itérés suivants. Nous ne gagnons donc presque rien sur le coût global si ce n'est que nous allons avoir à faire des calculs sur une dimension de moins. La mise à jour de rang 1 du processus entier de calcul de la décomposition de Schur ne semble pas possible.

Cependant, il reste possible de faire plusieurs mises à jour de rang 1 sur la matrice initiale en fonction du nombre d'observations que l'on veut supprimer, ce qui permet ensuite de faire un calcul à moindre coût de la nouvelle décomposition de Schur de cette matrice pour calculer sa racine carrée. Il est donc toujours possible d'avoir un gain, car une diminution  $k \in \mathbb{N}$  sur  $p, n$  se répercute en une diminution de coût de l'ordre de  $O(k^3)$  sur les factorisations QR qui suivront.

Ceci guide donc l'utilisation *a posteriori* de Wasserstein : pour ne pas payer un coût maximal lors du diagnostic, il faudra s'assurer au préalable du nombre d'observations à supprimer et de leurs indices pour ensuite recalculer les nouvelles racines carrées.

#### A.6.4 Algorithme de calcul à moindre coût pour les deux distances

Calcul de Wasserstein :

1. Calculer  $\hat{\mathbf{H}}$  ou SOBS pour l'observation actuelle,
2. Déterminer/choisir  $k$  le nombre d'observations à supprimer,
  - (a) Récupérer les indices  $\in I$  des observations correspondantes,
3. Effectuer  $k$  mises à jour de rang 1 selon l'algorithme de (BJÖRK, 1996, p. 137) à la matrice  $\mathbf{A} = \begin{bmatrix} \mathbf{B}^{-1/2} \\ \mathbf{R}^{-1/2} \mathbf{H} \end{bmatrix}$ . Donne  $\mathbf{A}_{|k} = \tilde{\mathbf{Q}} \tilde{\mathbf{R}}_A$ ,
4. Utiliser  $\tilde{\mathbf{Q}}, \tilde{\mathbf{R}}_A$  pour le calcul de  $\tilde{\mathbf{x}}_a$  et  $\tilde{\mathbf{P}}_a$  vu en sous-section (A.6.1),
5. Calculer  $\tilde{\mathbf{P}}_a^{1/2}$  par décomposition de Schur,
6. Calculer  $dW(\mathcal{N}(\tilde{\mathbf{x}}_a, \tilde{\mathbf{P}}_a), \mathcal{N}(\mathbf{x}_a, \mathbf{P}_a))$  avec les éléments des étapes (4) et (5).

1. Pour la preuve de la convergence, voir FRANCIS, 1961.

### Calcul de Kullback-Leibler :

Remplacer respectivement les étapes (5) et (6) précédentes par :

- Calculer  $\tilde{\mathbf{P}}_a^{-1} = \tilde{\mathbf{R}}_A^T \tilde{\mathbf{R}}_A$ ,
- Calculer  $dKL(\mathcal{N}(\tilde{\mathbf{x}}_a, \tilde{\mathbf{P}}_a), \mathcal{N}(\mathbf{x}_a, \mathbf{P}_a))$  avec les éléments de l'étape précédente et de l'étape (4) du paragraphe précédent.

### A.6.5 Conclusion

Nous avons vu dans cette partie comment procéder pour réduire le coût du calcul des distances probabilistes dans une utilisation pour diagnostic *a posteriori*. Cela passe avant tout par une mise à jour des matrices d'erreurs intervenant dans la formule des distances de Wasserstein et Kullback-Leibler. Pour cela, la factorisation QR et des mises à jour de rang 1 sur celle-ci sont utilisées.

Des gains sont possibles en proposant une bonne implémentation de ces algorithmes, mais nous choisissons de ne pas la mettre en place dans le cadre de la thèse pour nous concentrer sur le reste des travaux de recherche. Les systèmes que nous utiliserons ne sont pas limitatifs en terme de puissance calculatoire, nous n'avons pas besoin de procéder à une réduction des coûts.

Ces réductions sont marginales car ce n'est pas la formule de calcul des distances elle-même qui est mise à jour. Cependant, lors de l'application à un système de grande taille, tout gain est bon à prendre et l'avantage de ces algorithmes provient de leur simplicité de mise en œuvre. Ils ont été proposés en tant qu'ébauche pour une application où les coûts calculatoires deviennent prohibitifs et où une réduction de ceux-ci devient intéressante.

## A.7 Configurations des systèmes pour essais numériques

### A.7.1 Lorenz 63

#### Description du système

Il s'agit d'un modèle en 3D issue de la météorologie. Il a été présenté pour la première fois dans (LORENZ, 1963).

C'est un système d'équations différentielles couplées, non linéaire. Il est issu d'une simplification des équations de Navier-Stokes, décrivant le couplage de l'atmosphère terrestre avec l'océan. Cela aboutit à un système à 3 degrés de liberté, intégrable par les ordinateurs de l'époque de Lorenz.

Il est défini par 3 paramètres  $(\rho, \sigma, \beta) \in \mathbb{R}^3$ . En voici les équations :

$$\begin{cases} \dot{x}(t) = \sigma(y(t) - x(t)) \\ \dot{y}(t) = \rho x(t) - y(t) - x(t)z(t) \\ \dot{z}(t) = x(t)y(t) - \beta z(t) \end{cases}$$

C'est un système-test de référence en assimilation de données. Il est donc très documenté, et consiste en quelque sorte à un "benchmark" pour vérifier les méthodes d'assimilation. Propriétés du Lorenz3D :



- il est **non-linéaire**
- un point fixe "universel" en  $(0,0,0)$ ,  $\forall(\rho, \sigma, \beta) \in \mathbb{R}^3$ ,
- 2 points fixes symétriques en  $(-\sqrt{\beta(\rho-1)}, -\sqrt{\beta(\rho-1)}, \rho-1)$  et  $(\sqrt{\beta(\rho-1)}, \sqrt{\beta(\rho-1)}, \rho-1)$ , lorsque  $\rho > 1$ ,
- pour  $\rho = 28, \sigma = 10, \beta = 8/3$ , le système possède un attracteur en forme d'ailes de papillon. La trajectoire commence à s'enrouler sur une des ailes puis saute ensuite au bout d'un moment sur l'autre, etc, et ce quelques soient les conditions initiales tant qu'elles sont distinctes des points fixes. Il est donc **chaotique déterministe** pour ce jeu de paramètres. Le type de bifurcation observée (saut entre chaque aile du papillon) correspond à une **bifurcation de Hopf**.

Voici une première représentation de l'attracteur du système :

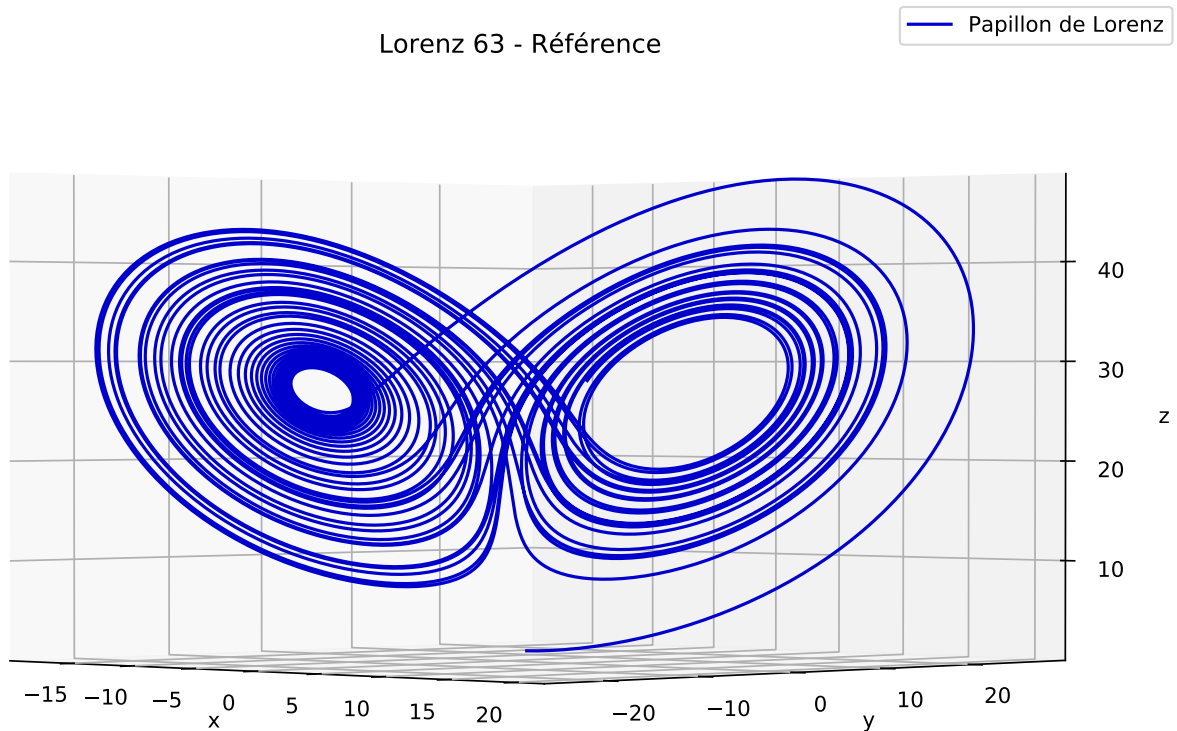


FIGURE A.2 – Attracteur du Lorenz 63 - Vue de face.

