# Quantification d'incertitudes et assimilation de données pour la modélisation hydrodynamique bidimensionnelle : application au modèle de prévision des hautes eaux de l'estuaire de la Gironde

Vanessya Laborie

▶ **To cite this version:**

Vanessya Laborie. Quantification d'incertitudes et assimilation de données pour la modélisation hydrodynamique bidimensionnelle : application au modèle de prévision des hautes eaux de l'estuaire de la Gironde. Ingénierie de l'environnement. Université Paris-Est, 2020. Français. NNT : 2020PESC1006 . tel-03020223

**HAL Id: tel-03020223**
**https://pastel.hal.science/tel-03020223**

Submitted on 23 Nov 2020

**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 PARIS-EST École Doctorale SIE n° 531 : Sciences, Ingénierie, Environnement

# THÈSE DE DOCTORAT

Spécialité : Sciences et Techniques de l'Environnement

Présentée par

**Vanessya LABORIE**

pour l'obtention du grade de docteure délivré par

L'UNIVERSITE PARIS-EST

# Quantification d'incertitudes et assimilation de données pour la modélisation hydrodynamique bidimensionnelle. Application au modèle Telemac2D de prévision des hautes eaux de l'estuaire de la Gironde

Soutenue le 13 / 01 / 2020 devant un jury composé de :

| | | |
|---|---|---|
| Pr. Olivier THUAL | Toulouse INP | Président du jury |
| Pr. Benoît LAIGNEL | Université de Rouen Normandie | Rapporteur |
| Dr. Emmanuel COSME | Université Grenoble Alpes | Rapporteur |
| Pr. Paul BATES | Université de Bristol | Examinateur |
| Dr. Nicole GOUTAL | EDF R&D/LHSV | Directrice de thèse |
| Dr. Sophie RICCI | CERFACS | Co-directrice de thèse |
| Dr. Maëlle NODET | Université Versailles-Saint Quentin | Invitée |
| Dr. Philippe SERGENT | Cerema | Invité |
| Dr. Gwenaële JAN | SHOM | Invitée |

Laboratoire d'Hydraulique Saint-Venant
(unité de recherche commune **Cerema**, **EDF R&D**, **École des Ponts Paris-Tech)**

Thèse effectuée au sein du **Laboratoire d'Hydraulique Saint-Venant**

de l'Université Paris-Est

6, quai Watier

BP 49

78401 Chatou Cedex

France

En collaboration avec le **CERFACS**

42 avenue Gaspard Coriolis 31057 Toulouse Cedex 01

# Résumé

Cette thèse s'inscrit dans le cadre du développement de techniques d'assimilation de données dans les codes de calcul bidimensionnels d'hydraulique à surface libre et, plus particulièrement, de l'évaluation des possibilités d'amélioration de la prévision des hautes eaux sur l'estuaire de la Gironde par le développement et l'utilisation de ces méthodes. Les caractéristiques principales de ces travaux concernent le caractère instationnaire de l'hydrodynamique, liée à la fois à la propagation de la marée et à l'événement hydrométéorologique considéré, et la corrélation entre les observations des différentes stations de mesures disponibles.

Les travaux réalisés pendant la première partie de la thèse ont porté sur la proposition d'une méthodologie de traitement des incertitudes par analyse globale de sensibilité par décomposition de variance pour un code d'hydrodynamique alimenté par des variables fonctionnelles. Cette méthodologie a permis la classification des sources d'incertitudes à partir d'un échantillon des réalisations du modèle Telemac2D de prévision des hautes eaux de l'estuaire de la Gironde, dans le but de hiérarchiser l'influence des forçages et des paramètres du modèle.

Dans un second temps, en lien avec les résultats de l'étude précédente, une méthodologie a été proposée pour développer une chaîne de simulation intégrant le code de calcul étudié et une méthode ensembliste d'assimilation de données fondée sur le Filtre de Kalman d'Ensemble, dans l'objectif de réduire les incertitudes du modèle liées à ses forçages fonctionnels et à sa paramétrisation par l'utilisation d'observations. Dans un premier temps, cette méthodologie a été appliquée au modèle Telemac2D de l'estuaire de la Gironde sous la forme d'un protocole d'expériences jumelles dont l'objectif est, d'une part, de procéder à sa validation et, d'autre part, de déterminer de manière optimale le vecteur de contrôle, le réseau de mesures et les paramètres liés à la méthodologie proposée. Les gains de la méthodologie proposée sont ensuite évalués en utilisant les observations disponibles pour plusieurs événements réels.

**Mots-clé :**

Hydrodynamique bidimensionnelle à surface libre, Telemac2D, surcotes, quantification des incertitudes, variables fonctionnelles, analyse de sensibilité, ANOVA, assimilation de données, filtre de Kalman d'Ensemble, estuaire de la Gironde, réseau de mesures.

# Uncertainty quantification and data assimilation for 2D shallow water based numerical models. Application to the Gironde estuary numerical model forecasting high water levels

# Abstract

The accuracy of predictions is a major issue in flood forecasting. This thesis analyzes the development of data assimilation techniques in bidimensional numerical codes based on the shallow water equations and, more specifically, the assessment of possible improvements in forecasting high water levels in the Gironde estuary through the development and use of these methods. The main characteristics of this work concern the non-stationarity of the hydrodynamics, which is linked to both the tidal propagation and hydrometeorological events, and the correlation between the observations of the different available observation stations.

The work carried out during the first part of the thesis focused on the proposal of a methodology for the treatment of uncertainties by a global sensitivity analysis and a variance decomposition for a code forced by time-dependent variables. This methodology made it possible to classify the sources of uncertainty based on a sample of realizations of the Telemac2D numerical model forecasting high water levels in the Gironde estuary, in order to prioritize the influence of the model's forcing terms and parameters. In the second part of the thesis, in connection with the results of the first study, a methodology was proposed to develop a simulation chain integrating the numerical code and an ensemblist data assimilation method based on an Ensemble Kalman Filter, in order to reduce the uncertainties of the model related to its time-dependent forcing terms and parameterization. First, this methodology was applied to the Telemac2D model of the Gironde estuary with a twin experimental protocol with the objective, on one hand, to validate it and, on the other hand, to determine optimally the control vector, the measurement network, and the parameters related to the proposed methodology. The advantages for forecasting high water levels in the Gironde estuary are then evaluated using the observations available for several real events.

**Keywords :**

Bidimensional free surface flows, Telemac2D, surge levels, uncertainty quantification, time-dependent variables, global sensitivity analysis, ANOVA, data assimilation, Ensemble Kalman Filter, Gironde estuary, observation network.

# Remerciements

C'est un grand plaisir pour moi que de clore la rédaction de ce mémoire de thèse en remerciant toutes celles et ceux qui m'ont accompagnée, encouragée, guidée et soutenue pendant ces 4 années. Cette thèse est le fruit d'un travail collectif au cours duquel j'ai énormément appris sur les plans humain et scientifique.

Je souhaite remercier chaleureusement et sincèrement le tandem formé par Nicole Goutal et Sophie Ricci qui ont accepté de diriger mon travail de thèse. Ces 4 années resteront parmi les plus enrichissantes de ma carrière professionnelle grâce à nos échanges extrêmement constructifs. J'ai tout à la fois bénéficié de leurs compétences et de leur complémentarité. Comme pour la montée vers la tête de l'Obiou, le parcours a été chaotique, constitué de retours en arrière, d'avancées inespérées, mais aussi d'un cheminement positif implacable avant l'arrête sommitale. Je les remercie de m'avoir supportée tout au long et jusqu'au bout de ce doctorat, début prometteur de collaborations fructueuses.

Je remercie Nicole, qui intègre à la fois des qualités humaines et des compétences multiples en hydraulique, en numérique, en quantification des incertitudes. Nicole, tu as rendu l'aventure possible en la soutenant administrativement dès ses prémisses, mais aussi scientifiquement et techniquement, en me permettant d'accéder, en particulier, aux moyens de calcul d'EDF R&D et en m'ouvrant ton réseau. Mes conditions matérielles et informatiques ont été exceptionnelles grâce à toi. Ton recul, nos échanges, ta disponibilité et ta confiance en moi m'ont été d'une aide inestimable. Je suis ravie de continuer mes activités de recherche au sein du Laboratoire d'Hydraulique Saint-Venant, dont tu es la Directrice, et de continuer notre collaboration.

Je remercie Sophie, si pétillante, disponible et dynamique. Sophie, je te remercie de m'avoir fait bénéficier pendant toute cette collaboration de tes idées et de ta compétence (et de celle de tes post-docs aussi) ! Les complémentarités dans nos compétences ont abouti à des échanges extrêmement riches que je souhaite voir continuer dans le futur. Tu n'as pas ménagé ton temps (au-delà des heures ouvrées) pour me permettre d'atteindre dans des conditions exceptionnelles le Saint-Graal de la publication, de la fin de la rédaction du mémoire, puis de la soutenance de thèse. Un immense merci à toi.

Je remercie l'ensemble des membres du jury de ma thèse, qui m'ont fait l'honneur de bien vouloir examiner mon travail.

Benoît Laignel, professeur à l'université de Rouen, et Emmanuel Cosme, maître de conférence à l'université Grenoble-Alpes, m'ont fait l'honneur d'être rapporteurs de ma thèse. Ils ont consacré du temps à la lecture de mon mémoire. Leurs remarques et nos échanges m'ont permis d'envisager mon travail sous un autre angle. Pour tout cela je les remercie et espère que nos collaborations

se poursuivront dans le futur.
J'exprime toute ma gratitude à Monsieur Paul Bates et à Monsieur Olivier Thual, respectivement Professeurs à l'Université de Bristol et Toulouse INP, qui ont bien voulu être examinateurs de mes travaux. Leur bienveillance et les pistes de recherche qu'ils m'ont apportées sont autant de perspectives intéressantes pour améliorer les résultats obtenus pendant ma thèse.
Philippe Sergent, directeur scientifique du Cerema EMF, co-construit mon parcours professionnel depuis mon arrivée dans la drim-team du CETMEF en 2005. Philippe, je te remercie pour avoir supporté mon projet de thèse intégré dans ma promotion à divi (que tu as soutenue également), mais aussi pour le temps que tu as consacré à me conseiller, à ouvrir mes horizons au travers de nos discussions scientifiques autour de mes résultats pendant ces années de thèse. Je te suis, enfin, très reconnaissante d'avoir accepté de faire partie du comité de pilotage scientifique et du comité restreint de suivi de ma thèse. Merci pour ton aide, nos échanges scientifiques et ta bienveillance.
Gwenaële Jan, ingénieure de recherche au SHOM, a successivement accepté de faire partie du comité de pilotage scientifique, puis du comité restreint de suivi, puis de mon jury de thèse. Je te remercie sincèrement pour ta bienveillance, Gwenaële, mais aussi pour tes apports précieux et nos discussions fructueuses sur la précision de la marée. Les sujets pour des travaux communs autour des estuaires et de la propagation de la marée ne manquent pas et nous permettront de continuer sur notre lancée.
Comme Gwenaële, Maëlle NODET, maitresse de conférences au sein de l'université Versailles-Saint-Quentin, fait partie des belles rencontres de ma thèse. Je te remercie, Maëlle, pour avoir accepté de contribuer par ton immense compétence (et ton pragmatisme) à mon jury de thèse et pour les échanges très fructueux qui s'en sont suivis sur la base de tes questionnements sur les méthodes d'assimilation implémentées, ainsi que la pertinence du réensemencement, entre autres.

Plusieurs personnes ont eu des contributions majeures dans la menée de ces travaux.
Matthias De Lozzo s'est montré patient et infiniment pédagogue face à mes interrogations sur les processus gaussiens, les indices de Sobol' et leurs usages dans le cas spatio-temporel de l'estuaire de la Gironde. Je n'oublie pas qu'une grande partie de l'originalité de mes résultats et de leur représentation graphique vient de lui. Un immense merci à toi, Matthias.
Pour la mise en œuvre du GENCI à l'IDRIS et des maquettes de quantification d'incertitudes et d'assimilation de données sur les super-calculateurs d'EDF R&D, rien n'aurait été possible informatiquement sans Yoann Audouin, qui allie compétence, disponibilité, gentillesse et efficacité. Je te l'ai souvent dit, Yoann, mais je le réaffirme (paraphrasant ainsi de nombreux collègues passés ou présents du LHSV) : je te remercie sincèrement pour le temps que tu as consacré à résoudre mes problèmes avec Telemac2D, ses api, le lancement des calculs, etc. Je te souhaite le meilleur pour tes projets à venir.
Sans données, difficile d'assimiler quoi que ce soit dans un modèle numérique. Je remercie chaleureusement Laurent Dieval, prévisionniste au Service de Prévision des Crues Gironde, Adour,

Dordogne et docteur en méca flu, pour sa disponibilité et sa réactivité dans l'envoi des données nécessaires. Merci pour nos nombreux échanges techniques concernant le travail opérationnel des prévisionnistes. L'estuaire de la Gironde offre tant de terrains de jeux à expérimenter que j'espère pouvoir longtemps poursuivre cette collaboration.
Je remercie enfin Patrick Ohl et Guillaume Morvan, de la division PREVIMAR de Météo-France, pour leur contribution majeure concernant les forçages météorologiques utilisés dans le cadre de ces travaux. Vous m'avez beaucoup aidée par votre réactivité et nos discussions techniques sur l'usage des faisceaux de données et les conditions aux limites des prévisions d'ensemble.

Je remercie le comité de pilotage scientifique de ma thèse (Denis Paradis, Bruno Janet, Yan Lacaze, Etienne Lepape, Nicolas Huybrechts et celles et ceux cités précédemment) pour les séances d'échanges passionnants et fructueux que nous avons pu mener au cours de ces 4 années. Je ne doute pas que notre collaboration ne s'arrêtera pas là.

Je remercie sincèrement pour leur aide régulière et efficace les chargé.e.s de la documentation de l'ENPC (Romain Boistel et Delphine Du Pasquier). Merci, Romain, pour m'avoir permis d'accéder aux précieuses ressources documentaires de l'ENPC. Merci à tous les deux pour votre réactivité exceptionnelle quand j'avais besoin d'un article "introuvable". Mon état de l'art vous doit beaucoup !
Un grand merci à Christelle Coloigner et Chantal Kasprzycki, assistantes de département au sein du Cerema EMF, et à Sergina Opou Nkouet, assistante du LHSV, qui ont indéniablement facilité mes missions, l'envoi des mémoires au jury, ainsi que l'organisation et la tenue de ma soutenance. Je salue avec une grande reconnaissance leur patience, leur efficacité et leur réactivité face à mes nombreuses sollicitations.

Mes travaux de thèse ont bénéficié de moyens informatiques exceptionnels au travers des ressources d'EDF R&D et de l'IDRIS.
Je remercie EDF R&D, la tutelle industrielle du laboratoire d'Hydraulique Saint-Venant, qui me permet d'utiliser ses super-calculateurs et de bénéficier de la gentillesse, de la compétence et de la réactivité des personnels "supports" de la DSI d'EDF. Un sincère merci à Cyril Baudry, Cristian Ruiz, Véronique Gouelle, Rachid Boulbaroud et tou.te.s celles et ceux qui gèrent et apportent des solutions aux nombreux tickets d'assistance que j'ouvre régulièrement.
Les deux projets soumis et acceptés par les Grands Équipements Nationaux de Calcul Intensif (GENCI), soumis en 2017 et 2018, m'ont permis d'accéder globalement à plus de 30 millions d'heures-coeur gratuitement sur l'un des super-calculateurs de l'IDRIS, centre de calcul national. Je remercie chaleureusement Isabelle Dupays et Sylvie Thérond pour leur réactivité, leur disponibilité et leur aide précieuses lors du déploiement et l'utilisation de Telemac2D sur Turing.
Le support informatique du Cerema Eau, Mer et Fleuves m'a énormément aidée afin que je puisse accéder à tous ces moyens de calcul depuis le/mon siège du Cerema EMF à Margny. Je remercie infiniment Michel Rogard et Michel Pennec pour leur patience et leur réactivité face à mes nombreuses demandes de "dérogation" aux pare-feux ministériels et autres systèmes de sécurité. Michel et Michel, vous m'avez facilité le travail jusqu'à ma soutenance de thèse en gérant avec

brio la retransmission (et les tests préalables) et en mettant en œuvre les conseils pratiques de Raphaël Le Guillou, que je remercie également pour sa réactivité et son aide efficace régulières pendant mes travaux. Merci également à Patrice Vincelet pour son aide lors de la retransmission. Je remercie enfin l'équipe de Données & Logiciels du Cerema Eau, Mer et Fleuves (Philippe, Christophe, Bruno et Guillaume) pour la gestion du cluster du Cerema et leur aide lors du crash de mes partitions.

Merci aussi aux prévisionnistes et météorologues de Météo-France (Carole Labadie, Pierrick Cebron, Laurent Descamps et Philippe Arbogast du GMAP), ainsi qu'aux océanographes de Toulouse (Florent Lyard, Nadia Ayoub) et de Rouen (Benoît Laignel et Imen Turki) et les chercheurs du Cerea pour l'intérêt qu'ils ont porté à mon travail, nos échanges techniques passionnants et leurs suggestions toujours intéressantes et pertinentes.

J'ai la chance d'évoluer dans deux univers professionnels, celui du Cerema et celui du LHSV (y.c EDF R&D), aussi riches et conviviaux l'un que l'autre et d'avoir, grâce à Sophie, été intégrée dans celui du Cerfacs, qui ne l'est pas moins.
Je remercie mes collègues du Cerema Eau, Mer et Fleuves qui m'ont soutenue pendant ce projet professionnel spécifique et, parmi eux, Nicolas Huybrechts, Sylvain Orseau avec qui les discussions dans le cadre de GirondeXL3D ont été très intéressantes, Céline Chouteau pour nos échanges constructifs et/ou révolutionnaires, François Ropert pour le bonjour amical du matin et le plaisir de parler statistique de manière intelligible et pratique.
Merci aussi à tou.te.s les chercheurs.ses et (post)-doctorant.e.s du LHSV (et ***in extenso*** à celles et ceux d'EDF R&D,) qui contribuent à l'émulation intellectuelle et à la bonne ambiance que j'apprécie tant quand je viens à Chatou, ainsi qu'à mes co-bureaux de longue durée (Teddy, Marina et Frédérique). Les petits moments d'échanges et de partage de confiseries (sans allergènes) ont été très réconfortants dans mes moments de doute.
J'ai également eu la chance de pouvoir échanger avec Mélanie Rochoux, ingénieure de recherche au Cerfacs, aux multiples qualités humaines et aux compétences scientifiques non moins nombreuses. La thèse de Sébastien Barthélémy, encadrée par Olivier Thual et Sophie Ricci au sein du Cerfacs, m'a beaucoup aidée pour dresser l'état de l'art des techniques d'assimilation de données. Je le remercie pour nos échanges techniques et le temps qu'il a consacré au défrichage de mon sujet au début de ma thèse. Je remercie enfin Isabelle Mirouze pour ses conseils éclairés et Andréa Piacentini pour ses encouragements face à l'ampleur de la tâche.
François Hissel se trouve à l'origine de mes travaux sur l'estuaire de la Gironde et en assimilation de données. Je te remercie, François, pour ton soutien pendant les années difficiles et ta présence amicale le jour J. Marissa Yates-Michelin fait partie des âmes bienveillantes qui ont largement contribué au succès de ma thèse. Merci, Marissa, pour ton soutien et ton aide précieuse pendant ces deux dernières années.

Le Ministère en charge du Développement Durable et le Cerema, au sein duquel je suis en Position Normale d'Activité, m'ont permis de m'investir à temps partiel afin d'approfondir mon projet de recherche en lien avec mon activité professionnelle. Je remercie la direction générale du Cerema,

la DSTREI, ainsi que la direction de EMF pour m'avoir permis de mener à bien ce projet, ainsi que le SCHAPI (et Bruno Janet, en particulier), qui a soutenu financièrement la réalisation de ces travaux.

Pour finir, je souhaite remercier ma famille (y compris celle qui est belle) et mes ami.e.s. Toutes celles et ceux qui me sont proches savent à quel point leur soutien et leurs encouragements ont contribué à la réussite de mon projet de thèse.
Je remercie Quentin, mon Bilou qui m'a tant soutenue et encouragée dans les moments de doute, et Clarisse, si câline et exigeante. Clarisse et Quentin, vous avez été le cadre qui m'a empêchée de me noyer et m'a permis de faire aboutir cette thèse.
Je remercie mes parents, toujours présents, qui m'ont transmis le sens du mot "travail" avec modestie et exemplarité. Leur force et leur générosité sont un modèle pour moi. Je leur adresse aussi un immense merci pour mon pot de thèse qui restera gravé dans ma mémoire. Un grand merci à tous les deux !
Merci aussi à ma sœur Magaly pour son soutien, son organisation infaillible et son dynamisme calme et efficace. Mag, je te remercie infiniment pour avoir été présente le jour de ma soutenance, pour avoir, avec papa, maman, tant aidé avant, pendant et après mon pot de thèse et pour avoir rédigé une synthèse en trois pages de mon mémoire plus intelligible que ce dernier.
Sylvie et Tita, ainsi que Marie-Thérèse, m'ont fait l'immense joie d'être présentes. Mes amies Hélène et Brigitte ont assisté Bertrand pour l'élaboration expresse et inattendue de mon pot de thèse et m'ont permis de passer une très agréable soirée de fin de thèse. Je vous remercie toutes pour votre présence, vos amicales attentions et vos encouragements.
Merci à ma sœur Sonya pour son soutien à distance et pour avoir cru en moi dès mon plus jeune âge, largement révolu à cette heure. Mme Van Hove (***Nounou***) et ma belle-famille (Marie-Thérèse, Sophie, Claire, Alice, Erik, Florence, Nico, etc.) ont beaucoup facilité l'avancée de mes travaux pendant toute leur durée, je les en remercie sincèrement.
Toute l'équipe de l'amap d'Ourscamp (Victor, Jean-François, Sylvie, Prunelle), mon maraîcher préféré, Philippe, mes amies (Véronique, Stéphanie, Hélène, Aurélie, Amélie), mais aussi Force Ouvrière et son cartel FO-Cerema (Gwénaël, Arno) m'ont encouragée et permis de prendre du recul en m'échappant de mes préoccupations professionnelles quand il le fallait. Je les en remercie vivement.

Je remercie enfin Bertrand pour m'avoir encouragée à réaliser mon projet et forcée, comme au cours de nos si nombreuses promenades en montagne, à me dépasser sur tous les plans. Tu m'as soutenue sans complaisance sur le chemin sinueux que j'avais choisi et c'est sans doute grâce à cela que je suis heureuse du beau chemin parcouru, sans y avoir sacrifié le reste. Merci pour tes incessants et ironiques "Alors qu'as-tu trouvé aujourd'hui ?" ou tes "le problème, c'est qu'il y a trop de chercheurs et peu de trouveurs"' qui m'ont permis de relativiser tout au long de mes travaux. Tu restes mon relecteur préféré… et le plus urticant.

*À Bertrand*
*et notre plus belle recherche expérimentale,*
*Quentin et Clarisse*
*À mes parents, Annick et Henry Laborie,*
*mes sœurs et*
*mon grand-père, Henri Laborie*

*Gutta cavat lapidem non vi, sed saepe cadendo.*

# Table des matières

**Principaux acronymes** **1**

**Notations principales** **7**

**Introduction** **9**

**1 Prévision des hautes eaux dans l'estuaire de la Gironde** **13**

1.1 L'estuaire de la Gironde . . . . . 14

1.2 Modélisations numériques de l'estuaire de la Gironde . . . . . 19

1.3 Conclusions . . . . . 38

**2 Quantification d'incertitudes avec variables fonctionnelles : application au modèle Telemac2D de prévision des hautes eaux de l'estuaire de la Gironde** **39**

2.1 Résumé . . . . . 40

2.2 Quantifying forcing uncertainties in the hydrodynamics of the Gironde estuary 41

2.3 Conclusions et perspectives . . . . . 71

**3 Méthodes d'assimilation de données : problématique et état de l'art** **77**

3.1 Introduction . . . . . 79

3.2 Formulation du problème dans un cadre général . . . . . 79

3.3 Formulation bayésienne du problème d'estimation de l'état du système . . . . . 81

3.4 Les méthodes séquentielles (ou filtres) . . . . . 83

3.5 Les méthodes variationnelles et variationnelles d'ensemble . . . . . 91

3.6 Application à la modélisation hydro-environnementale . . . . . 93

3.7 Conclusions . . . . . 100

**4 Assimilation de données ensembliste pour l'estimation conjointe des paramètres**

**et des forçages dans un espace réduit de Karhunen-Loève** **103**

4.1 EnKF-$\gamma$-KLBC . . . . . 104

4.2 Application de l'EnKF-$\gamma$-KLBC à l'estuaire de la Gironde . . . . . 109

4.3 Bilan et perspectives générales . . . . . 113

**5 Validation de l'EnKF-$\gamma$-KLBC-MEG dans un cadre synthétique** **119**

5.1 Introduction . . . . . 121

5.2 Dispositif Expérimental Synthétique . . . . . 122

5.3 Résultats : estimation des coefficients de Strickler et des forçages dans l'espace réduit de Karhunen-Loève . . . . . 137

5.4 Résultats : amélioration des prévisions de niveaux d'eau par l'EnKF-$\gamma$-KLBC-MEG 147

5.5 Optimisation des performances de l'EnKF-$\gamma$-KLBC-MEG . . . . . 152

5.6 Caractérisation des covariances croisées de l'EnKF-$\gamma$-KLBC-MEG . . . . . 158

5.7 Conclusions sur les performances de l'EnKF-$\gamma$-KLBC-MEG dans un cadre synthétique et perspectives . . . . . 162

**6 Évaluation de l'EnKF-$\gamma$-KLBC-MEG sur données réelles** **221**

6.1 Caractérisation d'une expérience réelle . . . . . 223

6.2 Description du dispositif d'expériences réelles . . . . . 225

6.3 Recalage par assimilation de données de l'événement de Février 2003 (Exp. $R1$) 230

6.4 Analyse des performances de l'EnKF-$\gamma$-KLBC-MEG sur un cycle de marée (Exp. $R2$) 233

6.5 Analyse des performances de l'EnKF-$\gamma$-KLBC-MEG sur un événement réel (Exp. $R3$) 236

6.6 Matrice Gain de Kalman : réseau de mesures idéal et localisation spatio-temporelle 239

6.7 Conclusions sur les performances de l'EnKF-$\gamma$-KLBC-MEG en réanalyse . . . . 242

**7 Conclusions et perspectives** **279**

**Conclusions et perspectives** **279**

7.1 Conclusions . . . . . 279

7.2 Perspectives d'évaluation dans un cadre opérationnel . . . . . 281

7.3 Perspectives générales . . . . . 285

**Bibliographie** **296**

**A Conséquences du rééchantillonnage sur la matrice de covariance croisée** **297**

**B Développement informatique de l'EnKF-$\gamma$-KLBC** **303**

B.1 Exemple de fichier de configuration . . . . . 303

B.2 Description du paramétrage . . . . . 305

B.3 Outils d'exploitation des résultats . . . . . 307

# Table des figures

**1 Prévision des hautes eaux dans l'estuaire de la Gironde** **13**

1.1 Localisation de l'estuaire de la Gironde en Europe et de son bassin versant (source : `https://fr.wikipedia.org/wiki/Bassin_Adour-Garonne`) . . . . . 16

1.2 Zone d'intervention du Service de Prévision des Crues Gironde-Adour-Dordogne (SPC-GAD). Source : `http:www.vigicrues.gouv.fr` . . . . . 19

1.3 Bulletin de vigilance sur l'estuaire de la Gironde du 11/02/2016. Le cercle rouge représente la zone de l'estuaire modélisée numériquement dans le cadre des projets Gironde et GirondeXL 3D (cf. section 1.2.1) (source : `http:www.vigicrues.gouv.fr`) . . . . . 20

1.4 Emprise du modèle de prévision de l'estuaire de la Gironde et localisation des points d'intérêt . . . . . 21

1.5 Emprise du modèle de prévision de l'estuaire de la Gironde et localisation des points d'intérêt et de mesures ***in situ*** . . . . . 22

1.6 RMSE sur tous les événements de référence du projet Gironde en chaque station de mesure de l'estuaire (source : Hissel [2010]) . . . . . 24

1.7 Emprise et localisation du modèle numérique de l'estuaire de la Gironde et délimitation des zones $A_1$ à $A_4$ des coefficients de Strickler. Les cercles représentent les nœuds du modèle numérique à partir d'un maillage construit en éléments finis (en noir). . . . . 29

1.8 Bathymétrie (m NGF69) du modèle numérique de l'estuaire de la Gironde. Les étoiles indiquent les 26 principales stations d'intérêt pour les prévisions des niveaux d'eau. . . . . 30

1.9 Évolution temporelle depuis le début de la tempête de 2003 à la condition à la limite : a) maritime : niveau d'eau total $CLMAR$ b) maritime : surcote instantanée ; c) fluviale des débits sur la Garonne ($QGAR$) et la Dordogne ($QDOR$). . . 32

1.10 Raffinement du maillage nécessaire pour une convergence sur les niveaux d'eau avec une précision de : a) 5 cm ; b) 10 cm . . . . . 37

**2 Quantification d'incertitudes avec variables fonctionnelles : application au modèle Telemac2D de prévision des hautes eaux de l'estuaire de la Gironde 39**

2.1 Values of the wind drag coefficient Cd according to formulation from Flather [1976], Wu [1980], Yelland and Taylor [1996] . . . . . 69

2.2 Convergence of the Sobol' indices with Ne at time step 50 for observing stations along the estuary for (a) Ks1, (b) Ks2, (c) Ks3, (d) Ks4, (e) Cd, (f) QDOR, (g) QGAR, (h) CLMAR. . . . . 73

2.3 Ranks diagrams for 3 stations along the Gironde estuary similar to station 0 Le Verdon. . . . . 74

2.4 Ranks diagrams for 8 stations along the Gironde estuary similar to station 5 bec d'Ambès. . . . . 75

2.5 Reliability diagram for all stations along the Gironde estuary with corresponding observed water level quantiles during the 2003 event. Mean water levels at each station are represented with hexagons. . . . . 76

**3 Méthodes d'assimilation de données : problématique et état de l'art 77**

3.1 Représentation de l'évolution de l'état du système pour un problème de filtrage (source : Carrassi et al. [2018]) . . . . . 83

**4 Assimilation de données ensembliste pour l'estimation conjointe des paramètres et des forçages dans un espace réduit de Karhunen-Loève 103**

4.1 Exemple d'enchaînement de deux cycles d'assimilation de données $\mathcal{F}_{k-1}$ et $\mathcal{F}_k$ pour l'EnKF-$\gamma$-KLBC pour $L = 6$ et $S = 2$ (adapté de Bocquet and Sakov [2014]). 105

4.2 Dual parameter and boundary condition estimation flowchart using the stochastic EnKF with reduced boundary conditions in KL space (EnKF-$\gamma$-KLBC) (adapté de Moradkhani et al. [2005]). . . . . 116

4.3 Fonction de transfert $\beta$ pour les stations de l'estuaire de la Gironde. . . . . 117

**5 Validation de l'EnKF-$\gamma$-KLBC-MEG dans un cadre synthétique 119**

5.1 Évolution temporelle (depuis le début de l'événement de 2003) pour l'expérience de référence $True2$ : a) de la condition à la limite maritime $CLMAR$; b) de la condition à la limite fluviale sur la Garonne. $CLMAR^{true}$ et $QGAR^{true}$ sont les valeurs "vraies" dont sont issues les observations synthétiques $\mathbf{y}^0$. $CLMAR^b$ et $QGAR^b$ sont les forçages d'ébauche ("background") du run de contrôle. . . . . 132

5.2 Évolution temporelle des coefficients de Strickler pour l'expérience $True2_b$ . . . 133

5.3 Évolution temporelle des coefficients de Strickler pour l'expérience $True2_c$ . . . 133

5.4 Évolution de l'écart-type de la distribution initiale (d'ébauche) des perturbations appliquées à la surcote sur la frontière maritime du modèle $CLMAR^b$. . . . . 136

5.5 Conditions aux limites du modèle corrigées à l'aide de l'EnKF-$\gamma$-KLBC-MEG pour l'Exp. $3-b$. Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. . . . . 144

5.6 Évolution de la différence entre la moyenne de l'ensemble après assimilation $Ks_i^a$ (resp. d'ébauche $Ks_i^b$) et la valeur "vraie" $Ks_i^{true}$ ($i \in \{1,2,3,4\}$) pour les sous-expériences Exp 1-a à 1-d au début de chaque cycle d'assimilation (L=1h - S=1h) - 1 pas de temps (time step) = 1 h - durée de l'Exp. 1 : 24 h. Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. (début) . . . . 165

5.6 Évolution de la différence entre la moyenne de l'ensemble après assimilation $Ks_i^a$ (resp. d'ébauche $Ks_i^b$) et la valeur "vraie" $Ks_i^{true}$ ($i \in \{1,2,3,4\}$) pour les sous-expériences Exp 1-a à 1-d au début de chaque cycle d'assimilation (L=1h - S=1h) - 1 pas de temps (time step) = 1 h - durée de l'Exp. 1 : 24 h. Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. (fin) . . . . . 166

5.7 Évolution de $Ks1$, $Ks2$ et $Ks3$ pour l'Exp 2-a. Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (début) . . . . . 167

5.7 Évolution de $Ks1$, $Ks2$ et $Ks3$ pour l'Exp 2-a. Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (fin) . . . . . 168

5.8 Évolution de $Ks1$, $Ks2$ et $Ks3$ avec réensemencement pour l'Exp. 2-b. Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (début) . . . . . 169

5.8 Évolution de $Ks1$, $Ks2$ et $Ks3$ avec réensemencement pour l'Exp. 2-b. Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (fin) . . . . . 170

5.9 Évolution de $\mathbf{P}^{\gamma}$ pour l'Exp. 2-a . . . . . 171

5.10 Évolution de $Ks1$, $Ks2$ et $Ks3$ pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (début) . . . . . 172

5.10 Évolution de $Ks1$, $Ks2$ et $Ks3$ pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (fin) . . . . . 173

5.11 Évolution des modes 1, 2, 3, 4 de CLMAR pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (début) . . . . . 174

5.11 Évolution des modes 1, 2, 3, 4 de CLMAR pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (fin) . . . . . 175

5.12 Évolution des modes 5, 6, 7 de CLMAR pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (début) . . . . . 176

5.12 Évolution des modes 5, 6, 7 de CLMAR pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (fin) . . . . . 177

5.13 Évolution des modes de QGAR pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (début) . . . . . 178

5.13 Évolution des modes de QGAR pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (fin) . . . . . 179

5.14 Évolution des modes de QDOR pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (début) . . . . . 180

5.14 Évolution des modes de QDOR pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (fin) . . . . . 181

5.15 Évolution temporelle (depuis le début de l'événement) pour l'expérience de référence $True2$ : a) de la condition à la limite maritime $CLMAR$ ; b) de la condition aux limites fluviale sur la Garonne. $CLMAR^{true}$ et $QGAR^{true}$ sont les valeurs "vraies" dont sont issues les observations synthétiques $\mathbf{y}^0$. $CLMAR^b$ et $QGAR^b$ sont les forçages d'ébauche ("background") du run de contrôle. . . . . 182

5.16 Matrice de covariance d'erreur d'ébauche de l'ensemble pour l'Exp. 3 . . . . . 183

5.17 Erreur quadratique moyenne (RMSE) sur la durée de l'Exp. 1 (24 h) le long de l'estuaire de la Gironde. (début) . . . . . 184

5.17 Erreur quadratique moyenne (RMSE) sur la durée de l'Exp. 1 (24 h) le long de l'estuaire de la Gironde. (fin) . . . . . 185

5.18 Évolution des niveaux d'eau pour l'Exp. 1-a au cours des cycles d'assimilation (L=1h - S=1h) (début) . . . . . 186

5.18 Évolution des niveaux d'eau pour l'Exp. 1-a au cours des cycles d'assimilation (L=1h - S=1h) (suite) . . . . . 187

5.18 Évolution des niveaux d'eau pour l'Exp. 1-a au cours des cycles d'assimilation (L=1h - S=1h) (fin) . . . . . 188

5.19 Évolution des niveaux d'eau pour l'Exp. 1-b au cours des cycles d'assimilation (L=1h - S=1h) (début) . . . . . 189

5.19 Évolution des niveaux d'eau pour l'Exp. 1-b au cours des cycles d'assimilation (L=1h - S=1h) (suite) . . . . . 190

5.19 Évolution des niveaux d'eau pour l'Exp. 1-b au cours des cycles d'assimilation (L=1h - S=1h) (fin) . . . . . 191

5.20 Évolution des niveaux d'eau pour l'Exp. 1-c au cours des cycles d'assimilation (L=1h - S=1h) (début) . . . . . 192

5.20 Évolution des niveaux d'eau pour l'Exp. 1-c au cours des cycles d'assimilation (L=1h - S=1h) (suite) . . . . . 193

5.20 Évolution des niveaux d'eau pour l'Exp. 1-c au cours des cycles d'assimilation (L=1h - S=1h) (fin) . . . . . 194

5.21 Évolution des niveaux d'eau pour l'Exp. 1-d au cours des cycles d'assimilation (L=1h - S=1h) (début) . . . 195

5.21 Évolution des niveaux d'eau pour l'Exp. 1-d au cours des cycles d'assimilation (L=1h - S=1h) (suite) . . . 196

5.21 Évolution des niveaux d'eau pour l'Exp. 1-d au cours des cycles d'assimilation (L=1h - S=1h) (fin) . . . 197

5.22 Évolution des niveaux d'eau sur les stations "assimilées" pour l'Exp. 2-b (début) 198

5.22 Évolution des niveaux d'eau sur les stations "assimilées" pour l'Exp. 2-b (fin) . . 199

5.23 Évolution des niveaux d'eau sur les stations de validation pour l'Exp. 2-b (début) 200

5.23 Évolution des niveaux d'eau sur les stations de validation pour l'Exp. 2-b (fin) . 201

5.24 Erreur quadratique moyenne le long de l'estuaire de la Gironde pour l'Exp 2-b. Les stations dont les observations synthétiques issues de $True2c$ sont assimilées sont : Le Verdon, Pauillac et Bordeaux PABx. . . . 202

5.25 Erreur quadratique moyenne le long de l'estuaire de la Gironde pour l'Exp 3-b . 202

5.26 Évolution des niveaux d'eau pour l'Exp 3-b (début) . . . 203

5.26 Évolution des niveaux d'eau pour l'Exp 3-b (suite) . . . 204

5.26 Évolution des niveaux d'eau pour l'Exp 3-b (fin) . . . 205

5.27 Diagramme de Talagrand après lissage en 3 stations (Le Verdon, Pauillac et Bordeaux PABX) pour les Exp. 2-c et Exp. 2-d . . . 206

5.28 Fonction de covariance de $Ks3$, $Cdz$ et $\alpha_{GAR,1}$ avec les niveaux d'eau de Telemac2D aux points de "mesures" pour l'Exp 2-c ($N_e = 100$) et l'Exp 2-d ($N_e = 200$) (début) . . . 207

5.28 Fonction de covariance de $Ks3$, $Cdz$ et $\alpha_{GAR,1}$ avec les niveaux d'eau de Telemac2D aux points de "mesures" pour l'Exp 2-c ($N_e = 100$) et l'Exp 2-d ($N_e = 200$) (suite) . . . 208

5.28 Fonction de covariance de $Ks3$, $Cdz$ et $\alpha_{GAR,1}$ avec les niveaux d'eau de Telemac2D aux points de "mesures" pour l'Exp 2-c ($N_e = 100$) et l'Exp 2-d ($N_e = 200$) (fin) . . . 209

5.29 Différence entre niveaux d'eau calculés par Telemac2D et observations selon le raffinement du maillage utilisé pour l'assimilation ou l'observation synthétique 210

5.30 Évolution de la variable de contrôle $Ks1$ pour l'Exp. 3-h ($L = 6\ h$, $S = 2\ h$) . . 211

5.31 Fonction temporelle de covariance entre $Ks2$ et $y$ à Pauillac pour : a) $N_e = 10$ membres sans rééchantillonnage (durée totale : 24 h) - b) $N_e = 10$ membres avec rééchantillonnage (durée totale : 24 h) - c) $N_e = 100$ membres avec rééchantillonnage (durée totale : 96 h) - d) $N_e$=100 membres sans rééchantillonnage et avec correction simultanée des autres $Ks$ et du forçage maritime dans l'espace réduit de KL (durée totale : 96 h). L'effondrement de l'ensemble a lieu autour de 18 h après le début de l'expérience. (début) . . . . . 212

5.31 Fonction temporelle de covariance entre $Ks2$ et $y$ à Pauillac pour : a) $N_e = 10$ membres sans rééchantillonnage (durée totale : 24 h) - b) $N_e = 10$ membres avec rééchantillonnage (durée totale : 24 h) - c) $N_e = 100$ membres avec rééchantillonnage (durée totale : 96 h) - d) $N_e$=100 membres sans rééchantillonnage et avec correction simultanée des autres $Ks$ et du forçage maritime dans l'espace réduit de KL (durée totale : 96 h). L'effondrement de l'ensemble a lieu autour de 18 h après le début de l'expérience. (fin) . . . . . 213

5.32 Fonction de covariance de $Ks1$, $Ks2$ et $Ks3$ avec les observations pour l'Exp. 3-b 214

5.33 Fonction de covariance de CLMAR avec les observations pour l'Exp. 3-b . . . . 215

5.34 Fonction de covariance de QGAR avec les observations pour l'Exp. 3-b . . . . . 216

5.35 Fonction de covariance de QDOR avec les observations pour l'Exp. 3-b . . . . . 217

5.36 Fonction de covariance croisée entre $\alpha 1$ et les stations d'observation de l'estuaire en l'absence de rééchantillonnage pour l'Exp. 3 . . . . . 218

5.37 Matrice de covariance entre vecteur de contrôle et niveaux d'eau calculés par Telemac2D dans l'espace des observations pour l'expérience 3-b au cours d'un cycle de marée . . . . . 219

**6 Évaluation de l'EnKF-$\gamma$-KLBC-MEG sur données réelles 221**

6.1 Débits de la Garonne ($QGAR$) et de la Dordogne ($QDOR$) pendant les événements de Janvier 2016 . . . . . 224

6.2 Évolution de $Ks1$, $Ks2$, $Ks3$ et $Ks4$ pour l'Exp. $R1$ au cours des cycles d'assimilation (L=6h - S=2h). L'ensemble est représenté par des cercles verts. (début) . 245

6.2 Évolution de $Ks1$, $Ks2$, $Ks3$ et $Ks4$ pour l'Exp. $R1$ au cours des cycles d'assimilation (L=6h - S=2h). L'ensemble est représenté par des cercles verts. (fin) . . 246

6.3 Évolution des niveaux d'eau pour l'expérience réelle Exp. $R1$ au cours des cycles d'assimilation (L=6h - S=2h) (début) . . . . . 247

6.3 Évolution des niveaux d'eau pour l'expérience réelle Exp. $R1$ au cours des cycles d'assimilation (L=6h - S=2h) (suite) . . . . . 248

6.3 Évolution des niveaux d'eau pour l'expérience réelle Exp. $R1$ au cours des cycles d'assimilation (L=6h - S=2h) (fin) . . . . . 249
6.4 Évolution de $Ks1$ et $Ks2$ pour l'Exp. $R2b$ au cours des cycles d'assimilation (L=4h - S=2h). L'ensemble est représenté par des cercles verts. . . . . 250
6.5 Évolution de $Ks1$, $Ks2$ et $Ks3$ pour l'Exp. $R2c$ au cours des cycles d'assimilation (L=4h - S=2h). L'ensemble est représenté par des cercles verts. . . . . 251
6.6 Conditions aux limites du modèle pour les expériences réelles $R2b$ et $R2c$ . . . 252
6.7 Évolution des niveaux d'eau sur les stations "assimilées" pour l'Exp. $R2b$ . . . . 253
6.8 Évolution des niveaux d'eau sur les stations "assimilées" pour l'Exp. $R2c$ (début) 254
6.8 Évolution des niveaux d'eau sur les stations "assimilées" pour l'Exp. $R2c$ (fin) . 255
6.9 Évolution des niveaux d'eau sur les stations de validation pour l'Exp. $R2b$ (début) 256
6.9 Évolution des niveaux d'eau sur les stations de validation pour l'Exp. $R2b$ (fin) . 257
6.10 Évolution des niveaux d'eau sur les stations de validation pour l'Exp. $R2c$ (début) 258
6.10 Évolution des niveaux d'eau sur les stations de validation pour l'Exp. $R2c$ (fin) . 259
6.11 Évolution de $Ks1$, $Ks2$ et $Ks3$ pour l'Exp. $R3$ au cours des cycles d'assimilation (L=3h - S=2h). L'ensemble est représenté par des cercles verts. . . . . 260
6.12 Évolution des modes 1, 2, 3 et 4 de CLMAR pour l'Exp. $R3$ au cours des cycles d'assimilation (L=3h - S=2h). L'ensemble est représenté par des cercles verts. . . 261
6.13 Évolution des modes 5, 6 et 7 de CLMAR pour l'Exp. $R3$ au cours des cycles d'assimilation (L=3h - S=2h). L'ensemble est représenté par des cercles verts. . . 262
6.14 Condition à la limite maritime $CLMAR$ du modèle pour l'Exp. $R3$ . . . . . 263
6.15 Correction des pics de Pleine Mer et de Basse Mer de la condition aux limites maritimes $CLMAR$ pour l'Exp. $R3$ . . . . . 264
6.16 Évolution des niveaux d'eau sur les stations "assimilées" pour l'Exp. $R3$ (début) 265
6.16 Évolution des niveaux d'eau sur les stations "assimilées" pour l'Exp. $R3$ (fin) . . 266
6.17 Évolution des niveaux d'eau pour les stations de validation au cours de l'Exp. $R3$ (L=3h - S=2h) (début) . . . . . 267
6.17 Évolution des niveaux d'eau pour les stations de validation au cours de l'Exp. $R3$ (L=3h - S=2h) (fin) . . . . . 268
6.18 Déphasages des pics de pleine mer le long de l'estuaire de la Gironde pour l'Exp. $R3$ avec (en bas) et sans assimilation de données (en haut) . . . . . 269
6.19 Erreur quadratique moyenne le long de l'estuaire de la Gironde pour l'Exp. $R3$ . 270
6.20 Erreur quadratique moyenne sur les pics de PM le long de l'estuaire de la Gironde pour l'Exp. $R3$ . . . . . 270