En voici une seconde :

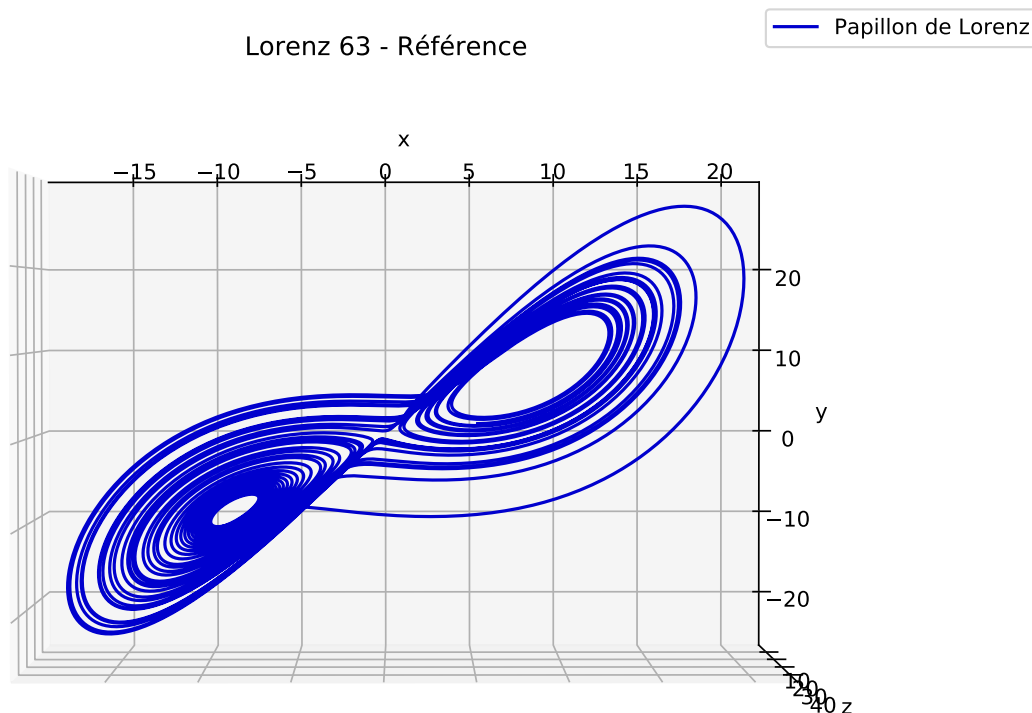


FIGURE A.3 – Vue de haut.

### Configuration des essais numériques

La configuration utilisée est la configuration standard décrite comme suit. La fenêtre d'assimilation va de  $t_0 = 0$  à  $t_f = 10$ , sauf mention explicite du contraire. Les variations peuvent être  $t_f = 25$  ou  $t_f = 250$ .

La méthode d'intégration numérique utilisée est celle de Runge-Kutta d'ordre 4.

L'ensemble est de taille  $N = 100$  membres, pour éviter tout problème de déficience de rang et être de taille similaire à ce qui peut se retrouver dans l'applicatif. Une erreur modèle additive suivant la loi  $\mathcal{N}(0, \mathbf{Q})$  est ajoutée à chaque pas d'intégration  $dt_{prev} = 10^{-2}$ , avec  $\mathbf{Q} = 10^{-1}\mathbf{I}_3$ . La matrice d'erreur d'observation vaut  $\mathbf{R} = 2\mathbf{I}_3$ , pour un pas de temps d'observation  $dt_{obs} = 0.25$ . L'opérateur d'observation vaut  $\mathbf{H} = \mathbf{I}_3$ . La matrice d'ébauche  $\mathbf{B}$  ainsi que le point de départ  $\mathbf{x}_0$  sont issus d'un run de chauffe d'une intégration pure d'une durée de l'ordre de 5000 cycles d'assimilation avec  $dt_{sim} = 5.10^{-3}$ , dont les valeurs finales ont ensuite été utilisées pour un run d'assimilation de 5000 cycles avec les mêmes statistiques d'erreur que la configuration décrite juste avant. L'ensemble de départ est échantillonné selon la loi  $\mathcal{N}(\mathbf{x}_0, \mathbf{B})$ . Cela suit la méthodologie utilisée dans (BOCQUET, 2011). La graine du générateur aléatoire est fixée.

Deux variantes possibles sont d'avoir :

- comme opérateur d'observation  $\mathbf{H} = \begin{pmatrix} \mathbf{I}_3 & & \\ 0 & 0 & 0 \end{pmatrix}$ ,  $\mathbf{R} = 2\mathbf{I}_4$  et  $y_4 = 0 + \epsilon_4$  où  $\epsilon_4$  représente le bruit associé à l'erreur sur cette observation,
- comme opérateur d'observation  $\mathbf{H} = \begin{pmatrix} \mathbf{I}_3 & & \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$ ,  $\mathbf{R} = 2\mathbf{I}_4$  et  $y_4 = \frac{1}{3}(y_1+y_2+y_3)+\epsilon_4$  où  $\epsilon_4$  représente le bruit associé à l'erreur sur cette observation.

La version avec  $p = 3$  observations correspond à la configuration 1, celle avec  $p = 4$  observations soit l'ajout de l'observation négligeable correspond à la configuration 2.

## A.7.2 Oscillateur de Van der Pol

### Description du système

L'oscillateur de Van der Pol est un modèle en 2D issu de l'électronique. Il décrit les oscillations d'une triode dans un circuit électrique. Il a été introduit pour la première fois en 1927 par Balthazar Van der Pol, contemporain d'Edward Lorenz. Il s'agit d'une équation différentielle à une dimension pouvant se reformuler comme un système dynamique à deux dimensions. Le système possède une non-linéarité marquée. Une description détaillée de l'oscillateur de Van der Pol est disponible dans (TSATSOS, 2006).

L'oscillateur original Van der Pol est régi par l'équation différentielle :

$$\begin{cases} \ddot{x}(t) - \varepsilon\omega_0(1 - x^2(t))\dot{x}(t) + \omega_0^2x(t) = 0 \\ \ddot{x}(t) - \varepsilon\omega_0(1 - x^2(t))\dot{x}(t) + \omega_0^2x(t) = \omega_0^2X \cos(\omega t), \quad \text{si régime forcé} \end{cases}$$

$\varepsilon$  est le coefficient de non-linéarité du système. Lorsque nous avons  $\varepsilon = 0$ , nous retrouvons l'équation d'un oscillateur harmonique.

Sa version bidimensionnelle est donnée par :

$$\begin{cases} \dot{y}_1(t) = y_2(t) \\ \dot{y}_2(t) = \mu(1 - y_1^2(t))y_2(t) - y_1(t) \end{cases}$$

$\mu$  correspond au  $\omega_0^2X$  dans l'équation en régime forcé précédente : ce paramètre va déterminer la période des oscillations du système, et notamment le caractère raide du système.

Propriétés du Van der Pol :

- il est **raide**. Malgré sa petite dimension, il peut mettre à mal certains intégrateurs pour systèmes raides,
- il est **chaotique déterministe** en régime forcé.

Voici une représentation de la première variable du système qui présente des oscillations régulières :

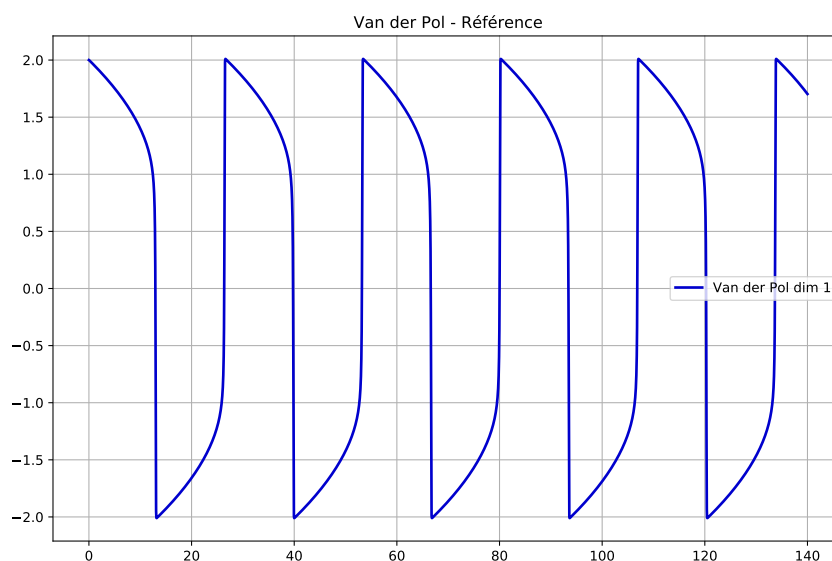


FIGURE A.4 – Référence pour les oscillations de la première dimension du Van der Pol

### Configuration des essais numériques (configuration 3)

La même méthodologie que pour le Lorenz 63 a été appliquée pour construire cette configuration.

La configuration utilisée est la configuration standard décrite comme suit. La fenêtre d'assimilation va de  $t_0 = 0$  à  $t_f = 150$ , sauf mention explicite du contraire. Les variations peuvent être  $t_f = 750$  ou  $t_f = 1500$ . Ceci permet de rajouter plus de changements de régime.

La méthode d'intégration numérique utilisée est celle de Runge-Kutta d'ordre 4. Le pas de temps a été choisi suffisamment petit pour ne pas être gêné par la raideur du système. L'ensemble est de taille  $N = 100$  membres, pour éviter tout problème de déficience de rang et être de taille similaire à ce qui peut se retrouver dans l'applicatif. Une erreur modèle additive suivant la loi  $\mathcal{N}(0, \mathbf{Q})$  est ajoutée à chaque pas d'intégration  $dt_{prev} = 5 \cdot 10^{-3}$ , avec  $\mathbf{Q} = 5 \cdot 10^{-4} \mathbf{I}_2$ . La matrice d'erreur d'observation vaut  $\mathbf{R} = 10^{-2} \mathbf{I}_2$ , pour un pas de temps d'observation  $dt_{obs} = 2$ . La matrice d'ébauche vaut  $\mathbf{B} = \mathbf{R}$ . Le point de départ est  $\mathbf{x}_0 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ . L'ensemble de départ est échantillonné selon la loi  $\mathcal{N}(\mathbf{x}_0, \mathbf{B})$ . La graine du générateur aléatoire est fixée.

### A.7.3 Système Xénon

Ce système est particulièrement intéressant car il est directement issu d'une problématique à laquelle fait face EDF tous les jours pour le pilotage des réacteurs nucléaires de ses centrales nucléaires pour la production d'électricité (CNPE), et pour garantir une production électrique sûre, stable, et adaptable. De plus amples détails sur la cinétique du Xénon-135 sont disponibles dans (REUSS, 2003) et (PONÇOT, 2008, Sec. 1.3).

#### Rapide introduction sur la fission au sein dans un réacteur nucléaire

Lors des réactions de fission au sein d'un réacteur nucléaire, de nombreux produits dit de *fission* sont créés. Ceux-ci étant majoritairement instables, ils vont par la suite se dégrader par radioactivité spontanée et conduire à la formation d'encore plus de produits intermédiaires. Ces derniers restent confinés à l'intérieur de la gaine entourant les pastilles de combustible. Les produits volatils et les gaz rares peuvent migrer au sein du combustible en restant à l'intérieur de la gaine, et vont s'accumuler dans les *chambres d'expansion* prévues pour cela.

Soumis à un flux intense de neutrons, comme c'est le cas des barres de combustible, ces noyaux sont susceptibles de capturer des neutrons : cela a pour conséquence, d'une part, de diminuer la réactivité du réacteur donc sa capacité à maintenir la réaction de fusion nécessaire à produire de l'énergie, d'autre part, de changer les concentrations des divers isotopes présents dans les barres de combustibles.

La diminution de réactivité due à la présence de produits de fission est appelée *empoisonnement du réacteur*. Il existe environ 800 produits de fission, nous ne nous intéresserons qu'à la cinétique de celui qui participe le plus à l'empoisonnement d'un réacteur : le *Xénon-135*.

#### Chaîne radioactive et cinétique du Xénon-135

Le Xénon-135 est issu à la fois directement de la dégradation de l'Uranium au sein et des réacteurs et comme produit intermédiaire d'un produit de fission : l'Iode, lui même issu du Tellure. Pour résumer, voici un schéma de la chaîne radioactive du Xénon-135 :

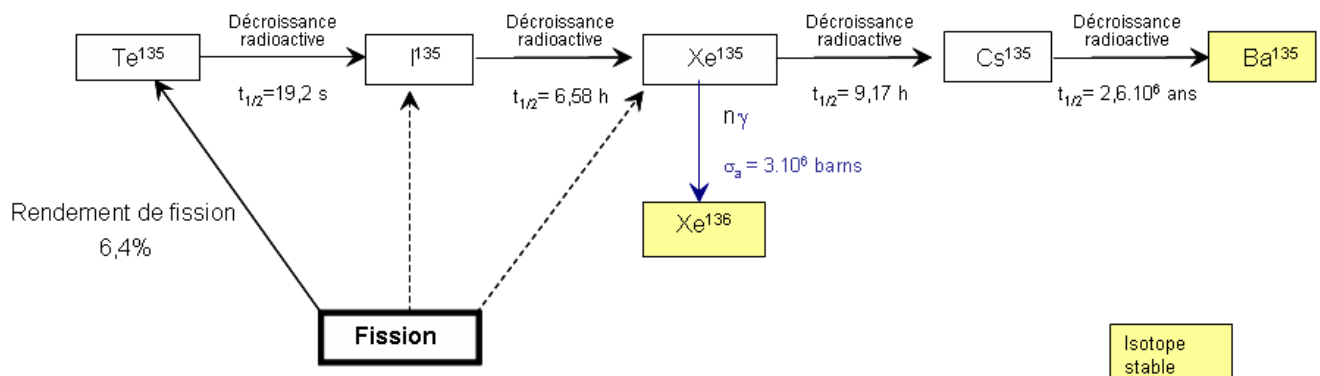


FIGURE A.5 – Chaîne radioactive du Xénon-135

Le rendement de la voie directe est de 0,1%. La durée de demi-vie du Tellure 135 étant

très faible, nous pouvons considérer que la fission donne directement de l'Iode-135, qui se dégradera lui-même en Xénon-135 au bout de quelques heures. Le rendement de fission pour l'Iode-135 étant de 6,4%, nous pouvons considérer que la formation de Xénon-135 n'est due qu'à cette voie là. De plus, nous ne nous intéressons pas aux nucléides situés au delà de celui-ci. Nous remarquons que le Xénon-135 peut capturer un neutron pour former un isotope stable, le Xénon-136.

Dans ces conditions, l'équation différentielle couplée d'évolution des concentrations d'Iode-135 (notée I) et de Xénon-135 (notée X) au sein du volume du réacteur sont :

$$\begin{cases} \dot{I} = \gamma_I \Sigma_f \Phi - \lambda_I I, \\ \dot{X} = \gamma_X \Sigma_f \Phi + \lambda_I I - (\lambda_X + \sigma_X \Phi) X. \end{cases} \quad (\text{A.43})$$

Les constantes sont :

$\Sigma_f \Phi$  : nombre de fissions par unité de volume et de temps,

$\gamma_{X,I}$  : les rendements de fission du Xénon-135 et de l'Iode-135,

$\lambda_{X,I}$  : les constantes de désintégration radioactive du Xénon-135 et de l'Iode-135,

$\sigma_X \Phi$  : nombre de captures de neutrons par le Xénon-135,

$\Phi$  : flux de neutrons au sein du réacteur.

Sa Jacobienne est définie  $\forall (x, y) \in \mathbb{R}^2$  par  $J_{X_e} = \begin{pmatrix} -\lambda_I & 0 \\ \lambda_I & -(\lambda_X + \sigma_X \Phi) \end{pmatrix}$ .

L'évolution de la concentration en Iode contient un terme source provenant de la fission de l'Uranium (membre de gauche), proportionnel à sa section efficace ainsi qu'au flux neutronique, et sa disparition par radioactivité spontanée, proportionnelle à sa constante radioactive.

L'évolution de la concentration de Xénon-135, le terme source provient directement de la fission de l'Uranium et de la désintégration de l'Iode-135 en Xénon-135, et le terme de disparition par radioactivité spontanée et par capture d'un neutron.

### Cinétique du Xénon et pilotage d'un réacteur

Supposons avoir au départ  $I = X = 0$  dans l'ensemble du réacteur. Ceci est physiquement plausible même si le réacteur a fonctionné au préalable avant son arrêt, car en quelques jours l'Iode-135 et le Xénon-135 se sont dégradés.

Pour un flux  $\Phi$  constant, l'équation (A.43) possède une solution analytique. En utilisant un système différentiel augmenté d'état  $A = \begin{pmatrix} I \\ X \end{pmatrix}$ , (A.43) se transforme en :

$$\frac{dA}{dt} = \begin{pmatrix} -\lambda_I & 0 \\ \lambda_I & -(\lambda_X + \sigma_X \Phi) \end{pmatrix} A + \begin{pmatrix} \gamma_I \Sigma_f \Phi \\ \gamma_X \Sigma_f \Phi \end{pmatrix} = J_{X_e} A + b \quad (\text{A.44})$$

Du fait des conditions initiales, la partie homogène de la solution est nulle, et après calcul de  $\int_0^t \exp(s J_{X_e}) b ds$ , les concentrations en Iode-135 et Xénon-135 sont :

$$\begin{cases} I(t) = \frac{\gamma_I \Sigma_f \Phi}{\lambda_I} (1 - e^{-\lambda_I t}), \\ X(t) = \frac{(\gamma_X + \gamma_I) \Sigma_f \Phi}{\lambda_X + \sigma_X \Phi} (1 - e^{-(\lambda_X + \sigma_X \Phi) t}) + \frac{\gamma_I \Sigma_f \Phi}{\lambda_I (\lambda_X + \sigma_X \Phi)} (1 - e^{-\lambda_I (\lambda_X + \sigma_X \Phi) t}), \end{cases} \quad (\text{A.45})$$

Ces résultats montrent qu'au bout d'environ 4 périodes radioactive de l'Iode-135, la concentration de Xénon-135 atteint une valeur asymptotique, proportionnelle au flux  $\Phi$ . Si l'on fixe bien la valeur de  $\Phi$ , l'empoisonnement est contrôlé et la réactivité du réacteur reste suffisamment élevée pour une production d'énergie acceptable. Les concentrations à l'équilibre sont obtenues en se plaçant en régime permanent :

$$I_{\text{eq}} = \frac{\gamma_I \Sigma_f \Phi}{\lambda_I}, \quad X_{\text{eq}} = \frac{(\gamma_X + \gamma_I) \Sigma_f \Phi}{\lambda_X + \sigma_X \Phi}, \quad (\text{A.46})$$

Il est à noter qu'une diminution même légère de  $\Phi$  va entraîner l'augmentation de la concentration en Xénon-135 au sein du réacteur.