6.21 Erreur sur le pic de tempête le long de l'estuaire de la Gironde pour l'Exp. $R3$ . 271

6.22 Évolution au cours des cycles d'assimilation des composantes de la matrice gain de Kalman **K** pour chaque variable de contrôle avec les niveaux d'eau calculés par Telemac2D dans l'espace des observations (de gauche à droite de 0 h d'échéance à $(L-1)$ h d'échéance après le début du cycle) pour l'Exp. R1 . . . . . 272

6.22 Évolution au cours des cycles d'assimilation des composantes de la matrice gain de Kalman (**K**) pour chaque variable de contrôle avec les niveaux d'eau calculés par Telemac2D dans l'espace des observations (de gauche à droite de 0 h d'échéance à $(L-1)$ h d'échéance après le début du cycle) pour l'Exp. R1 (suite) 273

6.23 Évolution au cours du temps des composantes de la matrice de covariance $\mathbf{P}^{\gamma,y}$ avec chaque observation (à 0 h d'échéance) et du gain de Kalman **K** pour chaque observation (de gauche à droite de 0 h d'échéance à $(L-1)$ h d'échéance après le début du cycle) et chaque variable de contrôle pour l'Exp. $R3$ (début) . . . . . 274

6.23 Évolution au cours du temps des composantes de la matrice de covariance $\mathbf{P}^{\gamma,y}$ avec chaque observation (à 0 h d'échéance) et du gain de Kalman **K** pour chaque observation (de gauche à droite de 0 h d'échéance à $(L-1)$ h d'échéance après le début du cycle) et chaque variable de contrôle pour l'Exp. $R3$ (suite) . . . . . 275

6.23 Évolution au cours du temps des composantes de la matrice de covariance $\mathbf{P}^{\gamma,y}$ avec chaque observation (à 0 h d'échéance) et du gain de Kalman **K** pour chaque observation (de gauche à droite de 0 h d'échéance à $(L-1)$ h d'échéance après le début du cycle) et chaque variable de contrôle pour l'Exp. $R3$ (fin) . . . . . 276

6.24 Évolution temporelle de $Ks1$ pour l'Exp. 3-c synthétique au cours des cycles d'assimilation (L=1h - S=1h) . . . . . 277

6.25 Évolution au cours des cycles d'assimilation de la composante $Ks3$ de la matrice Gain de Kalman **K** avec chaque observation simulée (de gauche á droite de 0 h d'échéance á $(L-1)$ h d'échéance après le début du cycle) pour l'Exp. $R2c$ . . . 278

**7 Conclusions et perspectives** **279**

7.1 Condition à la limite fluviale $QGAR$ du modèle pour le mode prévision de l'Exp. $R3$ 285

**A Conséquences d'un rééchantillonnage sur la matrice de covariance croisée entre le vecteur de contrôle et les observations simulées** **297**

A.1 Fonction de transfert $\beta$ pour les stations de l'estuaire de la Gironde. . . . . 298

**B Exemple de fichier de configuration pour l'environnement d'assimilation de données** **303**

# Liste des tableaux

**1 Prévision des hautes eaux dans l'estuaire de la Gironde** **13**

1.1 Niveaux de vigilance . . . . . 18

1.2 Caractéristiques des événements de calage et d'évaluation du modèle non-débordant de l'estuaire de la Gironde (modèle Gironde). Les lignes vertes (resp. rouges) indiquent les événements "non-débordants" (resp. "débordants") (d'après Hissel et al. [2006]) . . . . . 23

1.3 RMSE avec et sans prise en compte des forçages météorologiques pour le modèle GirondeXL 3D (d'après Huybrechts et al. [2018a]) . . . . . 26

1.4 Coefficients résultant du calage sur les Nash de PM rephasées ($\boldsymbol{Nash}_{HT}$) et sur la RMSE pour l'événement de 2003 (non-débordant) obtenus pour le modèle Gironde et le modèle actualisé MEG0V7P3 . . . . . 31

1.5 ***Nash*** de PM rephasées ($\boldsymbol{Nash}_{HT}$), ***RMSE*** et ***RMSE*** de PM rephasées ($\boldsymbol{RMSE}_{HT}$) pour les événements de 2003 (non-débordant) et de 1999 (débordant) le long de l'estuaire de la Gironde obtenus avec le modèle actualisé MEG0V7P3 . . . . . 33

1.6 Caractéristiques (nombre de nœuds et pas de temps du modèle) des modèles utilisés pour l'étude de l'influence du maillage sur les niveaux d'eau calculés par Telemac2D . . . . . 35

1.7 RMSE obtenu par le modèle le plus raffiné (MEG3V7P3) pour la période du 02 au 05/02/2003 et comparaison avec les modèles original (MEG0V7P3) et à raffinement médian (MEG1V7P3) (resp. (MEG2V7P3)) pour chaque station d'observation de l'estuaire de la Gironde . . . . . 36

**2 Quantification d'incertitudes avec variables fonctionnelles : application au modèle Telemac2D de prévision des hautes eaux de l'estuaire de la Gironde** **39**

**3 Méthodes d'assimilation de données : problématique et état de l'art** **77**

**4 Assimilation de données ensembliste pour l'estimation conjointe des paramètres et des forçages dans un espace réduit de Karhunen-Loève** **103**

**5 Validation de l'EnKF-$\gamma$-KLBC-MEG dans un cadre synthétique** **119**

5.1 Synthèse des expériences de référence $True1$, $True2a$, $True2b$ et $True2c$ dont les observations synthétiques sont extraites et assimilées au cours des expériences jumelles. Pour chaque expérience de référence ($True1$ ou $True2$), les coefficients de frottement utilisés sont indiqués dans la colonne $Ks^{true}$ (selon le zonage présenté dans le chapitre 1). Concernant les forçages temporels, une perturbation constante appliquée aux forçages réanalysés de l'événement de Février 2003 est indiquée dans la colonne "$d\{q\}^{true}$" ($q \in (sm, QGAR, QDOR)$). Une perturbation variable dans le temps dans l'espace réduit de KL est indiquée dans la colonne "$\alpha_{i,j}^{true}$" (pour $i \in (CLMAR, QGAR, QDOR)$ et $j \in (1, \cdots, p_i)$ où $p_i$ est le nombre de modes retenus pour le forçage $i$). Pour toutes les expériences de référence, $Cd$ est constant et est égal à $2.57 \times 10^{-6}$. . . . . 123

5.2 Observations synthétiques (à gauche), vecteur de contrôle (en haut à droite) et ébauche (en bas à droite) pour l'Exp. 1. Les observations synthétiques sont des niveaux d'eau à un pas de temps horaire. $CLMAR$, $Cd$, $QGAR$ et $QDOR$ ne sont ni perturbés ni contrôlés. "Exp." désigne le nom de l'expérience synthétique. La partie inférieure "Ébauche" des tableaux fournit la moyenne initiale de l'ensemble (ligne "Moy.") et sa distribution initiale (ligne "Loi") pour chaque variable de contrôle. $\mathcal{U}(x_1; x_2)$ désigne la loi uniforme sur l'ensemble support $[x_1 ; x_2]$. . 124

5.3 Observations synthétiques (à gauche), vecteur de contrôle (en haut à droite) et ébauche (en bas à droite) pour l'Exp. 2. Les observations synthétiques sont des niveaux d'eau à un pas de temps horaire. $Cd$, $Ks4$, $QGAR$ et $QDOR$ ne sont ni perturbés ni contrôlés. "Exp." désigne le nom de l'expérience synthétique. La partie inférieure "Ébauche" des tableaux fournit la moyenne initiale de l'ensemble (ligne "Moy.") et sa distribution initiale (ligne "Loi") pour chaque variable de contrôle. $\mathcal{U}(x_1; x_2)$ désigne la loi uniforme sur l'ensemble support $[x_1 ; x_2]$. . . . 125

5.4 Observations synthétiques (à gauche), vecteur de contrôle (en haut à droite) et ébauche (en bas à droite) pour l'Exp. 3. Les observations synthétiques sont des niveaux d'eau à un pas de temps horaire. $Ks4$ n'est ni perturbé ni contrôlé et est égal à 40. "Exp." désigne le nom de l'expérience synthétique. La partie inférieure "Ébauche" des tableaux fournit la moyenne initiale de l'ensemble (ligne "Moy.") et sa distribution initiale (ligne "Loi") pour chaque variable de contrôle. $\mathcal{U}(x_1; x_2)$ désigne la loi uniforme sur l'ensemble support $[x_1 ; x_2]$. . . . . 126

5.5 Autres paramètres d'entrée de la maquette pour les Exp. 1, Exp 2 et Exp. 3. . . . 127

5.6 Coefficients “vrais” associés aux modes de $CLMAR$, $QGAR$ et $QDOR$ pour l’expérience de référence $True2$ . . . . . 130

5.7 RMSE sans et avec EnKF-$\gamma$-KLBC-MEG pour l’Exp. 3 . . . . . 153

5.8 Influence du nombre de membres $N_e$ sur la RMSE dans un cadre synthétique. $N_e = 100$ (resp. $N_e = 200$) pour les Exp. 2-c et Exp. 3-a (resp. 2-d). $N_e = 50$ (resp. $N_e = 125$) pour l’Exp. 3-e (resp. 3-f). Les colonnes “(2-c) avec DA” (resp.“(2-c) sans DA”) et “(3-a) avec DA” indiquent les RMSE pour les Exp. 2-c et 3-a avec (resp. sans) l’EnKF-$\gamma$-KLBC-MEG. Les colonnes “(2-c)-(2d)” (resp. “(3-a)-(3-e)” ou “(3-a)-(3-f)”) indiquent la différence entre la RMSE de l’Exp. 2-c (resp. Exp. 3-a) et celle de l’Exp. 2-d (resp. Exp. 3-e ou Exp. 3-f). La différence est négative (resp. positive) si la RMSE de Exp. 2-c ou Exp. 3-a est plus petite (resp. grande). . . . . 156

5.9 Gains sur la RMSE pour différentes configurations synthétiques . . . . . 159

**6 Évaluation de l’EnKF-$\gamma$-KLBC-MEG sur données réelles 221**

6.1 Description des observations réelles utilisées, du vecteur de contrôle et de l’ensemble perturbé pour les expériences réelles. Toutes les observations sont des mesures réelles horaires de niveaux d’eau. . . . . 227

6.2 Description du vecteur de contrôle et de l’ensemble perturbé pour les expériences réelles . . . . . 228

6.3 Autres paramètres d’entrée de l’EnKF-$\gamma$-KLBC-MEG pour les expériences réelles. $N_e$=100 et le rééchantillonnage (persistance temporelle de $\mathbf{P}_0^{\gamma}$) est utilisé. Le pas de temps $\Delta t$ d’assimilation est égal à 1 h. . . . . 229

6.4 RMSE sans et avec EnKF-$\gamma$-KLBC-MEG pour l’Exp. $R1$ . . . . . 232

6.5 RMSE sans et avec EnKF-$\gamma$-KLBC-MEG pour l’Exp. $R2$ . . . . . 235

**7 Conclusions et perspectives 279**

**A Conséquences d’un rééchantillonnage sur la matrice de covariance croisée entre le vecteur de contrôle et les observations simulées 297**

A.1 Rapport $\frac{D_{np,np'}}{D_{np,np}}$ pour chaque coefficient modal $\alpha_{np}$ avec (np , np’) $\in \{1, .., n_{modes}\}^2$ . . . . . 302

**B Exemple de fichier de configuration pour l’environnement d’assimilation de données 303**

# Principaux acronymes

| | |
|---|---|
| AD | Assimilation de Données |
| ANOVA | Analyse par Décomposition de Variance |
| API | Interface de Programmation Applicative |
| BC | Conditions aux limites |
| BE | Basses Eaux |
| BM | Basses Mers |
| BLUE | Best Linear Unbiased estimator |
| CERFACS | Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique |
| CETMEF | Centre d'Études Techniques Maritimes et Fluviales |
| CFD | Mécanique des Fluides Numérique (Computational Fluid Dynamics) |
| CI | Conditions initiales |
| Cerema | Centre d'Expertise sur les Risques, l'Environnement, la Mobilité et l'Aménagement du Territoire |
| CNPE | Centre Nucléaire de Production d'Électricité |
| DREAL | Direction Régionale de l'Environnement, de l'Aménagement et du Logement |
| EDF R&D | Électricité de France, département Recherche et Développement |
| EKF | Filtre de Kalman Étendu |
| EnKF | Filtre de Kalman d'Ensemble |
| EnKF-$\gamma$-KLBC | EnKF avec correction des paramètres et des forçages dans l'espace réduit de KL |
| EnKF-$\gamma$-KLBC-MEG | EnKF-$\gamma$-KLBC appliqué au Modèle actualisé de l'estuaire de la Gironde |
| EnKS | Ensemble Kalman Smoother |
| Exp. 1 | Expérience synthétique 1 |
| Exp. 2 | Expérience synthétique 2 |
| Exp. 3 | Expérience synthétique 3 |
| Exp. $R1$ | Expérience réelle 1 |
| Exp. $R2$ | Expérience réelle 2 |
| Exp. $R3$ | Expérience réelle 3 |
| GENCI | Grands Équipements Nationaux de Calcul Intensif |
| GPMB | Grand Port Maritime de Bordeaux |
| GSA | Analyse de Sensibilité Globale |
| HPC | Calcul Haute Performance |
| IDRIS | Institut du Développement et des Ressources en Informatique Scientifique |
| IEnKF | Iterative Ensemble Kalman Filter |
| IEnKS | Iterative Ensemble Kalman Smoother |
| IPCC | The Intergovernmental Panel on Climate Change |
| KL | Karhunen-Loève |

| | |
|---|---|
| LHSV | Laboratoire d'hydraulique Saint-Venant |
| LNHE | Laboratoire National d'Hydraulique et d'Environnement |
| MEG$\{i\}$V7P3 | Modèle de l'Estuaire de la Gironde utilisant le raffinement du maillage $mesh\ \{i\}$ ($i \in (0, \cdots, 3)$ |
| $\text{Nash}_{\text{HT}}$ | Nash de Pleine Mer |
| PM | Pleine Mer |
| QoI | Quantité d'intérêt |
| RIG | Référentiel Inondation Gironde |
| RMSE | Erreur Quadratique moyenne (Root Mean Square Error) |
| pdf | Fonction densité de probabilité |
| REFMAR | Réseaux de Référence des Observations Marégraphiques |
| SA | Analyse de sensibilité |
| SCHAPI | Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations |
| SHOM | Service Hydrographique et Océanographique de la Marine |
| SPC | Service de Prévision des Crues |
| SPC-GAD | Service de Prévision des Crues Gironde-Adour-Dordogne |
| std | Standard deviation (écart-type) |
| SWE | Équations de Saint-Venant (Shallow Water Equations) |
| VNF | Voies Navigables de France |

# Notations principales

**Notations pour l'hydrodynamique**

| | | |
|---|---|---|
| $\nu$ : | Viscosité cinématique de l'eau | $[m^2.s^{-1}]$ |
| $\rho_w$ : | Densité de l'eau | $[kg.m^{-3}]$ |
| $\rho_{air}$ : | Densité de l'air | $[kg.m^{-3}]$ |
| $C_d$ : | Coefficient d'influence du vent | $[-]$ |
| $F_x, F_y$ : | Composantes horizontales des forces extérieures | $[m.s^{-2}]$ |
| $H$ : | Cote de la surface libre | $[mNGF69]$ |
| $h$ : | Hauteur d'eau | $[m]$ |
| $K_s$ : | Coefficient de frottement dans le lit mineur ou majeur de la rivière | $[m^{\frac{1}{3}}.s^{-1}]$ |
| $Ks\,\{i\}$ : | Coefficient de frottement de la zone $A_i$ | $[m^{\frac{1}{3}}.s^{-1}]$ |
| $Nash_{HT}$ : | Nash de Pleine Mer | $[-]$ |
| $P_{atm}$ : | Pression atmosphérique | $[Pa]$ |
| $RMSE$ : | Erreur Quadratique Moyenne | $[m]$ |
| $t$ : | Temps | $[s]$ |
| $u, v$ : | Composantes horizontales de la vitesse | $[m.s^{-1}]$ |
| $U_{w,x}, U_{w,y}$ : | Composantes horizontales de la vitesse | $[m.s^{-1}]$ |
| $z_f$ : | Cote du fond de la rivière | $[mNGF69]$ |

**Notations pour la quantification d'incertitude**

| | | |
|---|---|---|
| $\ell_{\text{CLMAR}}$ : | durée d'autocorrélation du signal à la frontière maritime | $[s]$ |
| $\ell_{\text{QDOR}}$ : | durée d'autocorrélation du signal de débit de la Dordogne | $[s]$ |
| $\ell_{\text{QGAR}}$ : | durée d'autocorrélation du signal de débit de la Garonne | $[s]$ |
| $\kappa$ : | solution de l'équation de Fredholm | |

$\lambda_i$ : i$^{\text{ème}}$ valeur propre de $\kappa$ $\in \mathbb{R}$

$\Phi_i$ : i$^{\text{ème}}$ vecteur propre de $\kappa$ sur le vecteur temporel discrétisé $\{t_1, t_{nt}\}$ $\mathbb{R}^{nt}$

$CLMAR$ : Condition à la limite maritime $[m\ NGF69]$

$d$ : Nombre de variables incertaines $[-]$

$f$ : Modèle d'évolution $\mathbb{R}^d \longrightarrow \mathbb{R}$

$h(t)$ : perturbation gaussienne sur le niveau d'eau $[m]$

$N$ : Nombre d'intégrations du modèle pour l'ANOVA par GSA $[-]$

$N_e$ : Nombre de membres $[-]$

$p_{\text{CLMAR}}$ : nombre de modes retenu pour la décomposition de KL du signal maritime $[-]$

$p_{\text{QDOR}}$ : nombre de modes retenu pour la décomposition de KL du débit de la Dordogne $[-]$

$p_{\text{QGAR}}$ : nombre de modes retenu pour la décomposition de KL du débit de la Garonne $[-]$

$q(t)$ : perturbation gaussienne sur le débit $[m^3.s^{-1}]$

$QDOR$ : Débit de la Dordogne $[m^3.s^{-1}]$

$QGAR$ : Débit de la Garonne $[m^3.s^{-1}]$

$S_i$ : Indice de Sobol' du 1$^{\text{er}}$ ordre $[-]$

$S_{i,j}$ : Indice de Sobol' du 2$^{\text{nd}}$ ordre $[-]$

$S_{Ti}$ : Indice de Sobol' total $[-]$

$V(Y)$ : Variance de la variable Y $[m^2]$

$X = (X_1, X_2, \ldots, X_d)$ : Vecteur des données d'entrée du modèle $\in \mathbb{R}^d$

$Y = f(X)$ : Résultat en sortie du modèle f $\in \mathbb{R}$

**Notations pour l'assimilation de données**

$\boldsymbol{\gamma}$ : Vecteur de contrôle contenant les paramètres et les forçages du modèle $\in \mathbb{R}^{n_{var} \times N_e}$

$\boldsymbol{\gamma}^a$ : Vecteur de contrôle après la phase d'analyse ($\boldsymbol{a}$ pour "analyse") $\in \mathbb{R}^{n_{var} \times N_e}$

$\boldsymbol{\gamma}^f$ : Vecteur de contrôle après la phase de prévision ($\boldsymbol{f}$ pour "forecast") $\in \mathbb{R}^{n_{var} \times N_e}$

$\mathcal{M}$ : Modèle d'évolution de l'état du système

$\Delta t$ : Fréquence d'assimilation des données $[s]$

$\{\boldsymbol{\alpha}_{CLMAR,j}\}_{j \in \{1,..,p_{CLMAR}\}}$ : coefficients modaux gouvernant la condition à la limite maritime dans l'espace réduit de KL, de dimension $p_{CLMAR}$ (aussi notés $(alpha\ \{j\})_{j \in \{1,..,p_{CLMAR}\}}$) $\in \mathbb{R}^{p_{CLMAR} \times N_e}$

$\{\boldsymbol{\alpha}_{i,j}\}_{(i,j)\in(\{DOR,GAR\},\{1,..,p_j\})}$ avec $p_j \in \{p_{DOR}, p_{GAR}\}$ : coefficients modaux appliqués aux composantes principales de la fonction d'autocorrélation des chroniques temporelles de débits de la Dordogne et de la Garonne obtenues dans l'espace réduit de KL de dimensions respectives $p_{DOR}$ et $p_{GAR}$ (aussi notés $(Coeff\{i\}\{j\})_{(i,j)\in(\{DOR,GAR\},\{1,..,p_j\})}$) $\in \mathbb{R}^{p_j\times N_e}$

**H** : Opérateur linéarisé d'observation $\mathbb{R}^m \longrightarrow \mathbb{R}^{n_{obs}}$

**M** : Modèle d'évolution linéarisé de l'état du système $\mathbb{R}^m \longrightarrow \mathbb{R}^m$

$\mathbf{P}_k^{\gamma}$ : Matrice de covariance d'erreur d'ébauche des paramètres contrôlés sur $\mathcal{F}_k \in \mathbb{R}^{n_{var}\times n_{var}}$

**R** : Matrice de covariance d'erreur d'observation $\in \mathbb{R}^{n_{obs}\times n_{obs}}$

**X** : Ensemble des anomalies $\in \mathbb{R}^{n_{var}\times N_e}$

**x** : Vecteur d'état du système $\in \mathbb{R}^{m\times N_e}$

$\mathbf{x}^a$ : État du système après la phase d'optimisation $\in \mathbb{R}^{m\times N_e}$

$\mathbf{x}^f$ : État du système après la phase de prévision $\in \mathbb{R}^{m\times N_e}$

**y** : Projection du vecteur de contrôle dans l'espace des observations $\in \mathbb{R}^{n_{obs}\times N_e}$

$\mathbf{y}^0$ : Vecteur d'observations $\in \mathbb{R}^{n_{obs}\times N_e}$

$\overline{\boldsymbol{\gamma}^{id}}$ : Moyenne sur les membres du vecteur de contrôle $\boldsymbol{\gamma}^{id}\ \forall\ id \in \{a, f\}$ $\in \mathbb{R}^{n_{var}}$

$dQDOR$ : Incrément sur le débit de la Dordogne $[m^3.s^{-1}], \in \mathbb{R}^{N_e}$

$dQGAR$ : Incrément sur le débit de la Garonne $[m^3.s^{-1}], \in \mathbb{R}^{N_e}$

$dsm$ : Incrément sur la cote de surface libre à la frontière maritime $[m], \in \mathbb{R}^{N_e}$

$L$ : Nombre de pas de temps $\Delta t$ dans une fenêtre d'assimilation $[-]$

$n_{obs}$ : Nombre total d'observations disponibles sur $\mathcal{F}_k$ $[-]$

$n_{var}$ : Nombre de variables de contrôle incluses dans le vecteur de contrôle $[-]$

$nb_{cycles}$ : Nombre de cycles d'assimilation de données d'une expérience $[-]$

$S$ : Nombre de pas de temps $\Delta t$ séparant deux fenêtres d'assimilation successives $[-]$

$T_0$ : Temps initial d'une fenêtre d'assimilation de données $[s]$

$T_L$ : Temps final d'une fenêtre d'assimilation de données $[s]$

$t_{k,deb}$ : Temps initial de la fenêtre d'assimilation de données $\mathcal{F}_k$ $[s]$

$t_{k,fin}$ : Temps final de la fenêtre d'assimilation de données $\mathcal{F}_k$ $[s]$

$\mathcal{F}_k$ : Cycle (ou fenêtre) d'assimilation $k$ avec $k \in \{1, \cdots, nb_{cycles}\}$ $[-]$

# Introduction

La hausse du niveau des mers projetée au cours du 21$^{ème}$ siècle aura un impact très important sur les systèmes côtiers, avec une multiplication des phénomènes de submersion, d'inondations et d'érosion (IPCC [2019]). La croissance démographique, le développement économique et l'urbanisation des zones côtières devraient accroître les impacts sur ces régions. Le niveau de la mer au niveau régional pourrait s'élever d'1 m environ dans l'Atlantique à l'horizon 2100.
Les impacts du changement climatique pourraient être substantiels en Europe (Jackson et al. [2015]). L'affaiblissement de la Circulation Méridienne de retournement Atlantique (AMOC), une composante importante de la circulation océanique mondiale, peut induire, d'une part, une augmentation des précipitations en Europe du Nord et une diminution en Europe du Sud.
D'autre part, l'AMOC aurait un impact sur la position et l'intensité de la trajectoire des tempêtes dans l'Atlantique Nord (Gastineau et al. [2016]), avec une potentielle augmentation du nombre de tempêtes hivernales frappant l'Europe (Woollings et al. [2012] ; Jackson et al. [2015]), bien qu'une certaine incertitude subsiste selon les modèles considérés (Peings et al. [2016]).

Il existe un consensus sur le fait que les politiques publiques pour la réduction des risques liés aux catastrophes naturelles sont rentables. Pappenberger et al. [2015] estime à 400 euros le bénéfice pour la collectivité par euro investi dans un système d'alerte et de prévision des inondations et submersions marines en Europe. Dans le cas où le réchauffement climatique se limiterait à 1,5℃ par rapport à l'ère préindustrielle, scénario le plus optimiste, l'Europe devra investir 15 milliards d'euros par an pour réparer les dégâts, soit plus que le double du budget actuel alloué. Le nombre de personnes affectées par les inondations, quant à lui, augmentera de 86%, soit presque 650.000 personnes par an (Alfieri et al. [2018]). Dans le scénario extrême à 3℃, cela représenterait 17 milliards d'euros par an et 780.000 personnes touchées.
A l'échelle de la France, les inondations sont sans l'ombre d'un doute les catastrophes les plus coûteuses. L'impact d'une crue centennale a été calculé à l'occasion de la stratégie de résilience annoncée par la ville de Paris le 4 octobre 2017 et les résultats sont effarants : 30 milliards d'euros de pertes directes en capital privé ou public, 60 milliards d'euros de pertes directes d'activités économiques, 430.000 emplois directement menacés.

À l'interface entre fleuves et mers, les estuaires présentent la particularité de subir les conséquences du changement climatique à deux titres : du fait des variations de niveau de la mer, d'une part, et à cause d'éventuelles variations des débits des rivières qui les alimentent d'autre part. Ces

deux facteurs font des estuaires des lieux particulièrement sensibles aux variations climatiques (Planton et al. [2015]). À la potentielle aggravation des impacts à la côte due à la conjonction et aux interactions entre surcotes et vagues, s'ajoute une transformation de l'onde de marée lors de son passage dans des régions littorales à géométrie particulière. C'est notamment le cas des estuaires, comme celui de la Gironde au sud-ouest de la France, dont la géométrie influe fortement sur la propagation de l'onde depuis l'embouchure, pouvant conduire à son atténuation comme à son amplification.

En France, le Service Central d'Hydrométéorologie et d'Appui à la prévision des Inondations (SCHAPI) et les Services de Prévision des Crues (SPC), au sein du Ministère en charge du Développement Durable, assurent la surveillance du réseau hydraulique et la prévision de l'aléa lié aux submersions et aux crues des fleuves et des rivières. Ils ont à leur disposition des outils numériques de modélisation hydrodynamique et des observations ***in situ*** issues des réseaux de mesures nationaux (REFMAR, services hydrométriques du Ministère, etc.) .
L'une des limites des modèles numériques utilisés est leur sensibilité aux paramètres d'entrée, aux forçages temporels appliqués notamment à leurs frontières liquides, mais aussi à leur emprise, ainsi qu'à la qualité des données bathymétriques et topographiques qui détermine leur maillage. En outre, un compromis doit souvent être trouvé entre la précision des résultats du modèle et les coûts de calcul. Les observations, quant à elles, sont à la fois imprécises et parcellaires dans le temps et dans l'espace. Elles servent principalement dans la phase de calibration des modèles afin d'évaluer ses performances.
Le défi actuel consiste donc à diminuer les incertitudes sur les entrées (paramètres, forçages, condition initiale) des modèles numériques et à utiliser des méthodes d'assimilation de données pour combiner de manière optimale le modèle numérique et les observations au fur et à mesure de leur disponibilité.

L'objectif de ce travail de thèse est le développement de techniques d'assimilation de données dans les modèles hydrodynamiques bidimensionnels avec application au modèle de prévision des hautes eaux de l'estuaire de la Gironde fondé sur le code de calcul Telemac2D (www.opentelemac.org). Financé par le Cerema et le SCHAPI, il a été mené au sein du Laboratoire d'Hydraulique Saint-Venant, commun entre le Cerema, EDF R&D et l'ENPC, en étroite collaboration avec le CERFACS. Il a, par ailleurs, bénéficié des ressources et des compétences informatiques de EDF R&D et de l'IDRIS dans le cadre de deux projets Grands Équipements de Calcul Intensif (GENCI) sous les allocations 2017-A0030110292 et 2018-A0050110292. Il a également été alimenté par les données et les échanges techniques avec les partenaires historiques du projet Gironde (METEO-FRANCE, SPC Gironde-Adour-Dordogne), ainsi que le SHOM.

Prévoir les niveaux d'eau dans les systèmes estuariens complexes est un enjeu majeur pour la sécurité des biens et des personnes à court et moyen termes, mais aussi à long terme dans un contexte de changement climatique. Le contexte de la modélisation numérique et des enjeux en termes de prévision des niveaux d'eau dans l'estuaire de la Gironde est présenté dans le chapitre 1.

Ce travail de thèse s'appuie sur le modèle bidimensionnel de l'estuaire de la Gironde issu du projet Gironde et actualisé ici. Malgré le recalage dont il a fait l'objet sur la tempête de 2003 et l'actualisation des données bathymétriques qu'il prend en compte, la qualité des résultats du modèle peut être améliorée. Celle-ci dépend des incertitudes sur les paramètres d'entrée (coefficients de frottement sur le fond et coefficient d'influence du vent) et les forçages spatio-temporels (débits des rivières en amont de la Garonne et de la Dordogne, niveaux d'eau imposés à la frontière liquide maritime et champs météorologiques de vent et de pression) du modèle. Il est donc nécessaire au préalable de mener une analyse de sensibilité globale (GSA) afin de déterminer les sources d'incertitudes du modèle les plus influentes parmi les forçages temporels (variables fonctionnelles) et les paramètres du modèle. Cette première étape doit permettre de déterminer quelles variables temporelles et paramètres du modèle corriger en assimilant des données d'observation pour mieux représenter les niveaux d'eau dans l'estuaire.

Le premier axe de travail, présenté dans le chapitre 2, a porté sur la proposition d'une méthodologie de traitement des incertitudes pour un code alimenté en particulier par des variables fonctionnelles et, pour ce qui concerne l'application, sur la classification des sources d'incertitudes du modèle de prévision des hautes eaux de l'estuaire de la Gironde et leur traitement.
Une méthodologie pour quantifier les incertitudes de variables aléatoires fonctionnelles à partir d'un échantillon de leurs réalisations a été proposée. Elle permet d'évaluer la contribution de chaque variable fonctionnelle et scalaire à la variabilité d'une autre variable qui peut être, par exemple, la sortie du code étudié, ainsi que leurs interactions. Des outils de visualisation spatio-temporelle des incertitudes et des caractéristiques des variables fonctionnelles et scalaires et de la variable de sortie ont été développés.
Cette méthodologie a été appliquée sur l'estuaire de la Gironde au travers de deux études de sensibilité, respectivement monodirectionnelle et par analyse globale de sensibilité par décomposition de variance. Ces études ont été mises en œuvre à l'aide du modèle numérique bidimensionnel non-débordant de l'estuaire de la Gironde actuellement opérationnel chez Météo-France et fondé sur le code de calcul Telemac2D, dans le but de hiérarchiser l'influence des forçages et des paramètres du modèle.
De manière pratique, l'une des caractéristiques de la réalisation de ces études est qu'il est nécessaire de perturber aussi bien les entrées scalaires du modèle (coefficients de frottement et d'influence du vent) que les variables fonctionnelles, c'est-à-dire dépendantes du temps (condition aux limites maritimes et débits des deux rivières alimentant l'estuaire) de telle sorte de prendre en compte le plus grand nombre de sources d'incertitudes et de pouvoir évaluer la part de variance de la variable de sortie (ici les niveaux d'eau) liée à chaque forçage ou paramètre.
La solution retenue pour perturber les variables fonctionnelles est fondée sur l'évaluation de la durée de corrélation physique du signal et sur l'analyse en composantes principales de la fonction d'autocorrélation considérée comme gaussienne de celui-ci. Le but est d'appliquer une perturbation corrélée dans le temps et cohérente avec les caractéristiques des processus physiques en jeu pendant les tempêtes et les crues dans l'estuaire.

La deuxième caractéristique de ces études concerne le coût de calcul élevé du code étudié qui limite le nombre de simulations possibles, ce qui a nécessité une analyse de la scalabilité du modèle numérique utilisé sur plusieurs architectures informatiques afin de favoriser l'utilisation optimale des ressources de calcul dans le cadre de la méthodologie proposée. Cette étude de quantification d'incertitudes a été publiée dans Laborie et al. [2019].

Après un rappel des principales méthodes d'assimilation de données dans le chapitre 3, le second axe de travail a consisté à développer une méthode ensembliste d'assimilation de données (filtre de Kalman d'ensemble) dans l'objectif est de réduire les incertitudes du modèle liées à ses forçages et à sa paramétrisation par l'utilisation de mesures. Cette méthodologie est présentée dans le chapitre 4. Le principal objectif de cette partie des travaux est de corriger les conditions aux limites (variables fonctionnelles) et les paramètres (variables scalaires) usuels des codes d'hydraulique à surface libre bidimensionnels au sein d'une plate-forme d'assimilation de données et de caractériser l'évolution des variables corrigées et de la variable de sortie du code, ainsi que l'évolution de leurs dépendances au fil du temps. Des outils de visualisation spatio-temporelle des caractéristiques des variables corrigées et de la variable de sortie du code ont été développés. Dans le chapitre 5, cette méthodologie a été appliquée sur l'estuaire de la Gironde sous la forme d'un protocole d'expériences jumelles dont l'objectif est, d'une part, de procéder à sa validation et, d'autre part, de déterminer de manière optimale le vecteur de contrôle, le réseau de mesures et les paramètres liés à la méthodologie proposée (fréquence d'assimilation, durée de la fenêtre d'assimilation, recouvrement entre les fenêtres, notamment). La solution retenue pour l'optimisation des variables fonctionnelles (conditions aux limites maritimes, débits des rivières) et des paramètres scalaires est cohérente avec les travaux de quantification d'incertitudes précédents. Par ailleurs, pour éviter l'effondrement de l'ensemble, une technique de réensemencement fondée sur sa dispersion initiale est proposée et appliquée. Les caractéristiques principales de ces travaux concernent, d'une part, le caractère instationnaire de l'hydrodynamique dans l'estuaire, liée à la fois à la propagation de la marée et aux caractéristiques de l'événement hydrométéorologique, et, d'autre part, la corrélation entre les observations des différentes stations de mesures disponibles. Dans le chapitre 6, les gains de la méthodologie proposée ont été évalués en utilisant les observations disponibles pour plusieurs événements réels.

Ce manuscrit de thèse se termine par un chapitre consacré aux conclusions et aux perspectives à envisager à l'issue de ce travail.

# Chapitre 1

# Prévision des hautes eaux dans l'estuaire de la Gironde

**Sommaire**


**1.1 L'estuaire de la Gironde . . . . . 14**
1.1.1 Situation géographique et bassin versant . . . . . 14
1.1.2 Enjeux de la prévision des niveaux d'eau dans l'estuaire de la Gironde 15
1.1.3 Organisation de la prévision des crues en France et sur l'estuaire de la Gironde . . . . . 17
1.1.3.1 Le réseau Vigicrues . . . . . 17
1.1.3.2 La prévision des niveaux d'eau sur l'estuaire de la Gironde . 18
**1.2 Modélisations numériques de l'estuaire de la Gironde . . . . . 19**
1.2.1 Contexte de mes travaux de thèse . . . . . 20
1.2.1.1 Le projet Gironde (2005-2010) . . . . . 20
1.2.1.2 Le projet GirondeXL 3D . . . . . 25
1.2.2 Présentation du modèle actualisé de prévision des hautes eaux de l'estuaire de la Gironde . . . . . 27
1.2.2.1 Équations de Saint-Venant dans Telemac2D . . . . . 27
1.2.2.2 Emprise du modèle . . . . . 28
1.2.2.3 Conditions aux limites du modèle . . . . . 28
1.2.2.4 Version du code de calcul Telemac2D . . . . . 29
1.2.3 Performances du modèle Gironde actualisé (MEG0V7P3) . . . . . 30
1.2.4 Convergence en maillage des résultats en niveaux d'eau . . . . . 34
**1.3 Conclusions . . . . . 38**

L'objet de ce travail de thèse a été, d'une part, de développer une méthode d'assimilation ensembliste permettant la correction des paramètres et des forçages temporels appliqués aux modèles bidimensionnels d'hydrodynamique à surface libre et, d'autre part, d'en évaluer les performances en l'appliquant au modèle de prévision des hautes eaux de l'estuaire de la Gironde.

La section 1.1 de ce chapitre présente la situation géographique, le bassin versant et les principales caractéristiques hydromorphologiques de l'estuaire de la Gironde. Les enjeux en termes de prévision des niveaux d'eau, ainsi que l'organisation de la prévision des crues et des inondations en France et sur le Territoire Gironde-Adour sont également décrits. La nécessité d'outils d'aide à la décision pour permettre aux services de l'État en charge de la sécurité des biens et des personnes d'y répondre, ainsi que les conséquences de la tempête de 1999, ont justifié le développement d'un modèle de prévision des hautes eaux de l'estuaire de la Gironde dans le cadre du projet Gironde. Le projet Gironde et ses conclusions sont synthétisés dans la section 1.2.1, alors que le projet GirondeXL 3D, dont l'objet est la construction d'un éco-pilote pour faciliter la navigation des bateaux dans l'estuaire, est présenté dans la section 1.2.1.2. Dans le cadre de ces travaux de thèse, le modèle Gironde issu du projet Gironde a été adapté et actualisé, notamment à l'aide des données bathymétriques recueillies dans le cadre du projet GirondeXL 3D. Le nouveau modèle, nommé MEG0V7P3, en référence à sa zone d'étude, au raffinement de son maillage et à la version du code Telemac2D qu'il utilise, est décrit dans la section 1.2.2. Ses performances, évaluées pour la tempête du 02 au 09 Février 2003, qui a permis de caler le modèle, ainsi que pour l'événement de 1999, justifient l'utilisation de techniques d'assimilation de données pour corriger l'incertitude sur les paramètres et les forçages les plus influents du MEG0V7P3. Ces derniers seront identifiés par une étude de quantification d'incertitudes décrite dans le chapitre 2 du manuscrit.

Par ailleurs, dans la mesure où il a été décidé de tester les gains obtenus par assimilation de données ***in situ*** le long de l'estuaire pour la prévision des niveaux d'eau en conservant le maillage originel ($mesh0$) du projet Gironde, une étude d'influence du raffinement du maillage sur l'estimation de ceux-ci a été réalisée. C'est ce que montre la section 1.2.4.

## 1.1 L'estuaire de la Gironde

### 1.1.1 Situation géographique et bassin versant

L'estuaire de la Gironde, situé dans le sud-ouest de la France, à proximité de la ville de Bordeaux et du Centre Nucléaire de Production d'Électricité (CNPE) du Blayais, est le plus grand estuaire macrotidal de France et d'Europe occidentale. Il s'étend depuis le golfe de Gascogne sur environ 170 km à l'intérieur des terres et couvre une surface de 635 km$^2$. L'estuaire de la Gironde naît du confluent de deux rivières, la Dordogne au Nord et la Garonne au Sud, qui se rejoignent au Bec d'Ambès. L'estuaire est représenté, ainsi que les bassins versants de ses affluents, sur la figure 1.1. Il atteint la côte atlantique 75 km environ plus en aval. Sa direction moyenne est orientée du sud-

est au nord-ouest et sa largeur varie de 1 km à proximité de Bordeaux à 12 km à son embouchure, c'est-à-dire au droit du Verdon. L'estuaire de la Gironde peut être divisé (Castaing and Allen [1981]) en trois sous-domaines : la zone amont du fleuve, la zone estuarienne centrale et la zone aval offshore.

La marée dans le golfe de Gascogne est une onde semi-diurne de période 12 h 25 mn dominée par la composante principale semi-diurne lunaire (M2) (Huybrechts et al. [2011]). Les hauts fonds dans l'estuaire, qualifié de macrotidal, induisent un fort marnage avec une amplitude de marée près de son embouchure allant de 2.2 à 5.4 m pendant le cycle de vives eaux. L'estuaire est également à morphologie convergente dite "hypersynchrone", c'est-à-dire que les rives convergent en allant vers l'amont, avec une forme caractéristique en entonnoir. L'estuaire, dont les principaux points d'intérêt sont représentés sur la figure 1.4, est ainsi soumis à des influences maritimes et l'amplitude de l'onde semi-diurne est amplifiée de l'aval vers l'amont du fait de sa morphologie hypersynchrone. L'onde de marée est asymétrique (4 h pour le flot et 8 h 25 mn pour le jusant). Les temps de propagation de la marée depuis le Verdon sont, pour la Pleine Mer (PM) (resp. basse mer (BM)), d'environ 20 mn (resp. 30 mn) à Richard, 50 mn (resp. 1 h 40 mn) à Pauillac, environ 2 h (resp. 3 h) au Marquis et environ 2 h 30 mn (resp. 4 h 40 mn) à Bordeaux.

A l'amont de l'estuaire, le débit de la Garonne (resp. Dordogne) varie de 50 (resp. 20) à 2 000 (resp. 1 000) $m^3.s^{-1}$. Lors d'inondations, le débit de la Garonne dépasse parfois 5 000 $m^3.s^{-1}$. Les débits moyens des rivières Dordogne et Garonne valent respectivement 380 $m^3/s$ et 630 $m^3/s$, alors que les apports d'eau par l'océan Atlantique sont compris entre 15 et 25000 $m^3$ par marée (Hissel et al. [2006]).

La combinaison de grandes amplitudes de marée et de débits fluviaux élevés peut entraîner de fortes inondations (Monbaliu et al. [2014]).

### 1.1.2 Enjeux de la prévision des niveaux d'eau dans l'estuaire de la Gironde

Avec 3800 $km^2$ de bassin versant, 185 communes de Gironde et de Charente-Maritime et près d'un million d'habitants, les enjeux environnementaux, économiques et sécuritaires sont importants sur l'estuaire de la Gironde. Son importance économique est confirmée par la présence de grandes villes et de ports, dont la ville de Bordeaux et le port de Bordeaux, ainsi que par la présence de diverses industries, telles que le CNPE du Blayais, la pêche et les activités touristiques. Malgré la forte densité de population et son fort développement économique, l'estuaire de la Gironde reste l'un des estuaires les plus naturels et est un site préservé pour de nombreuses espèces biologiques.