Lors de l'arrêt brusque du réacteur (chute des barres de contrôle en centrale, arrêt d'un sous-marin nucléaire), la production directe de Xénon-135 par fission s'arrête. Cependant, celle-ci ne compte que pour 1/65 de la quantité totale de Xénon-135 produite.

Les réserves d'Iode-135 présentes à l'équilibre vont alors se "déverser" dans les réserves de Xénon-135, par désintégration radioactive. Comme la "vidange" du Xénon-135 n'est plus assurée par la capture des neutrons issus de la fission, son niveau va monter. Cette augmentation est limitée dans le temps car les réserves d'Iode finissent par s'épuiser : la concentration en Xénon-135 va donc augmenter, passer par un maximum, puis diminuer jusqu'à devenir nulle (si pas de relance du réacteur avant). Ce phénomène est appelé "*pic de Xénon*", et il dure environ 24h.

Ce pic de Xénon peut être problématique pour la conduite d'un réacteur, car si les réserves de réactivité (barres de commandes ou poisons pouvant être extraits) sont trop faibles, le réacteur ne peut être redémarré avant la fin du surcroît d'empoisonnement induite par le pic, dont il faut attendre la baisse par décroissance radioactive.

Un exemple illustrant bien les contraintes de ceci est celui d'un sous-marin nucléaire devant effectuer une série de déplacements/arrêts, à chaque fois du Xénon-135 va être sur-produit à l'arrêt, et au bout d'un moment le redémarrage en sera compromis.

Plusieurs des constantes physiques qui décrivent le système dépendent de la valeur du flux, voire du temps depuis lequel le réacteur est en service. Les valeurs numériques fournies dans ce qui va suivre permettent de reproduire le même comportement que ce qui serait observé pour un véritable réacteur. Elles sont modifiées pour que les échelles de temps et de concentrations ne soient pas identiques à celles des réacteurs existants dans le parc d'EDF.

Cela mène à plusieurs régimes avec des propriétés différentes pour le système Xénon, nous allons en détailler trois.

#### Arrêt du réacteur (configuration 4)

La configuration décrite ici correspond à un arrêt brusque du réacteur, où la puissance passe de son niveau maximal en fonctionnement à zéro. Le flux neutronique  $\phi$  s'annule. Cela va déclencher le pic de Xénon mentionné précédemment.



Voici les valeurs approchées des constantes physiques du système :

$$\begin{cases} \lambda_I = 2.9 \times 10^{-5} \\ \lambda_X + \sigma_X \phi = 1.2 \times 10^{-2} \\ \gamma_I \Sigma_f \phi = 0 \\ \gamma_X \Sigma_f \phi = 0 \end{cases} \quad (\text{A.47})$$

Voici l'évolution des concentrations en Iode-135 et en Xénon-135 obtenues avec cette configuration :

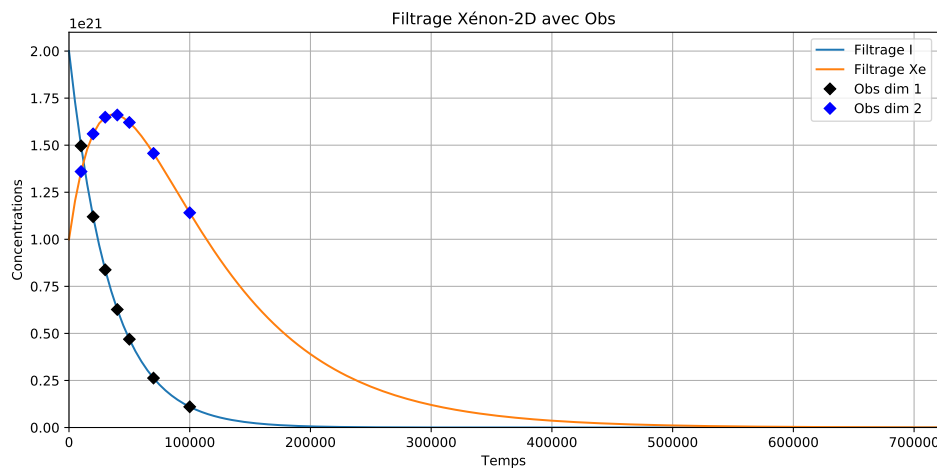


FIGURE A.6 – Evolution des concentrations à l'arrêt du réacteur. Le pic de Xénon représente le réservoir d'Iode-135 qui se "déverse" dans celui du Xénon-135 suite aux désintégrations radioactives.

### Démarrage du réacteur (configuration 5)

La configuration décrite ici suppose un réacteur chargé de combustible neuf, directement à puissance maximale de fonctionnement. Ce n'est pas un fonctionnement possible pour un réacteur opérationnel, il permet simplement d'illustrer l'établissement des concentrations d'équilibre de l'équation (A.46), qui est le phénomène qui nous intéresse.

Voici les valeurs approchées des constantes physiques du système :

$$\begin{cases} \lambda_I = 2.9 \times 10^{-5} \\ \lambda_X + \sigma_X \phi = 7.4 \times 10^{-5} \\ \gamma_I \Sigma_f \phi = 7.6 \times 10^{16} \\ \gamma_X \Sigma_f \phi = 3.9 \times 10^{15} \end{cases} \quad (\text{A.48})$$

Voici l'évolution des concentrations en Iode-135 et en Xénon-135 obtenues avec cette configuration :

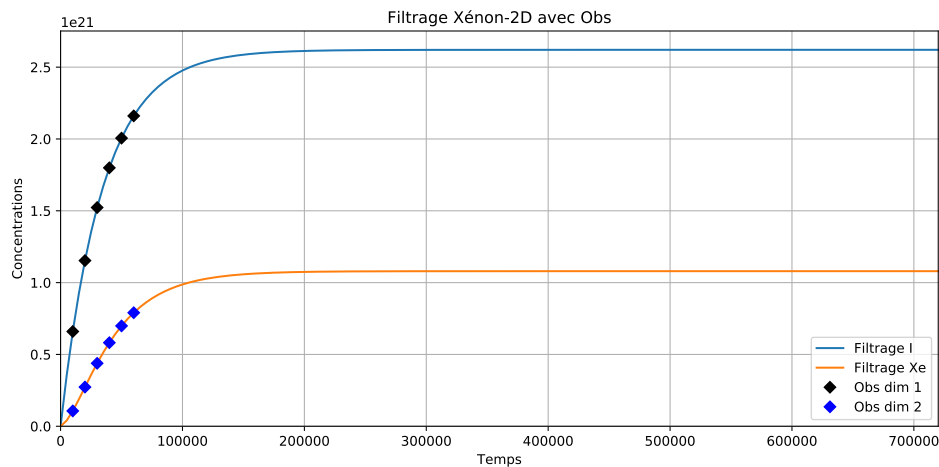


FIGURE A.7 – Evolution des concentrations à l'arrêt du réacteur. L'établissement des concentrations d'équilibre se fait à la même vitesse pour les deux entités chimiques.

### Puissance variable (configuration 6)

Dans ces conditions, les paramètres évoluent en fonction de la valeur de la puissance, pour passer des valeurs de la configuration 4 à celles de la configuration 5.

L'objectif de ce programme de puissance est de faire varier le système selon plusieurs régimes différents et de faire apparaître des oscillations de la concentration en Xénon-135. Ce phénomène n'apparaît pas en conditions opérationnelles car le pilote anticipe et étouffe les oscillations avant leur apparitions : elles sont un phénomène à éviter absolument lorsque le réacteur est en régime de fonctionnement.

Notons  $C_0$  les paramètres de la configuration 4 et  $C_{100}$  ceux de la configuration 5. La puissance du réacteur vaut  $P$ , décrite en pourcentage de sa puissance maximale de fonctionnement. L'évolution des paramètres en fonction de la puissance est régit la loi suivante :

$$C_P = \left(1 - \frac{P}{100}\right) C_0 + \frac{P}{100} C_{100}. \quad (\text{A.49})$$

Une représentation du système dans cette configuration est donnée à la page suivante.

Voici l'évolution de référence des concentrations en Iode-135 et en Xénon-135 obtenues avec cette configuration :

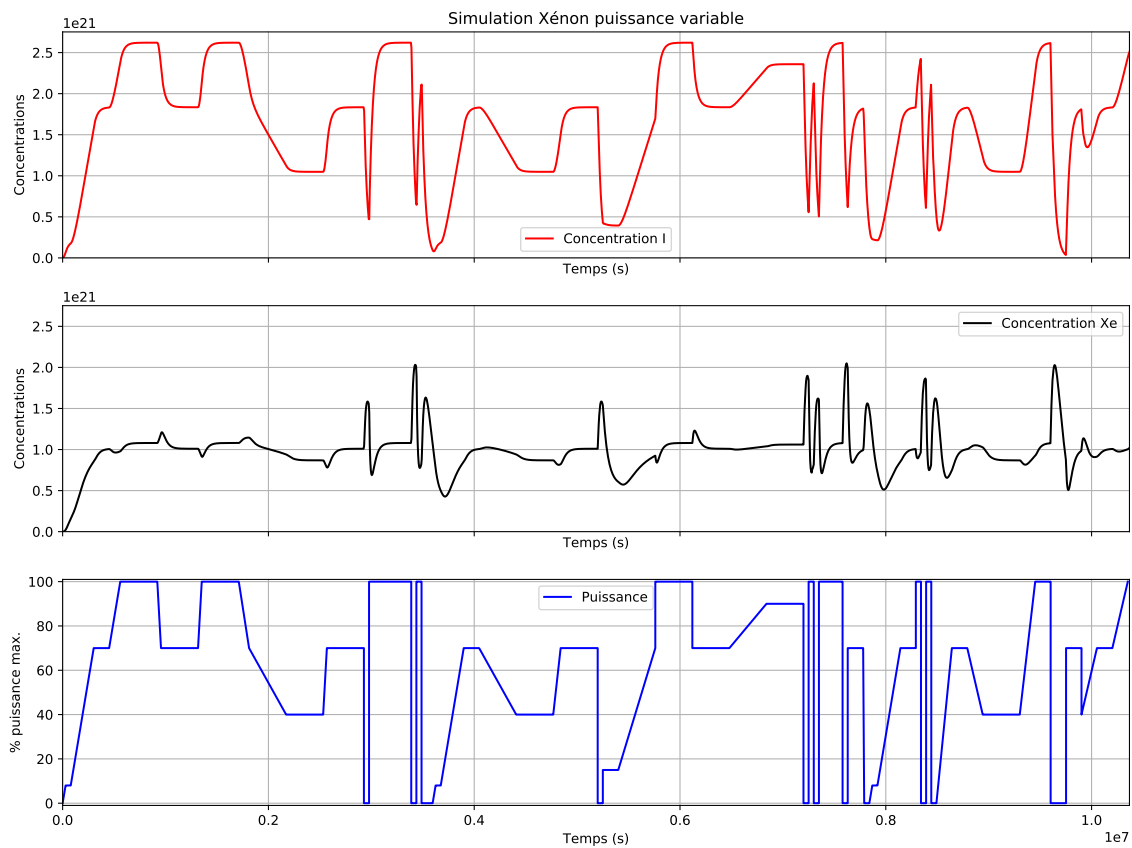


FIGURE A.8 – Référence pour le système Xénon avec forçage en puissance variable.

Une explication détaillée de cette figure est donnée pour la figure (7.29).

### Suivi de charge (configuration 7)

La puissance va évoluer selon la même loi que pour la configuration 6. C'est le programme de puissance qui va principalement changer. Celui-ci est représentatif de ce que l'on peut retrouver en conditions opérationnelles.

Voici l'évolution de référence des concentrations en Iode-135 et en Xénon-135 obtenues avec cette configuration :

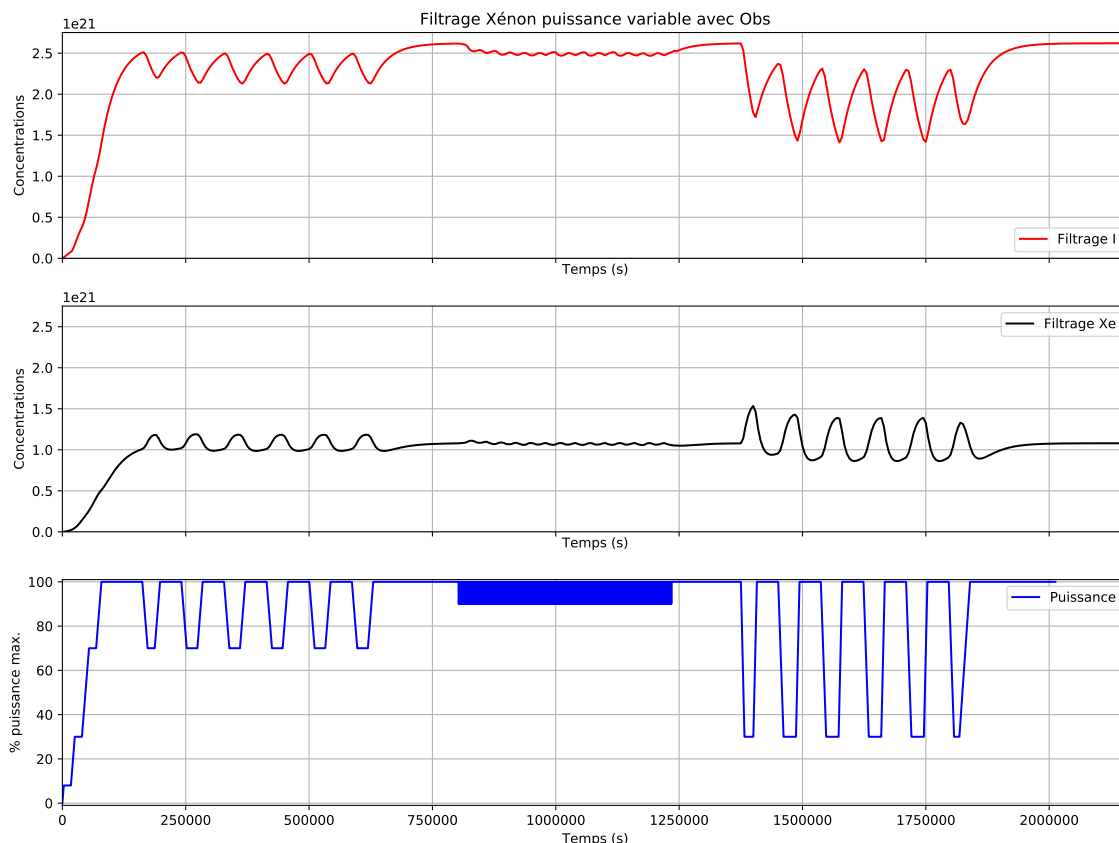


FIGURE A.9 – Référence pour le système Xénon avec forçage en puissance pour suivi de charge.

Il commence par un démarrage réaliste avec trois paliers de puissance dans un temps comparable à un démarrage réel. Viennent ensuite des oscillations journalières entre 100% de puissance et 70%, correspondant aux variations de la demande entre jour et nuit. La période bleue épaisse correspond à des oscillations entre 100% et 92% de puissance, avec une fréquence d'environ 30 min, typique d'un fonctionnement en suivi de charge. Vient enfin une période d'oscillations moins réalistes entre les paliers 100% et 30%, faisant apparaître quelques oscillations de Xénon-135.

### Configuration des essais numériques

La même méthode que pour la construction des configurations précédentes est utilisée.

La fenêtre d'assimilation va de  $t_0 = 0$  à :

- $t_f = 7,6 \cdot 10^5$  pour les configurations 4 et 5,
- $t_f = 1,0368 \cdot 10^7$  pour la configuration 6,
- $t_f = 2,16 \cdot 10^6$  pour la configuration 7.

La méthode d'intégration numérique utilisée est celle de Runge-Kutta d'ordre 4. Le pas de temps a été choisi suffisamment petit pour ne pas être gêné par la raideur du système de  $dt_{sim} = 10^3$ . L'unité de temps est la seconde.

L'ensemble est de taille  $N = 100$  membres, pour éviter tout problème de déficience de rang et être de taille comparable à ce qui peut se retrouver dans l'applicatif. Une erreur modèle additive suivant la loi  $\mathcal{N}(0, \mathbf{Q})$  est ajoutée à chaque pas d'intégration  $dt_{prev} = 5 \cdot 10^3$ , avec  $\mathbf{Q} = 10\mathbf{I}_2$ . La matrice d'erreur d'observation vaut  $\mathbf{R} = 10^3\mathbf{I}_2$ , pour un pas de temps d'observation  $dt_{obs} = 2 \cdot 10^4$ . La matrice d'ébauche vaut  $\mathbf{B} = 10^2\mathbf{I}_2$ . Le point de départ vaut :

- $\mathbf{x}_0 = \begin{bmatrix} 2 \cdot 10^{21} \\ 10^{21} \end{bmatrix}$  pour la configuration 4,
- $\mathbf{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$  pour les configurations 5, 6, et 7.

L'ensemble de départ est échantillonné selon la loi  $\mathcal{N}(\mathbf{x}_0, \mathbf{B})$ . La graine du générateur aléatoire est fixée.