Le risque d'inondation est depuis longtemps une préoccupation majeure des autorités de l'estuaire de la Gironde (Monbaliu et al. [2014]), comme en témoignent les documents historiques datant du XIII$^{ème}$ siècle. L'inondation la plus dévastatrice s'est produite en avril 1770, quand environ 24 000 $km^2$ ont été recouverts d'eau le long de la Garonne et de l'estuaire de la Gironde, causant environ 4 669 millions d'euros de dommages exclusivement dans la ville de Bordeaux (35

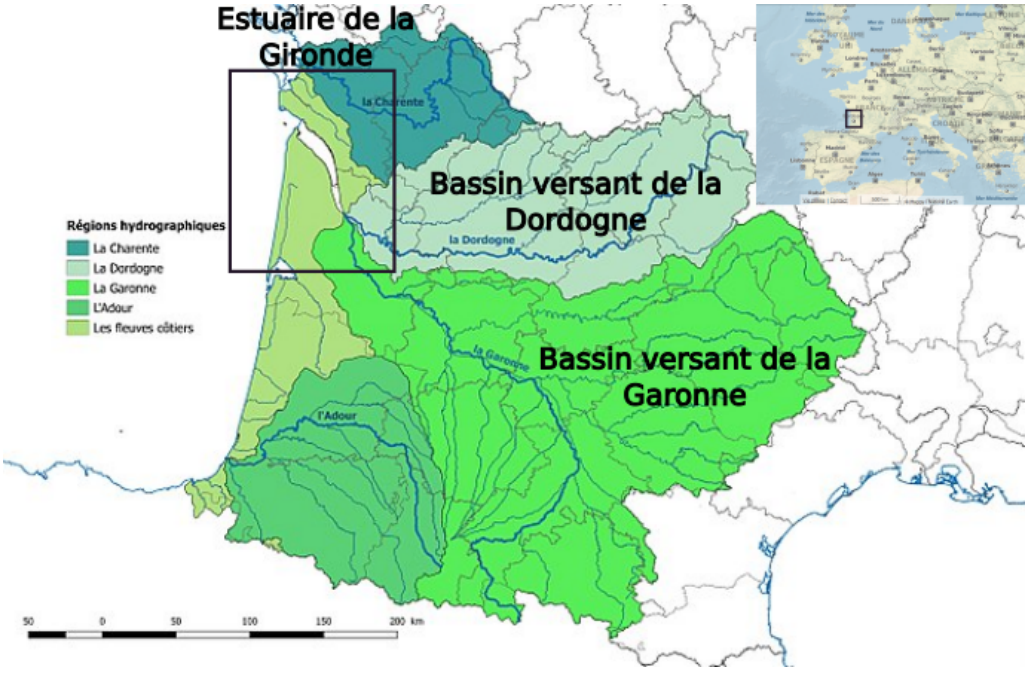



FIGURE 1.1 : Localisation de l'estuaire de la Gironde en Europe et de son bassin versant (source : https://fr.wikipedia.org/wiki/Bassin_Adour-Garonne)

millions d'euros pour les biens primaires) (Hissel et al. [2014]). Dès lors, des ouvrages de protection contre les inondations ont été construits pour limiter leurs conséquences, mais elles n'ont pas réussi à empêcher celles, fortes, de 1835, 1855 et 1856. En 1930, les inondations ont causé la destruction de 1 000 maisons et ont fait plus de 300 victimes. Au cours des dernières décennies, trois événements notables ont été observés : Décembre 1981, causé par de forts débits fluviaux, les tempêtes Lothar et Martin en 1999 et Xynthia en 2010.

Ainsi, lors de la tempête du 27 décembre 1999, de fortes dépressions et des vents très violents, d'intensité maximale atteignant 198 km/h, ont été observés sur le littoral atlantique. Dans l'estuaire de la Gironde, ces conditions climatologiques extrêmes ont provoqué des surcotes très importantes, près de 1.55 m en sortie d'estuaire au Verdon, et 2.25 m à Bordeaux (Salomon [2002]). À cette même date, le CNPE du Blayais, situé en rive droite de l'estuaire à proximité de Pauillac, a été partiellement inondé (Marano [2000]) du fait de la submersion des digues de protection de la centrale. Cet événement a démontré le besoin important de prévisions des niveaux d'eau dans l'estuaire pour anticiper les conséquences des événements intenses.

### 1.1.3 Organisation de la prévision des crues en France et sur l'estuaire de la Gironde

#### 1.1.3.1 Le réseau Vigicrues

Les services de l'État sont responsables de la sécurité des personnes et des biens. Ils s'appuient sur des systèmes d'aide à la décision pour prendre des mesures préventives afin d'alerter les populations et coordonner la gestion de crise.

En France, le réseau Vigicrues est chargé du suivi et de la prévision des niveaux d'eau et des débits sur 22 000 km de rivières et couvre 75 % de la population vivant en zone inondable. Ce réseau surveillé par l'État est piloté par le SCHAPI (Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations), Service à Compétence Nationale rattaché au Service des Risques Naturels et Hydrauliques de la Direction Générale de la Prévention des Risques du Ministère du Développement durable. Le SCHAPI coordonne les travaux de développement, homogénéïse les interventions des SPC et veille au partage de l'expertise. Au sein de ce réseau, les SPC sont chargés de la prévision des crues et de la mise en œuvre de la procédure vigilance crue. Ils apportent également leur appui aux élus locaux et aux préfets en charge de la sécurité des populations.

Depuis 2006, le réseau Vigicrues produit, sur le site http:\www.vigicrues.gouv.fr, deux fois par jour (10h et 16h) et davantage si nécessaire (en cas d'événement), une carte de vigilance. Celle-ci, disponible en ligne pour les services de l'Etat et le grand public, est accompagnée d'un bulletin d'information national et de bulletins locaux décrivant l'évolution des crues, de données en temps réel sur les débits et les hauteurs d'eau et de prévisions.

La vigilance est définie pour chaque tronçon du réseau surveillé et rend compte d'un niveau de risque caractérisé par une couleur. La vigilance crue, relayée dans le cadre de la vigilance météorologique au travers des pictogrammes "inondation" et "pluie-inondation", permet de prévenir le public et les autorités qu'il existe un risque de crue, plus ou moins important selon la couleur, et permet de se mettre en situation de réagir de manière appropriée si le danger se précise. Des conseils de comportement adaptés au niveau de vigilance et élaborés par les pouvoirs publics sont indiqués dans les bulletins d'information.

En fonction du niveau de vigilance, les pouvoirs publics en charge de la sécurité civile (préfets et maires) déclenchent l'alerte lorsque cela est nécessaire et mobilisent les moyens de secours adaptés. Les quatre couleurs de la vigilance sont définies en collaboration avec la Direction générale de la Sécurité Civile et de la Gestion de Crise (DGSCGC) du Ministère de l'Intérieur et sont identiques à celles utilisées par METEO-FRANCE pour la vigilance météorologique, comme indiqué dans le tableau 1.1.

La marque Vigicrues a été lancée le 23 février 2015 par Ségolène Royal. Le regroupement de toute l'information relative au risque de crue sous une identité commune répond à la volonté du gouvernement d'informer plus efficacement les Français sur les risques d'inondation.

TABLE 1.1 : Niveaux de vigilance

| | |
|---|---|
| Vert | Pas de vigilance particulière requise. |
| Jaune | Risque de crue génératrice de débordements et de dommages localisés ou de montée rapide et dangereuse des eaux, nécessitant une vigilance particulière notamment dans le cas d'activités exposées et/ou saisonnières. |
| Orange | Risque de crue génératrice de débordements importants susceptibles d'avoir un impact significatif sur la vie collective et la sécurité des biens et des personnes. |
| Rouge | Risque de crue majeure. Menace directe et généralisée de la sécurité des personnes et des biens. |

Les contraintes majeures de la prévision opérationnelle des niveaux d'eau résident principalement dans les contraintes de temps inhérentes à la prévision en temps réel. Les trois composantes majeures de la prévision sont :

- le réseau d'observation et les données d'observation,
- les modèles de prévision,
- le système d'alerte et d'information.

Pour créer les cartes de risques, le réseau Vigicrues s'appuie ainsi sur des modèles numériques de prévision et des mesures ***in situ*** (Golnaraghi [2012]). Les codes de calcul hydrodynamiques fondés sur les équations de Saint-Venant (SWE) sont des outils couramment utilisés pour aider à la gestion et à la protection des infrastructures urbaines situées près des rivières et des côtes. Ils sont également utilisés pour la prévision opérationnelle des crues.

#### 1.1.3.2 La prévision des niveaux d'eau sur l'estuaire de la Gironde

Le SPC Gironde-Adour-Dordogne (GAD), dont la zone d'intervention est représentée sur la figure 1.2, est en charge de l'émission des bulletins de vigilance sur l'estuaire de la Gironde, tel que celui de 16/02/2016 de la figure 1.3. Afin de répondre aux attentes opérationnelles, notamment pour les événements extrêmes, le SPC-GAD est passé d'un modèle statistique basé sur des données climatologiques à un modèle numérique qui résout les SWE fondé sur le logiciel hydraulique Telemac2D (Hervouet [2007]). L'historique de ces travaux, réalisés dans le cadre du projet Gironde, est indiqué en section 1.2.1.1.

Les points de prévision réglementaires à 3 marées d'horizon de prévision (environ 36 heures) sur l'estuaire de la Gironde, représentés sur la figure 1.4, sont les suivants :

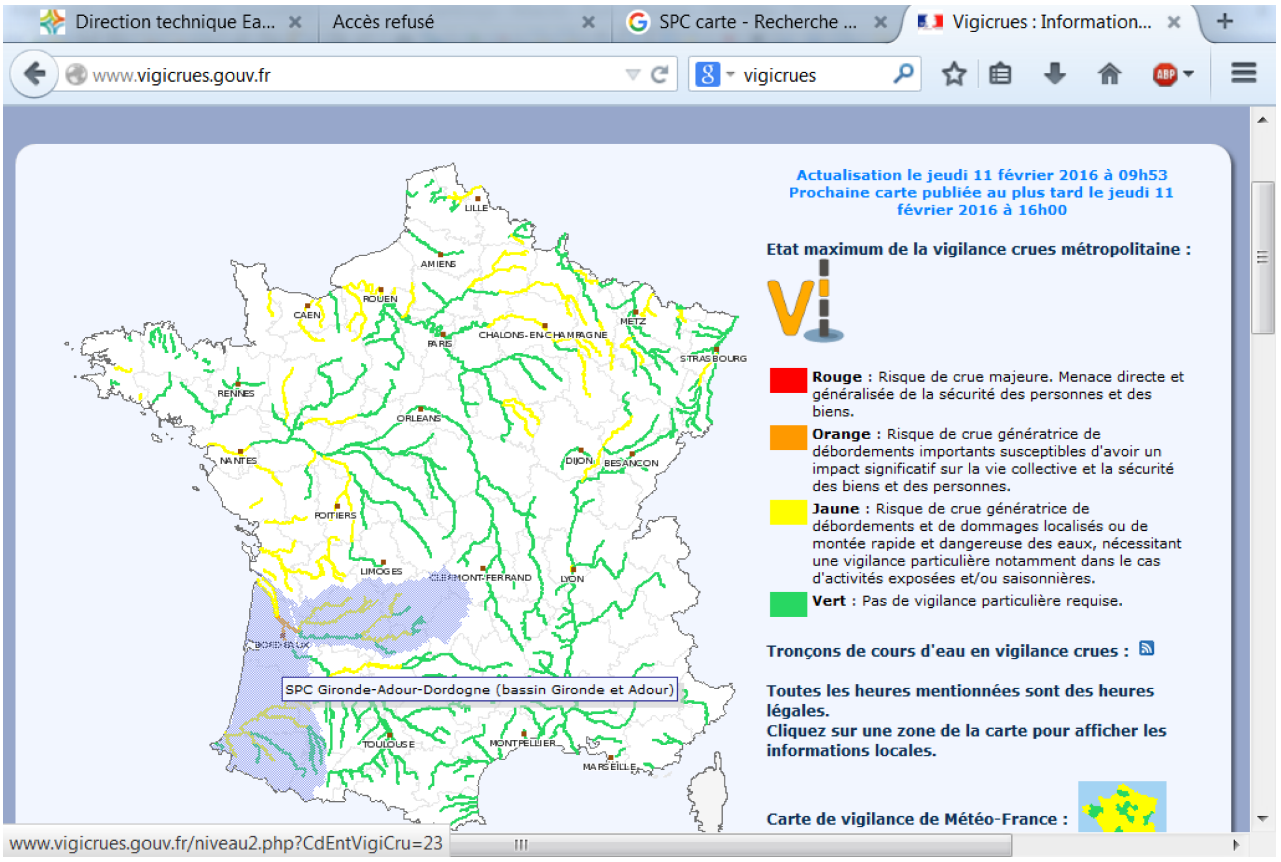



FIGURE 1.2 : Zone d'intervention du Service de Prévision des Crues Gironde-Adour-Dordogne (SPC-GAD). Source : http:\www.vigicrues.gouv.fr

- Le Verdon qui conditionne toutes les prévisions sur l'estuaire ;
- Pauillac situé en face du CNPE du Blayais ;
- la presqu'ile d'Ambès qui conditionne la prévision à Libourne et est donc très sensible ;
- La Réole, sur la Garonne ;
- Bordeaux qui est une zone de prévision très sensible du fait de sa vulnérabilité ;
- Libourne, à la confluence de l'Isle et de la Dordogne ;
- Cadillac et Langon à l'amont de la Garonne.

La figure 1.5 représente un schéma simplifié des stations de mesures (marégraphes ou limnigraphes) présentes le long de l'estuaire de la Gironde.

## 1.2 Modélisations numériques de l'estuaire de la Gironde

Après une description du projet Gironde (2005-2010) et du projet GirondeXL 3D (2016-2018) (section 1.2.1), le modèle de l'estuaire de la Gironde actualisé et utilisé dans le cadre de ces travaux de thèse est détaillé dans la section 1.2.2.

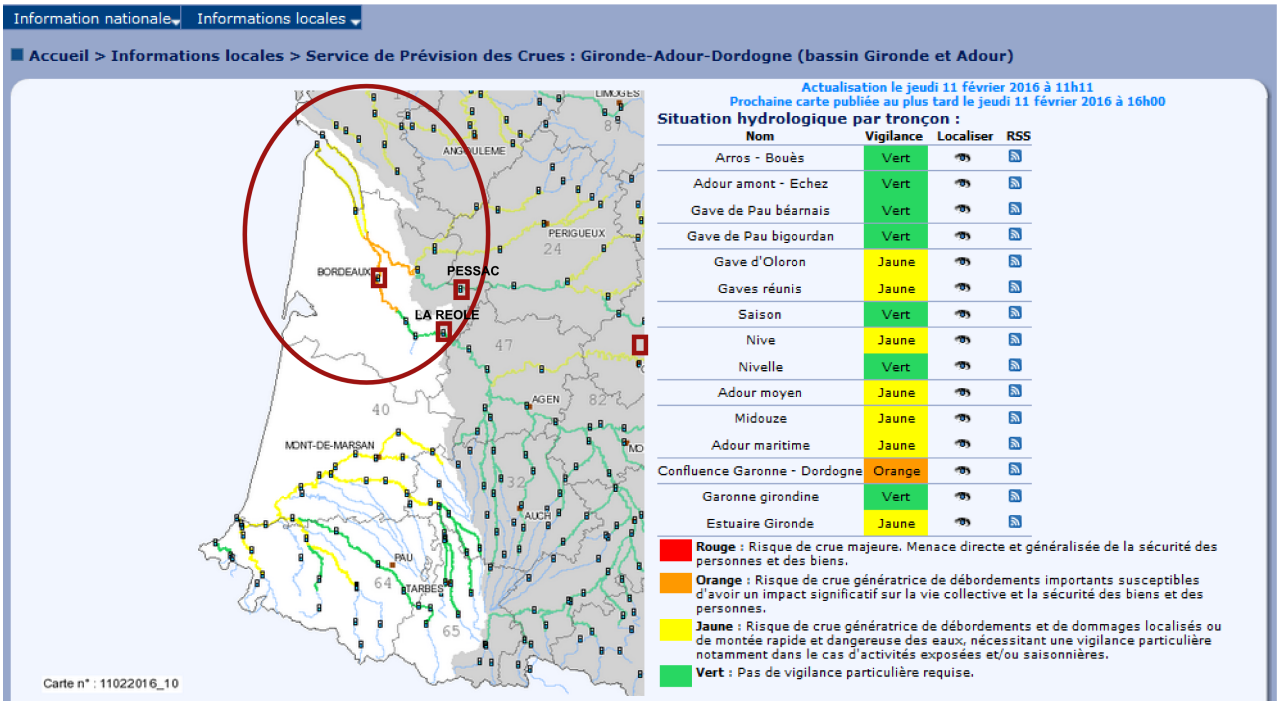



FIGURE 1.3 : Bulletin de vigilance sur l'estuaire de la Gironde du 11/02/2016. Le cercle rouge représente la zone de l'estuaire modélisée numériquement dans le cadre des projets Gironde et GirondeXL 3D (cf. section 1.2.1) (source : http:\www.vigicrues.gouv.fr)

## 1.2.1 Contexte de mes travaux de thèse

### 1.2.1.1 Le projet Gironde (2005-2010)

#### 1.2.1.1.a Contexte

Suite à la tempête du 27 décembre 1999 et à l'inondation partielle du CNPE du Blayais, le Laboratoire National d'Hydraulique et d'Environnement (LNHE) d'Électricité de France (EDF) a mené un ensemble d'études et de travaux à partir du début de l'année 2000 pour vérifier et mettre à niveau les ouvrages de protection et les procédures de surveillance et d'organisation contre le risque inondation (Hissel et al. [2006]).

Parmi les études réalisées au sein d'EDF/LNHE, des modèles hydrauliques, fondés sur le code de calcul Telemac2D développé par le laboratoire et simulant la marée et les surcotes météorologiques avec ou sans prise en compte des zones de débordement adjacentes à l'estuaire, ont été construits et calés. L'intérêt du modèle numérique qui prend en compte des zones de débordement réside dans la meilleure représentation de l'effet d'écrêtement observé sur le niveau d'eau de la Gironde pendant les événements exceptionnellement forts. Il ne pouvait néanmoins pas être utilisé dans une phase opérationnelle compte tenu des ressources de calcul qu'il nécessitait,

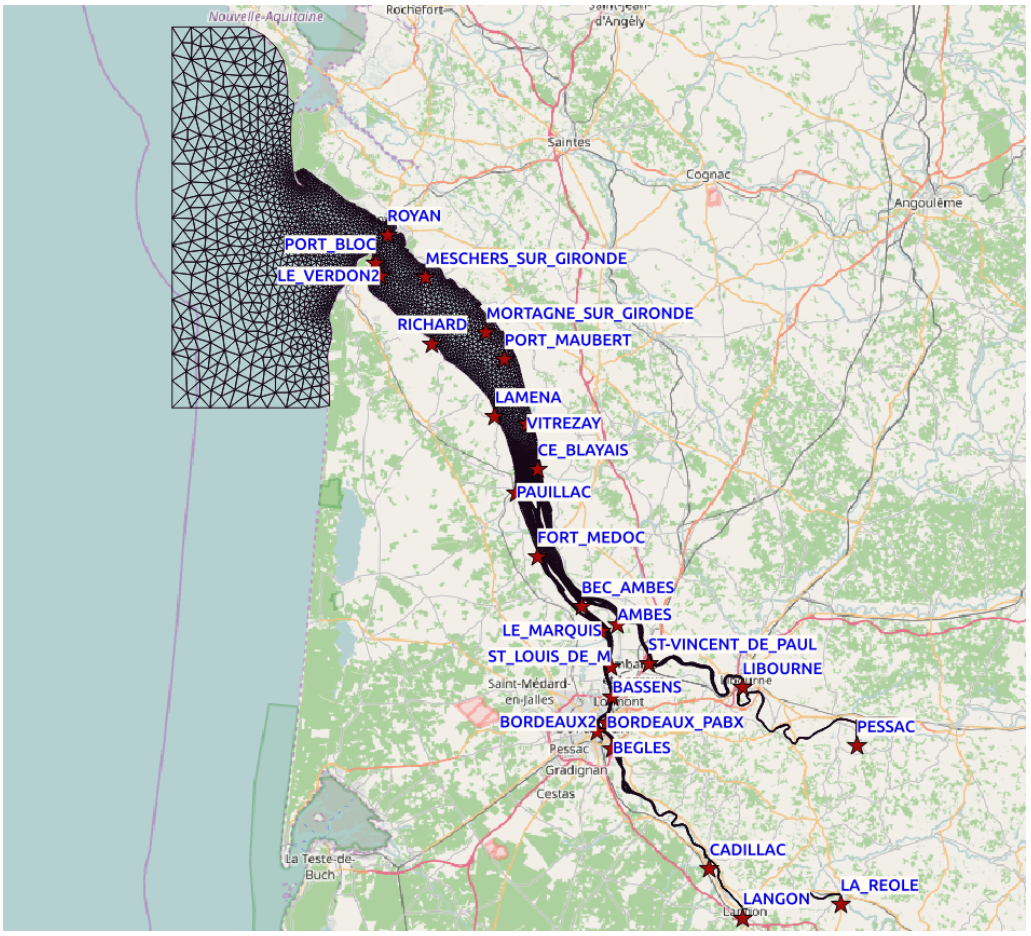



FIGURE 1.4 : Emprise du modèle de prévision de l'estuaire de la Gironde et localisation des points d'intérêt

incompatibles, à l'époque, avec les besoins du SPC. C'est pourquoi ce modèle avec débordement n'a pas été retenu par EDF R&D et METEO-FRANCE.

Dans le cadre d'une collaboration de recherche, le LNHE et METEO-FRANCE ont testé la faisabilité et les apports de modèles numériques de prévisions de surcotes et de vagues sur l'estuaire de la Gironde. Ce projet de Recherche et Développement visait à examiner et à quantifier leurs apports et leurs limitations pour éventuellement mettre en place des systèmes d'alerte pour les installations situées dans l'estuaire de la Gironde et, en particulier, pour le CNPE du Blayais. De fait, ces modèles n'ont pas fait l'objet d'efforts de calage poussés en amont. En outre, comme la zone étudiée était essentiellement sous influence maritime, les crues fluviales correspondant à des débits très élevés n'ont pas été testées.

En 2005, dans le cadre d'une convention quadri-partite entre le CETMEF (Cerema Eau, Mer et Fleuves depuis 2014), EDF R&D, METEO-FRANCE et le Ministère du Développement Durable incluant le SPC Littoral Atlantique (SPC-LA) (intégré dans le SPC-GAD depuis) et le SCHAPI,

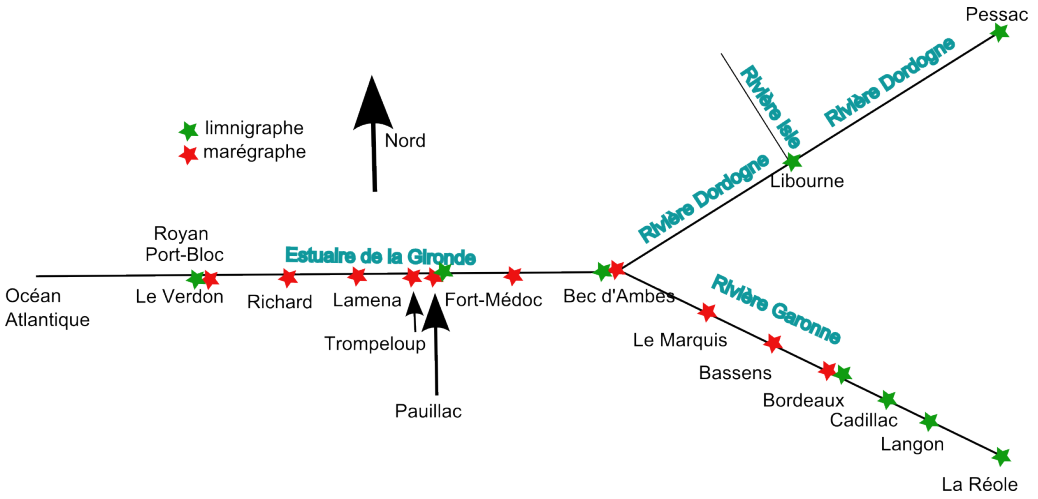



FIGURE 1.5 : Emprise du modèle de prévision de l'estuaire de la Gironde et localisation des points d'intérêt et de mesures ***in situ***

le modèle non-débordant (nommé modèle Gironde) choisi par METEO-FRANCE et EDF R&D à l'issue des phases précédentes du projet a fait l'objet de développements et d'interfaçage. Le CETMEF a évalué les performances et adapté le modèle non-débordant de l'estuaire de la Gironde construit par EDF R&D pour l'intégrer dans la plate-forme de calcul opérationnelle Synergie de METEO-FRANCE. Les différents travaux réalisés par le CETMEF ont consisté en :

1. l'évaluation du modèle LNHE-Gironde en termes de scores (Nash de Pleines Mers (PM) rephasées selon l'équation 1.1 et erreur quadratique moyenne (RMSE) selon l'équation 1.2) sur toutes les stations de mesure disponibles le long de l'estuaire de la Gironde pour 10 événements historiques dont 6 débordants (Hissel et al. [2006]). Les événements étudiés sont décrits dans le tableau 1.2.

$$Nash_{HT} = 1 - \frac{\sum_{i=1}^{n}(H_i - \widehat{H_i})^2}{\sum_{i=1}^{n}(\widehat{H_i} - \overline{H})^2}, \tag{1.1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(H_i - \widehat{H_i})^2} \tag{1.2}$$

avec $n$ l'indice correspondant au pas de temps considéré, $H_i$ (resp. $\widehat{H_i}$) les niveaux d'eau simulés (resp. observés) et $\overline{H}$ la moyenne temporelle des niveaux d'eau observés.

2. deux nouveaux calages du modèle Gironde, l'un fondé sur le Nash des PM rephasées, choisi ***in fine*** par les partenaires du projet et l'autre sur la RMSE, obtenus à l'aide des 4 événements non-débordants (2003, 2004, 2006-a et 2006-b) (Hissel [2007]) et une évaluation des performances du modèle sur les événements débordants (Hissel and Morellato [2009]). Le poids accordé à chacun des 4 événements non-débordants ayant servi au recalage est défini selon la fiabilité des données d'observation disponibles. L'événement dont les données

étaient les plus fiables est l'événement correspondant à la tempête qui s'est produite du 02/02 au 09/02/2003.

TABLE 1.2 : Caractéristiques des événements de calage et d'évaluation du modèle non-débordant de l'estuaire de la Gironde (modèle Gironde). Les lignes vertes (resp. rouges) indiquent les événements "non-débordants" (resp. "débordants") (d'après Hissel et al. [2006])

| Événement | Débits sur la Dordogne | Débits sur la Garonne | Plage de variation du coefficient de marée |
|---|---|---|---|
| 10 au 14/03/2006 | 850 à 1550 $m^3/s$ | 1700 à 4700 $m^3/s$ | 37 à 83 |
| 27 au 31/03/2006 | 500 à 950 $m^3/s$ | 650 à 1200 $m^3/s$ | 78 à 115 |
| 3 au 6/05/2004 | 450 à 700 $m^3/s$ | 1650 à 2450 $m^3/s$ | 86 à 105 |
| 2 au 9/02/2003 | 600 à 2200 $m^3/s$ | 1200 à 5900 $m^3/s$ | 43 à 90 |
| 25 au 31/12/1999 | 430 à 1830 $m^3/s$ | 400 à 3600 $m^3/s$ | 45 à 102 |
| 24 au 30/04/1998 | 300 à 1200 $m^3/s$ | 630 à 3100 $m^3/s$ | 70 à 105 |
| 3 au 10/02/1996 | 300 à 850 $m^3/s$ | 880 à 2550 $m^3/s$ | 71 à 87 |
| 20 au 26/12/1995 | 250 à 550 $m^3/s$ | 500 à 1200 $m^3/s$ | 79 à 107 |
| 17 au 20/03/1988 | 500 à 1800 $m^3/s$ | 1250 à 5500 $m^3/s$ | 99 à 115 |
| 12 au 17/12/1981 | 1000 à 2350 $m^3/s$ | 1700 à 7050 $m^3/s$ | 57 à 106 |

Les zones de débordement de l'estuaire de la Gironde, ainsi que l'influence du changement climatique sur leur emprise, sont cartographiées dans Laborie et al. [2012] à l'aide du modèle Référentiel Inondation Gironde (RIG, propriété du bureau d'études ARTELIA). Malgré une représentation des zones inondables très simplifiée et une emprise restreinte à l'intérieur de l'estuaire, le RIG est constitué de 21304 éléments et de 13621 nœuds, soit deux fois plus d'éléments que le modèle non-débordant retenu par les partenaires du modèle Gironde.

#### 1.2.1.1.b Principaux résultats et recommandations issus du projet Gironde

A l'issue du projet en 2010, le CETMEF a produit un rapport final (Hissel [2010]) sur les résultats de calage et d'évaluation du modèle de prévision des hautes eaux de l'estuaire de la Gironde, ainsi que des recommandations. Celles-ci ont constitué le point de départ du présent travail de thèse et sont présentées ici.

Tandis que la figure 1.6 représente les RMSE pour chacun des événements de référence du tableau 1.2, les résultats obtenus pour l'événement de calage non-débordant de Février 2003 en optimisant le Nash de PM avec le modèle Gironde sont récapitulés en termes de Nash de PM et de RMSE pour chaque station de mesures dans le tableau 1.5 (section 1.2.3 de ce chapitre). Hissel [2010] conclut de la figure 1.6 que les RMSE sont inférieures pour les événements ayant servi au calage et donc non-débordants, ce qui est tout à fait cohérent, puisque le modèle lui-même ne prend pas en compte les champs d'expansion de crues. Par ailleurs, la figure 1.6 et le tableau 1.5 montrent que les RMSE sont systématiquement plus grandes pour les stations de Pessac et de La

Réole, ce qui est également logique, compte-tenu du maillage peu raffiné en amont de l'estuaire et de l'artéfact numérique qui consiste à injecter le débit de l'Isle à Pessac, alors que cet affluent de la rivière Dordogne conflue à Libourne. Ensuite, les résultats au Verdon et à Port-Bloc sont meilleurs que sur les autres stations. Outre que la qualité des résultats en ces stations dépend principalement de la condition aux limites maritimes, ce constat peut être relié au caractère hypersynchrone de l'estuaire de la Gironde qui amplifie l'erreur sur les niveaux d'eau depuis le Verdon jusqu'aux stations les plus en amont des deux rivières. Enfin, les résultats sont particulièrement mauvais pour les événements débordants de 1988, 1998 et 1999. Pour ces deux derniers événements, Hissel [2010] évoque une condition aux limites de mauvaise qualité.

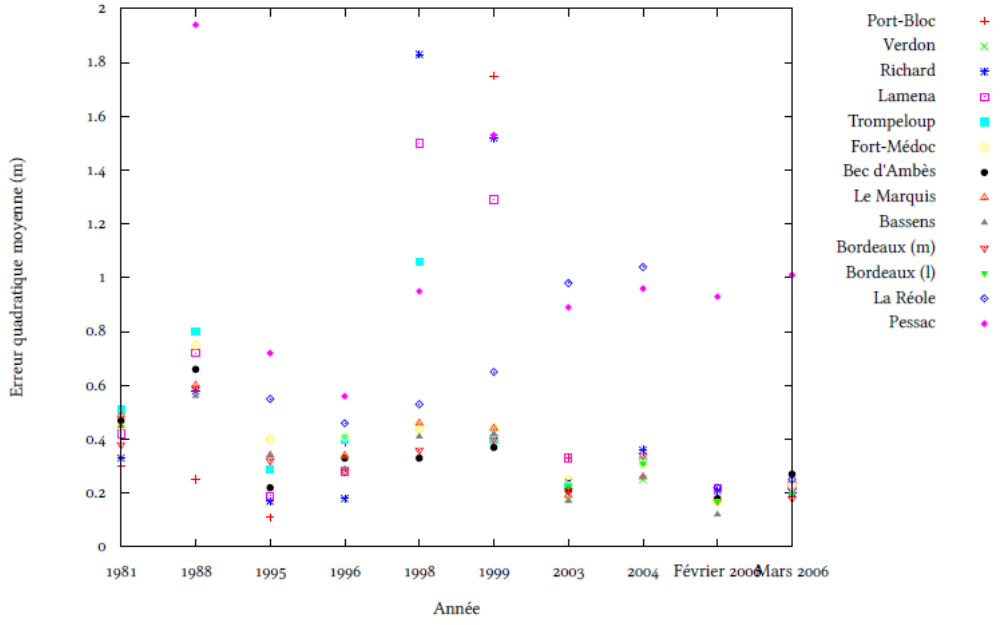



FIGURE 1.6 : RMSE sur tous les événements de référence du projet Gironde en chaque station de mesure de l'estuaire (source : Hissel [2010])

Sur la base de ces résultats, le projet Gironde a abouti à des recommandations (Hissel [2010]). L'une des principales limites du modèle Gironde est l'absence de prise en compte des champs débordants. Cependant, cette caractéristique résulte d'un choix de l'équipe-projet compte-tenu des temps d'exécution du modèle contraints par l'usage en prévision opérationnelle. De fait, si le choix d'un modèle non-débordant est maintenu, Hissel [2010] préconise plusieurs pistes d'amélioration : l'actualisation de la bathymétrie, l'amélioration des forçages maritimes et fluviaux, respectivement issus du modèle de prévision des surcotes de METEO-FRANCE et des courbes de tarage prises en compte à l'amont des rivières Garonne et Dordogne. Cette dernière piste inclut des perspectives à la fois sur le raffinement du maillage à utiliser en amont des rivières, mais aussi sur l'évaluation de l'incertitude sur les débits injectés en amont du modèle. Enfin, Hissel [2010] recommande le couplage du modèle Gironde avec des techniques d'assimilation de données afin

de tirer profit de ces nouvelles méthodes innovantes et à la pointe de la recherche en hydraulique qui permettent de faire communiquer le modèle avec les observations existantes pour diminuer l'incertitude sur les résultats du modèle.

#### 1.2.1.2 Le projet GirondeXL 3D

Le projet GirondeXL 3D, piloté par le Grand Port Maritime de Bordeaux (GPMB) et coordonné par le Cerema Eau, Mer et Fleuves (EMF), a réuni de 2017 à 2018 le Cerema EMF, l'université de Bordeaux et EDF R&D. L'objectif du projet était de construire un éco-pilote sur l'estuaire de la Gironde afin de mieux prévoir les pieds de pilote disponibles pour la navigation des bateaux. La zone d'intérêt du projet est située sur la partie navigable de l'estuaire, c'est-à-dire depuis l'embouchure jusqu'à Bordeaux.

Un modèle bidimensionnel non-débordant fondé sur Telemac2D, nommé modèle GirondeXL 3D dans la suite, a été construit dans ce cadre et calé dans des conditions de navigation, donc sans débordement.

Les domaines d'utilisation des deux modèles Gironde et GirondeXL 3D ne sont pas les mêmes. Le premier a pour objectif la prévision des hautes eaux de l'estuaire de la Gironde et a des contraintes opérationnelles en temps de calcul fortes. L'autre a pour objectif, dans un contexte opérationnel également contraint, la détermination des niveaux d'eau pour assurer le pied de pilote suffisant aux bateaux pour la navigation. Du point de vue numérique, les deux principales différences entre le modèle Gironde et le modèle GirondeXL 3D résident :

1. dans le raffinement du maillage, très raffiné dans le chenal de navigation pour le modèle GirondeXL 3D, alors que le maillage est plutôt grossier sur l'ensemble de l'estuaire et, en particulier, à l'amont des deux rivières.
2. dans la délimitation de 6 zones de frottement sur le modèle GirondeXL 3D, alors que les zones de la confluence (resp. en amont) des rivières sont les mêmes pour la Dordogne et la Garonne.

La méthodologie adoptée pour la calibration du modèle GirondeXL 3D a consisté à caler à la fois les coefficients de Strickler sur des zones prédéfinies le long de l'estuaire à l'aide de mesures ***in situ***, mais aussi le niveau moyen, la phase et l'amplitude de la marée appliquée à la frontière maritime du modèle, commune à celle du modèle Gironde. Le niveau moyen calibré peut être identifié à la composante moyennée sur la durée de l'événement de la surcote météorologique. Dans ce cas, l'intérêt de la prise en compte de la surcote météorologique fournie par le modèle HYCOM de surcotes sur la façade Atlantique de METEO-FRANCE à la frontière maritime du modèle GirondeXL 3D est atténué, puisqu'une partie de celle-ci est déjà reprise par le calage du niveau moyen sur la durée de l'événement. Pour une RMSE calculée sur 3 jours (Huybrechts et al. [2018a]), la prise en compte des forçages météorologiques et de la surcote météorologique fournie par METEO-FRANCE permet d'améliorer localement le calcul des niveaux d'eau.

Les travaux réalisés dans le cadre de GirondeXL 3D incluent le recensement auprès des différents opérateurs (GPMB, Voies Navigables de France (VNF), EDF, Service Hydrographique et Océanographique de la Marine (SHOM)) et services de l'État (Direction Régionale de l'Environnement, de l'Aménagement et du Logement (DREAL) Nouvelle Aquitaine) des données bathymétriques les plus récentes existant sur l'estuaire de la Gironde.

Les RMSE, récapitulées dans le tableau 1.3 pour des simulations avec et sans prise en compte des forçages météorologiques, sont calculées sur trois jours à Bordeaux, Pauillac et Le Verdon dans le cadre du projet GirondeXL 3D pour les périodes suivantes :

- du 06/08/2015 au 22/08/2015 (période d'étiage des deux rivières avec débits inférieurs à 300 $m^3/s$), avec une période de mortes-eaux (coefficients de marée inférieurs à 45) du 07/08/2015 au 10/08/2015 et de vives-eaux (coefficients de marée supérieurs à 95) du 14/08/2015 au 17/08/2015 ;
- du 03/05/2015 au 18/05/2015 (période de débits modérés sur la Garonne et la Dordogne, dont les débits sont respectivement inférieurs à 450 $m^3/s$ et 1040 $m^3/s$). La marée est de mortes-eaux du 10/05/2015 au 13/05/2015 et de vives-eaux du 14/05/2015 au 17/05/2015.

TABLE 1.3 : RMSE avec et sans prise en compte des forçages météorologiques pour le modèle GirondeXL 3D (d'après Huybrechts et al. [2018a])

| | | Avec forçages météo. | | Sans forçages météo | |
|---|---|---|---|---|---|
| | | Mortes-Eaux | Vives-eaux | Mortes-Eaux | Vives-eaux |
| Période d'étiage | Verdon | 7 | 11 | 11 | 9 |
| | Pauillac | 11 | 10 | 9 | 12 |
| | Bordeaux | 16 | 15 | 11 | 16 |
| Débits moyens | Verdon | 16 | 12 | 17 | 13 |
| | Pauillac | 20 | 22 | 20 | 22 |
| | Bordeaux | 24 | 16 | 19 | 22 |

En ajustant les coefficients de frottement et le niveau moyen du signal de marée à la frontière maritime du modèle, les RMSE au Verdon, calculées sur une durée de 3 jours et liées à la qualité du forçage maritime, sont comprises entre 12 cm et 17 cm pour des débits moyens des deux rivières. Les RMSE à Pauillac et Bordeaux sont égales à environ 20 cm. Elles sont du même ordre de grandeur que celles obtenues pour l'événement de calage de Février 2003 du modèle Gironde (resp. 18 cm, 23 cm et 21 cm au Verdon, à Trompeloup et à Bordeaux) (cf. tableau 1.5) pour des débits compris entre 600 et 2000 $m^3/s$ (resp. 1200 et 5900 $m^3/s$) sur la Garonne (resp. sur la Dordogne). Les travaux de Huybrechts et al. [2018a] permettent de conclure que, pour des débits modérés à forts, les niveaux d'eau au Verdon, à Pauillac et à Bordeaux sont très peu influencés par les débits des deux rivières, comme le confirmera l'étude de quantification d'incertitudes présentée dans le chapitre 2.

### 1.2.2 Présentation du modèle actualisé de prévision des hautes eaux de l'estuaire de la Gironde

Dans le cadre de ces travaux de thèse, une adaptation et une actualisation du modèle non-débordant issu du projet Gironde de la section 1.2.1.1 a été réalisée. Si l'emprise du modèle et la nature des paramètres du modèle (coefficients de frottement et d'influence du vent) et des forçages (forçages maritimes, débits de la Garonne et de la Dordogne et champs de vent et de pression) sont restés inchangés, la version du code Telemac2D et la bathymétrie du modèle ont été actualisés respectivement sur la dernière version du code de calcul Telemac2D disponible fin 2018 (V7P3) et sur les données bathymétriques les plus récentes recensées dans le cadre du projet GirondeXL 3D (cf. section 1.2.1.2 et Huybrechts et al. [2018a]). Le modèle adapté, nommé MEG0V7P3, est présenté dans cette section. Ses performances, à l'issue de son recalage sur la tempête de Février 2003, sont évaluées sur l'événement débordant de décembre 1999 et comparées avec celles du modèle Gironde dans la section 1.2.3.

#### 1.2.2.1 Équations de Saint-Venant dans Telemac2D

Les équations de Saint-Venant (SWE) sont couramment utilisées dans le domaine de la modélisation hydro-environnementale. Elles sont dérivées des équations de Navier-Stokes et expriment la conservation de la masse et du moment moyennés sur la dimension verticale. La forme non conservatrice des équations est écrite en fonction de la profondeur de l'eau ($h$ [m]) et des composantes horizontales de la vitesse ($u$ et $v$ [m.s$^{-1}$]).

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(hu) + \frac{\partial}{\partial y}(hv) = 0 \tag{1.3}$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = -g\frac{\partial H}{\partial x} + F_x + \frac{1}{h} div\left(h\nu\overrightarrow{grad}(u)\right) \tag{1.4}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = -g\frac{\partial H}{\partial y} + F_y + \frac{1}{h} div\left(h\nu\overrightarrow{grad}(v)\right) \tag{1.5}$$

avec :

$$\begin{aligned} F_x = &- \frac{g}{{K_s}^2}\frac{u\sqrt{u^2+v^2}}{h^{\frac{4}{3}}} - \frac{1}{\rho_w}\frac{\partial P_{atm}}{\partial x} \\ &+ \frac{1}{h}\frac{\rho_{air}}{\rho_w}\ C_d\ U_{w,x}\sqrt{U_{w,x}^2+U_{w,y}^2} \end{aligned} \tag{1.6}$$

$$\begin{aligned} F_y = &- \frac{g}{{K_s}^2}\frac{v\sqrt{u^2+v^2}}{h^{\frac{4}{3}}} - \frac{1}{\rho_w}\frac{\partial P_{atm}}{\partial y} \\ &+ \frac{1}{h}\frac{\rho_{air}}{\rho_w}\ C_d\ U_{w,y}\sqrt{U_{w,x}^2+U_{w,y}^2} \end{aligned} \tag{1.7}$$

et $\rho_w/\rho_{air}$[kg.m$^{-3}$] sont la densité eau/air, $P_{atm}$ [Pa] est la pression atmosphérique, $U_{w,x}$ et $U_{w,y}$ [m.s$^{-1}$] sont les composantes horizontales de la vitesse, $C_d$ est le coefficient d'influence du vent qui relie le vent à la surface libre à la contrainte de cisaillement, $K_s$ [m$^{\frac{1}{3}}$.s$^{-1}$] est le coefficient de frottement dans le lit mineur ou majeur de la rivière, d'après la formulation de Gauckler [1867]. $F_x$ et $F_y$ [m.s$^{-2}$] sont les composantes horizontales des forces extérieures (frottement, vent et forçages atmosphériques), $\boldsymbol{H}$ [m NGF69] est le niveau d'eau ($h = H - z_f$ si $z_f$ [m NGF69] est la cote du fond de la rivière) et $\nu$ [m$^2$.s$^{-1}$] est la viscosité cinématique de l'eau. $\boldsymbol{div}$ et $\overrightarrow{grad}$ sont respectivement les opérateurs divergence et gradient.

Pour résoudre le système d'équations Eq. (1.3) à Eq. (1.5), les conditions initiales $h(x, y, t = 0) = h_0(x, y)$ ; $u(x, y, t = 0) = u_0(x, y)$ ; $v(x, y, t = 0) = v_0(x, y)$ sont fournies. Les conditions aux limites (BC) à la fois sur la côte (conditions de glissement et d'imperméabilité) et aux frontières amont et maritime ($h(x_{BC}, y_{BC}, t) = h_{BC}(t)$) sont également imposées.

Dans cette étude, les SWE sont résolues à l'aide du code de calcul parallèle Telemac2D (www.opentelemac.org) en utilisant un schéma explicite du premier ordre pour la discrétisation temporelle et la méthode des éléments finis associée à la méthode itérative du gradient conjugué pour la discrétisation spatiale (Hervouet [2007]).

#### 1.2.2.2 Emprise du modèle

Le modèle MEG0V7P3, représenté sur la figure 1.7, a été adapté du modèle Gironde (Hissel et al. [2006]) et actualisé dans le cadre de ces travaux de thèse. Fondé sur Telemac2D et sur les données bathymétriques les plus récentes issues du projet GirondeXL 3D, telles que représentées sur la figure 1.8, ce modèle numérique hydrodynamique bidimensionnel de l'estuaire de la Gironde est utilisé pour calculer la hauteur d'eau et la vitesse de l'écoulement dans l'estuaire et sur les rivières Garonne et Dordogne.

La frontière maritime est située dans le golfe de Gascogne, à 35 km du Verdon, de manière à s'appuyer sur des nœuds du modèle de prévision des surcotes océaniques HYCOM de METEO-FRANCE. Les limites amont sont situées à La Réole sur la Garonne et à Pessac sur la Dordogne. Il faut noter que les apports de l'Isle et de la Dronne sont injectés artificiellement à Pessac. Les champs d'expansion de crue ne sont pas pris en compte. Le modèle numérique couvre environ 125 km d'est en ouest, comporte 12838 éléments finis et est composé de 7351 nœuds (mailles grossières). Des maillages raffinés (27546 nœuds, 106450 nœuds et 418314 nœuds) ont été construits pour une étude de convergence des résultats en termes de hauteurs d'eau du modèle. Cette étude est présentée dans la section 1.2.4.

#### 1.2.2.3 Conditions aux limites du modèle

Les forçages hydrologiques en amont de la Dordogne et de la Garonne sont fournis par la DREAL Nouvelle Aquitaine à un pas de temps horaire. La vitesse du vent de surface et les champs de

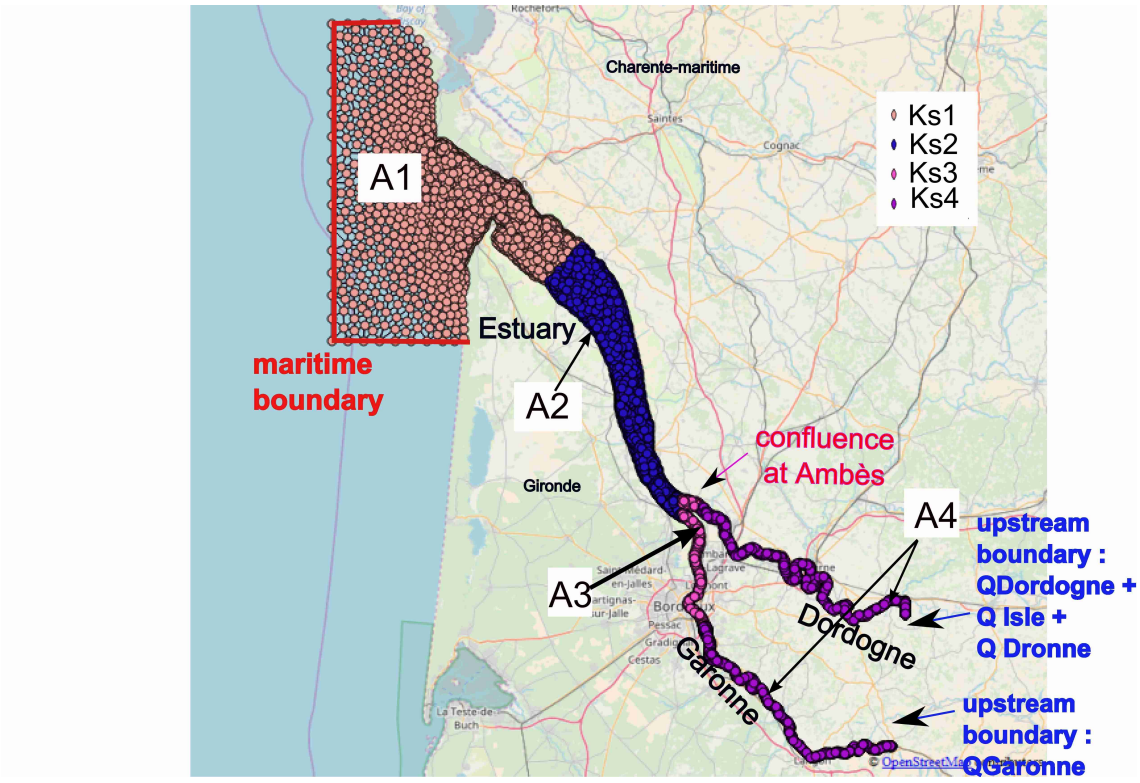



FIGURE 1.7 : Emprise et localisation du modèle numérique de l'estuaire de la Gironde et délimitation des zones $A_1$ à $A_4$ des coefficients de Strickler. Les cercles représentent les nœuds du modèle numérique à partir d'un maillage construit en éléments finis (en noir).

pression du modèle météorologique régional ALADIN sont fournis par METEO-FRANCE à un pas de temps de 3 heures. Les niveaux d'eau à la frontière maritime sont la somme des niveaux prévus des marées astronomiques fournies par le SHOM et des surcotes instantanées fournies par HYCOM. Ces données sont également fournies en temps réel par METEO-FRANCE toutes les 10 à 15 minutes.

#### 1.2.2.4 Version du code de calcul Telemac2D

Afin de tirer profit des dernières évolutions du code de calcul Telemac2D et, en particulier, du développement très récent des API (interface de programmation applicative) qui permettent de lancer des calculs et de modifier les paramètres du calcul par ligne de commande plutôt que par modification du fichier des paramètres, le modèle Gironde a chaque année depuis 2017 été actualisé dans des versions plus récentes du code. De fait, le modèle MEG0V7P3, utilisé dans le cadre de ce travail de thèse, n'est pas exactement le modèle qui tourne sur la chaîne de calcul opérationnelle Synergie de METEO-FRANCE. La prise en compte de la méthode de Thompson à la frontière maritime, basée sur la théorie des caractéristiques appliquées aux frontières liquides,

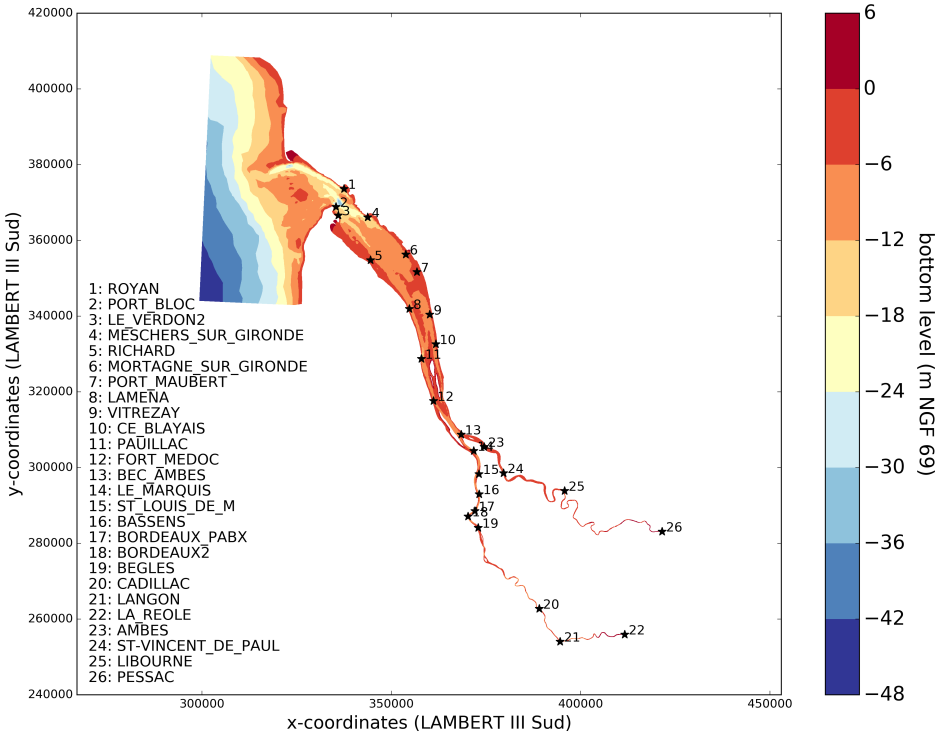



FIGURE 1.8 : Bathymétrie (m NGF69) du modèle numérique de l'estuaire de la Gironde. Les étoiles indiquent les 26 principales stations d'intérêt pour les prévisions des niveaux d'eau.

et des conditions aux limites fluviales par le code Telemac2D a, en particulier, évolué entre les versions V6P0 du modèle Gironde et V7P3 du MEG0V7P3 (Hervouet et al. [2011]).

### 1.2.3 Performances du modèle Gironde actualisé (MEG0V7P3)

Plusieurs événements de calage ont été utilisés lors du projet Gironde. La disponibilité des données d'entrée du modèle pour les tempêtes de Février 2003 et de Décembre 1999 a permis de recaler le nouveau modèle MEG0V7P3 sur le premier événement et d'évaluer ses performances sur le second. Comme pour le projet Gironde (Hissel et al. [2006]), les mesures bathymétriques les plus récentes, ultérieures à 2003 (Huybrechts et al. [2018a]), ont été intégrées dans le modèle. Ceci introduit un artifice dans le modèle MEG0V7P3 dans la mesure où celui-ci est calé sur la tempête de 2003 pour laquelle la bathymétrie prise en compte ne correspond pas forcément à celle de 2003. Il en est de même pour les événements d'évaluation du modèle. En revanche, la bathymétrie actualisée est plus proche de la bathymétrie réelle pour les événements les plus récents.

Le coefficient de frottement est défini sur 4 zones $A_1$ à $A_4$ comme indiqué sur la figure 1.7. Les valeurs correspondantes sont $(Ks\,\{i\})_{(i\in\{1,\cdots,4\})}$. La calibration du modèle a été réalisée à l'aide de l'événement non-débordant de 2003 en ajustant les coefficients de frottement de chaque zone et le coefficient d'influence du vent $Cd$ afin d'optimiser soit l'erreur quadratique moyenne du niveau d'eau (***RMSE***, Eq. (1.2)) soit les critères de Nash de PM (***Nash***$_{HT}$, Eq. (1.1)) calculés entre les niveaux d'eau simulés et observés disponibles aux 12 stations parmi les 26 stations d'intérêt montrées avec des étoiles rouges sur la figure 1.8.

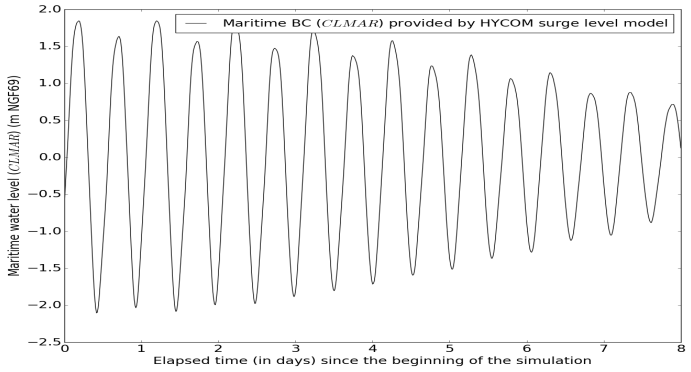
Les débits de la Garonne et de la Dordogne, ainsi que les conditions aux limites maritimes et la surcote instantanée sont représentés pour la période du 02/02/03 à 00:00 au 09/02/03 à 23:00 sur la figure 1.9 pour la tempête de 2003. Pendant cet événement, le débit de la Garonne (resp. Dordogne) atteint 5858 (resp. 2100) $m^3/s$. La surcote instantanée (sans rephasage du signal) est modérée et est comprise entre -10 cm et +20 cm.

Deux ensembles de coefficients de frottement ont été obtenus et sont présentés dans le tableau 1.4. D'une part, les coefficients obtenus pour le nouveau modèle MEG0V7P3 sur chaque zone sont très proches (à 5 $m^{\frac{1}{3}}.s^{-1}$ près) quel que soit le score optimisé (***Nash***$_{HT}$ ou ***RMSE***) et même sur les zones $A_3$ et $A_4$ respectivement situées à la confluence et à l'amont des deux rivières. La différence pour $Ks3$ et $Ks4$ entre le calage sur le ***Nash***$_{HT}$ et celui sur la ***RMSE*** était de 15 $m^{\frac{1}{3}}.s^{-1}$ pour le modèle Gironde. Ce résultat est positif, puisqu'il indique que le modèle actualisé MEG0V7P3, avec un même jeu de paramètres, parvient à reproduire les PM aussi bien que le reste du signal. D'autre part, le quadruplet [$Ks1$=65, $Ks2$=70, $Ks3$=65, $Ks3$=55] qui permet d'optimiser le Nash de PM du modèle MEG0V7P3 est à mi-chemin des deux jeux obtenus pour le modèle Gironde en calant ce dernier sur les Nash de PM ou les RMSE.

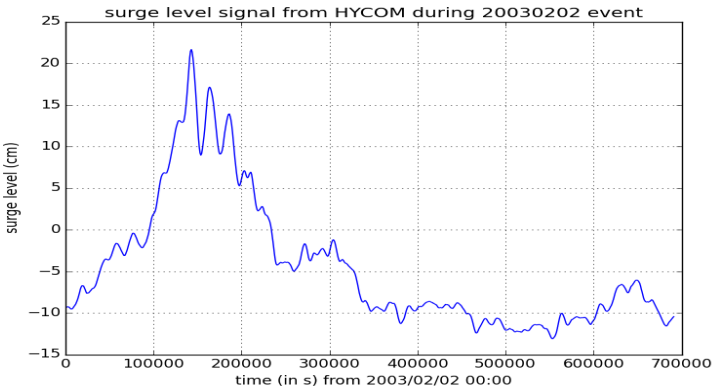
TABLE 1.4 : Coefficients résultant du calage sur les Nash de PM rephasées (***Nash***$_{HT}$) et sur la RMSE pour l'événement de 2003 (non-débordant) obtenus pour le modèle Gironde et le modèle actualisé MEG0V7P3

| Zone | Coefficient de frottement | Calage sur ***Nash***$_{HT}$ | | Calage sur la ***RMSE*** (m) | |
|---|---|---|---|---|---|
| | | Modèle Gironde | MEG0V7P3 | Modèle Gironde | MEG0V7P3 |
| $A_1$ | $Ks1$ | 70 | 65 | 60 | 60 |
| $A_2$ | $Ks2$ | 75 | 70 | 60 | 70 |
| $A_3$ | $Ks3$ | 70 | 65 | 50 | 75 |
| $A_4$ | $Ks4$ | 60 | 55 | 40 | 60 |

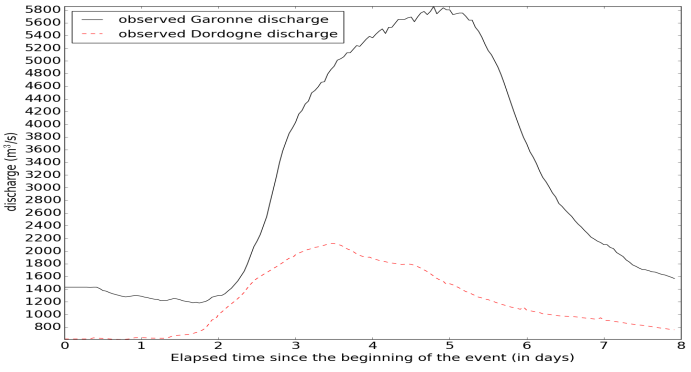
Les scores résultants pour l'événement de 2003 en utilisant le jeu de paramètres issu du calage sur les Nash de PM pour le modèle Gironde et le modèle actualisé MEG0V7P3 sont présentés sous forme de ***RMSE*** et ***Nash***$_{HT}$ dans le tableau 1.5.




(a)




(b)




(c)

FIGURE 1.9 : Évolution temporelle depuis le début de la tempête de 2003 à la condition à la limite : a) maritime : niveau d'eau total $CLMAR$ b) maritime : surcote instantanée ; c) fluviale des débits sur la Garonne ($QGAR$) et la Dordogne ($QDOR$).

TABLE 1.5 : ***Nash*** de PM rephasées (***Nash***$_{HT}$), ***RMSE*** et ***RMSE*** de PM rephasées (***RMSE***$_{HT}$) pour les événements de 2003 (non-débordant) et de 1999 (débordant) le long de l'estuaire de la Gironde obtenus avec le modèle actualisé MEG0V7P3

| Station | ***Nash***$_{HT}$ | | | | ***RMSE*** (m) | | | | $RMSE_{HT}$ (m) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1999 | | 2003 | | 1999 | | 2003 | | 1999 | 2003 |
| | modèle Gironde | modèle MEG0V7P3 | modèle Gironde | modèle MEG0V7P3 | modèle Gironde | modèle MEG0V7P3 | modèle Gironde | modèle MEG0V7P3 | modèle MEG0V7P3 | modèle MEG0V7P3 |
| Verdon | | -0.85 | 0.90 | 0.93 | | 2.0 | 0.17 | 0.16 | 0.44 | 0.12 |
| Richard | 0.53 | 0.45 | 0.72 | 0.8 | 1.51 | 1.78 | 0.26 | 0.25 | 0.24 | 0.21 |
| Laména | 0.77 | 0.89 | 0.86 | 0.88 | 0.80 | 1.49 | 0.23 | 0.36 | 0.11 | 0.18 |
| Pauillac | 0.81 | 0.72 | 0.92 | 0.93 | 0.40 | 0.23 | 0.36 | 0.29 | 0.16 | 0.14 |
| Fort Médoc | 0.83 | 0.72 | 0.92 | 0.93 | 0.44 | 0.27 | 0.40 | 0.31 | 0.16 | 0.14 |
| Bec d'Ambès | 0.89 | 0.75 | 0.95 | 0.95 | 0.40 | 0.17 | 0.29 | 0.18 | 0.16 | 0.13 |
| Le Marquis | 0.84 | 0.53 | 0.95 | 0.95 | 0.44 | 0.28 | 0.35 | 0.24 | 0.22 | 0.12 |
| Bassens | 0.81 | 0.54 | 0.95 | 0.96 | 0.42 | 0.28 | 0.28 | 0.17 | 0.22 | 0.12 |
| Bordeaux | 0.76 | 0.59 | 0.93 | 0.96 | 0.40 | 0.23 | 0.26 | 0.18 | 0.21 | 0.13 |
| Bordeaux2 | | | 0.88 | 0.91 | | - | 0.49 | 0.22 | - | 0.19 |
| La Réole | | | | | 0.46 | 0.51 | 1.24 | 0.93 | | |
| Pessac | | | | | 1.53 | 0.38 | 1.68 | 1.01 | | |
| Mean (without La Réole and Pessac) | 0.67 | 0.48 | 0.90 | 0.92 | 0.67 | 0.75 | 0.29 | 0.24 | 0.21 | 0.15 |

Les scores du modèle actualisé MEG0V7P3 sont meilleurs sur l'événement de 2003 que ceux obtenus par le modèle Gironde, ce qui est cohérent avec le fait que le modèle Gironde a été calé sur 4 événements non-débordants, même si une priorité supérieure a été attribuée à celui de 2003 (Hissel [2007]).

Les scores en termes de Nash de PM pour l'événement de calage sont supérieurs à 0.93 et donc très bons, en particulier sur la partie centrale de l'estuaire de Pauillac à Bordeaux sur la Garonne. Malgré cela, la $RMSE$ sur la partie aval de l'estuaire (resp. à l'amont de la confluence entre les rivières Garonne et Dordogne) est égale à 16 cm au Verdon et supérieure à 25 cm (resp. 18 cm) de Richard à Pauillac (resp. de Bec d'Ambès à Bordeaux).

L'évaluation du modèle MEG0V7P3 a été réalisée à l'aide de l'événement débordant de 1999. Les deux tempêtes de décembre 1999 se sont déroulées entre le 25 et le 31 décembre 1999. Les débits de la Garonne et de la Dordogne atteignent respectivement 3600 et 1830 $m^3/s$ au cours de l'événement. Les surcotes ont atteint près de 1.55 m en sortie d'estuaire au Verdon et 2.25 m à Bordeaux (Salomon [2002]) le 27 décembre 1999. Les scores pour cet événement débordant en termes de $RMSE$ et de Nash de PM sont récapitulés dans le tableau 1.5. La $RMSE$ moyenne est supérieure à 1.50 m à l'embouchure de l'estuaire et à 20 cm de Pauillac à Bordeaux.

Comme prévu, les scores ***RMSE*** et $\boldsymbol{Nash_{HT}}$ sont globalement meilleurs pour l'événement de calage non-débordant de 2003 par rapport à l'événement d'évaluation débordant de 1999. La ***RMSE*** atteint 16 cm au Verdon et 36 cm à Laména en 2003 et 2.0 m au Verdon et 1.49 m à Laména pour la crue de 1999. Ces erreurs restent supérieures à la précision cible de 10 cm attendue par le SPC-GAD.

Par ailleurs, ces résultats sont calculés sur des événements passés en mode réanalyse en utilisant des forçages météorologiques et hydrologiques parfaits. Des erreurs supplémentaires liées à des prévisions imparfaites des forçages et croissant avec l'échéance de prévision sont attendues en mode opérationnel (Hissel [2010]). Les scores obtenus pour les événements de calage et d'évaluation, en particulier en cas d'inondation dans le lit majeur de la Gironde, plaident donc en faveur d'une amélioration du modèle de l'estuaire de la Gironde par l'assimilation des niveaux d'eau observés qui permettrait de "rattraper" la "mauvaise modélisation" sans lit majeur mise en œuvre à l'aide du modèle Gironde actualisé.

### 1.2.4 Convergence en maillage des résultats en niveaux d'eau

Le maillage du modèle MEG0V7P3, qui contient 7351 nœuds et 12838 éléments et est identique à celui du modèle Gironde, peut être qualifié de "grossier", bien qu'il réponde à des contraintes opérationnelles en termes de coût de calcul.

Afin de tester l'influence du maillage sur les résultats en termes de niveaux d'eau calculés par le modèle MEG0V7P3, trois autres maillages ont été construits en divisant chaque élément du MEG0V7P3 en 4 triangles. Les caractéristiques des nouveaux modèles MEG1V7P3, MEG2V7P3

et MEG3V7P3 associés à ces maillages sont récapitulées dans le tableau 1.6.

TABLE 1.6 : Caractéristiques (nombre de nœuds et pas de temps du modèle) des modèles utilisés pour l'étude de l'influence du maillage sur les niveaux d'eau calculés par Telemac2D

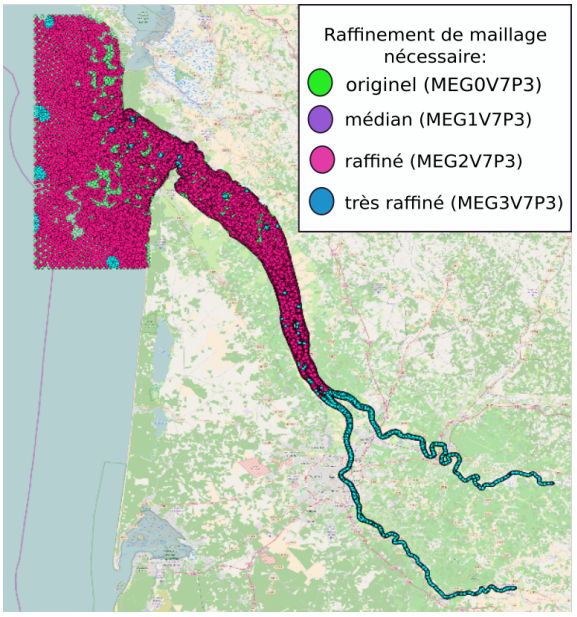
| Modèle | Solution | Nombre de nœuds | pas de temps de calcul de Telemac2D (s) |
|---|---|---|---|
| MEG0V7P3 | Originale | 7351 | 10.0 |
| MEG1V7P3 | Medium | 27546 | 10.0 |
| MEG2V7P3 | Raffinée | 106450 | 10.0 |
| MEG3V7P3 | Très Raffinée | 418314 | 5.0 |

Par hypothèse, le maillage MEG3V7P3 est la solution raffinée de référence à laquelle les résultats en termes de RMSE sur la période du pour l'événement de calage de 2003 pour chacun des modèles $(MEG\{i\}V7P3)_(i \in \{0, \cdots, 2\})$ sont comparés dans le tableau 1.7. Sur les stations considérées, la différence avec la RMSE du modèle le plus raffiné (MEG3V7P3) est inférieure à 5 cm sur la zone de la confluence, de Bec d'Ambès à Bassens pour le maillage médian (MEG1V7P3). En revanche, sur les parties fluviales, aucun modèle n'obtient une RMSE dont la différence avec celle du maillage le plus raffiné soit inférieure à 10 cm. Le maillage le plus raffiné est donc nécessaire sur les zones en amont de la confluence.

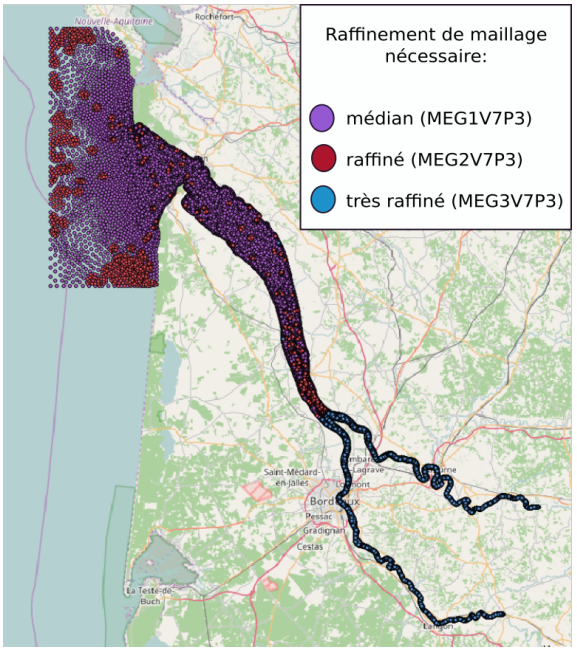
La figure 1.10 permet d'affiner ces résultats. La convergence de précision 5 cm (resp. 10 cm) est considérée comme atteinte lorsque la différence entre la RMSE calculée sur la durée de l'événement au nœud du maillage pour un modèle donné et celle calculée à l'aide du modèle dont le maillage est le plus raffiné (MEG3V7P3) est inférieure à 5 cm (resp. 10 cm).
La figure 1.10-a montre que, pour une convergence à 5 cm, il est nécessaire d'utiliser, d'une part, le maillage raffiné (MEG2V7P3) depuis la frontière maritime jusqu'à l'aval de la confluence entre les rivières Garonne et Dordogne et, d'autre part, le maillage de référence le plus raffiné (MEG3V7P3) sur chacune des branches fluviales (densité de 1.25 m tout à l'amont).
Pour une convergence à 10 cm, il est possible d'utiliser le raffinement médian (MEG1V7P3) jusqu'au droit du CNPE du Blayais, bien que certaines zones au large nécessitent le maillage raffiné (MEG2V7P3). A partir de la confluence, comme pour une convergence à 5 cm, il est nécessaire que le maillage soit le plus raffiné possible (MEG3V7P3). Finalement, pour la partie aval de l'estuaire jusqu'à la confluence entre les deux rivières Dordogne et Garonne, le maillage idéal contient environ 35000 nœuds et, pour les parties fluviales, environ 166000 nœuds, soit un total de 201000 nœuds prohibitif en conditions opérationnelles.

TABLE 1.7 : RMSE obtenu par le modèle le plus raffiné (MEG3V7P3) pour la période du 02 au 05/02/2003 et comparaison avec les modèles original (MEG0V7P3) et à raffinement médian (MEG1V7P3) (resp. (MEG2V7P3)) pour chaque station d'observation de l'estuaire de la Gironde

| station | RMSE du | Gain du MEG3V7P3 par rapport au | | |
|---|---|---|---|---|
| | MEG3V7P3 | MEG0V7P3 | MEG1V7P3 | MEG2V7P3 |
| Verdon | 0.15 | -0.02 | -0.01 | -0.01 |
| Richard | 0.20 | -0.01 | 0.00 | 0.00 |
| Laména | 0.24 | 0.02 | 0.01 | 0.01 |
| Pauillac | 0.20 | **0.05** | 0.02 | 0.01 |
| Fort Médoc | 0.23 | | 0.03 | 0.02 |
| Bec d'Ambès | 0.15 | **0.05** | 0.01 | 0.00 |
| Le Marquis | 0.18 | **0.11** | 0.04 | 0.01 |
| Bassens | 0.19 | **0.06** | 0.01 | 0.00 |
| Bordeaux | 0.23 | 0.02 | 0.00 | 0.01 |
| Bordeaux2 | 0.54 | -0.03 | -0.06 | 0.02 |
| La Réole | 0.41 | **1.04** | **0.40** | **0.13** |
| Pessac | 1.38 | **0.55** | **0.03** | **0.15** |




(a)




(b)

FIGURE 1.10 : Raffinement du maillage nécessaire pour une convergence sur les niveaux d'eau avec une précision de : a) 5 cm ; b) 10 cm

## 1.3 Conclusions

Le modèle de l'estuaire de la Gironde issu du projet Gironde a été limité au lit mineur, excluant ainsi la prise en compte des champs d'expansion de crue en raison de contraintes informatiques dans un contexte opérationnel. Malgré cette limitation, les travaux réalisés dans le cadre du projet Gironde ont permis d'améliorer la prévision des niveaux d'eau le long de l'estuaire et d'augmenter les échéances de prévision.

Bien que le modèle actualisé fournisse de bons résultats pour les événements passés en mode réanalyse, avec des erreurs inférieures à 10 cm pour les scénarios sans débordement, la simulation des périodes de marées hautes est plus difficile et, dans les situations de débordement, des erreurs de l'ordre de 30 cm existent près de Bordeaux. Le projet Gironde (Hissel [2010]) a recommandé des améliorations à apporter. D'une part, la prise en compte des zones de débordement peut être envisagée. D'autre part, la mise à jour de l'état et des paramètres du modèle avec des algorithmes d'assimilation de données est préconisée. La mise en œuvre de ces algorithmes, qui fait l'objet de ce travail de thèse, nécessite que les principales sources d'incertitude parmi les nombreuses données numériques incertaines (paramètres et forçages) soient identifiées, quantifiées et réduites.

Le chapitre 2 suivant présente l'étude par analyse de sensibilité (SA) dans le contexte de la prévision des crues sur l'estuaire de la Gironde. Il vise à identifier et à classer les principales sources d'incertitude qui limitent la capacité de prévision des niveaux d'eau et des débits des frontières fluviales en amont jusqu'à la côte atlantique en aval, en se concentrant sur les sites d'intérêt où la sécurité et les enjeux économiques sont les plus sensibles (Weerts et al. [2011]).

# Chapitre 2

# Quantification d'incertitudes avec variables fonctionnelles : application au modèle Telemac2D de prévision des hautes eaux de l'estuaire de la Gironde

**Sommaire**


**2.1 Résumé . . . . . 40**

**2.2 Quantifying forcing uncertainties in the hydrodynamics of the Gironde estuary . . . . . 41**

2.2.1 Article accepté dans la revue "Journal of Computational Geosciences" 41

2.2.2 Information brought during the reviewing process . . . . . 69

2.2.2.1 Ranges of the friction and the wind drag coefficients (reported in Table 1) . . . . . 69

2.2.2.2 Members convergence . . . . . 70

2.2.2.3 Limitations of the study . . . . . 70

2.2.2.4 Rank and fiability diagram for all stations . . . . . 70

**2.3 Conclusions et perspectives . . . . . 71**

## 2.1 Résumé

De forts coefficients de marée, combinés à des surcotes météorologiques et des débits élevés sur les rivières Garonne et Dordogne, peuvent provoquer des crues et des inondations près de la centrale nucléaire du Blayais et de la ville de Bordeaux, avec des impacts économiques et sociaux importants.

Une analyse de sensibilité globale (GSA) a été réalisée à l'aide du modèle numérique actualisé décrit dans le chapitre 1 fondé sur Telemac2D actuellement utilisé pour les prévisions opérationnelles des niveaux d'eau. Dans le cadre de cette étude, dont la méthodologie et les résultats sont présentés dans la section 2.2.1 , les principales sources d'incertitude ont été identifiées en calculant les indices de Sobol' pour les données incertaines à l'aide d'une approche par décomposition de variance (ANOVA) pour l'événement qui s'est déroulé du 02/02 au 09/02/2003, soit 7 jours physiques.

La génération de l'ensemble pour les simulations de la GSA consiste à échantillonner des variables aléatoires scalaires et fonctionnelles (dépendantes du temps) : les coefficients de frottement constants et uniformes par zone, ainsi que les forçages hydrologiques et maritimes variables dans le temps. Par hypothèse, la perturbation temporelle des forçages hydrologiques en amont et maritime à l'aval de l'estuaire est représentée par un processus gaussien caractérisé par une durée d'auto-corrélation calculée à partir des observations. Une décomposition de Karhunen-Loève est ensuite appliquée pour déterminer le nombre de modes propres à retenir pour représenter une part de variance suffisante du forçage.

La GSA, qui relève du calcul haute performance (HPC) pour ce qui concerne le parallélisme des tâches et la décomposition des domaines de calcul de Telemac2D, est réalisée pour 20 variables aléatoires à l'aide des ressources de calcul du GENCI (allocations 2017-A0030110292 et 2018-A0050110292). La GSA a nécessité 250 000 runs pour un temps de calcul de 101 jours sur 32768 cœurs.

La fiabilité de l'ensemble des niveaux d'eau en sortie du modèle a été évaluée à l'aide de diagrammes de rang et de courbes de fiabilité par rapport aux niveaux d'eau mesurés en 12 stations d'observation le long de l'estuaire pendant la durée de l'événement.

Cette étude de quantification d'incertitudes montre que, pour l'événement de 2003, les conditions à la frontière maritime et les coefficients de Strickler ont un rôle prédominant le long de l'estuaire et que l'évolution temporelle de leur influence dépend du cycle de marée. Dans les zones fluviales, en amont du modèle, le coefficient de frottement et les apports hydrologiques sont prédominants.

Ces analyses sont confirmées à la fois pour le maillage le plus grossier (MEG0V7P3) et pour le maillage le plus raffiné (MEG4V7P3). En outre, une étude de convergence en nombre de membres $N_e$ a été réalisée à l'aide du MEG4V7P3 en faisant varier $N_e$ dans l'intervalle $[100\,;\,10000]$ et a permis de mettre en lumière que les indices de sensibilité convergaient à partir de $N_e = 2000$. De fait, il a été montré que la GSA est convergente par rapport au maillage et par rapport au

nombre de membres $N_e$.

## 2.2 Quantifying forcing uncertainties in the hydrodynamics of the Gironde estuary

### 2.2.1 Article accepté dans la revue "Journal of Computational Geosciences"

Cette section est composée de l'article accepté pour publication dans le Journal of Computational Geosciences (DOI : 10.1007/s10596-019-09907-7) (consultable dans sa version finale sous : https://rdcu.be/b18JG).

**Computational Geosciences manuscript No.**
(will be inserted by the editor)

# Quantifying forcing uncertainties in the hydrodynamics of the Gironde estuary

**Vanessya LABORIE · Sophie RICCI · Matthias DE LOZZO · Nicole GOUTAL · Yoann AUDOUIN · Philippe SERGENT**

Received: 2018 07 07 / Accepted: 2019 10 14

**Abstract** High tide combined with high meteorological surge levels and discharges in the Garonne and Dordogne rivers in the Gironde estuary (south-west France), may lead to high water levels and flooding near the Blayais nuclear power plant and the city of Bordeaux, with significant economic and social impacts. A global sensitivity analysis (GSA) was performed with a Telemac2D numerical model currently used for operational water level forecasts. The major sources of uncertainties were identified by computing the Sobol' indices for uncertain inputs with an analysis of variance (ANOVA) approach for a 7-day storm event in 2003. The generation of the GSA ensemble of simulations consists of sampling scalar and field random variables: constant and uniform friction coefficients, as well as time-varying hydrological and maritime forcings. The temporal perturbation of time-dependent upstream hydrological and downstream maritime forcings is assumed to be represented by a Gaussian Process characterized by a correlation time scale calculated from observations. A Karhunen-Loève decomposition was then applied to retain a limited number of eigenmodes. The GSA is performed for 20 random variables using GENCI HPC computational resources for task parallelism and domain decomposition. This requires the use of 250 000 runs for an elapsed simulation time of 101 days on 32768 cores. The performance of the ensemble was assessed with a rank diagram and a reliability curve in comparison to a set of measured water levels at 12 observing stations along the estuary. It was shown that, for this event, the maritime boundary conditions and the Strickler coefficients have a predominant role along the length of the estuary with an influence driven by the tidal cycle. In the upstream fluvial areas, the friction coefficient and hydrological inputs are predominant.

Vanessya LABORIE
Saint-Venant Hydraulics laboratory, Cerema, 134, route de Beauvais, CS 60039, 60280 Margny Lès Compiègne, France
Tel.: +33-3-44416508
E-mail: vanessya.laborie@cerema.fr

Sophie RICCI
Ceci, Cerfacs/Cnrs, 42 avenue Gaspard Coriolis, 31057 Toulouse Cedex 01, France

Matthias DE LOZZO
Ceci, Cerfacs/Cnrs, 42 avenue Gaspard Coriolis, 31057 Toulouse Cedex 01, France

Nicole GOUTAL
Saint-Venant Hydraulics laboratory, Edf R&D, LNHE, 6 quai Watier, 78401 Chatou, France

Yoann AUDOUIN
Edf R&D, LNHE, 6 quai Watier, 78401 Chatou, France

Philippe SERGENT
Cerema, 134, route de Beauvais, CS 60039, 60280 Margny Lès Compiègne, France

**Keywords** uncertainty quantification · global sensitivity analysis · time-dependent forcings · Karhunen-Loève decomposition · Sobol' indices · Gironde estuary

**Acknowledgements** We would like to thank the service in charge of flood forecasting in the Garonne, Adour and Dordogne watersheds (SPC GAD), as well as METEO-FRANCE and the Great Maritime Port Councils of Bordeaux (GPMB) for the bathymetric and observational data they provided for this study. The sea level observations along the Gironde estuary are the property of GPMB and the French Ministry in charge of sustainable development (MTES). This work was granted access to the HPC resources of IDRIS under the allocation 2017-A0030110292. The authors sincerely thank Sylvie THEROND and Isabelle DUPAYS at IDRIS, for technical support, as well as Marissa YATES-MICHELIN for helping out with the proofreading of this article.

## 1 Introduction

Environmental, economical and security issues are at stake in the Gironde estuary catchment located in South-West France near the city of Bordeaux and the nuclear power plant of Blayais. The Dordogne and Garonne rivers meet at Bec d'Ambès and the estuary reaches the Atlantic Ocean coastline about 75 km further downstream. The estuary is subject to maritime influences and the combination of strong tidal amplitudes with high fluvial inflows can lead to strong flood events [1]. Governmental agencies are responsible for the safety of people and property. They rely on Decision Support Systems to take preventive measures to alert communities and to coordinate crisis management. Since the most severe flood event on record in the Gironde estuary in 1770, infrastructures were built to limit the consequences of flooding, yet these failed to protect fully the area during strong events such as Lothar and Martin in 1999 and Xynthia in 2010. The overflowing of dikes protecting the Blayais nuclear power plant in 1999 demonstrated the strong need for water level forecasts in the estuary. In France, the SCHAPI (Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations) and Flood Forecast Services (FFS) are in charge of monitoring and forecasting water levels and discharge over 22 000 km of rivers. They produce a twice-daily vigilance colored-risk map available online for governmental authorities and the general public (http://www.vigicrues.gouv.fr). To create these risk maps, they rely on numerical models and in-situ measurements [2].

Hydrodynamic numerical software packages based on the Shallow-Water Equations (SWE) are commonly used tools to aid in the management and protection of urban infrastructures located near rivers and coasts. They are also used for operational flood forecasting. Yet, these numerical codes remain imperfect as uncertainties in the models and in the inputs (model parameters, boundary conditions, geometry, etc.) propagate into uncertainties in the outputs (water levels, discharge). Quantifying physical uncertainties goes beyond the limits of deterministic forecasting and represents a significant challenge. The FFS Garonne-Adour-Dordogne (GAD) is in charge of the Gironde estuary area. In order to meet operational expectations, especially for extreme events, FFS GAD moved from a statistical model based on climatological data to a numerical model that solves the SWE based on the hydraulic software Telemac2D [3]. The Gironde estuary model was limited to the non-overflowing area, excluding the floodplains owing to computational constraints in an operational context. Despite this limitation, it resulted in improving water level forecast skill and increasing alert lead-times. While this model provides good results for past events in reanalysis mode, with errors of less than 10 cm for non-overflowing scenarios, simulating high tides periods is more challenging and in overflowing situations, errors of the order of 30 cm remain near Bordeaux. The 'Gironde project' [4] recommended areas for improvement including updating the model state and parameters with data assimilation algorithms. To do so, the major sources of uncertainty amongst the numerous uncertain numerical data, input data and forcing data should be identified, quantified and reduced.

This paper presents a Sensitivity Analysis (SA) study in the context of flood forecasting in the Gironde estuary. It aims at identifying and classifying the major sources of uncertainties that limit the predictive capability of water levels and discharge from the fluvial upstream boundaries to the downstream Atlantic coastline, focusing on locations of interest where safety and economical assets are at stake [5].

A wide range of SA methods are proposed in the literature [6] to estimate the contribution of uncertain model parameters and inputs to the uncertainty in the model Quantities of Interest (QoI). On the one hand, local SA approaches provide the sensitivity of the model outputs with respect to the model inputs around a reference value using the tangent linear of the model (when available) or finite differences techniques. On the other hand, Global SA approaches (GSA) provide the contribution to the QoI's uncertainty from the uncertain input parameter when varying over the whole input parameter space. The ANOVA (ANalysis Of VAriance) method consists in estimating the QoI variance decomposition in terms of elementary variances associated with the different parameters and their interactions [7]. This decomposition is obtained from a Proper Orthogonal Decomposition of the uncertain QoI over the probabilized parameter space [8]. Sensitivity indices, called Sobol' indices, represent the contribution of each parameter and their interactions to the model output variance. The ANOVA approach is suited even for non linear and non monotonic models ([9], [10]) provided that the uncertain parameters are uncorrelated and independent [11].

GSA requires the integration of a large number of simulations with a direct model. In spite of increasing High Performance Computing (HPC) resources, the ensemble's computational cost is still incompatible with operational or industrial applications. Monte Carlo random sampling techniques are often used as they are generic, robust, and easily portable on massively parallel architectures. Yet, they remain computationally expensive due to their slow convergence rate, which scales

as the inverse of the square root of the number of members [12]. Using a Sobol' sequence space-filling strategy is prefered to compute first and total Sobol' sensitivity indices [10] and to assess the contribution of each uncertain variable and their interactions to the total variability of the system. When uncertainties are associated with field inputs, for instance discretized over time and/or space, the cost of Monte Carlo-based GSA becomes untrackable and advanced solutions should be adopted. One approach consists in replacing the direct solver with a meta (or surrogate) model built from a limited number of integrations of the direct solver ([13], [14], [15], [16], [17]). The quality of the surrogate depends on the complexity of the physics, for instance the non-linearity between inputs and outputs of the model, the size of the learning sample, and the surrogate strategy. Various metamodeling algorithms are proposed in the literature. [18] gives an overview of metamodeling approaches adapted to uncertainty propagation and GSA. This approach is beyond the scope of the present study. When the dimension of the uncertain space is large, the GSA is carried out in a reduced space with a Proper Orthogonal Decomposition [19]. The Proper Orthogonal Decomposition (also refereed to as Principal Component Analysis or Karhunen-Loève (KL) decomposition) is a procedure for extracting a basis of a modal decomposition from an ensemble of signals [20]. Its power lies in the mathematical properties that suggest that it is the appropriate basis, as it is a linear procedure and makes no assumptions on the linearity of the problem to which it applies. The KL decomposition minimizes in the mean squared sense the representation error.

GSA is often used to classify the sources of uncertainties for large scale hydrology as well as for hydrodynamics at the scale of French rivers, assuming the uncertainties originate from scalar inputs. A multivariate GSA based on the ANOVA technique was applied in the Amazon River basin by [21] to highlight the major sources of errors in the river water level and discharge simulated by the river-routing scheme Total Runoff Integrating Pathways (TRIP). By assuming a particular input uncertainty distribution, it was shown that geomorphological parameters explain most of the water level variance with purely additive contributions from the river Manning coefficients, riverbed slope and river width. [22] presented a Sobol' SA for flow simulation by a SWAT (Soil And Water Assessment Tool) model of river, for a complex environmental system controlled by a large number of parameters (about 30). It was shown that even with a limited number of direct solver evaluations, the GSA identified the most significant parameters and improved the understanding of the model behavior. [23] presents an overview of SA studies for the Garonne river in steady flow conditions with both local SA and GSA. The details for the local SA are given in [24]. The 1.5D hydraulic model SIC (Irstea) is used within a variational framework implying its tangent linear and adjoint codes to acknowledge the impact of the geometry, friction coefficient and upstream discharge on the water level and discharge discretized along the 50-km river reach. The GSA was carried out with both 1D and 2D solvers Mascaret-Telemac (Electricité de France) with a Monte Carlo approach as well as with a polynomial surrogate model. When uncertainties stem from field data, such as meteorological (wind, pressure, temperature) fields or river discharges for example, the cost of Monte Carlo-based GSA drastically increases, and the computation of the surrogate becomes challenging. [25] presents solutions to perform GSA with field inputs and outputs to the numerical code NOE which is a spatial model for cost-benefit analysis of flood risk management plans and which is used to assess, through simplified equations, the economic impact of flood risk [26]. [27] uses empirical orthogonal functions and analyzes the combined impact of uncertainties in initial conditions and wind forcing fields in ocean general circulation models using polynomial chaos (PC) expansions.