# Bibliographie

- [1] Jeffrey L. ANDERSON. « An adaptative covariance inflation error correction algorithm for ensemble filters ». In : *Tellus* 59A (2007), p. 201-224. DOI : [10.1111/j.1600-0870.2006.00216.x](https://doi.org/10.1111/j.1600-0870.2006.00216.x).
- [2] Jeffrey L. ANDERSON. « Spatially and temporally varying adaptive covariance inflation for ensemble filters ». In : *Tellus* 61A.1 (2009), p. 72-83. DOI : [10.1111/j.1600-0870.2008.00361.x](https://doi.org/10.1111/j.1600-0870.2008.00361.x).
- [3] Mark ASCH, M. BOCQUET et M. NODET. *Data Assimilation, methods, algorithms, applications*. Sous la dir. de SIAM. SIAM, 2016.
- [4] L. BERTINO et al. *Advanced data assimilation in oceanography*. 2012.
- [5] Craig H. BISHOP, Brian J. ETHERTON et Sharanya J. MAJUMDAR. « Adaptive Sampling with the ETKF. Part 1 : Theoretical Aspects ». In : *Monthly Weather Review* 129 (2001), p. 420-436.
- [6] Äke BJÖRK. *Numerical methods for Least squares Problems*. Sous la dir. de SIAM. SIAM, 1996.
- [7] M. BOCQUET, P. N. RAANES et A. HANNART. « Expanding the validity of the ensemble Kalman filter without the intrinsic need for inflation ». In : *Nonlinear Processes in Geophysics* 22 (2015), p. 645-662. DOI : [10.5194/npg-22-645-2015](https://doi.org/10.5194/npg-22-645-2015).
- [8] M. BOCQUET et P. SAKOV. « An iterative ensemble Kalman smoother ». In : *Quarterly Journal of the Royal Meteorological Society* 140 (2014), p. 1521-1535. DOI : [10.1002/qj.2236](https://doi.org/10.1002/qj.2236).
- [9] Marc BOCQUET. « Ensemble Kalman Filtering without the intrinsic need for inflation ». In : *Nonlinear Processes in Geophysics* 18 (2011), p. 735-750. DOI : [10.5194/npg-18-735-2011](https://doi.org/10.5194/npg-18-735-2011).
- [10] Marc BOCQUET. « Localization and the Iterative Ensemble Kalman Smoother ». In : *Quarterly Journal of the Royal Meteorological Society* 142 (2016), p. 1075-1089. DOI : [10.1002/qj.2711](https://doi.org/10.1002/qj.2711).
- [11] Marc BOCQUET et Pavel SAKOV. « Combining Inflation-free and Iterative ensemble Kalman filters for Strongly nonlinear systems ». In : *Nonlinear Processes in Geophysics* 19 (2012), p. 383-399. DOI : [10.5194/npg-19-383-2012](https://doi.org/10.5194/npg-19-383-2012).
- [12] F. BOUTTIER et P. COURTIER. *Data assimilation concepts and methods*. 1999. URL : <https://www.ecmwf.int/sites/default/files/elibrary/2002/16928-data-assimilation-concepts-and-methods.pdf>.
- [13] Gerrit BURGERS, Peter J. VAN LEUWEEN et Geir EVENSEN. « Analysis Scheme in the Ensemble Kalman Filter ». In : *Monthly Weather Review* 126 (1998), p. 1719-1724.

- [14] C. CARDINALI. « Forecast sensitivity observation impact with an observation-only based objective function ». In : *Quarterly Journal of the Royal Meteorological Society* (2018). DOI : [10.1002/qj.3305](https://doi.org/10.1002/qj.3305).
- [15] C. CARDINALI. « Influence-matrix diagnostic of a data assimilation system ». In : *Quarterly Journal of the Royal Meteorological Society* 130 (2004), 2767–2786. DOI : [10.1256/qj.03.205](https://doi.org/10.1256/qj.03.205).
- [16] C. CARDINALI. « Monitoring the Observation impact on the Short-range forecast ». In : *Quarterly Journal of the Royal Meteorological Society* 135 (2009), p. 239-250. DOI : [10.1002/qj.366](https://doi.org/10.1002/qj.366).
- [17] C. CARDINALI. *Observation influence diagnostic of a data assimilation system*. 2013.
- [18] A. CARRASSI, S. VANNITSEN et C. NICOLIS. « Model error and Sequential data assimilation : a Deterministic formulation ». In : *Quarterly Journal of the Royal Meteorological Society* 134 (2008), p. 1297-1313. DOI : [10.1002/qj.284](https://doi.org/10.1002/qj.284).
- [19] A. CARRASSI et al. « Data Assimilation in the Geosciences ». In : *Arxiv* (2018), p. 1-79.
- [20] B. CHAPNIK et al. « Diagnosis and tuning of observational error in a quasi-operational data assimilation setting ». In : *Quarterly Journal of the Royal Meteorological Society* 132 (2006), p. 543-565.
- [21] P. COURTIER. « Dual formulation of four-dimensional Variational assimilation ». In : *Quarterly Journal of the Royal Meteorological Society* 123 (1997), p. 2449-2461.
- [22] P. COURTIER, J.-N. THÉPAUT et A. HOLLINGSWORTH. « A strategy for operational Implementation of 4D-Var using an incremental approach ». In : *Quarterly Journal of the Royal Meteorological Society* 120 (1994), p. 1367-1387.
- [23] D. N. DAESCU. « On the Sensitivity Equations of Four-Variational (4 D-Var) data assimilation ». In : *Monthly Weather Review* 136 (2008), p. 3050-3065. DOI : [10.1175/2007MWR2382.1](https://doi.org/10.1175/2007MWR2382.1).
- [24] Dacian N. DAESCU. « Forecast sensitivity to the observation error covariance in variational data assimilation ». In : *Procedia Computer Science* 1 (2012), p. 1277-1285. DOI : [doi:10.1016/j.procs.2010.04.142](https://doi.org/10.1016/j.procs.2010.04.142).
- [25] Nicoles DAGET. « Estimation d'ensemble des paramètres des covariances d'erreur d'ébauche dans un système d'assimilation variationnelle de données océaniques ». Thèse de doct. Université Toulouse III - Paul Sabatier, 2008.
- [26] Cécile L. DEFFORGE et al. « Improving CFD atmospheric simulation at local scale for wind resource assessment using the iterative ensemble Kalman smoother ». In : *Journal of Wind Engineering & Industrial Aerodynamics* 189 (2019), p. 243-257.
- [27] J. E. DENNIS et R. B. SCHNABEL. *Numerical methods for unconstrained optimization and Nonlinear equations*. Sous la dir. de SIAM. SIAM, 1996.
- [28] J. DERBER et F. BOUTTIER. « A reformulation of the background error covariance in the ECMWF global data assimilation system ». In : *Tellus* 51A.2 (1999), p. 195-211.

- [29] G. DESROZIERS et al. « Diagnosis of observation, background and analysis-error statistics in observation space ». In : *Quarterly Journal of the Royal Meteorological Society* 131 (2005), p. 3385-3396. DOI : [10.1256/qj.05.108](https://doi.org/10.1256/qj.05.108).
- [30] Gerald DESROZIERS et Serguei IVANOV. « Diagnosis and Adaptive tuning of Observation-error parameters in a variational assimilation ». In : *Quarterly Journal of the Royal Meteorological Society* 127 (2001), p. 1433-1452.
- [31] Ronald M. ERRICO. « Interpretations of an adjoint-derived observational impact measure ». In : *Tellus* 59A (2007), p. 273-276. DOI : [10.1111/j.1600-0870.2006.00217.x](https://doi.org/10.1111/j.1600-0870.2006.00217.x).
- [32] Erin EVANS et al. « RISE undergraduate find that regime Changes in Lorenz's model are predictable ». In : *Bulletin of the American Meteorological Society* (2004).
- [33] G. EVENSEN. *Data Assimilation, the Ensemble Kalman Filter*. Sous la dir. de SPRINGER. 2nd. Springer-Verlag, 2009.
- [34] G. EVENSEN. « The Ensemble Kalman Filter for Combined State and Parameter Estimation ». In : *IEEE Control Systems* (2009). DOI : [10.1109/MCS.2009.932223](https://doi.org/10.1109/MCS.2009.932223).
- [35] G. EVENSEN et P. J. VAN LEUWEN. « An Ensemble Kalman Smoother for Non-linear Dynamics ». In : *Monthly Weather Review* 128 (2000), p. 1852-1867.
- [36] Geir EVENSEN. « Sequential data assimilation with a nonlinear quasi-Geostrophic model using Monte Carlo methods to forecast error statistics ». In : *Journal of Geophysical Research* 99.C5 (1994), p. 10,143-10,162.
- [37] Geir EVENSEN. « The Ensemble Kalman Filter : theoretical formulation and practical implementation ». In : *Ocean Dynamics* 53 (2003), p. 343-367. DOI : [10.1007/s10236-003-0036-9](https://doi.org/10.1007/s10236-003-0036-9).
- [38] Geir EVENSEN. « Using the Extended Kalman Filter with a Multilayer Quasi-Geostrophic Ocean Model ». In : *Journal of Geophysical Research* 97.C11 (1992), p. 17,905-17,824.
- [39] Alban FARCHI et al. « Using the Wasserstein distance to compare fields of pollutants; application to the radionuclide atmospheric dispersion of the Fukushima-Daiichi accident ». In : *Tellus* 68 (2016), p. 1-20.
- [40] Nelson FEYEU, A. VIDARD et Maëlle NODET. « Optimal transport for variational data assimilation ». In : *Nonlinear Processes in Geophysics* 25 (2018), p. 55-66. DOI : <https://doi.org/10.5194/npg-25-55-2018>.
- [41] J. G. F. FRANCIS. « The QR transformation : a unitary analogue to the LR transformation - part 1 ». In : *The Computer Journal* 4 (1961), p. 265-271. DOI : <https://doi.org/10.1093/comjnl/4.3.265>.
- [42] Li GAOMING et A.C. REYNOLDS. « An Iterative Ensemble Kalman Filter for Data Assimilation ». In : *Society of Petroleum Engineers* (2007), p. 1-18.
- [43] Surykant GAUTAM et al. « Sensors incipient fault detection and isolation using Kalman filter and Kullback-Leibler divergence ». In : *IEEE Transactions on Nuclear Science* (2018). DOI : [10.1109/TNS.2019.2907753](https://doi.org/10.1109/TNS.2019.2907753).
- [44] Ronald GELARO, Yanqiu ZHU et R. M. ERRICO. « Examination of various-order adjoint-based approximations of observation impact ». In : *Meteorologische Zeitschrift* 16 (2007), p. 685-692. DOI : [10.1127/0941-2948/2007/0248](https://doi.org/10.1127/0941-2948/2007/0248).