In the context of the Gironde estuary study, the uncertainties depend on the time-varying hydrological upstream forcing and on maritime boundary conditions, the space and time-varying bathymetry, and wind and pressure forcing, the space-varying bottom friction coefficients (4 scalars defined by uniform areas), and the wind drag coefficient (scalar). Only uncertainties that relate to friction and wind drag coefficients, hydrological and maritime forcing are taken into account here. The SA with respect to surface forcing is beyond the scope of the present study. A KL expansion is used to reduce the size of the time-dependent uncertain inputs, assuming that each perturbation of the observed forcing is a gaussian process characterized by a correlation function and a correlation time scale. The originality of the present study is that a GSA is performed on unsteady flows simulations using the Telemac2D solver, while accounting for the maritime influence of the Gironde estuary. Beyond the central objective of the GSA, the scope of the study is to build a reference case which will enable future work with data assimilation techniques, but also on metamodels or multi-fidelity simulations applied to the Gironde estuary site. Moreover, the hydrodynamics of the Gironde estuary results from varying power balancing between the space- and/or time-dependent friction, geometry and both upstream and downstream boundary conditions, which complicate the

immediate and intuitive interpretation of the complex hydrodynamics of the estuary. Quantifying uncertainties both in space and time, and identifying the most influential variables can thus help to understand the dominant physical processes in the estuary.
The structure of the paper is as follows. Section 2 presents the materials and methods for the GSA. Section 2.1 presents the Gironde estuary, SWE and the hydrodynamic model implemented with Telemac2D. Section 2.2 describes the computation of sensitivity indices. The ensemble generation and sensitivity analysis with field inputs techniques are described in Sect. 2.3. Details on the HPC resources and performance of the ensemble are also detailed. Statistics drawn from the ensemble are presented in Sect. 3: the water level probability density function is shown at specific locations along the estuary and the Sobol' indices are plotted as a function of time and space. The focuses on the sensitivity to the maritime boundary forcing eigenmodes, conclusions and perspectives for the study are finally presented.

## 2 Framework

### 2.1 Hydrodynamic model of the Gironde estuary

#### *2.1.1 Presentation of the Gironde estuary*

The Gironde estuary is the largest macrotidal estuary in France and Western Europe. The Gironde estuary extends from the Bay of Biscay approximately 170 km inland and covers a surface of 635 km$^2$. It is located in southwest France. Created from the confluence of two rivers (Garonne and Dordogne rivers), it extends 75 km to reach the mouth of the estuary at the Atlantic coastline (Fig. 1). On average, it is oriented from south-east to north-west in a valley, and its width ranges from 1 km near Bordeaux to 15 km at the coast. The Gironde estuary can be divided [28] into three subdomains: the upstream river area, the central estuary area and the downstream offshore area [29]. The estuary can be classified as macrotidal, hypersynchronous, and with an asymmetric tide (4 h for the flood and 8 h 25 min for the ebb). The tide in the Bay of Biscay is semidiurnal, with a period of 12 h 25 min and is dominated by the principal semidiurnal lunar (M2) component. The open ocean induces a strong tidal forcing with a tidal amplitude near the mouth of the estuary ranging from 2.2 to 5.4 m over the spring-neap cycle. The Garonne (resp. Dordogne) river discharge typically ranges from 50 (resp. 20) up to 2 000 (resp. 1 000) m$^3$.s$^{-1}$. During flood events, the Garonne flow rate occasionally exceeds 5 000 m$^3$.s$^{-1}$. The economic importance of the Gironde estuary is evidenced by the presence of large cities and ports, including the city of Bordeaux and the Harbor of Bordeaux, as well as by the presence of various industries, such as the Blayais nuclear power plant, fisheries and tourism activities. The flooding risk has been a major concern for authorities along the Gironde estuary [1] for a long time, as shown by historical documents dating from the XIII$^{\text{th}}$ century. The most damaging flood occurred in April 1770, when about 24 000 km$^2$ were covered by water along the Garonne river and the Gironde estuary, causing about 4,669 million euros in damage exclusively in the city of Bordeaux (35 million euros using purchasing power parity for primary goods). From this point, flood protection structures were built to limit the consequences of flooding. However, this did not prevent strong floods in 1835, 1855 and 1856. In 1930, floods caused the destruction of 1 000 houses and more than 300 casualties. In recent decades, three notable events were observed: the December 1981 event, which was caused by strong river discharges during high tides, and the Lothar and Martin storms in 1999 and Xynthia in 2010 [30].

#### *2.1.2 Shallow water equations in Telemac2D*

The SWE are commonly used in environmental hydrodynamics modelling. They are derived from the Navier-Stokes equations and express mass and momentum conservation averaged in the vertical dimension. The non-conservative form of the equations are written in terms of the water depth (h [m]) and the horizontal components of velocity (u and v [m.s$^{-1}$]).

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(hu) + \frac{\partial}{\partial y}(hv) = 0 \tag{1}$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = -g\frac{\partial H}{\partial x} + F_x + \frac{1}{h} div\left(h\nu_e \overrightarrow{grad}(u)\right) \tag{2}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = -g\frac{\partial H}{\partial y} + F_y + \frac{1}{h} div\left(h\nu_e \overrightarrow{grad}(v)\right) \tag{3}$$

where:

$$\begin{aligned} F_x = &- \frac{g}{{K_s}^2}\frac{u\sqrt{u^2+v^2}}{h^{\frac{4}{3}}} - \frac{1}{\rho_w}\frac{\partial P_{atm}}{\partial x} \\ &+ \frac{1}{h}\frac{\rho_{air}}{\rho_w}\ C_d\ U_{w,x}\sqrt{U_{w,x}^2 + U_{w,y}^2} \end{aligned} \tag{4}$$

$$F_y = -\frac{g}{{K_s}^2}\frac{v\sqrt{u^2+v^2}}{h^{\frac{4}{3}}} - \frac{1}{\rho_w}\frac{\partial P_{atm}}{\partial y} + \frac{1}{h}\frac{\rho_{air}}{\rho_w}\ C_d\ U_{w,y}\sqrt{U_{w,x}^2+U_{w,y}^2} \quad (5)$$

and $\rho_{air}/\rho_w$ [kg.m$^{-3}$] are the air/water density, $P_{atm}$ [Pa] is the atmospheric pressure, $U_{w,x}$ and $U_{w,y}$ [m.s$^{-1}$] are the horizontal wind velocity components, $C_d$ [-] is the wind drag coefficient that relates the free surface wind to the shear stress, $K_s$ [m$^{\frac{1}{3}}$.s$^{-1}$] is the river bed and floodplain friction coefficient, using the Strickler formulation [31]. $F_x$ and $F_y$ [m.s$^{-2}$] are the horizontal components of external forces (friction, wind and atmospheric forces), $H$ [m NGF69] is the water level ($h = H - z_f$ if $z_f$ [m NGF69] is the bottom level) and $\nu_e$ [m$^2$.s$^{-1}$] is the water diffusion coefficient. *div* and $\overrightarrow{grad}$ are respectively the divergence and gradient operators. To solve the system of equations Eq. (1) to Eq. (3), initial conditions $h(x,y,t=0) = h_0(x,y)$; $u(x,y,t=0) = u_0(x,y)$; $v(x,y,t=0) = v_0(x,y)$ are provided. Boundary conditions (BC) both at the coastline (slip and impermeability conditions) and at the upstream and downstream boundaries ($h(x_{BC}, y_{BC}, t) = h_{BC}(t)$) are also given.
In the present study, the SWE are solved with the parallel numerical solver Telemac2D (www.opentelemac.org) with an explicit first-order time integration scheme, a finite element scheme and an iterative conjugate gradient method [3]. In the following, we will take into account parametric uncertainties, that are due to the stochastic nature of the atmosphere-surface system. These include the forcing fields and epistemic uncertainties that come from a lack of knowledge concerning the physical processes of the hydrodynamic system, leading to a simplified parametrization, such as friction or turbulence, but also the imperfect description of the system, such as the geometry of the river. These uncertainties can be represented by independent scalars (friction coefficients) or time and/or space correlated discretized fields (fluvial, atmospheric and maritime forcings).

### *2.1.3 The Gironde estuary numerical model*

A hydrodynamic numerical model of the Gironde estuary (Fig. 1), based on Telemac2D and on a bathymetry / topography field (Fig. 2), is used operationally to compute the water depth and velocity in the estuary and along the Garonne and Dordogne rivers. The maritime boundary is located in the Gascogne Gulf, 35 km away from le Verdon. The upstream boundaries are located at La Réole on the Garonne River and at Pessac on the Dordogne River. It should be noted that inflows from the Isle and Dronne rivers are artificially injected at Pessac [4] and that floodplains are not taken into account. The operational numerical model used by FFS GAD covers about 125 km from east to west, features 12838 finite elements and is composed of 7351 nodes (coarse mesh). Refined meshes (27546 nodes, 106450 nodes and 418314 nodes) were built for a convergence study. It was shown that mesh convergence is obtained for water levels (using a 5 cm error threshold) in the entire estuary except in the fluvial areas for meshes with more than 106450 nodes. The 418314 node fine mesh was thus used in the following for SA. Hydrological upstream forcings for the Dordogne and Garonne rivers are provided by the DREAL (Direction Régionale de l'Environnement, de l'Aménagement des Territoires et du Logement) Nouvelle Aquitaine at a 1-hour time step. Surface wind velocity and pressure fields from the regional meteorological model ALADIN [32] are provided by Meteo-France at a 3-hour time step. Water levels at the maritime boundary are the sum of the predicted astronomical tide and surge levels; these data are also provided in real-time by Meteo-France every 10 to 15 min. The friction coefficient is uniformly defined over 4 areas as shown in Fig. 1. The model calibration was achieved during the non-flooding 2003 event to optimize either the water level Root Mean Square Error (*RMSE*, Eq. (6)) or the Nash criteria at high tide ($Nash_{HT}$, Eq. (7)) computed between simulated and observed water levels available at the 12 stations among the 26 stations of interest shown with red stars in Fig. 2. Two sets of friction coefficients were obtained from the calibration and are presented in Table 1. The resulting *RMSE* and $Nash_{HT}$ scores for the 2003 event are presented in Table 2 along with the evaluation scores for the 1999 flooding event. The $Nash_{HT}$ (Eq. (7)) is evaluated at high tide and the *RMSE* (Eq. (6)) (resp. $RMSE_{HT}$) is evaluated by summing over the entire flood event (resp. at high tide):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(H_i - \widehat{H_i})^2} \quad (6)$$

$$Nash_{HT} = 1 - \frac{\sum_{i=1}^{n}(H_i - \widehat{H_i})^2}{\sum_{i=1}^{n}(\widehat{H_i} - \overline{H})^2}, \quad (7)$$

where $n$ is the time index, $H_i$ and $\widehat{H_i}$ are the simulated and observed water levels and $\overline{H}$ the time-averaged observed water levels. As expected, the *RMSE* and $Nash_{HT}$ scores are significantly better for the 2003 non-flooding calibration event than for the 1999 flooding evaluation event. The water level *RMSE* reaches 16 cm at Le Verdon and 36 cm at Laména in 2003 and 2.0 m

at Le Verdon and 1.49 m at Laména for the overflowing event in 1999. These errors remain higher than the target 10 cm precision expected by FFS GAD. These results are computed over past events in reanalysis mode using perfect meteorological and hydrological forcing. Additional errors due to imperfect forecasting are expected in operational mode, especially with increasing lead time [4]. The scores obtained for calibration and evaluation events, especially in flooding conditions, advocate further improvement of the Gironde model with the assimilation of observed water levels.

The wind drag coefficient formulates the wind shear stress at the free surface from the wind velocity [33]. A uniform and constant value was chosen here ($C_d = 2.14\ 10^{-3}$) consistently with the calibration of the surge level numerical model for the Atlantic Ocean, English Channel and North Sea [34].

## 2.2 Computation of sensitivity indices

### *2.2.1 Variance decomposition and Sobol' indices*

Sobol' indices [35] measure the contributions of the different independent inputs $X_1, X_2, \ldots, X_d$ and their interactions in the variance V(Y) of the output $Y = f(X)$ with $X = (X_1, X_2, \ldots, X_d)$, $E((f(X))^2) < \infty$ and $f$ the model. $S_i = \frac{V_i}{V(Y)}$ is the first-order Sobol' index of $X_i$, representing the normalized elementary contribution of $X_i$ to $V(Y)$. $S_{i,j} = \frac{V_{i,j}}{V(Y)}$ is the second-order Sobol' index of $X_i$ and $X_j$, representing the normalized contribution of the interaction between $X_i$ and $X_j$ to $V(Y)$, and so on. As described in Appendix A, using Eq. (8):

$$1 = \sum_{i \subseteq I_d} S_i + \sum_{\substack{\{i,j\} \subseteq I_d^2 \\ j>i}} S_{i,j} + \ldots + S_{1,2,\ldots,d} = \sum_{u \subseteq I_d} S_u \quad (8)$$

where: $I_d = \{1, \ldots, d\}$ is the set of input indices, the total Sobol' index $S_{T_i}$ gathering all contributions related to $X_i$ is then defined as:

$$S_{Ti} = S_i + \sum_{\substack{j \in I_d \\ j>i}} S_{i,j} + \ldots + S_{1,2,\ldots,d} = \sum_{\substack{u \subseteq I_d \\ u \ni i}} S_u \quad (9)$$

It should be noted that $\sum_i S_i = 1$ if there is no interaction between the input parameters.

### *2.2.2 Implementation of Sobol' indices computation*

The main steps for the stochastic estimation of the Sobol' indices of the different independent inputs $X_1, X_2, \ldots, X_d$ with $X = (X_1, X_2, \ldots, X_d)$ are described in Appendix B according to [10], which the reader can refer to for more details.

It should be noted that the calculation of the first and total Sobol' indices for an ensemble of size $N$ requires the integration of two independent samples. Given $d$ uncertain variables and $N_e$ perturbed members for each variable, the total number of simulations is thus $N = N_e(d+2)$.

If $(X_i)_{i=k,,k+m}$ correspond to the $m$ uncertain modes of a field input, the contribution of each mode can be estimated separately with the methodology described in Appendix B, but also the whole contribution of the field input. If $k_m$ denotes the number of uncertain field variables whose contribution is estimated in addition to the contribution of each mode, then the total number of simulations is $N = N_e(d + 2 + k_m)$.

## 2.3 Ensemble generation with field inputs

### *2.3.1 Uncertain space for SA in the Gironde estuary*

The GSA study was carried for the 7-day February 2003 event with a tide coefficient in the range [43 ; 90], Dordogne upstream discharge (resp. Garonne) in the range [600 ; 2200] $m^3.s^{-1}$ (resp. [1200 ; 5900] $m^3.s^{-1}$).

The present study has $d = 8$ uncertain sources: the 4 zone-distributed scalar friction coefficients (Ks1, Ks2, Ks3, Ks4), the scalar wind drag coefficient Cd, and the time-dependent boundary conditions at the hydrological limits (QDOR and QGAR for the Dordogne and Garonne rivers respectively) and at the maritime limit (CLMAR). Additional sources of uncertainty, such as space- and time-varying meteorological forcing or bathymetry, exist and have not been considered in the study. The uncertain input vector is denoted by $X = (X_i)_{i \in I_8}$ where $X_1$ = Ks1, $X_2$ = Ks2, $X_3$ = Ks3, $X_4$ = Ks4, $X_5$ = Cd, $X_6$ = QDOR, $X_7$ = QGAR and $X_8$ = CLMAR. The QoI is the water level defined over the simulation domain at a given time, with 26 stations that are of particular interest (see Fig. 2). The water level at a given location is a scalar denoted by $Y$ in the following.

The wind drag coefficient and friction coefficients are supposed to follow uniform distributions with ranges described in Table 1. [33] presents a review of parametric formulations of the wind drag coefficient based on [36], [37], [38]. These parametric formulations were used for the Gironde case using climatological wind intensities (Climate Forecast System Reanalysis from NOAA) to define the range for the Cd uniform distribution. The ranges of the friction coefficients are chosen so as to include the calibration values for the 2003 event using both the $Nash_{HT}$ and $RMSE$ criteria. The interval

corresponds to the commonly accepted uncertainty on friction for engineering studies.

The time-dependent hydrological forcing is assumed to be perturbed by an additive centered Gaussian process $q(t)$. The covariance of $q(t)$ is defined by a squared exponential kernel $\kappa(t,t') = \sigma^2 \exp(-\ell^{-2}(t-t')^2)$, where $\ell$ is the correlation time scale estimated from observations over the 1981-2016 period, and the standard deviation $\sigma$ is the amplitude of the perturbation. The time dependent maritime forcing is also assumed to be perturbed by an additive Gaussian Process $h(t)$ with a gaussian covariance function. A truncated form $q_p(t)$ (resp. $h_p(t)$) of $q(t)$ (resp. $h(t)$) is formulated with a KL decomposition [20] of $q(t)$ (resp. $h(t)$).

Numerically, as explained in Appendix C, the eigenfunction problem defined by the Fredholm equation is approximated by the eigenvector problem defined over the discretized time series $\{t_1, t_2, \ldots, t_N\}$:

$$K\Phi_i = \lambda_i \Phi_i \tag{10}$$

where $\Phi_i = (\phi_i(t_1) \ldots \phi_i(t_N))^T$, $K = (\kappa(t_i,t_j))_{1\leq i,j\leq p}$ and $(\lambda_i, \phi_i)$ are the $i^{\text{th}}$ eigenvalue and eigenfunction of $\kappa$, the solution of the Fredholm equation [39].
The solution $(\lambda_i, \Phi_i)$ of Eq. (10) is obtained from a singular value decomposition (SVD) and contributes to the truncated expansion of $q(t)$ (and $h(t)$) discretized over the discretized time series:

$$q_p = (q_p(t_1), \ldots, q_p(t_N))^T = \sum_{i=1}^{p} \sqrt{\lambda_i} \Phi_i \alpha_i. \tag{11}$$

where $\alpha_i$ are independent standard Gaussian variables. Last but not least, based on the property $\sum_{i=1}^{p} \lambda_i = \sigma^2$, the degree of truncation $p$ is chosen such that the fraction of the total variance $\sum_{i=1}^{p} \lambda_i \sigma^{-2}$ exceeds a threshold the closest to one.

Sampling the perturbation $q(t)$ (resp. $h(t)$) associated with the time-dependent hydrological (resp. maritime) forcing over the discretized time series $\{t_1, t_2, \ldots, t_N\}$ sums up sampling the random vector $q_p$ (resp. $h_p$), as sampling $p$ independent standard Gaussian variables. Fig. 3-a and Fig. 3-b present a set of 7 perturbations of the boundary condition $q(t)$ at the Garonne upstream location and $h(t)$ at the maritime boundary.

The boundary condition resulting from the aforementioned perturbed signal is displayed in Fig. 3-c for the Garonne and in Fig. 3-d for the maritime boundary. The autocorrelation time scale was estimated for about 10 discharge signals during major flood events on the Garonne and Dordogne rivers. The correlation time scales $\ell_{\text{QGAR}}$ and $\ell_{\text{QDOR}}$ are thus set to 3 days for the 2003 event and the correlation time scale $\ell_{\text{CLMAR}}$ is set to 6 hours (approximately half a tidal cycle) for the maritime boundary. The amplitudes $\sigma_{\text{QDOR}}$ and $\sigma_{\text{QGAR}}$ of the perturbations of the upstream discharges are set proportional (20 %) to the observed discharges as the uncertainties of the rating curves used to translate the water levels into discharges are larger for high flow time series. The amplitude $\sigma_{\text{CLMAR}}$ of the maritime boundary condition perturbation is set to 50 cm, representing the sum of the uncertainties in surge levels and in the predicted offshore tide, at the maritime boundary of the numerical model. The KL decompositions of the hydrological and maritime forcings are respectively truncated to $p_{\text{QDOR}} = p_{\text{QGAR}} = 4$ and $p_{\text{CLMAR}} = 7$ modes, retaining respectively 90 % and 45 % of the Gaussian process variability. This value, associated to $\ell_{\text{CLMAR}}$, will be discussed in Sect. 3.4. Moreover, the KL modes will be either aggregated or treated separately in the following for Sobol' indices formulation. The GSA is thus carried out in an uncertain space described by 20 variables: Ks1, Ks2, Ks3, Ks4, Cd, 4 modes for each hydrological boundary condition (QGAR and QDOR) and 7 modes for the maritime boundary condition CLMAR.

### *2.3.2 High Performance Computing resources*

The GSA for the 2003 event was performed using both coarse and fine meshes. Due to computational constraints, it was only carried out over one tidal cycle with the fine mesh. Yet the conclusions drawn for the GSA over this period with the fine mesh are similar to those drawn for the coarse mesh over the 7-day event. As a consequence, in the following, illustrations for the GSA are given for the coarse mesh model. Computational resources for the GSA are given in Table 3. For the fine mesh, the convergence of the GSA results was investigated with increasing number of members $N_e$ ranging from 100 to 10000. The Sobol' indices reach convergence for all variables over the simulation time period (not shown) from $N_e = 2000$. The simulations were performed on the HPC resources from GENCI-IDRIS (grant 2017-A0030110292).
For coarse and fine meshes (Table 3 - Col. 1), the simulated period is indicated in Table 3 - Col. 2. For each HPC architecture (Table 3 - Col. 9), a scalability study was performed for a single run with domain decomposition: the optimal number of cores is shown in Table 3 - Col. 3 along with the elapsed time in Table 3 -Col. 4. The GSA involves both domain decomposition and task parallelism. The total number of available cores for the GSA is shown in Table 3 - Col. 5. As Sobol' indices are computed for field inputs considering the contribution of each mode only and also the whole contribution of the uncertain variable, the ensemble size $N_e$ and the

total number of simulations $N$ are given in Table 3 - Col. 6 and Col. 7 respectively. Finally, the elapsed time for the GSA is given in Table 3 - Col. 8.

## 3 GSA results

The GSA study was carried for the 7-day February 2003 event with a tide coefficient in the range [43 ; 90], Dordogne (resp. Garonne) upstream discharge in the range of [600 ; 2200] $m^3.s^{-1}$ (resp. [1200 ; 5900] $m^3.s^{-1}$).

### 3.1 Performance of the ensemble

#### *3.1.1 Description of the criteria*

The performance of the ensemble forecasts is commonly assessed with criteria such as consistency and reliability [40]. The former measures the average spread of an ensemble compared to observations, whereas the latter generally reflects the accuracy of a forecast model.
The consistency criterion characterizes the coherence between the distribution of the ensemble members and a set of observations through the use of the rank histogram [41] for a given simulation time or over a simulation period. The ensemble values are ranked in classes, and the occurrence of the observed value within these classes is computed and represented with a rank histogram. Fig. 4 displays the rank histogram at stations Le Verdon, Richard and Bec d'Ambès for all time steps during the 2003 event. It is expected to be flat when the ensemble members and the observations follow similar distributions, U-shaped when the ensemble is under-dispersive, and bump-shaped when the ensemble is over-dispersive. Here, the occurrence is normalized by the ensemble size and is displayed in Fig. 5 with a blue-red color bar along the curvilinear abscissa (x-axis) of 12 observing stations as a function of time (y-axis). This time and space distributed representation allows the determination of when and where the ensemble is consistent with the observations (ranks uniformly distributed from 0.1 to 0.9) or not (extreme rank values equal to 0 when the ensemble over estimates water levels, and rank values equal to 1 when the ensemble under estimates water levels).
The reliability criterion evaluates the coherence between the forecasted and the observed probabilities of an event [42]. An event is defined as $Z >= Z_T$, where $Z$ is the random value simulated in the ensemble and $Z_T$ is the threshold value. The reliability plot represents the observation probability for the events with respect to the simulated probability. The ensemble is reliable if the relation follows the first bisector line; it is under or over dispersive otherwise. The reliability criterion for 4 observing stations (Le Verdon, Richard, Bec d'Ambès and La Réole) is represented in Fig. 6. Over each curve, the observation quantiles ($q_{10}$ to $q_{90}$) are represented with circles ($q_{50}$ represented by a larger symbol). Reliability curves under (resp. above) the perfect reliability diagonal curve reveal an over-predictive (resp. under-predictive) ensemble.

#### *3.1.2 Interpretation of water level ensemble performance*

Considering both performance criteria in Fig. 4, Fig. 5 and Fig. 6, it appears that the performance of the ensemble is closely related to the tidal cycle and the location in the estuary. At Le Verdon (resp. Richard), at the mouth of the estuary, the reliability curve (blue (resp. green) curve) shows that the ensemble is slightly under- (resp. over-) predictive for all quantiles with water levels that are under- (resp. over-) estimated. Considering the corresponding locations of Fig. 5 for each of them shows respectively a predominance of red and blue nuances independent of the tidal cycle. For water levels lower than the mean tide water level, this can be explained at Le Verdon by a truncated observed signal during the 2003 event. Moreover, calibration results (Table 2) have shown the trend of the numerical model to under-estimate water levels at Le Verdon and to over-estimate them at Richard during flood tide, as the mean error for high tides are respectively negative (-12 cm at Le Verdon) and positive (+21 cm at Richard). In the middle part of the estuary, from Laména to Bordeaux, Fig. 4 shows a tide-dependent rank. The ensemble is over-predictive at low tide and under-predictive at high tide, as suggested in Fig. 6 by the reliability curve at Bec d'Ambès (red line). The Bec d'Ambès reliability curve is under the perfect diagonal reliability curve for water quantiles lower than the mean tide level and above this curve for quantiles higher than the mean tide level. This reflects calibration choices for a better representation of high tides leading to under-dispersive behavior. Indeed, the reliability curve exhibits smaller distances to the diagonal for higher water levels than for lower ones. At the upstream part of Garonne and Dordogne rivers, Fig. 4, Fig. 5 and Fig. 6 show a strong under-predictive signature of the ensemble at La Réole and an over-predictive signature at Pessac (for the end of the storm). This highlights the need for a more refined mesh in the fluvial part of the model.
The ensemble performance has been assessed. It shows that the over/under predictive signature or non-reliability is directly linked to the numerical model calibration or to the mesh refinement in the fluvial areas.

### 3.2 Water level probability density function

The water level mean and standard deviation are displayed with a blue-red color bar along the curvilinear abscissa (x-axis) of the 26 stations of interest as a function of time (y-axis) in Fig. 7. Fig. 7-a shows the ensemble mean water level; it illustrates the propagation of tides from the mouth of the estuary to the upstream part of the Garonne and Dordogne rivers. The hypersynchronic characteristic of the Gironde estuary due to a funnel effect leads to the amplification of high tides in the estuary from the decreasing water depth in bathymetry. The absolute differences of the median with 95%-quantile and 5%-quantile (not shown) shows a nearly perfect symmetry of the distribution of QoIs between its extreme quantiles. The standard deviation plotted in Fig. 7-b increases from 20 cm at Royan to 45 cm at Pessac and La Réole for high tides.
Fig. 8 displays the PDF of water level anomalies (with respect to the ensemble mean) at the storm peak (a), ebb tide (b), low tide (c) and flood tide (d) for 14 observing stations (blue curves for the maritime boundary, green for the estuary, yellow, orange and red for the Garonne and Dordogne rivers). Upstream of the fluvial areas, the PDF is asymmetric for all times with a median value of about -40 cm, a mean value of 10 cm, and a fat tail for positive anomalies at La Réole, Langon and Pessac. Upstream of La Réole, Langon and Pessac, the PDF is symmetric on the Garonne and Dordogne rivers with a standard deviation of about 1.8 m for the flood rise and storm peak, leading to anomalies up to 1 m and a standard deviation of about 0.6 m for ebb tide and low flow. Along the estuary and in the maritime area, the PDF is rather symmetric. The standard deviation decreases from the confluence at Bec d'Ambès to the mouth of the estuary. The standard deviation over the estuary, except in the fluvial area, is larger for flood tide and storm peak than for ebb tide and low flow. This behavior can be explained by the hypersynchronous shape of the estuary which amplifies perturbations from the maritime boundary conditions along the estuary at flood tide. Moreover, it reflects respectively the predominance of the gaussian perturbation of the maritime signal and the uniform perturbation of the Strickler coefficient with a non-linear behavior in the fluvial part.

### 3.3 Global sensitivity analysis indices for aggregated modes

#### *3.3.1 Temporal analysis*

Sobol' indices are displayed with a blue-red color bar for the 8 uncertain inputs $X_i$ (Sect. 2.3.1) in Fig. 9, along the curvilinear abscissa (x-axis) and over time (y-axis). The total Sobol' indices are plotted for each input variable. Blue/red means small/large Sobol' indices for $X_i$.
For the 2003 event studied here, Fig. 9 clearly shows the predominance of the maritime boundary conditions and the dependency of all variables on the tidal signal. The wind drag coefficient, the hydrological boundary conditions and the friction coefficient in area $A_4$ have no influence on the water level variability except at the upstream location of the Garonne and Dordogne rivers. Over the maritime area $A_1$, the water level variance is explained by the variance in the maritime boundary condition with Sobol' indices close to 1 from Royan (station 1) to Richard (station 5) near the mouth of the estuary. It should be noted that the fraction of the water level variance that is not explained by CLMAR is explained by Ks1. In estuarian area $A_2$, CLMAR and Ks2 are the most significant sources of uncertainty, with a large predominance of CLMAR at high tide. At low tide, the Sobol' index for Ks2 reaches 0.8. The influence of the maritime boundary condition decreases from Fort-Médoc (in $A_2$) to the confluence at Bec d'Ambès (in $A_3$) and the influence of Ks2 and Ks3 increases. For this event, in area $A_4$, the influences of CLMAR and Ks4 are alternatively predominant in coherence with the tidal signal. At La Réole and Pessac, Sobol' index for Ks4 reaches 0.8. It should be noted that the difference between total and first order indices, representing interactions of $X_i$ with other uncertain variables, has been computed for all uncertain variables. As it is nearly equal to 0, it can be concluded that, for the 2003 storm, no interactions are detected between any input variable $X_i$. Total Sobol' indices and ranks are time-averaged and plotted in Fig. 10 along the curvilinear abscissa (x-axis) and for all $X_i$ (y-axis). As expected, for the 2003 event, the maritime boundary conditions have a predominant impact along the entire domain, with the significant influence of friction coefficients, while hydrological boundary conditions have a limited impact to the upstream locations of the Garonne and Dordogne rivers.

#### *3.3.2 Global sensitivity analysis space mapping*

Figs 11-a (resp. -c) and 11-b (resp. -d) display maps of total Sobol' indices over the entire mesh of the variable

$X_8$ = CLMAR (resp. $X_1$ = Ks1), at the storm peak and at low tide respectively, for the 2003 event. These maps show the homogeneity of the total Sobol' indices geographical repartition and confirm that the conclusions drawn previously for the 26 locations of interest can be generalized for all points located in $A_1$, $A_2$, $A_3$ and $A_4$. The same maps have been obtained for all uncertain variables for the storm peak, ebb tide, low tide and flood tide, but are not shown here. As formerly observed, for the 2003 event, the wind drag coefficient (Cd), Dordogne river discharge (QDOR) and Garonne river discharge (QGAR) have a very small influence over the tidal cycle along the estuary. As 2003 event is a flood event for both rivers, further investigation is needed to confirm this conclusion. The uncertainty in the maritime boundary condition is predominant, especially during the flood tide and storm peak (at Royan) (Fig 11-a) along the estuary except at the upstream part of the Garonne and Dordogne rivers (in $A_4$), where friction coefficient Ks3 is the most dominant uncertain variable. Ks1 (displayed in panels c- and d-), Ks2 and Ks4 show no influence during the storm peak. During ebb tide, the influence of CLMAR propagates along the Garonne and Dordogne rivers, while the upstream parts of the rivers remain fully under the influence of Ks3. At low tide, Fig. 11-b shows that the influence of CLMAR returns to the middle part of the estuary in $A_2$, whereas Ks1 (resp. Ks2) is the most dominant uncertain variable farther downstream (resp. upstream), consistent with the tidal signal. The influence of Ks3 in the upstream part of the Dordogne and Garonne rivers propagates to $A_4$. The Gironde Estuary hydrodynamics are the result of interplays between time and/or space dependent processes, such as friction, convergence of the estuary and maritime and fluvial usptream boundary conditions, but also of the memory of the state of the system. UQ and GSA help to identify and understand the evolution of complex physical processes that drive the Gironde estuary and are not entirely intuitive, in particular, when considering their phase with the tidal cycle.

### 3.4 Focus on the eigenmodes of the maritime boundary forcings

A GSA study dedicated to the maritime boundary forcing was performed considering each eigenmode of the KL decomposition for the 7-day 2003 event. The other parameters (Ks1, Ks2, Ks3, Ks4, QGAR and QDOR) are not perturbed and their associated values are their mean values. Table 4 shows the variance associated with each orthogonal function of the KL decomposition of the Gaussian correlation function with its time scale set to 6 hours (half-tide). No clear predominance of the first modes is observed with about 6 % of the input variance explained by the first 4 modes and 1 % for the following 19 modes. The accumulated variance explained by the first 7 modes kept for the GSA is about 45 %. The Sobol' indices associated with these modes are plotted in Fig. 12. Their contributions are linked to the tide cycle but are asynchronous. The influence of the eigenmodes decreases from the mouth of the estuary to the upstream part of the rivers with respect to the periodic tide signal. Modes 2, 3, 4 and 5 have major contributions, their respective Sobol' indices reaching 30 %, whereas modes 6 and 7 have less influence during the second part of the storm event and mode 1 has no influence on water levels except at the end of the event when its total Sobol' index reaches 20 %. The uncertainty in the tide and surge level was estimated by [43] to be respectively 5 cm and 40 cm for a 36-hour lead time. This suggests dividing the maritime signal uncertainty into a deterministic tidal contribution and a stochastic surge level contribution caused by the meteorological forcings [44] as well as the tide/surge level interactions [45] through surge levels. Further work should thus focus on separating the tidal signal from the surge level signal. The correlation time scale calculated from the surge level time series provided by Meteo-France is about 2.1 days. Only 14 eigenmodes (resp. 4) are required to represent 99 % (resp. 90 %) of the total variance of the signal, as shown in Fig. 13.

## 4 Conclusions and perspectives

The Telemac2D numerical model used for operational flood forecasting by FFS GAD in the Gironde estuary was used for a GSA based on ANOVA to estimate the Sobol' indices over a 7-day storm event in 2003. It was shown that the maritime boundary conditions drive the dynamics of the estuary.

Moving from the mouth of the estuary to the upstream part of the Garonne and Dordogne rivers, the influence of the friction coefficient increases, and the hydrological forcing has a very local influence upstream in the rivers. This GSA study allows the identification of the most significant sources of uncertainty. Once identified, these uncertainties should be reduced, for instance with a data assimilation algorithm such as Ensemble Kalman Filter, in order to improve the water level in the estuary in simulation and forecast mode. A perspective for this study is to take into account uncertainties in the surface forcing atmospheric fields (wind and pressure). Further work should also focus on the formulation of a surrogate model for the 2D Gironde hydraulics model

for high flow discharges in order to reduce the cost of the GSA and the ensemble assimilation. Gaussian Process or Polynomial Chaos surrogates have been used for UQ and GSA in the context of hydraulics ([46], [23]). Both methods proved efficient for metamodelling of the SWE with the 1D solver Mascaret ([47], [48]) for GSA with Sobol' indices and covariance matrix estimation. A polynomial surrogate was also implemented for Telemac2D on the Garonne river and allowed for a cost-reduced computation of the sensitivity indices [23]. This work should be extended to the Gironde estuary taking into account scalar and field uncertainties. The proposed surrogate modeling and data assimilation approaches would meet this need and would result in a reduced-cost ensemble based data assimilation algorithm to reduce major sources of uncertainties and improve water level forecasting in the Gironde estuary.

## A Appendix A

Sobol' indices [35] measure the contributions of the different independent inputs $X_1, X_2, \ldots, X_d$ and their interactions in the variance V(Y) of the output $Y = f(X)$ with $X = (X_1, X_2, \ldots, X_d)$, $E((f(X))^2) < \infty$ and $f$ the model. They are based on the Hoeffding decomposition of $f$ [49],

$$\begin{aligned} f(X) =& f_\emptyset + \sum_{i \subseteq I_d} f_i(X_i) \\ &+ \sum_{\substack{\{i,j\} \subseteq I_d^2 \\ j>i}} f_{i,j}(X_i, X_j) \\ &+ \ldots + f_{1,2,\ldots,d}(X_1, X_2, \ldots, X_d) \\ &= \sum_{u \subseteq I_d} f_u(X_u) \end{aligned} \tag{12}$$

where $I_d = \{1, \ldots, d\}$ is the set of input indices, $f_\emptyset = E(Y)$ is the expectation of Y (here the average of all values Y can take), $f_i(X_i) = E(Y|X_i) - f_\emptyset$ is the elementary contribution of $X_i$ to $f(X)$, and $f_{i,j}(X_i, X_j) = E(Y|X_i, X_j) - f_i(X_i) - f_j(X_j) - f_\emptyset$ is the contribution of the interaction between $X_i$ and $X_j$ to $f(X)$.

From Eq. (12) and using the orthogonality condition $E(f_i f_j) = 0$ if $i \neq j$, the variance of Y is:

$$V(Y) = \sum_{i \subseteq I_d} V_i + \sum_{\substack{\{i,j\} \subseteq I_d^2 \\ j>i}} V_{i,j} + \ldots + V_{1,2,\ldots,d} = \sum_{u \subseteq I_d} V_u \tag{13}$$

where $V_i = V(f_i(X_i))$ is the elementary contribution of $X_i$ to $V(Y)$, $V_{i,j} = V(f_{i,j}(X_i, X_j))$ is the contribution of the interaction between $X_i$ and $X_j$ to $V(Y)$, and so on. Dividing Eq. (13) by $V(Y)$ leads to Eq. (8).

## B Appendix B

This appendix describes the main steps for the stochastic estimation of the Sobol' indices of the different independent inputs $X_1, X_2, \ldots, X_d$ with $X = (X_1, X_2, \ldots, X_d)$ according to [10].

1. Step 1: generation of two ensembles of size $N_e$ for the normalized input parameters set $(x_1, x_2, \ldots, x_d)$. The first (resp. second) $(N_e, d)$ matrix is denoted by $A$ (resp. $B$) (Eq. (14) (resp. (Eq. (15)))). The space filling strategy is carried out with a Sobol' sequence rather than a classical Monte-Carlo strategy:

$$A = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \ldots & x_i^{(1)} & \ldots & x_d^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \ldots & x_i^{(2)} & \ldots & x_d^{(2)} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ x_1^{(Ne-1)} & x_2^{(Ne-1)} & \ldots & x_i^{(Ne-1)} & \ldots & x_d^{(Ne-1)} \\ x_1^{(Ne)} & x_2^{(Ne)} & \ldots & x_i^{(Ne)} & \ldots & x_d^{(Ne)} \end{pmatrix} \tag{14}$$

$$B = \begin{pmatrix} x_1^{(Ne+1)} & x_2^{(Ne+1)} & \ldots & x_i^{(Ne+1)} & \ldots & x_d^{(Ne+1)} \\ x_1^{(Ne+2)} & x_2^{(Ne+2)} & \ldots & x_i^{(Ne+2)} & \ldots & x_d^{(Ne+2)} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ x_1^{(2Ne-1)} & x_2^{(2Ne-1)} & \ldots & x_i^{(2Ne-1)} & \ldots & x_d^{(2Ne-1)} \\ x_1^{(2Ne)} & x_2^{(2Ne)} & \ldots & x_i^{(2Ne)} & \ldots & x_d^{(2Ne)} \end{pmatrix} \tag{15}$$

2. Step 2: definition of $d$ matrices $C_i$ formed by all columns of $A$ except the $i^{\text{th}}$ column taken from $B$ [49].

3. Step 3: computation of the model output (and QoI) $y_A$ (resp. $y_B$ and $y_{C_i}$ with $i = 1, , d$), for all the input values in the sample matrix $A$ (resp. $B$ and the $d$ matrices $C_i$), obtaining $(d+2)$ vector outputs of dimension $N_e$: $y_A = f(A)$, $y_B = f(B)$, $y_{C_i} = f(C_i)$ with $i = 1, , d$.

[49], [50] and [51] describe the best approach to compute simultaneously $S_i$ and $S_{T_i}$ for each input variable.
In this study, the following estimators have been chosen for:
$V_{X_i}(E_{X_i}(Y|X_i)) = \frac{1}{N}\sum_{j=1}^{N} y_B^{(j)}\left(y_{C_i}^{(j)} - y_A^{(j)}\right)$,
$E_{X_i}(V_{X_i}(Y|X_i)) = \frac{1}{2N}\sum_{j=1}^{N}\left(y_A^{(j)} - y_{C_i}^{(j)}\right)^2$ and
$V(Y) = \frac{1}{N}\sum_{j=1}^{N}\left(y_A^j\right)^2 - f_0^2$
with $f_0^2 = \frac{1}{N}\sum_{j=1}^{N} y_A^j y_B^j$.
Anomalies due to an uncertain variable $X_i$ corresponding to the normalized uncertain variable $x_i$ are defined as $y_{C_i} - y_A$ and represent the part of the QoIs output resulting from $X_i$ only.
If $(X_i)_{i=k,,k+m}$ correspond to the $m$ uncertain modes of a field input, the contribution of each mode can be estimated separately with the methodology described above. The whole contribution of the field input can also be estimated by processing the uncertain variable modes in one block and by defining a matrix $C_k$ formed by all columns of $A$ except the $k^{\text{th}}$ to $(k+m)^{\text{th}}$ columns taken from $B$.

## C Appendix C

A truncated form $q_p(t)$ (resp. $h_p(t)$) of $q(t)$ (resp. $h(t)$) is formulated with a KL decomposition [20] of $q(t)$ (resp. $h(t)$):

$$q_p(t) = \sum_{i=1}^{p} \sqrt{\lambda_i}\phi_i(t)\,\alpha_i \tag{16}$$

where $\alpha_i$ are independent standard Gaussian variables and $(\lambda_i, \phi_i)$ are the $i^{\text{th}}$ eigenvalue and eigenfunction of $\kappa$, the solution of the Fredholm equation [39]:

$$\int \kappa(t,\tau)\,\phi_i(\tau)\,d\tau = \lambda_i \phi_i(t) \tag{17}$$

with $\int \phi_i(t)\phi_j(t)dt = \delta_{i,j}$.