- [45] Clark R. GIVENS et Rae Michael SHORTT. « A class of Wasserstein metric for probability distributions ». In : *Michigan Mathematic Journal* 31 (1984), p. 231-240.
- [46] Gene H. GOLUB et Charles F. VAN LOAN. *Matrix computations*. Sous la dir. de The John Hopkins University PRESS. The John Hopkins University Press, 2013.
- [47] S. GRATTON, A. S. LAWLESS et N. NICHOLS. « Approximate Gauss-Newton methods for nonlinear least squares problems ». In : *SIAM Journal on Optimization* 18.1 (2007), p. 106-122. DOI : [10.1137/050624935](https://doi.org/10.1137/050624935).
- [48] Serge GRATTON et al. « Observation Thinning in Data Assimilation Computations ». In : *EURO Journal on Computational Optimization* 3 (2015), p. 31-51. DOI : [10.1007/s13675-014-0025-4](https://doi.org/10.1007/s13675-014-0025-4).
- [49] Yaqing GU et Dean S. OLIVER. « An iterative Ensemble Kalman Filter for multiphase fluid flow data assimilation ». In : *Society of Petroleum Engineers* (2007), p. 438-446.
- [50] E. HAIRER et G. WANNER. *Solving ordinary differential equations II : Stiff problems*. Springer, 1996.
- [51] Ernst HAIRER, Syvert Paul NØRSETT et Gerhard WANNER. *Solving ordinary differential equations I : Nonstiff problems*. Springer, 2000.
- [52] G. HALL et J.M. WATT. *Modern numerical methods for ordinary differential equations*. Sous la dir. d'Oxford PRESS. Oxford press, 1976.
- [53] J. HARLIM. « Model error in data assimilation ». In : *Arxiv* (2015).
- [54] Daisuke HOTTA et al. « EFSR : Ensemble Forecast Sensitivity to observation error covariance ». In : *Monthly Weather Review* 145 (2017), p. 5015-5031. DOI : [10.1175/MWR-D-17-0122.1](https://doi.org/10.1175/MWR-D-17-0122.1).
- [55] P. L. HOUTEKAMER et H. L. MITCHELL. « Model error Representation in an Operational EnKF ». In : *Monthly Weather Review* 137 (2008), p. 2126-2144.
- [56] Chen HU et al. « Kullback-Leibler divergence based distributed cubature Kalman filter and its application in cooperative space object tracking ». In : *Entropy* 20.116 (2018), p. 1-20. DOI : [10.3390/e20020116](https://doi.org/10.3390/e20020116).
- [57] B. HUNT et al. « 4DEnKF Filtering for assimilation of Asynchronous Observations ». In : (2003), p. 1-6.
- [58] Brian R. HUNT, Eric J. KOSTELICH et Istvan SZUNYOGH. « Efficient data Assimilation for spatiotemporal chaos : a local Ensemble Kalman Filter ». In : *Physica D* 230 (2007), p. 112-126. DOI : [10.1016/j.physd.2006.11.008](https://doi.org/10.1016/j.physd.2006.11.008).
- [59] Kayo IDE et al. « Unified notation for data assimilation : operational, sequential and variational ». In : *Journal of the Meteorological Society of Japan* 75.1B (1997), p. 181-189.
- [60] T. JANJIC et al. « On the Representation error in data assimilation ». In : *Quarterly Journal of the Royal Meteorological Society* 2017 (2017), p. 1-17. DOI : [10.1002/qj.3130](https://doi.org/10.1002/qj.3130).
- [61] Andrew H. JAZWINSKI. *Stochastic Processes and Filtering theory*. Sous la dir. de DOVER. Dover, 1970.

- [62] Lei JU et al. « An adaptive Gaussian process-based Iterative Ensemble Smoother for data assimilation ». In : *Advances in Water Resources* 115 (2018), p. 125-133.
- [63] R. E. KALMAN. « A New Approach to Linear Filtering and Prediction Problems ». In : *Journal of Basic Engineering* 82 (1960), p. 34-45.
- [64] E. KALNAY et al. « A simple formulation of forecast sensitivity to observations : application to ensemble Kalman filters ». In : *Tellus A* 64 (2012), p. 1-9. DOI : <http://dx.doi.org/10.3402/tellusa.v64i0.18462>.
- [65] Eugenia KALNAY et Shu-Chih YANG. « Accelerating the spin-up of the Ensemble Kalman Filter ». In : *Monthly Weather Review* 136 (2010), p. 1644-1651.
- [66] Jeffrey D. KEPERT. « On ensemble representation of the Observation-error covariance in the Ensemble Kalman Filter ». In : *Ocean Dynamics* 54 (2004), p. 561-569. DOI : [10.1007/s10236-004-0104-9](https://doi.org/10.1007/s10236-004-0104-9).
- [67] Sung-Min KIM et Hyun Mee KIM. « Forecast Sensitivity Observation Impact in the 4D-Var and Hybrid 4D-Var data assimilation systems ». In : *Journal of Atmospheric and Oceanic Technology* 36 (2019), p. 1563-1575. DOI : [10.1175/JTECH-D-18-0240.1](https://doi.org/10.1175/JTECH-D-18-0240.1).
- [68] Richard KLEEMAN. « Information theory and dynamical system predictability ». In : *Entropy* 13 (2011), p. 612-649.
- [69] Daryl KLEIST et al. *Hybrid 4D ENVar for the NCEP GFS*. 2014.
- [70] Shunji KOTSUKI, Kenta KUROSAWA et T. MIYOSHI. « On the properties of ensemble forecast sensitivity to observations ». In : *Quarterly Journal of the Royal Meteorological Society* 145 (2019), p. 1897-1914. DOI : [10.1002/qj.3534](https://doi.org/10.1002/qj.3534).
- [71] Brett ØKSENDAL. *Stochastic Differential Equations*. Sous la dir. de SPRINGER. Springer-Verlag, 2016.
- [72] Patrick LALOYEAUX et Massimo BONAVIDA. « Improving the handling of model bias in data assimilation ». In : *ECMWF Newsletter* 163 (2020), p. 18-22.
- [73] Rolf H. LANGLAND et Nancy L. BAKER. « Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system ». In : *Tellus* 56A (2004), p. 189-201.
- [74] F.-X. LE DIMET, I. M. NAVON et D. N. DAESCU. « Second-order information in data assimilation ». In : *Monthly Weather Review* 130 (2002), p. 629-648.
- [75] F.-X. LE DIMET et O. TALAGRAND. « Variational algorithms for analysis and assimilation of meteorological observations : theoretical aspects ». In : *Tellus* 38A (1986), p. 97-110.
- [76] Long LI et al. « Topological data assimilation using Wasserstein distance ». In : *Inverse Problems* (2018). DOI : <https://doi.org/10.1088/1361-6420/aae993>.
- [77] Ruoxia LI et al. « Constrained multimodal Ensemble Kalman filter based on Kullback-Leibler divergence ». In : *Journal of Process Control* 79 (2019), p. 16-28. DOI : [10.1016/j.jprocont.2019.03.012](https://doi.org/10.1016/j.jprocont.2019.03.012).
- [78] Junjie LIU et E. KALNAY. « Estimating Observation Impact without adjoint model in an Ensemble Kalman filter ». In : *Quarterly Journal of the Royal Meteorological Society* 134 (2008), 1327-1335. DOI : [10.1002/qj.280](https://doi.org/10.1002/qj.280).

- [79] A. C. LORENC. « The potential of the ensemble Kalman filter for NWP – a comparison with 4D-Var ». In : *Quarterly Journal of the Royal Meteorological Society* 129 (2003), p. 3183-3203.
- [80] E. N. LORENZ. « A study of the predictability of a 28-variable Atmospheric model ». In : *Tellus XVII* (1965), p. 321-333.
- [81] E. N. LORENZ. « Atmospheric predictability Experiments with a large numerical model ». In : *Tellus* 34 (1982), p. 505-513.
- [82] E. N. LORENZ. « Predictability : a problem partially solved ». In : *ECMWF Workshop on predictability*. 2002.
- [83] E. N. LORENZ. « The predictability of a flow which possesses many scales of motion ». In : *Tellus XXI* (1969), p. 289-307.
- [84] Edward N. LORENZ. « Deterministic Nonperiodic Flow ». In : *Journal of the atmospheric sciences* 20 (1963), p. 130-141.
- [85] Cristina LUPU, Pierre GAUTHIER et Stéphane LAROCHE. « Evaluation of the impact of observations on analyses in 3D- and 4D-Var based on information content ». In : *Monthly Weather Review* 139 (2011), p. 726-737.
- [86] A. MAJDA, R. KLEEMAN et D. CAI. « A mathematical framework for quantifying predictability through relative entropy ». In : *Methods and Applications of Analyses* 9.3 (sept. 2002), p. 425-444.
- [87] Majdi MANSOURI, Hazem NOUNOU et Mohamed NOUNOU. « Kullback-Leibler divergence-based improved particle filter ». In : *IEEE 11th International Multi-Conference on Systems, Signals and Devices, SSD 2014*. Sous la dir. d'IEEE. 68087932014. IEEE. IEEE, 2014.
- [88] Serge MARGUET. *La physique des réacteurs nucléaires*. Sous la dir. de Tec & DOC. Lavoisier, 2011.
- [89] Matthias MORZFELD, Daniel HODYSS et Chris SNYDER. « What the collapse of the Ensemble Kalman filter tells us about particle filters ». In : *Tellus* 69 (2017), p. 1-14. DOI : [10.1080/16000870.2017.1283809](https://doi.org/10.1080/16000870.2017.1283809).
- [90] C. NICOLIS. « Dynamics of model error : some generic features ». In : *Journal of the atmospheric sciences* 60 (2003), p. 2208-2218.
- [91] Lipeng NING et al. « Coping with model error in variational data assimilation using optimal mass transport ». In : *Water Resources Research* (2014), p. 5817-5830. DOI : [10.1002/2013WR014966](https://doi.org/10.1002/2013WR014966).
- [92] Angélique PONÇOT. « Assimilation de données pour la dynamique du Xénon dans les coeurs de centrale nucléaire ». Thèse de doct. INP Toulouse, 2008.
- [93] Paul REUSS. *Précis de Neutronique*. Sous la dir. d'EDP SCIENCES. INSTN, 2003.
- [94] P. SAKOV, Dean S. OLIVER et Laurent BERTINO. « An Iterative EnKF for Strongly Nonlinear Systems ». In : *Monthly Weather Review* 140 (2012), p. 1988-2004. DOI : [10.1175/MWR-D-11-00176.1](https://doi.org/10.1175/MWR-D-11-00176.1).
- [95] P. SAKOV et al. « TOPAZ4 : an ocean-sea ice data assimilation system for the North Atlantic and Arctic ». In : *Ocean Science* 8 (2012), p. 633-656. DOI : [10.5194/os-8-633-2012](https://doi.org/10.5194/os-8-633-2012).