## References


1. Jaak Monbaliu, Zhongyuan Chen, Didier Felts, Jianzhong Ge, Francois Hissel, Jens Kappenberg, Siddharth Narayan, Robert J Nicholls, Nino Ohle, Dagmar Schuster, et al. Risk assessment of estuaries under climate change: lessons from western europe. Coastal Engineering, 87:32–49, 2014.
2. Maryam Golnaraghi. Institutional partnerships in multi-hazard early warning systems: a compilation of seven national good practices and guiding principles. Springer Science & Business Media, 2012.
3. Jean-Michel Hervouet. Hydrodynamics of free surface flows: modelling with the finite element method. John Wiley & Sons, 2007.
4. François Hissel. Projet gironde – rapport final d'évaluation du modèle gironde. Technical report, CETMEF, 2010.
5. AH Weerts, HC Winsemius, and JS Verkade. Estimation of predictive hydrological uncertainty using quantile regression: examples from the national flood forecasting system (england and wales). Hydrology and Earth System Sciences, 15(1):255–265, 2011.
6. Bertrand Iooss and Paul Lemaître. A review on global sensitivity analysis methods. In Uncertainty management in simulation-optimization of complex systems, pages 101–122. Springer, 2015.
7. Bradley Efron. Linear statistical inference and its applications., 1967.
8. Bradley Efron and Charles Stein. The jackknife estimate of variance. The Annals of Statistics, pages 586–596, 1981.
9. Toshimitsu Homma and Andrea Saltelli. Importance measures in global sensitivity analysis of nonlinear models. Reliability Engineering & System Safety, 52(1):1–17, 1996.
10. William Becker and Andrea Saltelli. Design for sensitivity analysis. In Handbook of design and analysis of experiments, pages 647–694. Chapman and Hall/CRC, 2015.
11. Chonggang Xu and George Zdzislaw Gertner. Uncertainty and sensitivity analysis for models with correlated parameters. Reliability Engineering & System Safety, 93(10):1563–1573, 2008.
12. Jia Li and Dongbin Xiu. On numerical properties of the ensemble kalman filter for data assimilation. Computer Methods in Applied Mechanics and Engineering, 197(43-44):3574–3583, 2008.
13. Bertrand Iooss and Andrea Saltelli. Introduction to sensitivity analysis. Handbook of uncertainty quantification, pages 1103–1122, 2017.
14. Bertrand Iooss, Loïc Boussouf, Vincent Feuillard, and Amandine Marrel. Numerical studies of the metamodel fitting and validation processes. arXiv preprint arXiv:1001.1049, 2010.
15. Matieyendou Lamboni, Hervé Monod, and David Makowski. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. Reliability Engineering & System Safety, 96(4):450–459, 2011.
16. Olivier Le Maître and Omar M Knio. Spectral methods for uncertainty quantification: with applications to computational fluid dynamics. Springer Science & Business Media, 2010.
17. Andrea Saltelli and Michaela Saisana. Settings and methods for global sensitivity analysis–a short guide. In PAMM: Proceedings in Applied Mathematics and Mechanics, volume 7, pages 2140013–2140014. Wiley Online Library, 2007.
18. Géraud Blatman. Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis. PhD thesis, Clermont-Ferrand 2, 2009.
19. C Audouze, F De Vuyst, and PB Nair. Reduced-order modeling of parameterized pdes using time–space-parameter principal component analysis. International journal for numerical methods in engineering, 80(8):1025–1057, 2009.
20. Gal Berkooz, Philip Holmes, and John L Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. Annual review of fluid mechanics, 25(1):539–575, 1993.
21. Charlotte M Emery, Sylvain Biancamaria, Aaron Boone, Pierre-André Garambois, Sophie Ricci, Mélanie C Rochoux, and Bertrand Decharme. Temporal variance-based sensitivity analysis of the river-routing component of the large-scale hydrological model isba–trip: Application on the amazon basin. Journal of Hydrometeorology, 17(12):3007–3027, 2016.
22. Jiri Nossent, Pieter Elsen, and Willy Bauwens. Sobol' sensitivity analysis of a complex environmental model. Environmental Modelling & Software, 26(12):1515–1525, 2011.
23. Nicole Goutal, Cedric Goeury, Riadh Ata, Sophie Ricci, Nabil El Mocayd, M Rochoux, Hind Oubanas, Igor Gejadze, and Pierre-Olivier Malaterre. Uncertainty quantification for river flow simulation applied to a real test case: The garonne valley. In Advances in Hydroinformatics, pages 169–187. Springer, 2018.
24. Hind Oubanas, Igor Gejadze, Pierre-Olivier Malaterre, and Franck Mercier. River discharge estimation from synthetic swot-type observations using variational data assimilation and the full saint-venant hydraulic model. Journal of Hydrology, 559:638–647, 2018.
25. Amandine Marrel, Nathalie Saint-Geours, and Matthias De Lozzo. Sensitivity analysis of spatial and/or temporal phenomena. Handbook of Uncertainty Quantification, pages 1–31, 2016.
26. Katrin Erdlenbruch, Éric Gilbert, Frédéric Grelot, and Chritophe Lescoulier. Une analyse coût-bénéfice spatialisée de la protection contre des inondations. application de la méthode des dommages évités à la basse vallée de l'orb. Ingénieries-EAT, (53):p–3, 2008.
27. Guotu Li, Mohamed Iskandarani, Matthieu Le Hénaff, Justin Winokur, Olivier P Le Maître, and Omar M Knio. Quantifying initial and wind forcing uncertainties in the gulf of mexico. Computational Geosciences, 20(5):1133–1153, 2016.
28. Patrice Castaing and George P Allen. Mechanisms controlling seaward escape of suspended sediment from the gironde: a macrotidal estuary in france. Marine Geology, 40(1-2):101–118, 1981.
29. Nicolas Huybrechts, Catherine Villaret, and Florent Lyard. Optimized predictive two-dimensional hydrodynamic model of the gironde estuary in france. Journal of Waterway, Port, Coastal and Ocean Engineering, 138(4):312–322, 2011.
30. François Hissel, Gilles Morel, Gianluca Pescaroli, Herman Graaff, Didier Felts, and Luca Pietrantoni. Early warning

and mass evacuation in coastal cities. Coastal Engineering, 87:193–204, 2014.
31. Ph Gauckler. Etudes Théoriques et Pratiques sur l'Ecoulement et le Mouvement des Eaux. Gauthier-Villars, 1867.
32. M Janišková. Study of the systematic errors in aladin associated to the physical part of the model. Note Aladin, (7), 1995.
33. Florence Levy. Construction d'un modéle de surcotes sur la façade atlantique. Technical report, CETMEF, 2013. rapport provisoire.
34. Florence Levy and Antoine Joly. Modélisation des surcotes avec telemac2d. Technical report, EDF/LHSV, 2013. rapport EDF à accessibilité restreinte.
35. Andrea Saltelli and Il'ya Meerovich Sobol'. Sensitivity analysis for nonlinear mathematical models: numerical experience. Matematicheskoe Modelirovanie, 7(11):16–28, 1995.
36. Roger A Flather. Results from a storm surge prediction model of the north-west european continental shelf for april, november and december, 1973. 1976.
37. Jin Wu. Wind-stress coefficients over sea surface near neutral conditions—a revisit. Journal of Physical Oceanography, 10(5):727–740, 1980.
38. Margaret Yelland and Peter K Taylor. Wind stress measurements from the open ocean. Journal of Physical Oceanography, 26(4):541–558, 1996.
39. Siegfried Prössdorf and Bernd Silbermann. Numerical analysis for integral and related operator equations. Operator theory, 52:5–534, 1991.
40. Allan H Murphy. A new vector partition of the probability score. Journal of Applied Meteorology, 12(4):595–600, 1973.
41. O Talagrand, R Vautard, and B Strauss. Evaluation of probabilistic prediction systems, paper presented at ecmwf workshop on predictability, eur. cent. for med. range weather forecasts. Reading, UK, 1997.
42. Guillem Candille and Olivier Talagrand. Evaluation of probabilistic prediction systems for a scalar variable. Quarterly Journal of the Royal Meteorological Society, 131(609):2131–2150, 2005.
43. Loren Carrère, Florent Lyard, M Cancet, A Guillot, and Laurent Roblou. Fes 2012: a new global tidal model taking advantage of nearly 20 years of altimetry. In 20 Years of Progress in Radar Altimatry, volume 710, 2013.
44. Riccardo Mel, Daniele Pietro Viero, Luca Carniello, Andrea Defina, and Luigi D'Alpaos. Simplified methods for real-time prediction of storm surge uncertainty: The city of venice case study. Advances in water resources, 71:177–185, 2014.
45. D Idier, H Muller, R Pedreros, J Thiébot, M Yates, R Créach, G Voineson, F Dumas, F Lecornu, L Pineau-Guillou, et al. Système de prévision de surcotes en manche/atlantique et méditerranée: Amélioration du système existant sur la façade manche/gascogne [d4]. Technical report, Rapport BRGM/RP-61019-FR, 2012.
46. Amandine Marrel. Mise en oeuvre et utilisation du métamodèle processus gaussien pour l'analyse de sensibilité de modèles numériques: application à un code de transport hydrogéologique. PhD thesis, Toulouse, INSA, 2008.
47. Nabil El Moçayd. La décomposition en polynôme du chaos pour l'amélioration de l'assimilation de données ensembliste en hydraulique fluviale. PhD thesis, 2017.
48. Pamphile T Roy, Nabil El Moçayd, Sophie Ricci, Jean-Christophe Jouhaud, Nicole Goutal, Matthias De Lozzo, and Mélanie C Rochoux. Comparison of polynomial chaos and gaussian process surrogates for uncertainty quantification and correlation estimation of spatially distributed open-channel steady flows. Stochastic Environmental Research and Risk Assessment, 32(6):1723–1741, 2018.
49. Andrea Saltelli and Paola Annoni. How to avoid a perfunctory sensitivity analysis. Environmental Modelling & Software, 25(12):1508–1517, 2010.
50. Michiel JW Jansen. Analysis of variance designs for model output. Computer Physics Communications, 117(1-2):35–43, 1999.
51. Andrea Saltelli. Making best use of model evaluations to compute sensitivity indices. Computer physics communications, 145(2):280–297, 2002.

Table 1: Calibrated Strickler (K) coefficients computed from the $Nash_{HT}$ and $RMSE$ criteria.

| Input variable | Updated Strickler coefficients with $Nash_{HT}$ criterion | Updated Strickler coefficients with $RMSE$ criterion | Range of uniform distribution |
|---|---|---|---|
| Ks1 | 55 | 70 | [50 ; 70] |
| Ks2 | 70 | 70 | [45 ; 75] |
| Ks3 | 75 | 65 | [25 ; 75] |
| Ks4 | 50 | 55 | [40 ; 80] |
| Cd | 2.57.10-6 | 2.57.10-6 | [0.678.10-6 ; 3.016.10-6] |

Table 2: *Nash* at high tides ($Nash_{HT}$), $RMSE$ and $RMSE$ at high tides ($RMSE_{HT}$) for the 1999 and 2003 events along the Gironde estuary

| Station | $Nash_{HT}$ | | $RMSE$ (m) | | $RMSE_{HT}$ (m) | |
|---|---|---|---|---|---|---|
| | 1999 | 2003 | 1999 | 2003 | 1999 | 2003 |
| Verdon | -0.85 | 0.93 | 2.0 | 0.16 | 0.44 | 0.12 |
| Richard | 0.45 | 0.8 | 1.78 | 0.25 | 0.24 | 0.21 |
| Laména | 0.89 | 0.88 | 1.49 | 0.36 | 0.11 | 0.18 |
| Pauillac | 0.72 | 0.93 | 0.23 | 0.29 | 0.16 | 0.14 |
| Fort Médoc | 0.72 | 0.93 | 0.27 | 0.31 | 0.16 | 0.14 |
| Bec d'Ambès | 0.75 | 0.95 | 0.17 | 0.18 | 0.16 | 0.13 |
| Le Marquis | 0.53 | 0.95 | 0.28 | 0.24 | 0.22 | 0.12 |
| Bassens | 0.54 | 0.96 | 0.28 | 0.17 | 0.22 | 0.12 |
| Bordeaux | 0.59 | 0.96 | 0.23 | 0.18 | 0.21 | 0.13 |
| La Réole | | | 0.51 | 0.93 | | |
| Pessac | | | 0.38 | 1.01 | | |
| Mean (without La Réole and Pessac) | 0.48 | 0.92 | 0.75 | 0.24 | 0.21 | 0.15 |

Table 3: Computational costs for the 2003 event GSA.

| Mesh (number of nodes) | Physical elapsed time (s) | Optimal nb. of cores for one run | Single run elapsed time (min) | Available nb. of cores for GSA | Ensemble size (Ne) | Total nb. of simulations for GSA (N) | GSA elapsed time (days) | Architecture |
|---|---|---|---|---|---|---|---|---|
| 7351 | 250000 | 1 | 25 | 96 | 1000 | 25000 | 8 | CentOS6 - Fedora<br>2 nodes<br>48 cores per node |
| 7351 | 250000 | 1 | 25 | 96 | 5000 | 125000 | 40 | CentOS6 - Fedora |
| 418314 | 83400 | 112 | 13 | 2688 | 5000 | 125000 | 47 | Intel Broadwell (2.4 GHz)<br>2 processors per node<br>14 cores per processor |
| 418314 | 83400 | 32 | 1440 | 32768 | 7000 | 175000 | 71 | BlueGene-Q<br>64 processors per node<br>16 cores per processor |
| 418314 | 83400 | 32 | 1440 | 32768 | 10000 | 250000 | 101 | BlueGene-Q<br>64 processors per node<br>16 cores per processor |

Table 4: Eigenvalues and cumulated variances for the tide signal correlation function

| mode rank | eigenvalue | cumulated variance | variance due to i | $\frac{(variance\ due\ to\ i)}{(variance\ due\ to\ the\ first\ mode}$ |
|---|---|---|---|---|
| 0 | 42829.68 | 6.69 % | 6.69 % | 1.0 |
| 1 | 41746.64 | 13.21 % | 6.52 % | 97.47 % |
| 2 | 40054.17 | 19.46 % | 6.26 % | 93.52 % |
| 3 | 37895.49 | 25.38 % | 5.92 % | 88.48 % |
| 4 | 35429.63 | 30.92 % | 5.53 % | 82.72 % |
| 5 | 32807.55 | 36.04 % | 5.12 % | 76.60 % |
| 6 | 30156.19 | 40.75 % | 4.71 % | 70.41 % |
| 7 | 27571.53 | 45.06 % | 4.31 % | 64.37 % |
| 8 | 25118.82 | 48.98 % | 3.92 % | 58.65 % |
| 9 | 22836.84 | 52.55 % | 3.57 % | 53.32 % |
| 10 | 20743.96 | 55.79 % | 3.24 % | 48.43 % |
| 11 | 18844.02 | 58.73 % | 2.94 % | 44.00 % |
| 12 | 17131.37 | 61.40 % | 2.68 % | 40.00 % |
| 13 | 15594.70 | 63.84 % | 2.44 % | 36.41 % |
| 14 | 14219.79 | 66.06 % | 2.22 % | 33.20 % |
| 15 | 12991.35 | 68.09 % | 2.03 % | 30.33 % |
| 16 | 11894.18 | 69.95 % | 1.86 % | 27.77 % |
| 17 | 10913.85 | 71.65 % | 1.70 % | 25.48 % |
| 18 | 10037.04 | 73.22 % | 1.57 % | 23.43 % |
| 19 | 9251.70 | 74.66 % | 1.44 % | 21.60 % |
| 20 | 8547.08 | 76.00 % | 1.33 % | 19.96 % |
| 21 | 7913.65 | 77.24 % | 1.24 % | 18.48 % |
| 22 | 7343.06 | 78.38 % | 1.15 % | 17.14 % |
| 23 | 6827.94 | 79.45 % | 1.07 % | 15.94 % |

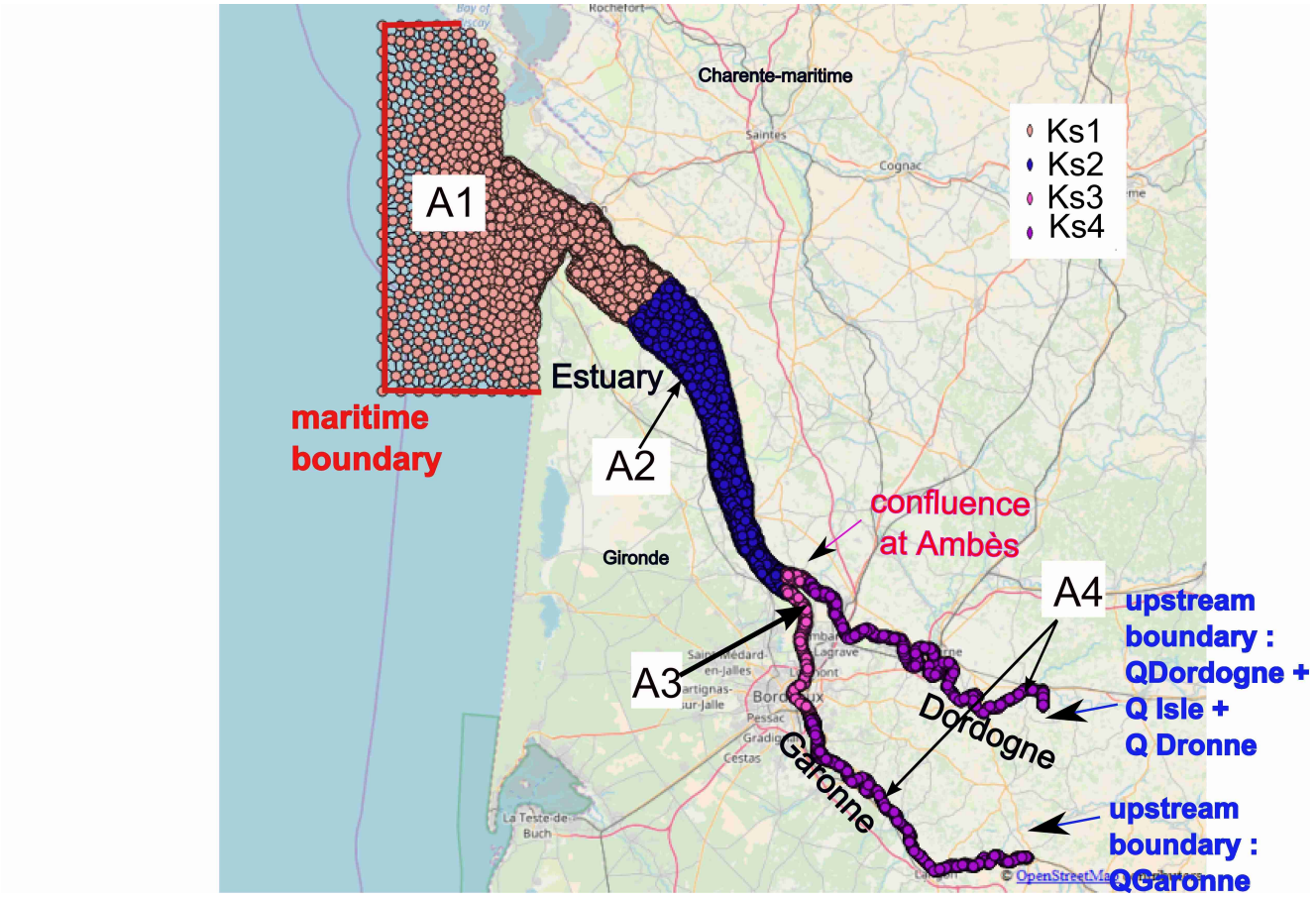



Fig. 1: Extension and location of the numerical model of the Gironde estuary and delimitation of the Strickler coefficient areas 1 to 4. Circles represent the nodes of the numerical model based on a mesh built with finite elements (in black).

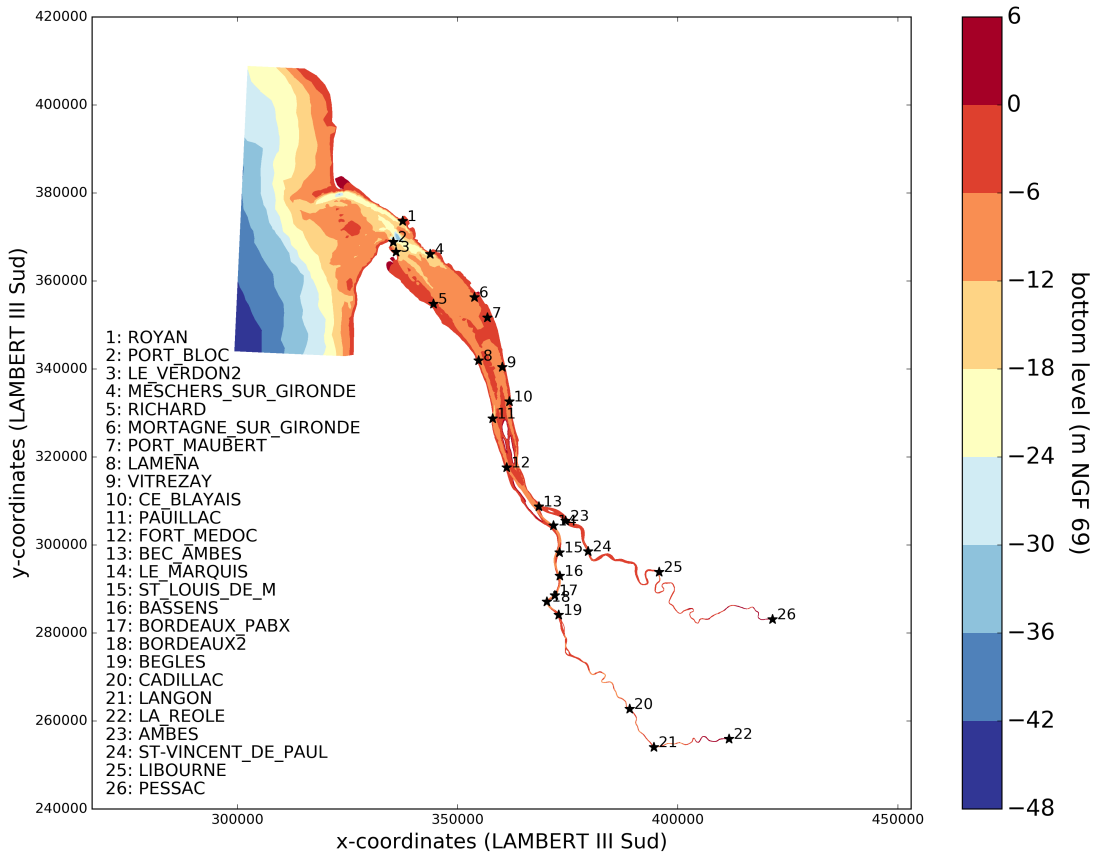



Fig. 2: Bathymetry (m NGF69) of the numerical model of the Gironde estuary. Stars show the 26 main stations of interest for the water level forecasts.

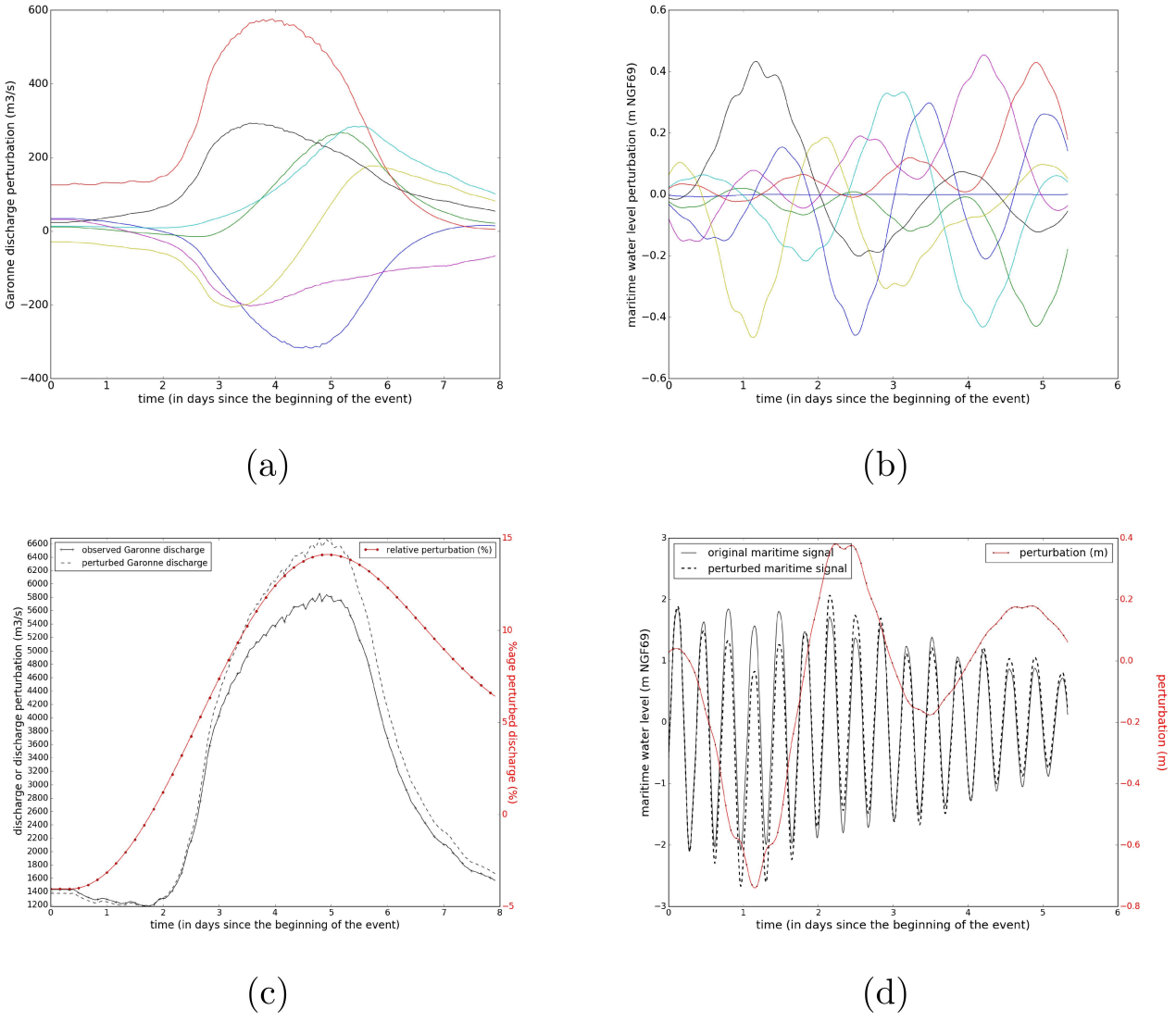



Fig. 3: Representation with time (since the beginning of the storm) during the February 2003 event of a sample of perturbations applied to the (a) Garonne river discharge and (b) maritime boundary conditions. Representation of (c) the reference Garonne river discharge (black line with crosses and left y-axis), of one perturbed member (black dotted line and left y-axis) and of the relative perturbation (red line marked with circles and right y-axis) and (d) the reference water level signal (black line and left y-axis) at one node of the maritime boundary, of one perturbed member (black dotted line and left y-axis) and of the perturbation (red line and right y-axis)

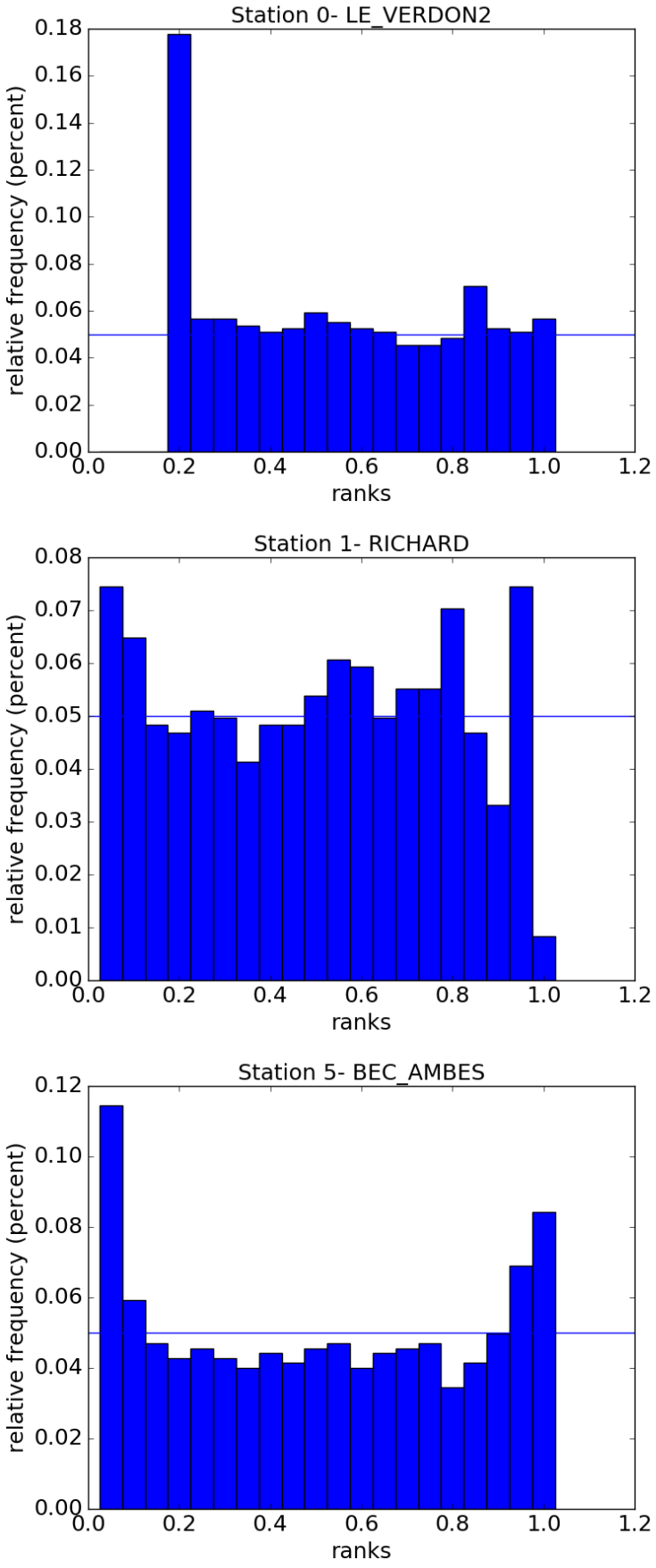



Fig. 4: Ranks diagrams for 3 stations along the Gironde estuary (Le Verdon at the mouth (top), Richard (middle) and Bec d'Ambès (bottom)).

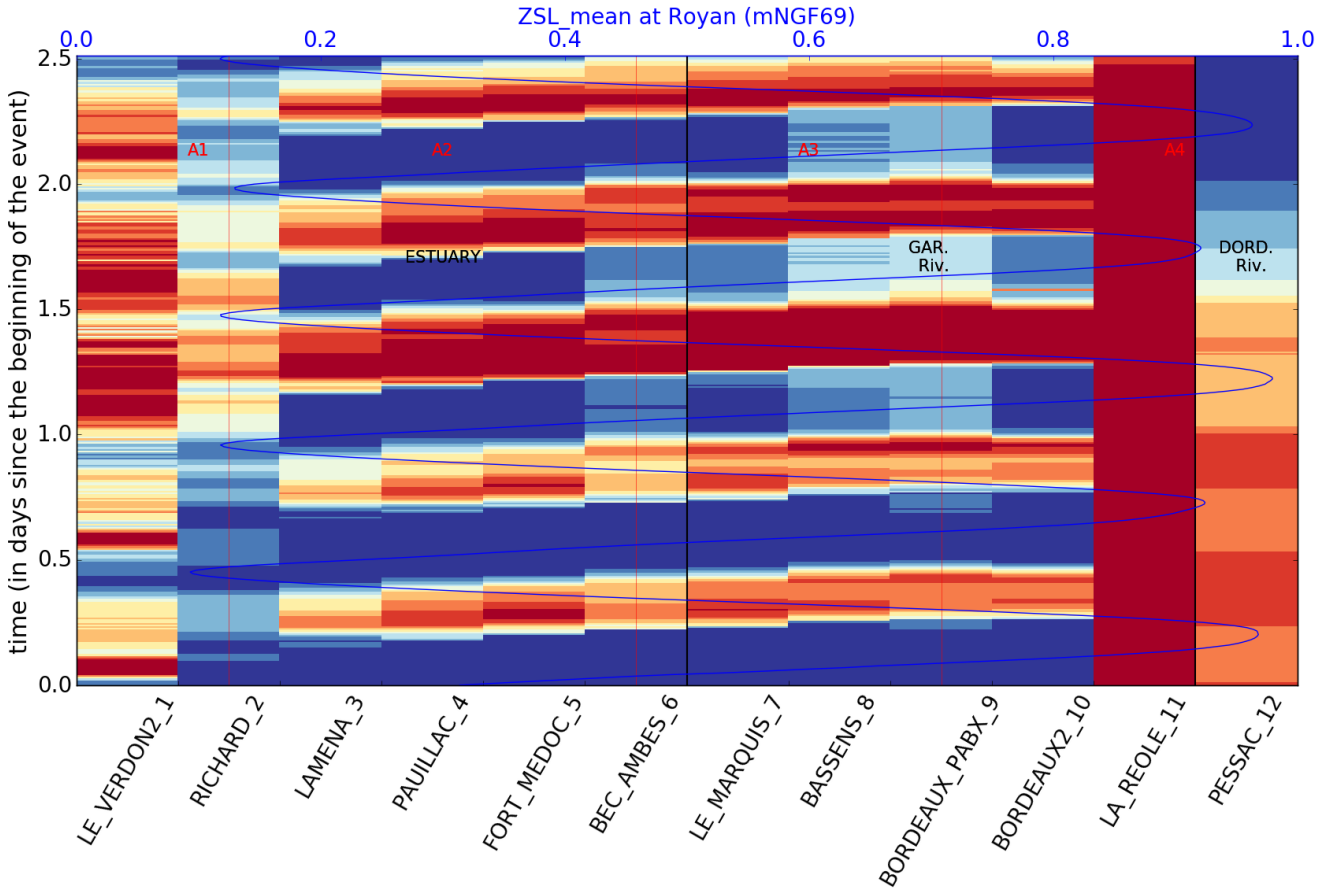



Fig. 5: Surface distribution of ranks along the Gironde estuary during the February 2003 storm. The black vertical lines represent the limits between the estuary, the confluence and the Garonne and Dordogne rivers. The x-axis displays the 12 observing stations classified from left to right from downstream to upstream, from Royan to the confluence at bec d'Ambès (1$^{st}$ black vertical line), from the confluence to La Réole (Garonne river area located between both black vertical lines), and from the confluence to Pessac on the Dordogne river (Dordogne river area on the right of the 2$^{nd}$ black vertical line). The red vertical lines represent the limits between the 4 friction coefficients areas (denoted by $A_1$, $A_2$, $A_3$ and $A_4$).

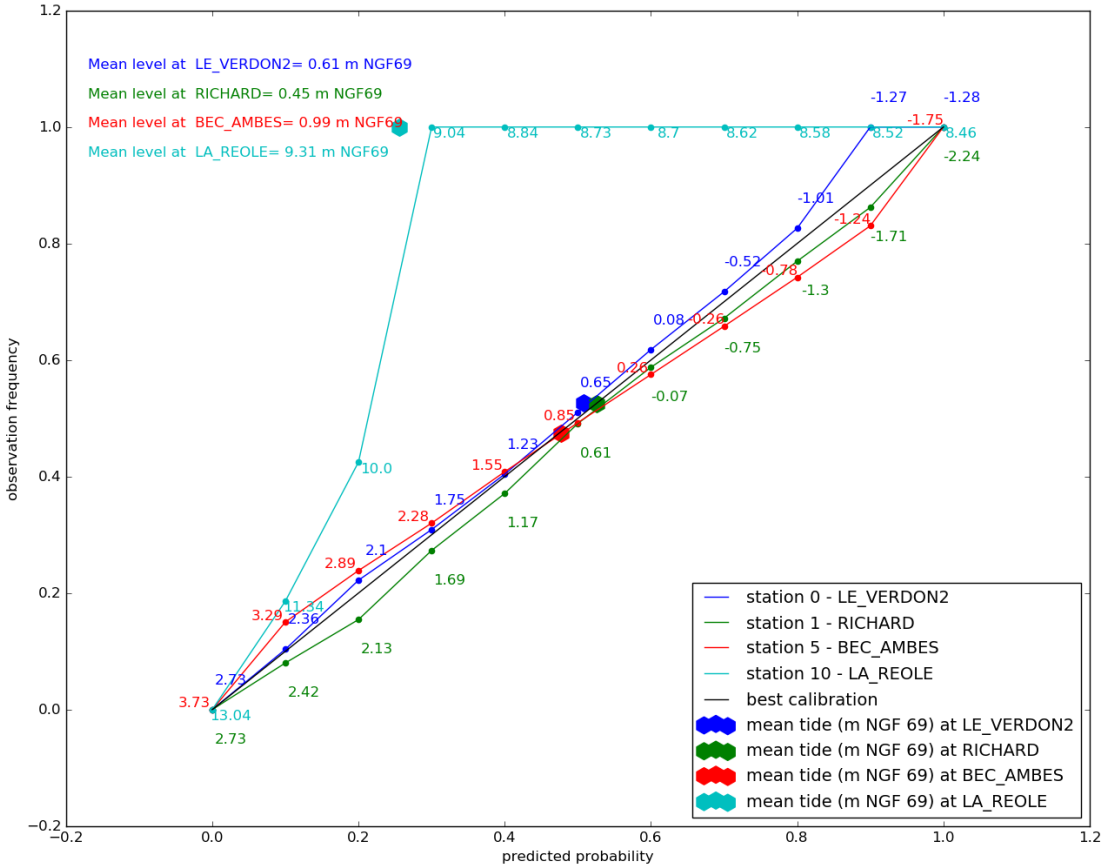



Fig. 6: Reliability diagram for 4 stations along the Gironde estuary (Le Verdon at the mouth (deep blue), Richard (green), Bec d'Ambès (red) and La Réole (cyan)) with corresponding observed water level quantiles during the 2003 event. Mean water levels at each station are represented with hexagons.

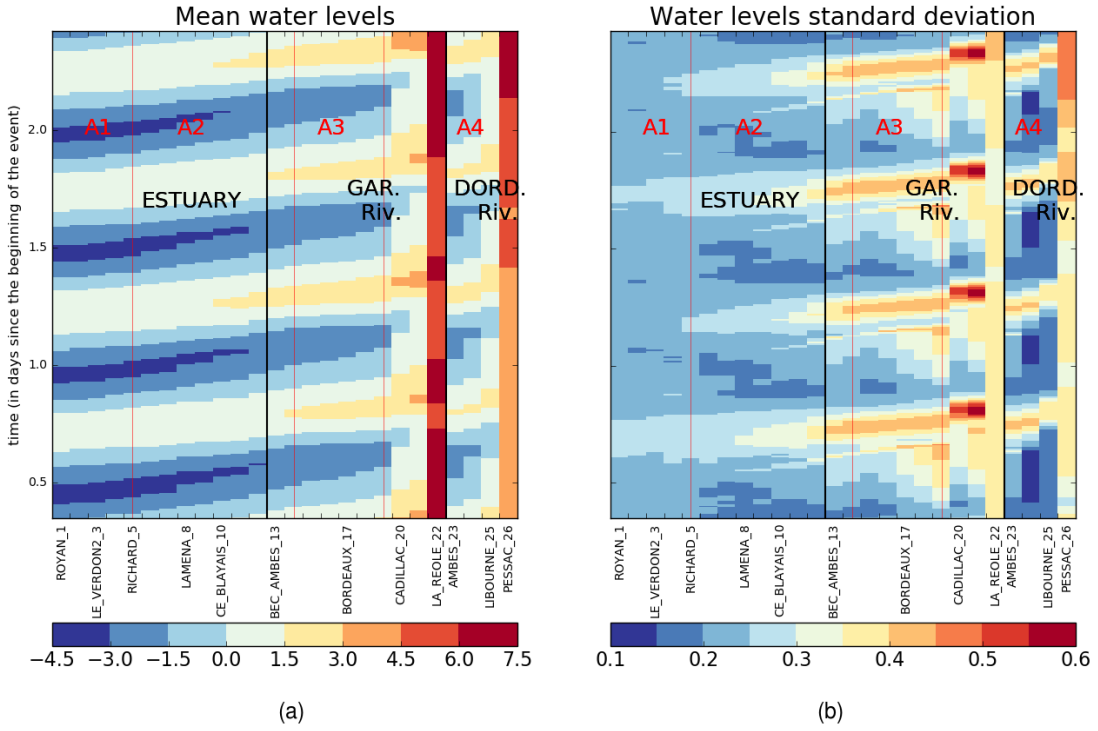



Fig. 7: Median (left) and standard deviation (right) of the water level outputs ensemble along the Gironde estuary during the February 2003 storm. The black vertical lines represent the limits between the estuary, the confluence and the Garonne and Dordogne rivers. The x-axis displays the 26 stations of interest classified from left to right for downstream to upstream, from Royan to the confluence at bec d'Ambès (1$^{st}$ black vertical line), from the confluence to La Réole (Garonne river area located between both black vertical lines) and from the confluence to Pessac on the Dordogne river (Dordogne river area on the right of the 2$^{nd}$ black vertical line). The red vertical lines represent the limits between the 4 friction coefficients areas (denoted by $A_1$, $A_2$, $A_3$ and $A_4$).

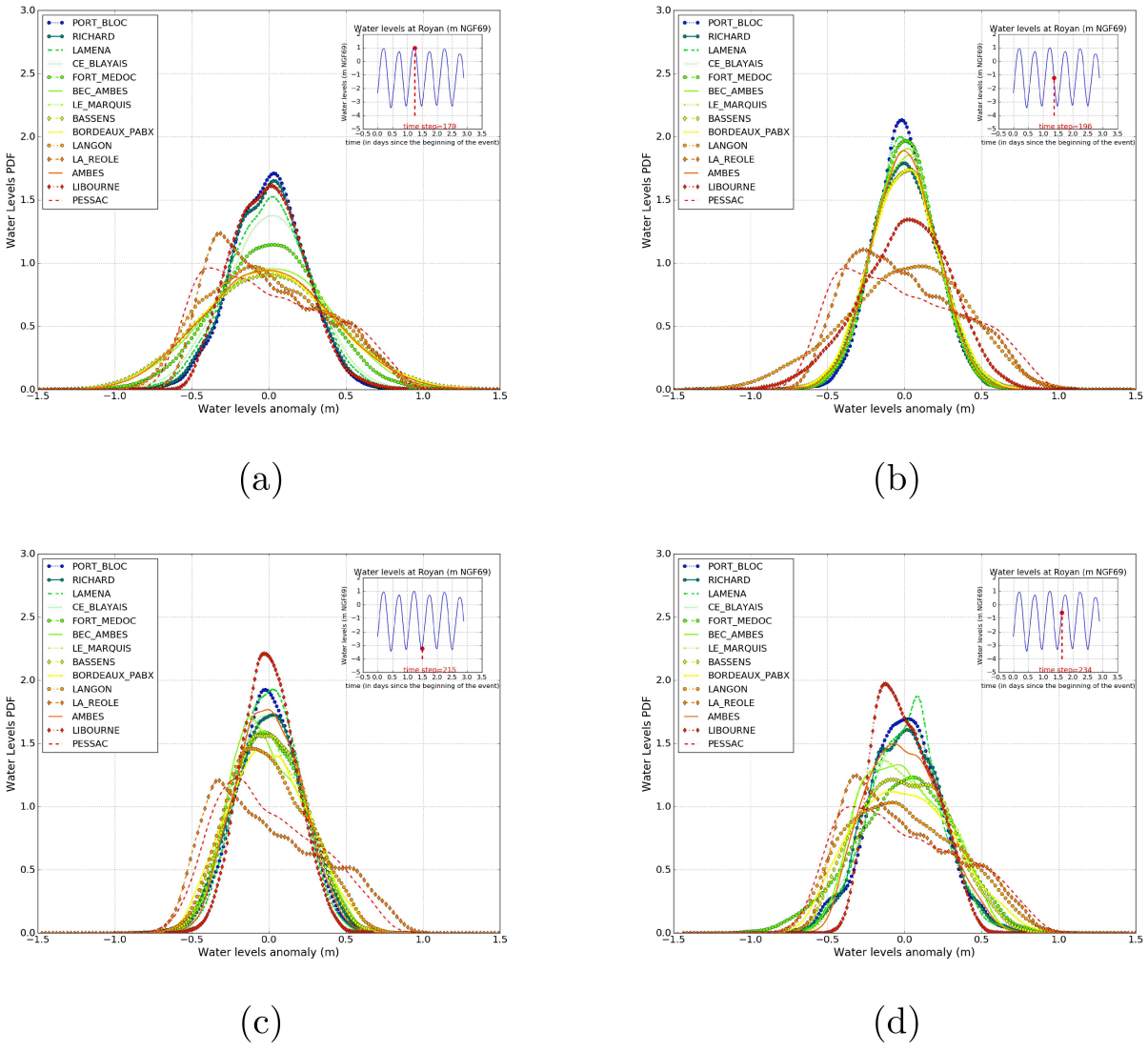



Fig. 8: Water level anomaly probability density functions along the Gironde estuary at the (a) storm peak, (b) ebb tide, (c) low tide, and (d) flood tide during the February 2003 storm.