- [96] Pavel SAKOV et Laurent BERTINO. « Relation between two common localisation methods for the EnKF ». In : *Computational Geoscience* (2010). DOI : [10.1007/s10596-010-9202-6](https://doi.org/10.1007/s10596-010-9202-6).
- [97] Pavel SAKOV, Geir EVENSEN et L. BERTINO. « Asynchronous data Assimilation with the EnKF ». In : *Tellus* 62 (2010), p. 24-29. DOI : [10.1111/j.1600-0870.2009.00417.x](https://doi.org/10.1111/j.1600-0870.2009.00417.x).
- [98] V. SHUTYAEV et al. « On model error in variational data assimilation ». In : *Russian Journal of Numerical Analysis and Mathematical Modelling* 31.2 (2016), p. 105-113. DOI : [10.1515/rnam-2016-0011](https://doi.org/10.1515/rnam-2016-0011).
- [99] V. SHUTYAEV et al. « Optimal solution error quantification in variational data assimilation involving imperfect models ». In : *International Journal for Numerical Methods in Fluids* (2016). DOI : [10.1002/flid.4266](https://doi.org/10.1002/flid.4266).
- [100] Joram SOCH et Carsten ALLEFELD. « Kullback-Leibler divergence for the Normal-Gamma distribution ». In : *Arxiv* (2016), p. 1-11.
- [101] Sergei SOLDATENKO et Rafael YUSUPOV. « Predictability in Deterministic Dynamical Systems with applications to weather forecasting and climate modelling ». In : IntechOpen, 2017. Chap. 5, p. 101-122. DOI : [10.5772/66752](https://doi.org/10.5772/66752).
- [102] Matthias SOMMER et Martin WEISSMANN. « Ensemble-based approximation of observation impact using an observation-based verification metric ». In : *Tellus* 68 (2016), p. 1-13. DOI : <http://dx.doi.org/10.3402/tellusa.v68.27885>.
- [103] Sagar K. TAMANG et al. « Regularized variational data assimilation for bias treatment using the Wasserstein metric ». In : *Quarterly Journal of the Royal Meteorological Society* (2020), p. 1-15. DOI : [10.1002/qj.3794](https://doi.org/10.1002/qj.3794).
- [104] R. TODLING. « Comparing two approaches for assessing observation impact ». In : *Monthly Weather Review* 141 (2013), p. 1484-1501. DOI : [10.1175/MWR-D-12-00100.1](https://doi.org/10.1175/MWR-D-12-00100.1).
- [105] Vu Duc TRAN. « Assimilation de données : les propriétés asymptotiques du filtre de Kalman d'ensemble ». Thèse de doct. Université de Bretagne Sud, 2009.
- [106] Y. TRÉMOLET. « Accounting for an imperfect model in 4D-Var ». In : *Quarterly Journal of the Royal Meteorological Society* 132.621 (2006), p. 2483-2504. DOI : [10.1256/qj.05.224](https://doi.org/10.1256/qj.05.224).
- [107] Y. TRÉMOLET. « Computation of Observation Sensitivity and Observation impact in incremental variational data assimilation ». In : *Tellus* 60A (2008), p. 964-978.
- [108] Y. TRÉMOLET. « Corrigendum for 'Accounting for an imperfect model in 4D-Var' ». In : *Quarterly Journal of the Royal Meteorological Society* 132.621 (2006), p. 3127.
- [109] Y. TRÉMOLET. « Weak Constraints 4D-Var for the Stratosphere ». In : *SPARC Workshop*. 2010.
- [110] Yannick TRÉMOLET. « First-order and high-order approximations of observation impact ». In : *Meteorologische Zeitschrift* 16.6 (2007), p. 693-694.
- [111] Marios TSATSOS. « Theoretical and Numerical Study of the Van der Pol equation ». Dissertation. Aristotle University of Thessaloniki, 2006.

- [112] P. A. VIDARD, A. PIACENTINI et F.-X. LE DIMET. « Variational data analysis with control of the forecast bias ». In : *Tellus* 56A (2004), p. 177-188.
- [113] Cédric VILLANI. *Optimal transport old and new*. Sous la dir. de SPRINGER. Springer, 2009.
- [114] Grace WHABA et James WENDELBERG. « Some new Mathematical Methods for Variational Objective Analysis using Splines and Cross Validation ». In : *Monthly Weather Review* 108 (1980), p. 1122-1143.
- [115] Grace WHABA et al. « Adaptive tuning of Numerical Weather Prediction Models : Randomized GCV in Three- and Four-Dimensional Data Assimilation ». In : *Monthly Weather Review* 123 (1995), p. 3358-3369.
- [116] J. S. WHITAKER et Thomas M. HAMILL. « Ensemble data assimilation without perturbed observations ». In : *Monthly Weather Review* 130 (2002), p. 1913-1924.
- [117] Jiping XIE et al. « Evaluation of arctic ocean surface salinities from the soil moisture and ocean salinity (SMOS) mission against a regional reanalysis and in situ data ». In : *Ocean Science* 15 (2019), p. 1191-1206.
- [118] Qin XU. « Measuring information content from observations for data assimilation : relative entropy versus shannon entropy difference ». In : *Tellus A* 59 (2007), p. 198-209. DOI : [10.1111/j.1600-0870.2006.00222.x](https://doi.org/10.1111/j.1600-0870.2006.00222.x).

**Doctorat de l'Université de Toulouse**  
**Délivré par l'Institut National Polytechnique de Toulouse**  
**Ecole doctorale : MITT**  
**Spécialité : Mathématiques appliquées**  
**Date de soutenance : A définir**  
**Auteur : Dimitri Mottet**  
**Directeur de thèse : Serge GRATTON**

---

### **Résumé**

Le filtrage par filtre de Kalman d'ensemble (EnKF) pour des systèmes dynamiques non-linéaires nécessite de raffiner l'algorithme initial pour obtenir de bonnes performances. Les indicateurs de qualité de prévision « Forecast Sensitivity Observation Impact » (FSOI) permettent ces améliorations. En suivant cette voie, cette thèse propose d'utiliser et comparer de nouveaux indicateurs inspirés des FSOI, pour formuler des stratégies d'assimilation consistant à sélectionner les instants des étapes d'analyse et le nombre d'observations assimilées pour chacune. Les indicateurs *a priori* se calculent à l'arrivée d'une observation et les indicateurs *a posteriori* après l'étape d'analyse. Leurs coûts numériques sont calculés et discutés, montrant une utilisation possible pour des algorithmes type ETKF. Les indicateurs sont testés sur des systèmes emblématiques de petite dimension et un système industriel. Nous montrons la possibilité de sélectionner pour l'assimilation les observations les plus importantes, et les instants d'assimilation les plus opportuns pour un système fortement observé. Cela valide les indicateurs pour des essais sur des systèmes industriels plus complexes. Des perspectives sont données pour leur insertion dans les filtres d'ensemble itératifs, la réduction de leur coût numérique, et des applications industrielles possibles.

**Mots-clés :** assimilation de données, EnKF, diagnostics a priori, diagnostics a posteriori, impact des observations, qualité de prévision.

---

**Title :** Adaptive refinement of the data assimilation process by ensemble Kalman methods for non-linear problems.

### **Abstract**

Data assimilation with ensemble Kalman filters (EnKF) for non-linear dynamical systems implies some refinements of the base algorithm to get good overall filtering performances, even more when the system's dimension increases. Forecast quality indicators such as « Forecast Sensitivity Observation Impact » ones (FSOI) allows for improvements of the base algorithm. By following their insights this thesis aims to use and compare new FSOI-inspired indicators to establish assimilation strategies that consist in dynamically managing analysis steps instants as well as the number of observations assimilated at each analysis step. *A priori* indicators are computed at an observation's arrival while *a posteriori* ones are computed after the analysis step. Numerical costs of those indicators are computed and discussed showing their pertinence for ETKF-based algorithms and they are tested on emblematic small dimension systems and an industrial system. Results show the possibility to select for assimilation only the most important observations as well as the best analysis step instants when the system is heavily observed. It validates these indicators for further tests on more complex industrial systems. Perspectives for their use in iterative ensemble filters are given, as well as for further cost reductions and possible industrial applications.

**Keywords :** data assimilation, EnKF, a priori diagnostics, a posteriori diagnostics, observation impact, forecast quality.

---

### **Intitulé et adresse du laboratoire :**

IRIT-ENSEEIH  
Bâtiment F  
2, Rue Charles Camichel  
31000 TOULOUSE