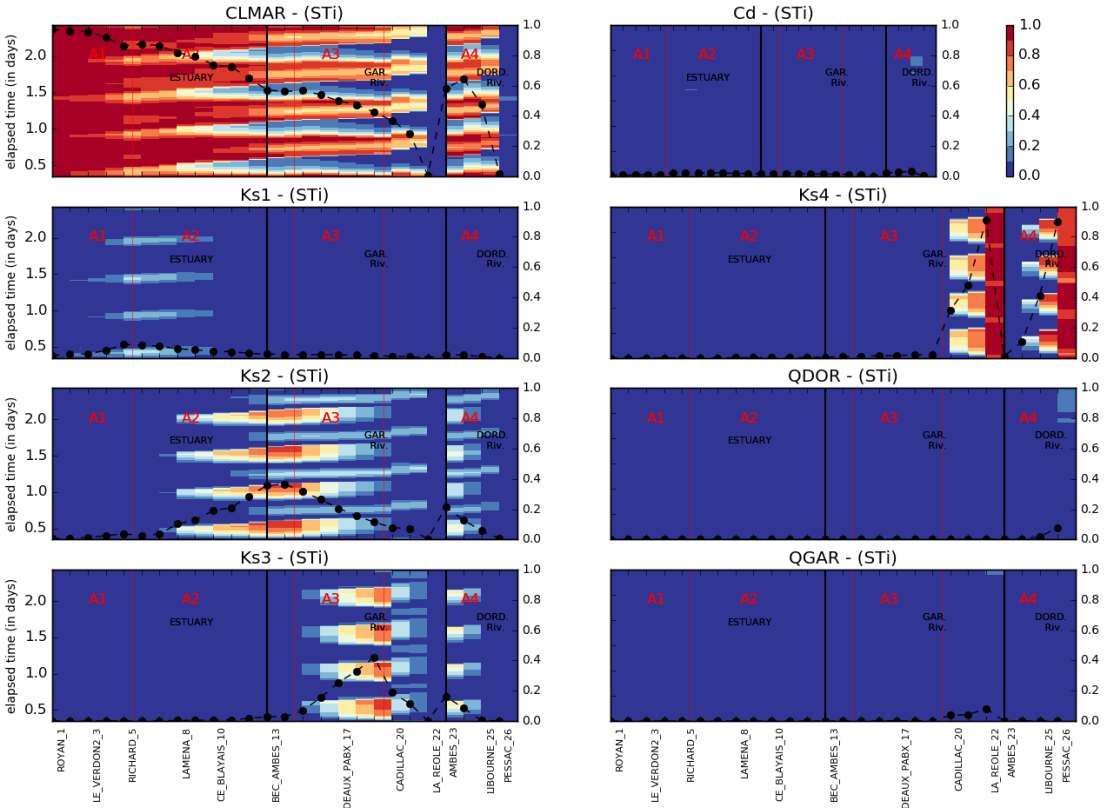



Fig. 9: Total Sobol' indices (STi) along the Gironde estuary during the February 2003 storm for 8 uncertain input variables (maritime boundary conditions (CLMAR), friction coefficients (Ks1, Ks2, Ks3, Ks4), the wind drag coefficient Cd and hydrological boundary conditions (Dordogne river discharge: QDOR; Garonne river discharge: QGAR). The x-axis is similar to that of Fig. 7, with the 26 stations of interest classified from left to right from downstream to upstream. The red vertical lines represent the limits between the 4 friction coefficient areas. The black vertical lines represent the limits between the estuary, the confluence, and the Garonne and Dordogne rivers. The thick dashed lines represent the time-averaged values for Sobol' indices along the Gironde estuary.

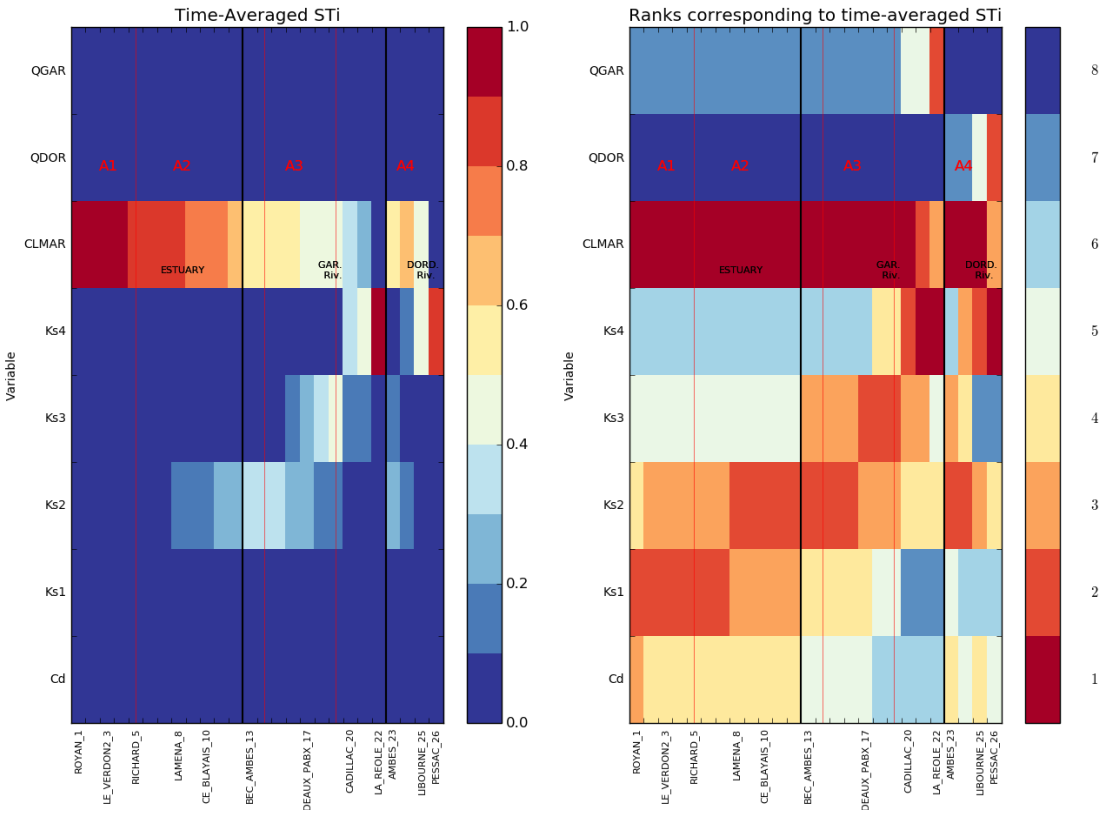



Fig. 10: representation of time averaged STi (left) and of the rank of the input variable with respect to STi (right) along the Gironde estuary during the February 2003 storm.

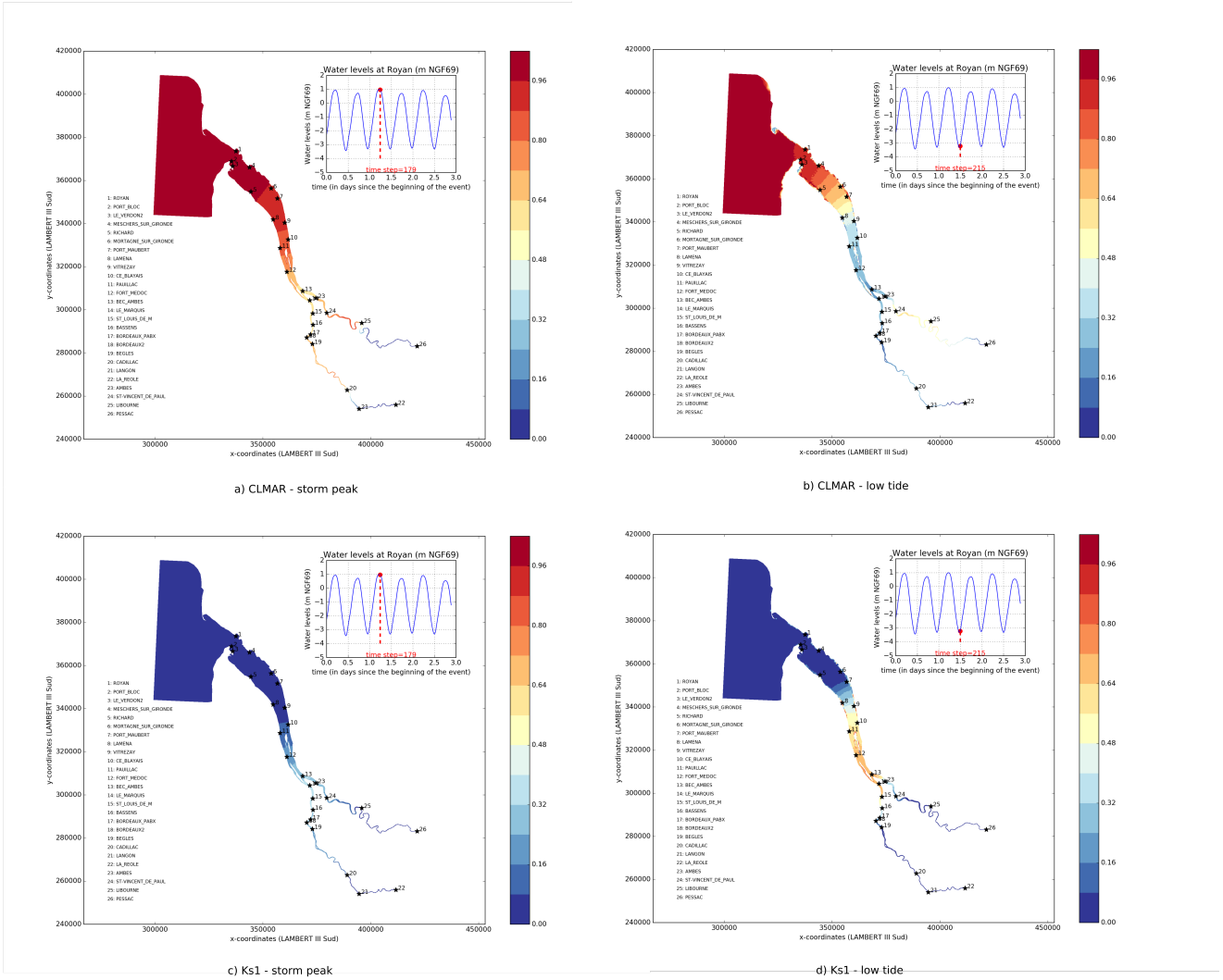



Fig. 11: Total Sobol' indices (STi) in the Gironde estuary at (a) (resp. (c)) storm peak and (b) (resp. (d)) low tide for uncertain variable maritime boundary condition CLMAR (resp. friction coefficient Ks1).

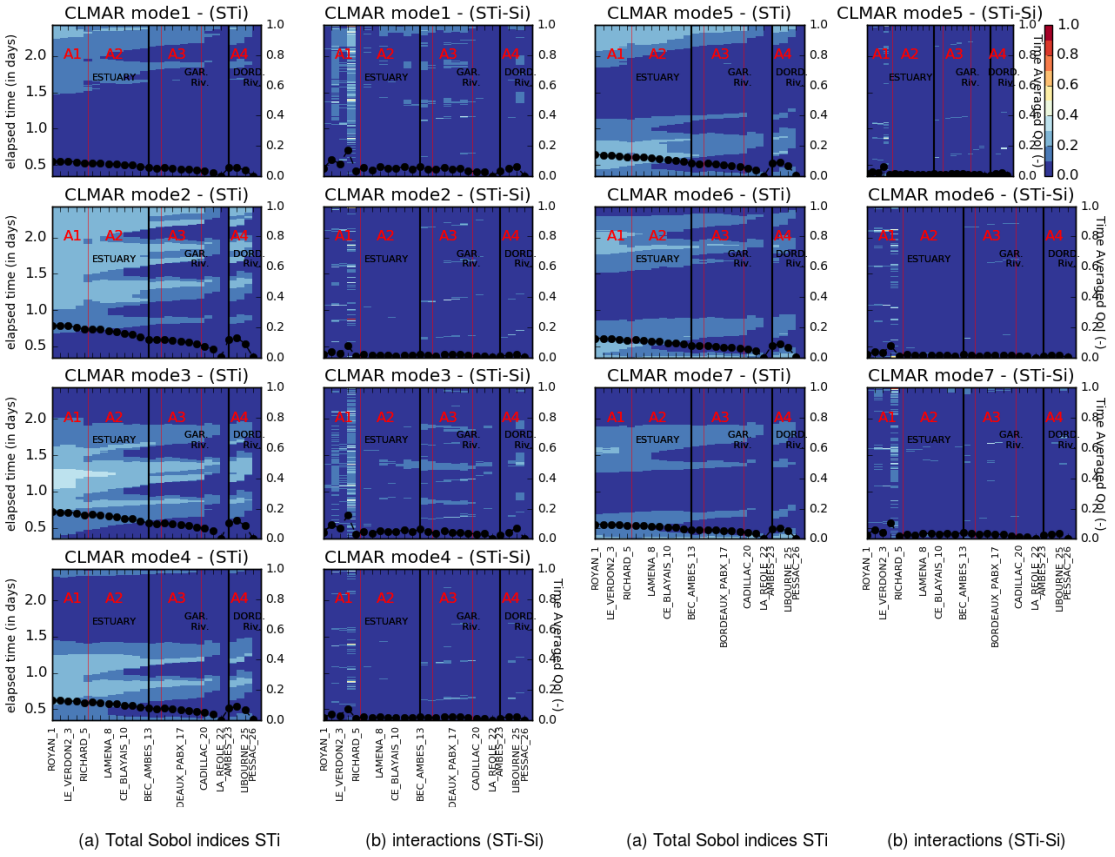



Fig. 12: Total Sobol' indices (STi) in panels a- (resp. interactions STi - Si in panels b-) along the Gironde estuary during the February 2003 storm for 7 uncertain input variables (maritime boundary (CLMAR) eigenmodes : mode1, mode2, mode3, mode4, mode5, mode6, mode7)

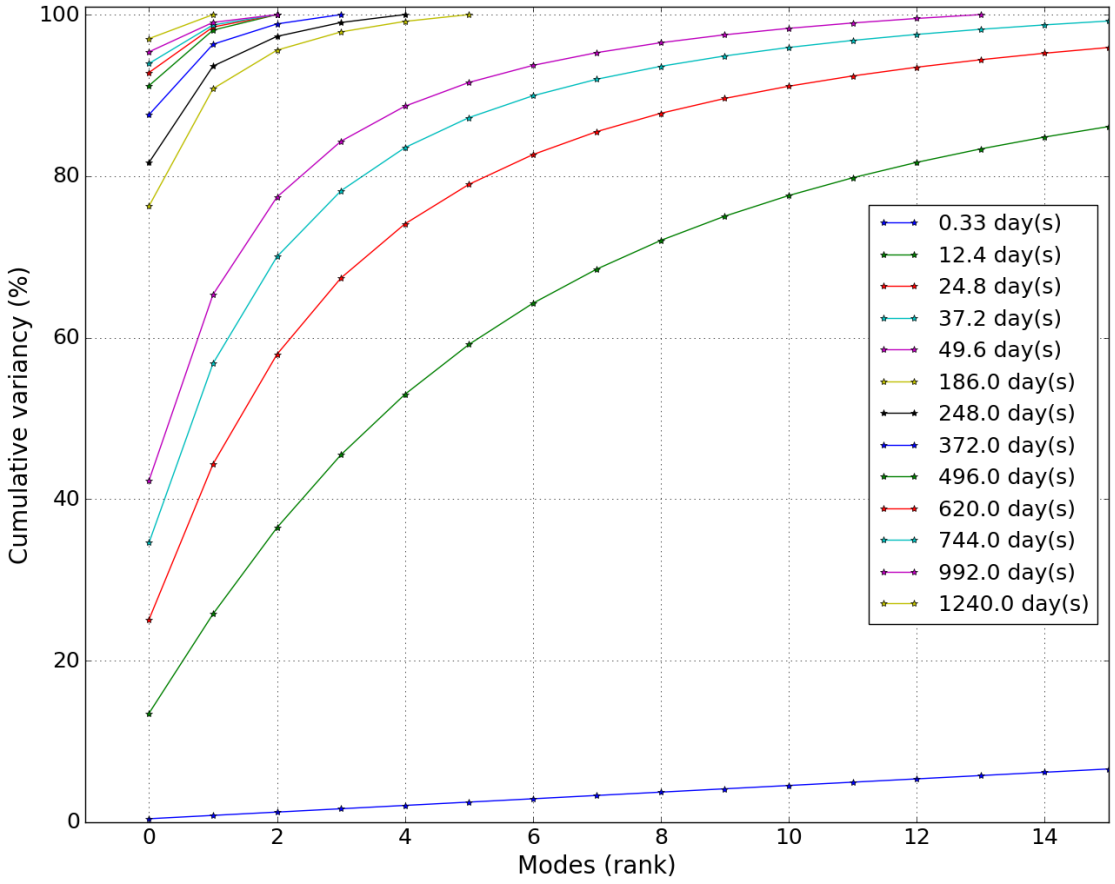



Fig. 13: Cumulated variance for the 14 first modes for several correlation time scales.

### 2.2.2 Information brought during the reviewing process

A couple of points were raised during the review process and are brought here as complementary material to the article.

#### 2.2.2.1 Ranges of the friction and the wind drag coefficients (reported in Table 1)

The ranges of the friction coefficients are chosen so as to include the calibration values for the 2003 event using both $Nash_{HT}$ and $RMSE$ criteria. The interval corresponds to the commonly accepted uncertainty on friction for engineering studies.
Concerning the wind drag coefficient $Cd$, a large bibliography was achieved in Levy (2003). Several parametric formulations are given by Flathers (1976), Wu (1980), Yelland and Taylor (1996). These parametric values were computed by the author for the Gironde case using climatological wind intensities (Climate Forecast System Reanalysis from NOAA) as presented in Fig. 2.1. It should first be noted that the calibration value issued from the Telemac2D calibration for $Cd$ ($Cd_{calib} = 2.14 * 10^{-6}$) is included within the parametric ranges. The parametric ranges are used to define the ranges for the uniform distribution of $Cd$.

Levy [2013] presents a review of parametric formulations of the wind drag coefficient based on Flather [1976], Wu [1980], Yelland and Taylor [1996]. These parametric formulations were used for the Gironde case using climatological wind intensities (Climate Forecast System Reanalysis from NOAA) to define the range for the $Cd$ uniform distribution. The ranges of the friction coefficients are chosen so as to include the calibration values for the 2003 event using both the $\boldsymbol{Nash_{HT}}$ and ***RMSE*** criteria. The interval corresponds to the commonly accepted uncertainty on friction for engineering studies.

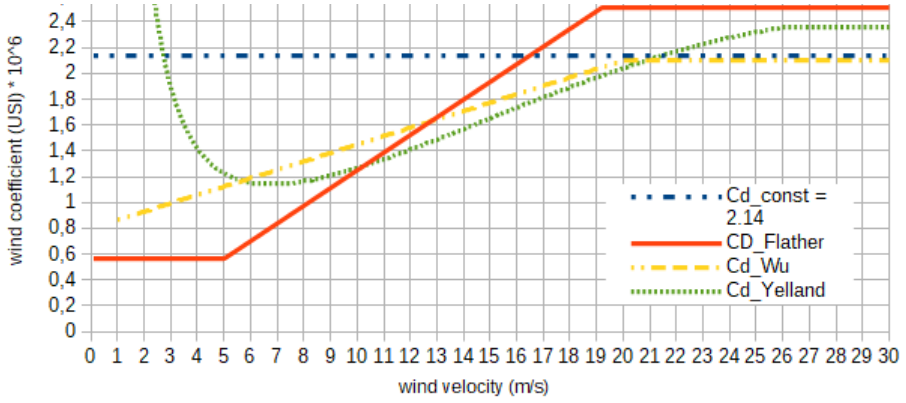



FIGURE 2.1 : Values of the wind drag coefficient Cd according to formulation from Flather [1976], Wu [1980], Yelland and Taylor [1996]

#### 2.2.2.2 Members convergence

The convergence of the Sobol' sensitivity indices was investigated with an increasing number of members $Ne$ ranging from 100 to 10000 as shown in Fig. 2.2 for all random variables at time step 50 for observing stations along the estuary. The same diagnostic was carried out over the entire simulation period. It should be noted that the Sobol' indices reach convergence for $Ne$ = 2000.

#### 2.2.2.3 Limitations of the study

One should notice the importance of the limited number of events for the simulations (one seven-day event) as it cannot physically be interpreted as the complete set of all events that can occur in this area. Even if the study was carried out over a single event, yet, this flood event was used for calibration and validation of the numerical model, not for nuclear plant design, and, however, the methodology for sensitivity analysis may apply to any event. Indeed, the concepts and methodology can also be applied to general design specifications by using the adequate calibrated numerical model and adapting the ranges for numerical parameters such as the Strickler and wind drag coefficients or river discharges and surge levels' uncertainties. In case of additional forcing or parameters (dike failure time duration, length and width, for example), the GSA could easily be adapted to additional sources of uncertainties.

The operation of a nuclear power plant strategy differs from that for design. The present paper aims at providing solutions for flood monitoring and alert dedicated to security of infrastructures and people. The Telemac2D model is integrated in real-time to forecast water level that are used for flood monitoring at Blayais. Nuclear power plant design is beyond the scope of the study.
The numerical model is here dedicated to flood forecast and was calibrated on the 2003 event on which the Global Sensitivity analysis is applied.
The 2003 event, on which the concept has been applied here, is characterized by a tide coefficient equal to 85 and instantaneous storm surges reaching 20 cm, leading to observed water levels around 3 m NGF69 at Pauillac, near the Blayais nuclear plant. In the middle part of the Gironde estuary, this water level is consequently regularly reached (Laborie et al. [2014]). Nevertheless, return periods for Garonne and Dordogne discharges for this event respectively correspond to (10-17) years and (2-5) years and the 2003 event is a flood event for both rivers without overflowing. The 2003 event is thus a normal routing event against which the Blayais nuclear plant is protected.

#### 2.2.2.4 Rank and fiability diagram for all stations

The validation results for the 12 stations whose ranks are represented on Fig. 5 have been investigated so that the rank (resp. reliability) diagrams are available for all stations as shown on Fig. 2.3 and Fig. 2.4 (resp. Fig. 2.5).

Concerning rank diagrams, Fig. 2.4 shows, as emphasized in section 3.1.2, that stations not plotted on Fig. 4 show similar results as bec d'Ambès, except Pessac and La Réole, whose rank diagrams plotted on Fig. 2.3 are closed to that observed at Le Verdon.

Concerning reliability diagrams, Fig. 2.5 shows that reliability diagrams for stations located in the central part of the Gironde estuary are also similar to bec d'Ambès one, while the reliability diagram for Pessac is similar to Richard's one for high tides and to La Réole for low tides.

## 2.3 Conclusions et perspectives

Le modèle numérique Telemac2D utilisé pour la prévision opérationnelle des crues par le SPC-GAD dans l'estuaire de la Gironde actualisé (MEG0V7P3 et MEG3V7P3) a été utilisé pour réaliser une analyse de sensibilité globale (GSA) par décomposition de variance (ANOVA) afin d'estimer les indices de Sobol' de chaque variable pendant la tempête de 2003. Il a été démontré que la condition à la frontière maritime détermine la dynamique de l'estuaire.

L'influence du coefficient de frottement augmente de l'embouchure de l'estuaire vers l'amont de la Garonne et de la Dordogne. Les forçages hydrologiques ont une influence très locale en amont de chaque rivière. Cette GSA permet d'identifier les sources d'incertitude les plus significatives et leur évolution temporelle sur plusieurs cycles de marée. Une fois identifiées, ces incertitudes devraient être réduites, par exemple avec un algorithme d'assimilation de données tel que le filtre de Kalman d'Ensemble, afin d'améliorer le niveau d'eau dans l'estuaire en modes simulation et prévision. Le développement d'une méthodologie d'assimilation de données pour la correction des forçages fonctionnels dans l'espace réduit de Karhunen-Loève et des paramètres est développée dans le chapitre 3. Son application au modèle Telemac2D actualisé de l'estuaire de la Gironde, ainsi que sa validation dans un cadre synthétique et l'évaluation des performances obtenues, sont décrites dans les chapitres 4 à 6.

L'une des perspectives de cette étude de quantification d'incertitudes (UQ) sera de prendre en compte les incertitudes des champs atmosphériques (vent et pression) qui sont des forçages spatio-temporels bidimensionnels. D'autres travaux devraient également porter sur la formulation d'un modèle de substitution (métamodèle) pour le modèle hydrodynamique 2D de l'estuaire de la Gironde afin de réduire le coût de la GSA et de la méthode d'assimilation de données. Des métamodèles fondés sur les processus gaussiens ou les polynômes du chaos ont été utilisés pour des études de UQ et de GSA dans le domaine de la modélisation hydro-environnementale (Marrel [2008], Goutal et al. [2018]). Les deux méthodes se sont avérées efficaces pour la métamodélisation des SWE avec le solveur 1D Mascaret (El Moçayd [2017], Roy et al. [2018]), de même que pour l'estimation par une GSA des indices de Sobol' et de la matrice de covariance. Un métamodèle a également été mis en place pour Telemac2D sur la rivière Garonne (en amont de l'emprise du modèle de l'estuaire de la Gironde) et a permis un calcul à coût réduit des indices de sensibilité (Goutal et al. [2018]). Ces travaux devraient être étendus à l'estuaire de la Gironde en

prenant en compte compte les incertitudes scalaires et spatio-temporelles. La combinaison d'une méta-modélisation et d'une méthode d'assimilation des données pourrait permettre de répondre à ce besoin en obtenant un algorithme d'assimilation des données ensembliste à coût de calcul plus faible pour réduire les principales sources d'incertitude et améliorer la prévision des niveaux d'eau dans l'estuaire de la Gironde.

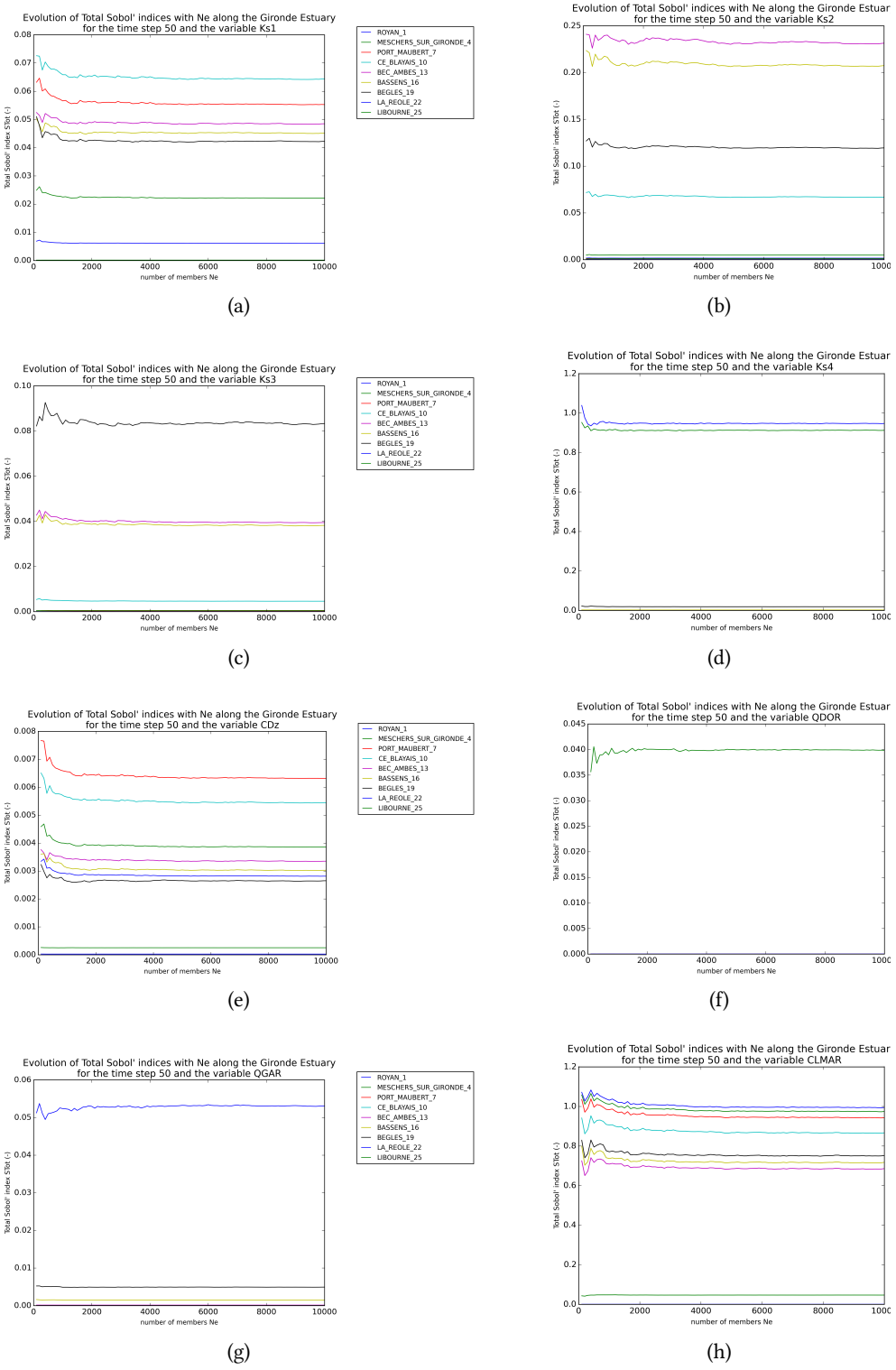



FIGURE 2.2 : Convergence of the Sobol' indices with Ne at time step 50 for observing stations along the estuary for (a) Ks1, (b) Ks2, (c) Ks3, (d) Ks4, (e) Cd, (f) QDOR, (g) QGAR, (h) CLMAR.

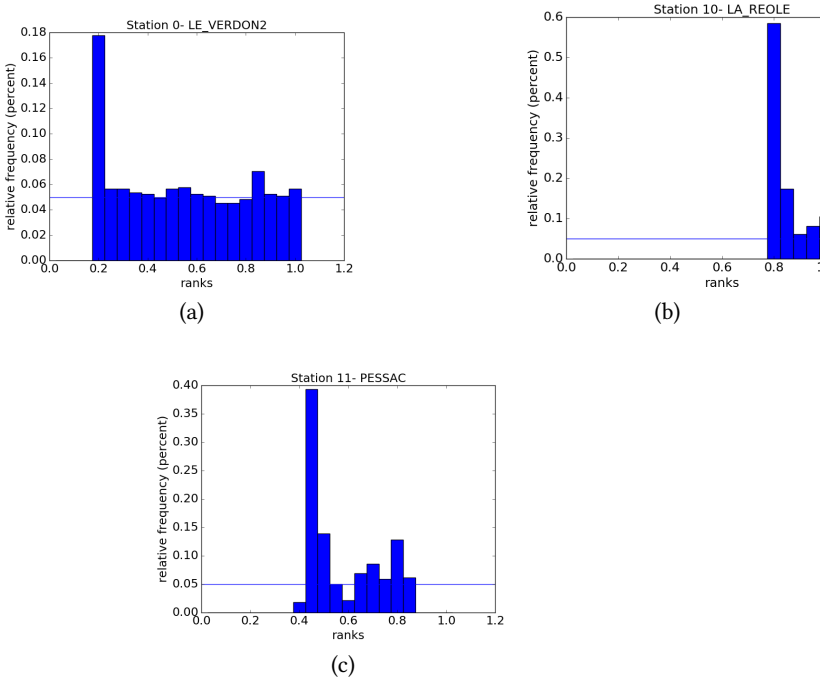



FIGURE 2.3 : Ranks diagrams for 3 stations along the Gironde estuary similar to station 0 Le Verdon.

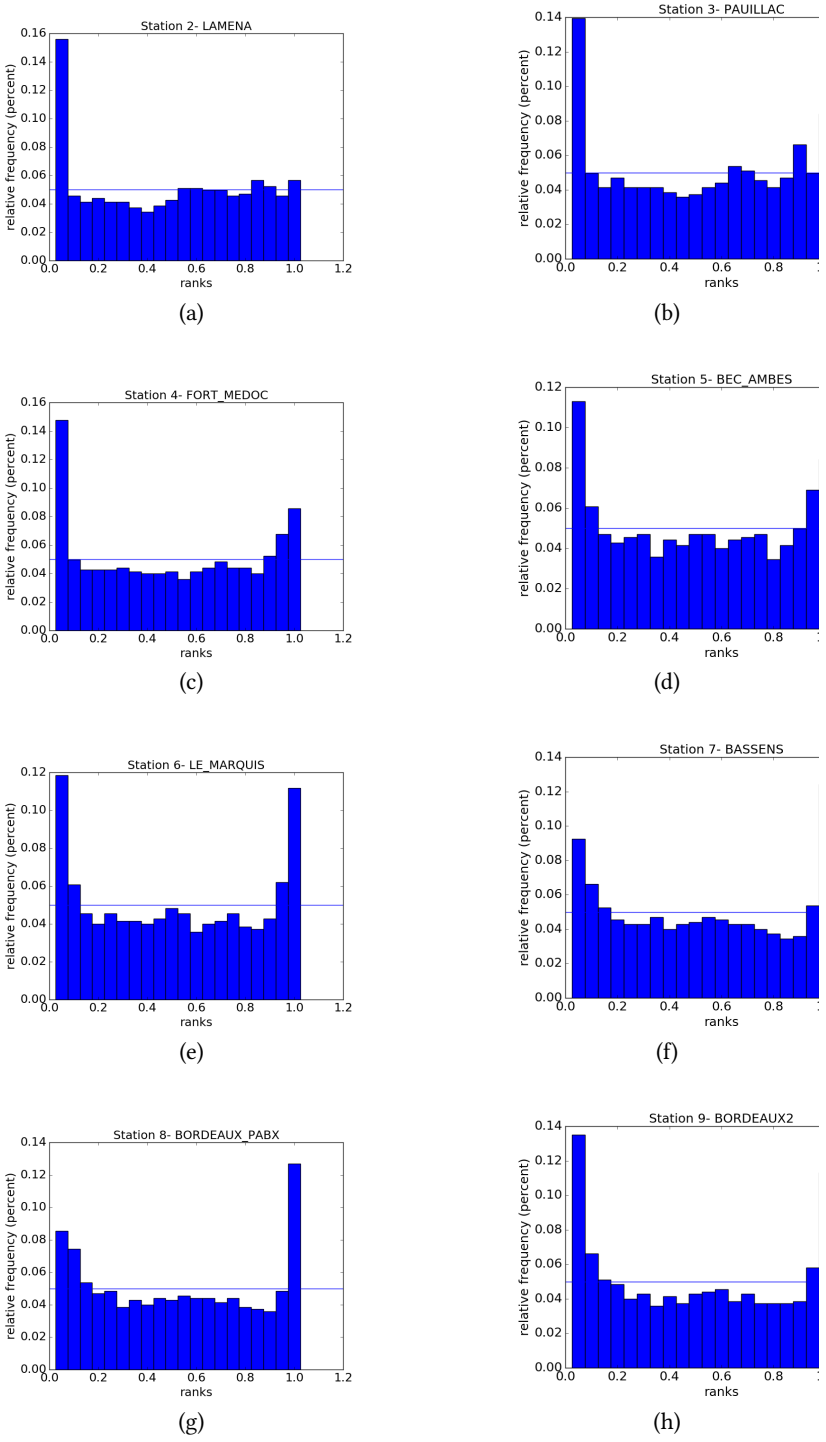
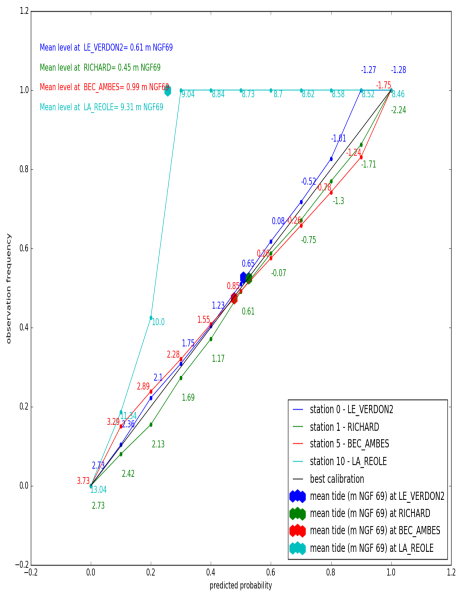
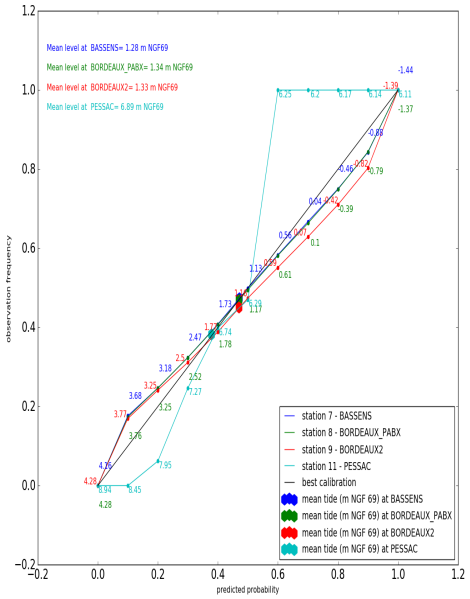


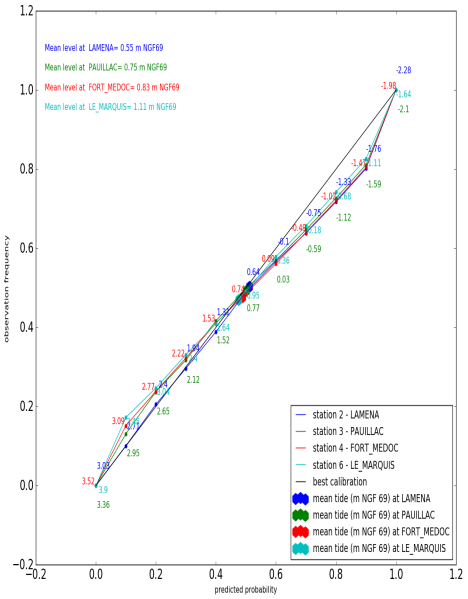

FIGURE 2.4 : Ranks diagrams for 8 stations along the Gironde estuary similar to station 5 bec d'Ambès.




(a)




(b)




(c)

FIGURE 2.5 : Reliability diagram for all stations along the Gironde estuary with corresponding observed water level quantiles during the 2003 event. Mean water levels at each station are represented with hexagons.

# Chapitre 3

# Méthodes d'assimilation de données : problématique et état de l'art

## Sommaire


**3.1 Introduction . . . . . 79**
**3.2 Formulation du problème dans un cadre général . . . . . 79**
3.2.1 Modèle d'évolution dynamique du système . . . . . 80
3.2.2 Le modèle d'observation . . . . . 80
**3.3 Formulation bayésienne du problème d'estimation de l'état du système 81**
**3.4 Les méthodes séquentielles (ou filtres) . . . . . 83**
3.4.1 Le filtre de Kalman dans le cadre linéaire et de l'approximation gaussienne . . . . . 84
3.4.2 Le filtre de Kalman étendu . . . . . 85
3.4.3 Le filtre de Kalman d'ensemble stochastique . . . . . 86
3.4.3.1 Cas général où le vecteur de contrôle est le vecteur d'état du système . . . . . 86
3.4.3.2 Estimation conjointe des paramètres du modèle et de l'état du système . . . . . 89
**3.5 Les méthodes variationnelles et variationnelles d'ensemble . . . . . 91**
**3.6 Application à la modélisation hydro-environnementale . . . . . 93**
3.6.1 Panorama dans le domaine des géosciences . . . . . 93
3.6.2 Correction de l'état hydraulique (niveau d'eau, vagues ou surcotes) . . 94
3.6.2.1 En milieu océanique et côtier . . . . . 94
3.6.2.2 En milieu estuarien . . . . . 95
3.6.2.3 Prise en compte de l'incertitude sur les conditions aux limites dans l'ébauche de l'ensemble . . . . . 96
3.6.3 Correction des conditions aux limites et de l'état du système . . . . . 97
3.6.4 Estimation de l'état et des paramètres pour la correction des niveaux d'eau . . . . . 97

3.6.4.1 En rivière . . . . . 97
3.6.4.2 En milieu estuarien . . . . . 98
3.6.5 Nature des observations . . . . . 99
3.6.6 Prévision dans un cadre opérationnel . . . . . 99
**3.7 Conclusions . . . . . 100**

## 3.1 Introduction

Les méthodes d'assimilation de données (AD) cherchent à répondre au problème d'estimation de l'état d'un système (qu'il soit atmosphérique, océanique, estuarien) suivant : comment combiner la modélisation numérique imparfaite de l'évolution de l'état d'un système et des observations ponctuelles dans le temps et dans l'espace, elles-mêmes entachées d'incertitudes, pour estimer l'état "vrai" de l'ensemble du système et, mieux, le prévoir ?

Les méthodes d'AD, initialement plutôt réservées à la communauté météorologique, se sont répandues dans tous les domaines de la géoscience, y compris l'hydraulique de surface et souterraine, depuis une dizaine d'années. Elles sont généralement divisées en deux groupes : les méthodes variationnelles et les méthodes séquentielles, toutes deux fondées sur l'hypothèse d'approximation Gaussienne des erreurs du modèle et d'observation. Elles sont équivalentes dans un cadre linéaire. Asch et al. [2016] et Carrassi et al. [2018] dressent un panorama de l'état de l'art des deux types de méthodes d'AD dans le domaine des géosciences, ainsi que des algorithmes sous-jacents en lien avec ses applications pratiques. Les nouvelles méthodes hybrides qui associent des concepts variationnels et ensemblistes, ainsi que le filtre particulaire et les méthodes de couplage, sont également présentés dans ces deux références.

La section 3.2 s'appuie sur Carrassi et al. [2018] pour formuler le problème dans un cadre général d'estimation de l'état du système. Elle présente successivement le modèle d'évolution dynamique du système (section 3.2.1) et l'opérateur d'observation (section 3.2.2). La formulation bayésienne du problème d'estimation est décrite dans la section 3.3.
La section 3.4 détaille les méthodes séquentielles par filtrage. Ainsi, la section 3.4.1 présente la solution optimale obtenue à l'aide du filtre de Kalman (KF) dans un cadre linéaire sous hypothèse d'approximation gaussienne des erreurs d'observation et du modèle. La section 3.4.3 présente la méthode du filtre de Kalman d'Ensemble (EnKF) dans un cadre général (section 3.4.3.1), puis dans le cas particulier de l'estimation conjointe de l'état et des paramètres du système (section 3.4.3.2).
Les méthodes variationnelles et variationnelles d'ensemble sont abordées dans la section 3.5.
La section 3.6 fait, enfin, l'état de l'art de l'utilisation des travaux utilisant le KF dans le domaine de l'hydrodynamique côtière, estuarienne et fluviale, pour ce qui concerne l'inférence des paramètres d'un modèle hydroenvironnemental et la correction des conditions aux limites (BC), de manière conjointe ou non avec l'état du système.

## 3.2 Formulation du problème dans un cadre général

Cette section est fondée sur Carrassi et al. [2018].

### 3.2.1 Modèle d'évolution dynamique du système

Supposons l'existence d'un modèle d'évolution d'un système dynamique discrétisé dans le temps tel que :

$$\mathbf{x}_{k+1} = \mathcal{M}_{k+1:k}(\mathbf{x}_k, \boldsymbol{\gamma}) + \boldsymbol{\eta}_{k+1}, \tag{3.1}$$

où $\mathbf{x}_k = \mathbf{x}(t_k) \in \mathbb{R}^m$ représente l'état du système au temps $t_k$ et $\boldsymbol{\gamma} \in \mathbb{R}^p$ l'espace des paramètres du modèle, indépendant ou non du temps. $\mathcal{M}_{k+1:k}$ est non-linéaire par rapport à $\mathbf{x}_k$ et $\boldsymbol{\gamma}$ et permet de propager l'état du système de $t_k$ à $t_{k+1}$.
$\boldsymbol{\eta}_{k+1} \in \mathbb{R}^m$ est l'erreur du modèle représentée par un terme additif stochastique. $\boldsymbol{\gamma}$, l'espace des paramètres, est susceptible d'inclure les forçages extérieurs et les conditions aux limites du modèle. Dans ce cas, $\boldsymbol{\eta}$ dépend du temps. Le modèle d'évolution "vrai" du processus est inconnu. $\boldsymbol{\eta}_{k+1}$ représente l'erreur de $\mathcal{M}_{k+1:k}$ et prend en compte l'effet cumulé, sur l'intervalle $[t_k\,;t_{k+1}]$, des erreurs sur les paramètres et forçages, mais aussi celles liées à la modélisation mathématique et numérique utilisée pour intégrer l'équation (3.1), ainsi que les effets d'échelle, de dimension et de discrétisation.

### 3.2.2 Le modèle d'observation

Des observations "bruitées" de l'état du système $\mathbf{x}_k$ sont disponibles au temps $t_k$. Elles sont représentées par les composantes du vecteur d'observation $\mathbf{y}_k^0 \in \mathbb{R}^d$ où $d$ est le nombre total d'observations disponibles à $t_k$. En supposant le bruit additif, alors le vecteur d'observation est relié au vecteur d'état du système par :

$$\mathbf{y}_k^0 = \mathcal{H}_k(\mathbf{x}_k, \boldsymbol{\gamma}) + \boldsymbol{\epsilon}_k. \tag{3.2}$$

L'équation (3.2) fait intervenir l'opérateur d'observation, $\mathcal{H}_k : \mathbb{R}^m \longrightarrow \mathbb{R}^d$, qui permet de passer de l'espace de l'état du modèle à l'espace des observations. Cet opérateur peut contenir lui-même des interpolations spatiales ou des produits de convolution. On suppose par la suite que l'espace des observations est de dimension constante $d$, alors que ce n'est pas forcément le cas dans la réalité, puisque le nombre d'observations disponibles peut différer d'un temps à l'autre.
De la même manière que pour l'erreur du modèle $\boldsymbol{\eta}_k$, l'erreur sur l'opérateur d'observation $\boldsymbol{\epsilon}_k$ est représentée comme un terme additif stochastique. Il prend en compte l'erreur instrumentale des dispositifs d'observation, l'incertitude dans la formulation de l'opérateur d'observation lui-même et l'erreur de représentativité. Cette dernière découle de la présence d'échelles non résolues et représente leur effet sur les échelles résolues par le modèle. Elle est difficile à estimer en raison du manque d'observations à petite échelle, mais omniprésente en géosciences. Il faut noter que l'hypothèse de bruit "additif" relève également de la simplification puisqu'une formulation plus générale du type $\mathbf{y}_k^0 = \mathcal{H}_k(\mathbf{x}_k, \boldsymbol{\eta}, \epsilon_k)$ pourrait être retenue.

En géosciences, le nombre d'observations $d$ est généralement très inférieur à la taille $m$ du système et est insuffisant pour décrire de manière complète son état. De fait, le modèle sert à remplir les lacunes spatio-temporelles laissées par le réseau des observations en propageant dans le temps et dans l'espace l'information des zones "observées" aux zones qui ne le sont pas. Malgré la disponibilité croissante de mesures, qu'elles soient ***in situ*** ou satellitaires, les méthodes d'AD permettent d'utiliser au mieux les données disponibles afin d'atteindre un rapprochement efficace des données et du modèle dans un sens optimal de minimisation de la variance (méthodes séquentielles par filtrage ) ou de maximum de vraisemblance (méthodes variationnelles). Cet objectif est un véritable challenge tant en termes de formulation mathématique que de ressources informatiques.

## 3.3 Formulation bayésienne du problème d'estimation de l'état du système

Du fait du caractère aléatoire de l'erreur du modèle et de l'erreur sur l'observation, il est possible de décrire l'état du système et l'observation en termes de densité de probabilité (pdf). L'approche bayésienne offre alors un cadre à la formulation du problème d'AD pour l'estimation de l'état du système. Le théorème de Bayes' permet d'écrire la probabilité conditionnelle de l'état du système inconnu $\mathbf{x}$ connaissant la mesure $\mathbf{y}$ à l'aide des probabilités marginales de $\mathbf{x}$ et de $\mathbf{y}$ :

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \tag{3.3}$$

L'équation (3.3) met en lumière les trois composantes principales de l'inférence bayésienne. $p(\mathbf{x})$ est la probabilité ***a priori*** de $\mathbf{x}$ qui rassemble l'information sur $\mathbf{x}$ avant l'assimilation de nouvelles observations. $p(\mathbf{y}|\mathbf{x})$ est la fonction de vraisemblance de $\mathbf{y}$ connaissant $\mathbf{x}$. $p(\mathbf{y})$ est la probabilité marginale ou ***a priori*** de l'observation avec $p(\mathbf{y}) = \int d\mathbf{x}\, p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x})$. Cette dernière quantité, indépendante de $\mathbf{x}$, est un facteur de normalisation.
En supposant les erreurs discrétisées sur le vecteur temporel $[t_1\,;\,t_L]$ du modèle, $\{\boldsymbol{\eta}_k\}_{k\in(1,\cdots,L)}$, et d'observation, $\{\boldsymbol{\epsilon}\}_{k\in(1,\cdots,L)}$, indépendantes, décorrélées dans le temps et distribuées selon les pdfs $p_{\boldsymbol{\eta}}$ et $p_{\boldsymbol{\epsilon}}$, en s'affranchissant pour l'instant des paramètres $\boldsymbol{\gamma}$ et en réintroduisant le temps, il vient :

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}) = p_{\boldsymbol{\eta}}\left[\mathbf{x}_k - \mathcal{M}_{k:k-1}\left(\mathbf{x}_{k-1}\right)\right], \tag{3.4}$$

$$p(\mathbf{y}_k|\mathbf{x}_k) = p_{\boldsymbol{\epsilon}}\left[\mathbf{y}_k - \mathcal{H}_k\left(\mathbf{x}_k\right)\right]. \tag{3.5}$$

Dans un intervalle $[t_0,t_L]$, les états séquentiels du système et les observations sont respectivement notés $\mathbf{x}_{L:0} = \{\mathbf{x}_L, \mathbf{x}_{L-1}, \cdots, \mathbf{x}_0\}$ et $\mathbf{y}_{L:1} = \{\mathbf{y}_L, \mathbf{y}_{L-1}, \cdots, \mathbf{y}_1\}$. Il existe alors trois

problèmes d'estimation de l'état du système selon la période des observations utilisée :

1. le problème de **filtrage** : il s'agit d'estimer $p(\mathbf{x}_k|\mathbf{y}_{k:1})$ ;

2. le problème de **lissage** : il s'agit d'estimer $p(\mathbf{x}_{L:0}|\mathbf{y}_{L:1})$ ou les probabilités marginales de la pdf, telle que $p(\mathbf{x}_k|\mathbf{y}_{L:1})$ avec $0 \leq k < L$ ;

3. le problème de **prévision** : il s'agit d'estimer $p(\mathbf{x}_l|\mathbf{y}_{k:1})$ avec $l > k$.

La figure 3.1 est extraite de Carrassi et al. [2018]. Elle représente schématiquement les problèmes de filtrage et de prévision cités précédemment. Le signal vrai (resp. analysé et prédit) est représenté par une ligne continue bleue (resp. rouge et verte). Les observations sont représentées par des étoiles bleues et l'état analysé par des carrés rouges. Les incertitudes sur les observations, la prévision et l'analyse sont respectivement représentées par des ellipses de taille proportionnelle à l'incertitude représentée.

Le filtrage, très courant en géosciences, consiste, dans une chaîne séquentielle, à utiliser les observations dès qu'elles sont disponibles. La méthode consiste alors à alterner une phase d'analyse dont l'objet est d'actualiser $p(\mathbf{x}_k|\mathbf{y}_{k:1})$ en utilisant la dernière observation, $\mathbf{y}_l$, avec une phase de prévision qui consiste à propager la pdf de l'état du système au pas de temps suivant jusqu'à ce qu'une nouvelle mesure soit disponible. La phase d'analyse est fondée sur la formule de Bayes' (équation (3.3)). Le processus est alors répété de manière séquentielle, le résultat de l'équation précédente fournissant la distribution marginale pour le pas de temps suivant dont il devient l'ébauche (ou background).
Ainsi, dans le cas du filtrage tel que représenté sur la figure 3.1, l'état prédit $\mathbf{x}^f_{k+1}$ est obtenu à l'aide de l'état analysé $\mathbf{x}^a_k$ propagé jusqu'à l'observation suivante à $t_{k+1}$. Un état analysé $\mathbf{x}^a_{k+1}$ est alors calculé à l'aide de l'état prédit $\mathbf{x}^f_{k+1}$ et de l'observation $\mathbf{y}_{k+1}$. L'état analysé est statistiquement plus proche (dans un sens optimal à préciser) de l'état "vrai" $\mathbf{x}^t_{k+1}$ et la distance entre l'état analysé et l'état vrai diminue au cours du temps. Il est alors possible de renouveler l'étape de prédiction à partir de l'état analysé $\mathbf{x}^a_{k+1}$ (carré rouge) à $t_{k+1}$. Ainsi, l'état prédit $\mathbf{x}^f_{k+2}$ est obtenu en propageant $\mathbf{x}^a_{k+1}$. S'ensuit une étape d'analyse en utilisant l'observation suivante $\mathbf{y}_{k+2}$. L'information est apportée par le passé et actualisée par les données d'observation du pas de temps courant.

Le lissage consiste à estimer la pdf conditionnelle de l'état du système à tout pas de temps $t_k$, $0 \leq k \leq L$, en utilisant toutes les observations (passées, présentes et futures par rapport à $k$). Il s'agit de procéder, une fois les observations collectées, à une analyse "rétrospective" qui peut être utile pour mettre en œuvre une réanalyse de l'état du système aussi bien qu'une estimation des paramètres $\boldsymbol{\gamma}$. Les réanalyses sont ainsi courantes dans le domaine de la climatologie (Dee et al. [2011]).

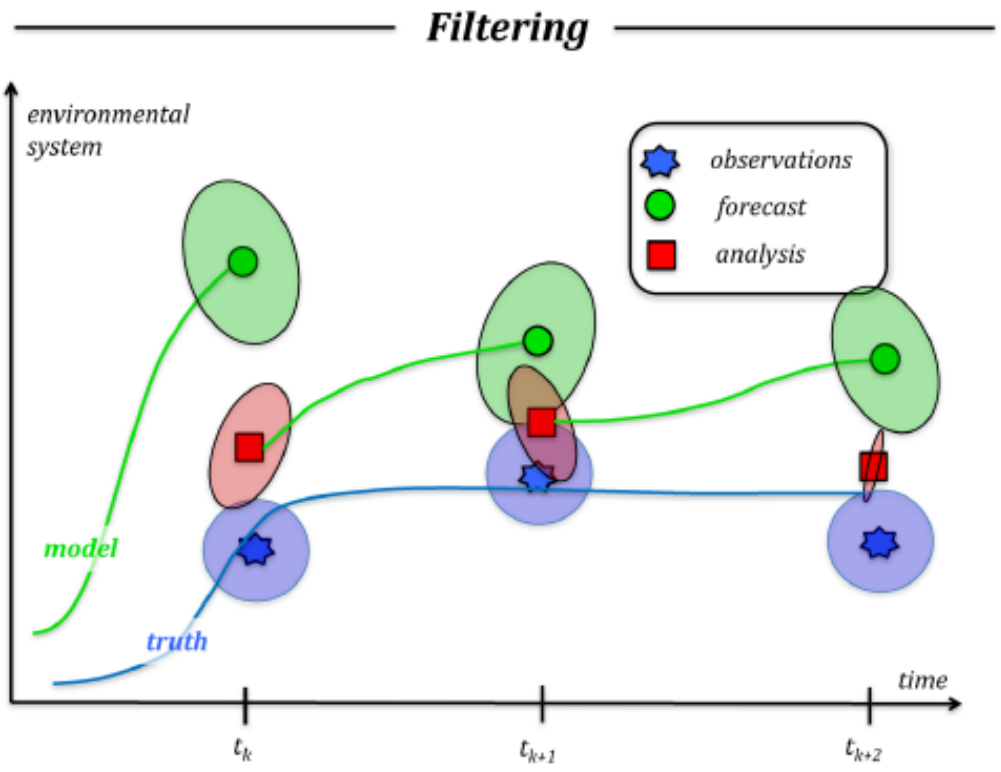



FIGURE 3.1 : Représentation de l'évolution de l'état du système pour un problème de filtrage (source : Carrassi et al. [2018])

## 3.4 Les méthodes séquentielles (ou filtres)

La méthode d'assimilation utilisée dans le cadre de ce travail est séquentielle et fondée sur le KF (Kalman [1960]). Le KF, décrit dans la section 3.4.1, permet d'estimer l'état optimal d'un modèle en combinant de manière séquentielle le modèle et les observations tout en prenant en compte leurs incertitudes respectives. Une description complète du KF peut être trouvée dans Jazwinski [2007] et Maybeck [1982]. Les difficultés d'utilisation du KF résident dans l'hypothèse de linéarité et le problème de dimensions. L'EnKF (Evensen [2003]), décrit dans la section 3.4.3, permet de résoudre ces problèmes et peut être développé facilement, au prix cependant d'un coût de calcul qui croît avec la dimension de l'ensemble. L'EnKF procède, en effet, à une approximation des équations du KF par une approche de type Monte Carlo pour la représentation des incertitudes. Il repose sur l'estimation des deux premiers moments statistiques, c'est-à-dire la moyenne et l'écart-type (ou la covariance) de l'ensemble plutôt que sur l'estimation complète de la loi de probabilité de la variable de sortie du modèle, dont le coût en ressources informatiques serait prohibitif.

### 3.4.1 Le filtre de Kalman dans le cadre linéaire et de l'approximation gaussienne

Les pdfs issues de la résolution des problèmes de lissage, de filtrage et de prévision permettent de décrire de manière complète la probabilité de tous les états possibles du système. Il importe alors de choisir un estimateur de ces pdfs afin d'estimer l'état du système. La moyenne et le mode sont deux estimateurs possibles. Les méthodes dont le but est d'estimer la moyenne de l'état du système sont fondées sur la minimisation de l'erreur quadratique moyenne. L'état analysé doit être aussi proche que possible de l'état vrai dans le sens du minimum de variance. Le mode, quant à lui, représente le pic de la distribution et l'état du système le plus probable. De fait, ces méthodes sont des méthodes qui cherchent le maximum de vraisemblance ***a posteriori***.
Dans la pratique, il n'est pas possible de mettre en œuvre l'approche bayésienne complète du fait des ressources informatiques que les dimensions des problèmes traités nécessiteraient. Pour pallier ces difficultés, l'hypothèse utilisée couramment consiste à supposer que les incertitudes sur les observations, le modèle et les marginales sont gaussiennes. De fait, cette hypothèse conduit à des simplifications substantielles, dans la mesure où les pdfs peuvent être alors complètement décrites par leurs premier et second moments, c'est-à-dire la moyenne et la matrice de covariance.

Supposons que $\mathbf{x}$ est le vecteur d'état du système confondu ici avec le vecteur de contrôle du KF et que le modèle dynamique $\mathcal{M}$ et l'opérateur d'observation $\mathcal{H}$ sont linéaires, les équations (3.1) et (3.2) s'écrivent :

$$\mathbf{x}_k = \mathbf{M}_{k:k-1}\mathbf{x}_{k-1} + \boldsymbol{\eta}_k, \quad \boldsymbol{\eta}_k \sim \mathcal{N}(0, \mathbf{Q}_k), \tag{3.6}$$

$$\mathbf{y}_k^0 = \mathbf{H}_k\mathbf{x}_k + \boldsymbol{\epsilon}_k, \quad \boldsymbol{\epsilon}_k \sim \mathcal{N}(0, \mathbf{R}_k), \tag{3.7}$$

où $\mathbf{M}_{k:k-1}$ et $\mathbf{H}_k$ sont respectivement des matrices de dimension $m \times m$ et $d \times m$. Les erreurs du modèle et d'observation sont supposées décorrélées entre elles et en temps, sans biais et gaussiennes de matrices de covariance respectives $\mathbf{Q}_k \in \mathbb{R}^{m,m}$ et $\mathbf{R}_k \in \mathbb{R}^{d,d}$.

Dans ces hypothèses gaussiennes et linéaires, le KF est la solution exacte. Elle est optimale pour la minimisation de la variance de l'erreur d'analyse, qui permet d'estimer de manière récursive la moyenne conditionnelle de la pdf du filtrage $p(\mathbf{x}_k|\mathbf{y}_{k:0})$, ainsi que l'incertitude associée, c'est-à-dire sa covariance conditionnelle. Selon un processus séquentiel, les observations sont assimilées dans l'ordre chronologique. Comme illustré par la figure 3.1, le KF alterne une étape de prévision, pendant laquelle la moyenne et la covariance de la pdf conditionnelle de la prévision, $p(\mathbf{x}_k|\mathbf{y}_{k-1:0})$, est propagée, avec une phase d'analyse au cours de laquelle la moyenne et la covariance de la pdf conditionnelle, $p(\mathbf{x}_k|\mathbf{y}_{k:0})$, est actualisée. Du fait de l'hypothèse gaussienne, la moyenne et la covariance suffisent à décrire les pdfs et le KF fournit alors un système fermé d'équations. Les formules classiques du KF sont fournies ci-dessous. Il en existe de nombreuses formes dérivées décrites de manière non exhaustive dans les sections 3.4.2 et 3.4.3.

En notant respectivement $\mathbf{x}^{f/a}$ et $\mathbf{P}^{f/a}$ les états moyens et les matrices de covariance des états prédits/analysés, les équations du KF s'écrivent :

1. pour l'étape de prévision :

$$\mathbf{x}_k^f = \mathbf{M}_{k:k-1}\mathbf{x}_{k-1}^a, \tag{3.8}$$

$$\mathbf{P}_k^f = \mathbf{M}_{k:k-1}\mathbf{P}_{k-1}^a\mathbf{M}_{k:k-1}^T + \mathbf{Q}_k, \tag{3.9}$$

2. pour l'étape d'analyse :

$$\mathbf{K}_k = \mathbf{P}_k^f\ \mathbf{H}_k^T\ (\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k)^{-1}, \tag{3.10}$$

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{y}_k^0 - \mathbf{H}_k\mathbf{x}_k^f), \tag{3.11}$$

$$\mathbf{P}_k^a = (\mathbf{I}_k - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f. \tag{3.12}$$

La matrice $\mathbf{K}_k \in \mathbb{R}^{m\times d}$ est la matrice gain de Kalman et contient les coefficients de régression de la combinaison linéaire optimale, au sens du minimum de variance, entre le ***prior*** $\mathbf{x}_k^f$, obtenu par l'étape de prévision, et les observations. La variance de l'erreur de l'état analysé $\mathbf{x}_k^a$ est minimale et non-biaisée.

### 3.4.2 Le filtre de Kalman étendu

Dans le KF classique, le modèle d'évolution et l'opérateur d'observation sont supposés linéaires par rapport à l'état. Cependant, il arrive souvent que l'hypothèse de linéarité ne soit pas valide. Le KF est alors généralisé en utilisant des formes linéarisées de l'opérateur d'observation et du modèle d'évolution pour les équations (3.9), (3.10) et (3.12) et la forme non-linéaire pour les équations (3.8) et (3.11). Ce filtre est appelé filtre de Kalman étendu (EKF). L'EKF et ses dérivés sont décrits dans le chapitre 6 de Asch et al. [2016].

L'EKF présente deux inconvénients majeurs. D'une part, il est inutilisable pour des problèmes de grande dimension pour des raisons de calcul de la matrice de covariance d'erreur. Si $n$ est la taille du vecteur de contrôle et que le modèle linéaire tangent est calculé par différences décentrées, il requiert $2 \times n$ intégrations de ce dernier pour le calcul de la matrice de covariance d'erreur en phase de prévision. D'autre part, si son utilisation fournit de bonnes analyses dans un cadre faiblement non-linéaire, ce n'est plus le cas pour des modèles d'évolution fortement non-linéaires.

C'est pourquoi l'EKF a été simplifié afin de réduire le nombre d'intégrations du modèle linéaire-tangent et en ne propageant la matrice de covariance d'erreur que suivant certaines directions (Evensen [1994] ; Verlaan and Heemink [2001] ; Evensen [2003]). Un sous-espace de dimension

réduite est recherché et seule la projection de la matrice de covariance dans ce sous-espace, et non la matrice complète, est alors propagée. Le RRSQRT (Reduced-Rank Square Root Filter), le SEEK (Pham et al. [1998]) et sa variante, le SEIK (Pham [2001]), figurent parmi les variantes les plus connues de l'EKF. Elles sont décrites dans le chapitre 5 de Asch et al. [2016].

### 3.4.3 Le filtre de Kalman d'ensemble stochastique

#### 3.4.3.1 Cas général où le vecteur de contrôle est le vecteur d'état du système

L'EnKF (Evensen [1994] ; Burgers et al. [1998] ; Evensen [2003]) fait partie des méthodes d'AD ensemblistes les plus populaires actuellement. Fondées sur la méthode de Monte-Carlo et sur l'approximation Gaussienne, ses étapes d'analyse et de prévision fournissent une estimation de la moyenne de l'état $\overline{\mathbf{x}}$ et de la covariance sur la base d'un ensemble de réalisations du ***prior*** $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$. La matrice de covariance d'erreur de l'état prédit du système de l'EnKF ($\mathbf{P}^f$), la matrice de covariance croisée entre $\mathbf{x}$ et l'observation simulée ($\mathbf{P}^f\mathbf{H}^T$) et la covariance d'erreur des observations simulées $\mathbf{H}\mathbf{P}^f\mathbf{H}^T$ des équations (3.10) et (3.11) sont alors estimées à l'aide de l'ensemble de taille $N_e$. L'EnKF utilise une représentation ensembliste de la statistique des erreurs, à l'aide d'un ensemble de réalisations de l'état du système obtenues à l'aide du modèle. Quand la taille $N_e$ de l'ensemble augmente, l'erreur d'échantillonnage de Monte-Carlo décroît en $\frac{1}{\sqrt{N_e}}$. Les équations ci-dessous supposent un opérateur d'observation linéaire. Cependant, les méthodes ensemblistes peuvent être utilisées en présence d'un opérateur d'observation faiblement non-linéaire (Evensen [2009a]). Il existe deux familles principales d'EnKF, l'une stochastique et l'autre déterministe. La version stochastique dans un cadre général est décrite ici. Carrassi et al. [2018] propose une description plus détaillée des deux familles et les références afférentes.

##### 3.4.3.1.a Équations

Étant donné un ensemble de réalisations du modèle à un temps donné, la matrice de l'ensemble des états du système, dont chaque colonne correspond à l'un de ses $N_e$ membres, est notée :

$$\mathbf{E} = [\mathbf{x}_1, \cdots, \mathbf{x}_{N_e}]\,. \tag{3.13}$$

Les matrices de covariance pour les états prédits et analysés $\mathbf{P}^{f/a}$ sont calculées à l'aide de la matrice normalisée $\mathbf{X}$ de l'anomalie de l'ensemble autour de sa moyenne $\overline{\mathbf{x}}$ de la manière suivante :

$$\mathbf{X} = \frac{1}{\sqrt{N_e - 1}}\left[\mathbf{x}_1 - \overline{\mathbf{x}}, \cdots, \mathbf{x}_{N_e} - \overline{\mathbf{x}}\right] \in \mathbb{R}^{m \times N_e}, \tag{3.14}$$

avec

$$\overline{\mathbf{x}} = \frac{1}{N_e}\sum_{n=0}^{N_e} \mathbf{x}_n \in \mathbb{R}^{m \times N_e}. \tag{3.15}$$

Les matrices de covariance d'erreur $\mathbf{P}^f$ et $\mathbf{P}^a$ pour les ensembles prédits et analysés s'écrivent alors :

$$\mathbf{P}^f = (\mathbf{X}^f)(\mathbf{X}^f)^T \ \in \mathbb{R}^{m\times m}, \tag{3.16}$$

$$\mathbf{P}^a = (\mathbf{X}^a)(\mathbf{X}^a)^T \ \in \mathbb{R}^{m\times m}, \tag{3.17}$$

avec :

$$\mathbf{X}^f = \frac{1}{\sqrt{N_e-1}}\left[\mathbf{x}_1^f - \overline{\mathbf{x}^f}, \cdots, \mathbf{x}_{N_e}^f - \overline{\mathbf{x}^f}\right] \ \in \mathbb{R}^{m\times N_e}, \tag{3.18}$$

et

$$\mathbf{X}^a = \frac{1}{\sqrt{N_e-1}}\left[\mathbf{x}_1^a - \overline{\mathbf{x}^a}, \cdots, \mathbf{x}_{N_e}^a - \overline{\mathbf{x}^a}\right] \ \in \mathbb{R}^{m\times N_e}. \tag{3.19}$$

Si $\mathbf{y}^0$ ($\in \mathbb{R}^d$) est le vecteur des observations, un ensemble d'observations perturbé par un bruit blanc $\epsilon$ est défini de la manière suivante :

$$\mathbf{y}_n^0 = \mathbf{y}^0 + \boldsymbol{\epsilon}_n \quad avec\, n \ \in \ \{1, \cdots, N_e\} \tag{3.20}$$

et stocké dans la matrice d'observation :

$$\mathbf{Y}^0 = \left[\mathbf{y}_1^0, \cdots, \mathbf{y}_{N_e}^0\right] \ \in \mathbb{R}^{d\times N_e}. \tag{3.21}$$

La matrice des anomalies d'observation $\mathbf{Y'}_0$ s'écrit alors :

$$\mathbf{Y'}^0 = \frac{1}{\sqrt{N_e-1}}\left[\mathbf{y}_1^0 - \mathbf{y}^0, \cdots, \mathbf{y}_{N_e}^0 - \mathbf{y}^0\right] = \frac{1}{\sqrt{N_e-1}}\left[\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_{N_e}\right]. \tag{3.22}$$

La matrice de covariance d'erreur d'observation s'écrit :

$$\mathbf{R} = \mathbf{Y'}^0\left(\mathbf{Y'}^0\right)^T. \tag{3.23}$$

La matrice de covariance d'erreur de $\mathbf{x}$ dans l'espace des observations est $\mathbf{HPH}^T$. Si :

$$\mathbf{Y} = \mathbf{Hx} = \left[\mathbf{y}_1, \cdots, \mathbf{y}_{N_e}\right] \ \in \mathbb{R}^{d\times N_e}, \tag{3.24}$$

et

$$\mathbf{Y'} = \frac{1}{\sqrt{N_e-1}}\left[\mathbf{y}_1 - \overline{\mathbf{y}}, \cdots, \mathbf{y}_{N_e} - \overline{\mathbf{y}}\right], \tag{3.25}$$

alors :

$$\mathbf{HPH}^T = \mathbf{Y}'(\mathbf{Y}')^T. \tag{3.26}$$

L'étape d'analyse de l'EnKF consiste alors à actualiser chaque membre de l'ensemble selon :

$$\mathbf{x}_n^a = \mathbf{x}_n + \mathbf{PH}^T\ (\mathbf{HPH}^T + \mathbf{R})^{-1}(\mathbf{y}_n^0 - \mathbf{y}_n), \quad avec\ n\ \in\ \{1, \cdots, N_e\}\,. \tag{3.27}$$

L'équation (3.27) permet d'écrire l'équation d'évolution de la moyenne de l'ensemble suite à l'étape d'analyse :

$$\overline{\mathbf{x}^a} = \overline{\mathbf{x}} + \mathbf{PH}^T\ (\mathbf{HPH}^T + \mathbf{R})^{-1}(\mathbf{y}^0 - \overline{\mathbf{y}}). \tag{3.28}$$

Il est alors possible de démontrer (Evensen [1994]) que la matrice de covariance de l'ensemble à l'issue de la phase d'analyse s'écrit :

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}, \tag{3.29}$$

où $\mathbf{K} = \mathbf{PH}^T\ (\mathbf{HPH}^T + \mathbf{R})^{-1}$ est la matrice gain de Kalman obtenue à l'aide des réalisations de l'ensemble.

L'ensemble actualisé hérite des propriétés non-gaussiennes de l'ensemble prédit, dans la mesure où seul le dernier membre de l'équation (3.28) est linéaire et que, par conséquent, si le prior n'est pas gaussien, le posterior ne le sera pas non plus. Par ailleurs, l'utilisation d'un ensemble de taille finie $N_e$ introduit des erreurs d'approximation des matrices de covariance "vraies". En pratique, la taille $N_e$ de l'ensemble est critique, ce qui incite à trouver le meilleur compromis entre la précision à atteindre et la taille de l'ensemble.

#### 3.4.3.1.b Localisation

Le nombre de membres $N_e$ de l'ensemble étant limité, les matrices de covariance croisées (entre le vecteur de contrôle et sa projection dans l'espace des observations) ne sont que des approximations et ne sont pas de rang plein. Ceci implique des erreurs d'échantillonnage importantes, car, en d'autres termes, la matrice de covariance d'erreur est échantillonnée en utilisant un nombre limité de modes. Il peut alors apparaître dans cette dernière des covariances entre variables qui n'ont pas lieu d'être sur de longues distances, ce qui diminue la performance du KF. Dans le cas général décrit en section 3.4.3.1, la localisation consiste à faire l'hypothèse que les covariances tendent vers 0 quand deux éléments du vecteur de contrôle sont distants en espace ou en temps et ainsi à transformer les matrices de covariance des erreurs d'ébauche ou d'observation pour faire disparaître les covariances intempestives. Pour ce faire, il est possible soit de réduire les valeurs des coefficients loin de la diagonale de la matrice de covariance (localisation des covariances),

soit de faire une analyse autour de chaque point de grille, en ne gardant que les observations et variables de l'état les plus proches de celui-ci (localisation de domaine). Ces deux méthodes sont décrites dans le chapitre 6.5.1 de Asch et al. [2016].
En particulier, la localisation des covariances (ou localisation par produit de Schur) modifie la matrice de covariance pour lisser les corrélations erronées loin de la diagonale. Pour ce faire, un produit de Schur (ou de Hadamard) avec une matrice de corrélation $\rho \in R^{m \times m}$ est appliqué à la matrice de covariance des erreurs d'état du système. Ainsi, si $\rho$ est choisie telle que les corrélations décroissent avec la distance, les corrélations de longue distance intempestives de la matrice de covariance sont filtrées.
Le Local Ensemble Transform Kalman Filter (LETKF), décrit par Hunt et al. [2007], est une méthode dérivée de l'EnKF emblématique du succès de l'EnKF associé à la localisation de la covariance pour des modèles de très grandes dimensions.

#### 3.4.3.1.c Inflation

Des erreurs d'échantillonnage apparaissent lors du calcul des matrices de covariances. Elles sont liées au nombre limité de membres de l'ensemble et peuvent s'accumuler au fur et à mesure des cycles d'assimilation, malgré l'utilisation de techniques de localisation (cf. section 3.4.3.1.b), et entraîner l'effondrement de l'ensemble. Pour ré-injecter de la dispersion au sein de l'ensemble, les matrices de covariance sont transformées par des méthodes d'inflation. Deux méthodes d'inflation, l'une multiplicative et l'autre additive, sont décrites dans le chapitre 6.5.2 de Asch et al. [2016]. L'inflation "multiplicative" consiste à multiplier soit la matrice de covariance des variables de contrôle, soit l'anomalie du vecteur de contrôle (inflation de l'ensemble) par un coefficient $\lambda > 1$. L'inflation additive consiste à contrebalancer la sous-estimation des erreurs liées à la présence d'erreur du modèle en appliquant à chaque membre de l'ensemble une perturbation corrélée selon la matrice de covariance d'erreur du modèle $\mathbf{Q}$.

### 3.4.3.2 Estimation conjointe des paramètres du modèle et de l'état du système

L'estimation conjointe de l'état d'un système et des paramètres du modèle est désormais l'objectif de nombreuses études (Evensen [2009b], Bocquet and Sakov [2013]). L'EnKF décrit dans le cadre général (section 3.4.3.1) montre la méthode d'estimation de l'état d'un système dynamique et stochastique dont les paramètres et autres données d'entrée sont considérés comme parfaits et/ou invariants dans le temps. Les variables de contrôle sont les variables d'état du système. Or, il est légitime de remettre en cause le fait que les paramètres sont constants, mais aussi de vouloir corriger les conditions initiales (CI) et/ou aux limites (BC) qui sont variables dans le temps et différentes du vecteur d'état. Le vecteur de contrôle est donc généralement différent du vecteur d'état et, tout du moins, ne s'y limite pas.
Dans la pratique, le vecteur de contrôle peut donc inclure l'état du système, mais également les CI, BC ou les paramètres du modèle d'évolution dont une valeur ***a priori***, appelée "background",

est connue. Les méthodes d'AD combinent alors le background, les observations, le modèle numérique, etc. pour obtenir une meilleure représentation de l'état du système. Deux approches existent : l'une consiste à intégrer les paramètres dans le vecteur d'état du système et est connue sous le nom de "state augmentation" ou "joint-EnKF". Elle est utilisée dans le domaine des géosciences (Bocquet and Sakov [2013]). L'autre approche consiste en l'estimation de l'état du système et des paramètres du modèle à l'aide de deux filtres interagissant, l'un estimant l'état du système, l'autre les paramètres du modèle (Moradkhani et al. [2005]). Elle est nommée "dual estimation". Ces deux types d'estimation, conjointe ou duale, de l'état du système et des paramètres du modèle sont explicitement décrites dans Sun et al. [2016].

Le modèle d'évolution du système dynamique dans un cadre non-linéaire général est défini par l'équation (3.1). Le modèle d'évolution $\mathbf{M}_{k+1:k}$ linéarisé entre deux pas de temps successifs $k$ et $k+1$ dépend à la fois de l'état du système $\mathbf{x}_k$ et des paramètres (au sens large, incluant éventuellement les forçages) $\boldsymbol{\gamma}_k$ au pas de temps précédent $t_k$. Par ailleurs, l'opérateur d'observation $\mathcal{H}_{k+1}$ de l'équation (3.2) est également linéarisé et devient $\mathbf{H}_{k+1}$.

Il est alors nécessaire de considérer les paramètres comme des variables d'état à la différence près que le modèle de "pseudo-"propagation des paramètres ($\Upsilon$) doit être choisi arbitrairement. Si $\gamma$ est la condition aux limites, $\Upsilon$ n'est pas persistant, mais n'est pas connu. Dans le cas d'un modèle d'évolution reflétant la persistance de l'ensemble $\boldsymbol{\gamma}$ de $t_k$ à $t_{k+1}$, alors $\Upsilon$ est l'identité. Il vient :

$$\boldsymbol{\gamma}^f_{k+1} = \boldsymbol{\gamma}^a_k + \boldsymbol{\tau}_k \quad avec\ \boldsymbol{\tau}_k \sim \mathcal{N}(0, \mathbf{P}^{\gamma}_k). \tag{3.30}$$

$\mathbf{P}^{\gamma}_k$ est la matrice de covariance des paramètres contrôlés $\boldsymbol{\gamma}_k$ à $t_k$.

Le vecteur de contrôle ne contient pas forcément l'état du système $\mathbf{x}$. Cependant, le fait que $\boldsymbol{\gamma}$ soit contrôlé modifie l'état $\mathbf{x}$ et produit une "pseudo"-analyse de $\mathbf{x}$, notée $\mathbf{x}^a_k$. En utilisant les paramètres prédits $\boldsymbol{\gamma}^f_{k+1}$ à l'aide de l'équation (3.30), l'ensemble de l'état du système prédit $\tilde{\mathbf{x}}^f_{k+1}$ et l'ensemble des prévisions $\tilde{\mathbf{y}}^f_{k+1}$ simulées dans l'espace d'observation à l'aide des paramètres issus de l'analyse au temps $t_k$ s'écrivent :

$$\tilde{\mathbf{x}}^f_{k+1} = \mathbf{M}_{k+1:k}(\mathbf{x}^a_k, \boldsymbol{\gamma}^f_{k+1}), \tag{3.31}$$

$$\tilde{\mathbf{y}}^f_{k+1} = \mathbf{H}_{k+1}(\tilde{\mathbf{x}}^f_{k+1}, \boldsymbol{\gamma}^f_{k+1}). \tag{3.32}$$

L'actualisation ou phase d'analyse des paramètres du modèle est réalisée à l'aide de l'équation classique de l'EnKF (équation (3.27)) :

$$\boldsymbol{\gamma}^a_{k+1} = \boldsymbol{\gamma}^f_{k+1} + \mathbf{K}^{\gamma}_{k+1}\ (\mathbf{y}^0_{k+1} - \tilde{\mathbf{y}}^f_{k+1}). \tag{3.33}$$

$\mathbf{K}^{\gamma}_{k+1}$ est la matrice gain de Kalman, obtenue à l'aide de l'équation suivante :

$$\mathbf{K}^{\gamma}_{k+1} = \mathbf{P}^{\gamma,y}_k \left[\mathbf{P}^{\gamma,\gamma}_k + \mathbf{R}_{k+1}\right]^{-1}. \tag{3.34}$$

L'opérateur d'observation généralisé et non-linéaire $\mathcal{G}$ qui relie l'espace des paramètres $\boldsymbol{\gamma}$ à l'espace des observations $\mathbf{y}^0$ est : $\mathcal{G} = \mathcal{H} \circ \mathcal{M}$. $\mathcal{G}$ est la combinaison de l'opérateur d'extraction et d'interpolation $\mathcal{H}$ et du modèle d'évolution $\mathcal{M}$.

$\mathbf{P}_{k+1}^{\gamma,y}$ est la matrice de covariance croisée entre l'ensemble des paramètres $\boldsymbol{\gamma}_{k+1}^f$ et l'ensemble des observations simulées $\tilde{\mathbf{y}}_{k+1}^f$.

En utilisant les paramètres analysés $\boldsymbol{\gamma}^a$, il est possible de calculer le nouvel état prédit du système :

$$\mathbf{x}_{k+1}^f = \mathbf{M}_{k+1:k}(\mathbf{x}_k^a, \boldsymbol{\gamma}_{k+1}^a), \tag{3.35}$$

$$\mathbf{y}_{k+1}^f = \mathbf{H}_{k+1}(\mathbf{x}_{k+1}^f, \boldsymbol{\gamma}_{k+1}^a). \tag{3.36}$$

Dans le cas d'une estimation conjointe de l'état du système et des paramètres du modèle à l'aide de deux EnKF combinés, la phase d'analyse de l'état du système s'écrit alors :

$$\mathbf{x}_{k+1}^a = \mathbf{x}_{k+1}^f + \mathbf{K}_{k+1}\ (\mathbf{y}_{k+1}^0 - \mathbf{y}_{k+1}^f), \tag{3.37}$$

où $\mathbf{K}_{k+1}$ est la matrice gain permettant de corriger l'état du système obtenue à l'aide de :

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1}^{x,y}\ \left[\mathbf{P}_{k+1}^{y,y} + \mathbf{R}_{k+1}\right]. \tag{3.38}$$

$\mathbf{P}_{k+1}^{x,y}$ (resp. $\mathbf{P}_{k+1}^{y,y}$) est la matrice de covariance croisée entre l'état du système et la prévision simulée (resp. la matrice de covariance de la prévision simulée).

Dans le cas de l'estimation des paramètres seuls et de la mise en œuvre d'un unique EnKF dédié aux paramètres et forçages regroupés au sein de $\boldsymbol{\gamma}$, les équations (3.30) à (3.36) permettent de mettre en place un aller-retour entre l'optimisation des paramètres vis-à-vis de l'observation et l'ajustement de l'état du système au temps $t_{k+1}$. La formulation des états analysés $\mathbf{x}_{k+1}^a$ au temps $t_{k+1}$, qui remplace l'équation (3.37), et la projection $\mathbf{y}_{k+1}^a$ de $\boldsymbol{\gamma}_{k+1}^a$ dans l'espace des observations s'écrivent alors :

$$\mathbf{x}_{k+1}^a = \mathbf{M}_{k+1:k}(\mathbf{x}_k^a, \boldsymbol{\gamma}_{k+1}^a), \tag{3.39}$$

$$\mathbf{y}_{k+1}^a = \mathbf{H}_{k+1}(\mathbf{x}_{k+1}^a, \boldsymbol{\gamma}_{k+1}^a). \tag{3.40}$$

Les équations (3.33) à (3.36) et (3.39) à (3.40) sont les équations du KF stochastique pour l'estimation des paramètres $\gamma$ d'un modèle, noté EnKF-$\gamma$.

## 3.5 Les méthodes variationnelles et variationnelles d'ensemble

Les méthodes variationnelles ont été introduites au début des années cinquante par Sasaki [1955] et sont fondées sur la théorie de l'optimisation. L'approche variationnelle traite le problème glo-

balement sous la forme de la minimisation d'une fonctionnelle (fonction objectif ou coût) qui "optimise" l'écart aux observations et l'éloignement à la solution du modèle. Par ailleurs, dans le cas linéaire et en cas d'hypothèse gaussienne, la solution est optimale à la fois par approche variationnelle (maximum de vraisemblance) et par méthode séquentielle (minimisation de la variance).

La suite de cette section s'appuie sur Asch et al. [2016]. Dans le mesure où seul l'EnKF est utilisé dans le cadre de ce travail de thèse, les méthodes variationnelles ne sont pas détaillées.

Les approches variationnelles les plus classiques sont le 3DVar et le 4DVar. Elles sont décrites dans le chapitre 2.4 de Asch et al. [2016]. Le 3DVar est utilisé lorsque toutes les observations sont disponibles au même temps et les assimile en ce seul pas de temps. Sa différence avec le 4DVar réside dans le fait que cette dernière méthode prend en compte les observations d'une fenêtre (ou cycle) d'assimilation complète et recherche la trajectoire du vecteur de contrôle qui s'ajuste au mieux au background et à toutes les observations. Le 4DVar recherche une estimation non-linéaire du maximum ***a posteriori*** de la pdf à l'aide de techniques d'optimisation non-linéaires, alors que l'EnKF réalise une actualisation linéaire de l'état du système. L'EnKF propage les statistiques des erreurs et améliore leur estimation au fur et à mesure des cycles d'assimilation, alors que le 4DVar traditionnel repose sur une statistique des erreurs invariante dans le temps. L'une des faiblesses du 4DVar, d'un point de vue technique, réside dans la nécessité d'estimer le gradient de la fonction coût. Il est possible d'estimer ce gradient soit en utilisant une méthode de calcul par différences finies, soit en utilisant le modèle adjoint du modèle dynamique. La première méthode est facile à implémenter, mais devient inefficace lorsque la taille du vecteur de contrôle augmente. Par ailleurs, l'utilisation du modèle adjoint requiert la maintenance du linéaire tangent et de l'adjoint du modèle d'évolution, ce qui reste un challenge informatique, alors que l'EnKF estime le linéaire tangent de l'opérateur d'observation et de son adjoint dans le sous-espace de l'ensemble.

La comparaison des deux types de méthode dans le domaine de la météorologie (Bocquet and Sakov [2013]) n'a pas permis de définir une supériorité de l'EnKF sur le 4DVar, mais a permis de cerner des domaines optimaux d'application (Asch et al. [2016], Defforge [2019]). L'EnKF est ainsi plus performant que le 4DVar sur des fenêtres d'assimilation courtes, alors que sur des fenêtres d'assimilation plus longues caractérisées par des processus non-linéaires, le 4DVar donne de meilleurs résultats que l'EnKF (Asch et al. [2016]). C'est pourquoi des travaux récents ont développé des méthodes variationnelles d'ensemble, regroupées sous la dénomination de méthodes EnVar4D, dans le but de bénéficier simultanément des avantages des méthodes variationnelles de type 4DVar et ensemblistes telles que l'EnKF. Une revue des méthodes variationnelles d'ensemble EnVar4D est disponible dans Asch et al. [2016]. Le chapitre 7 décrit les méthodes hybrides, les ensembles de méthodes variationnelles (EDA pour "Ensemble of [variational] data assimilation methods"), le 4DEnVar et l'IEnKS (Iterative Ensemble Kalman Smoother). En particulier, l'IEnKS est une méthode qui utilise un ensemble de trajectoires 4D pour faire une analyse variationnelle non-linéaire tout en propageant les erreurs (Bocquet and Sakov [2014]). Il s'agit d'une exten-

sion de l'IEnKF (Iterative Ensemble Kalman Filter) développée par Sakov et al. [2012] qui permet d'assimiler des observations asynchrones. La spécificité de l'IEnKS réside dans la connaissance parfaite des approximations et des sources de suboptimalité liées aux non-linéarités des systèmes. Elle permet, par ailleurs, d'estimer les modèles linéaire tangent et adjoint du modèle d'évolution, s'ils ne sont pas connus, mais aussi de les utiliser s'ils sont déjà disponibles. Pour ces raisons, l'IEnKS est plus performant que le 4DVar et l'EnKF, aux coûts de calcul près. Les méthodes variationnelles d'ensemble ci-dessus ont été développées en l'absence d'erreur du modèle. Sakov et al. [2018] présente une nouvelle méthode variationnelle d'ensemble en présence d'une erreur du modèle, nommée IEnKF-Q.

## 3.6 Application à la modélisation hydro-environnementale

### 3.6.1 Panorama dans le domaine des géosciences

Déjà largement utilisées par la communauté météorologique, les méthodes d'AD, dont Evensen [2009b] cite des exemples d'utilisation dans le domaine des géosciences, sont désormais couramment utilisées en hydrodynamique océanique, côtière et fluviale. De nombreux travaux de recherche ont ainsi été menés pour l'amélioration des modèles hydrodynamiques de surface, à l'aide de méthodes ensemblistes ou variationnelles, dans le but de caractériser, notamment, l'aléa vis-à-vis du risque d'inondation ou de submersion marine ou encore d'améliorer la prévision des crues ou des niveaux d'eau. Le chapitre 8 de Asch et al. [2016] présente quelques exemples d'utilisation des méthodes d'AD dans le domaine des sciences environnementales et, en particulier, en hydrologie et dans le domaine de la prévision des crues.

L'incertitude, que les méthodes d'AD cherchent à réduire, peut être considérée en incluant l'état du système et/ou les conditions aux limites du modèle et/ou les paramètres du modèle dans le vecteur de contrôle. Elle peut également être déclinée, dans le cadre des méthodes ensemblistes, en perturbant les conditions aux limites et/ou les paramètres du modèle pour constituer l'ébauche des membres de l'ensemble. Enfin, pour les forçages aux limites et les méthodes de type EKF ou ses dérivés de rang réduit (RRSQRT, SEEK, SEIK), ainsi que pour les méthodes variationnelles, elle apparaît au travers de la construction de la matrice de covariance d'erreur d'ébauche de l'état du système.

Pour un même objectif, qui reste l'amélioration de la simulation des niveaux d'eau et/ou des vitesses du système modélisé, le choix du vecteur de contrôle et l'ébauche de l'ensemble permettent donc de discriminer les méthodes usuelles mises en œuvre.

Alors que Liu et al. [2012] dresse l'état de l'art de l'utilisation des méthodes d'AD pour la prévision en hydrologie, Moore et al. [2019] réalise un état des lieux et évoque les perspectives et les défis à relever pour l'AD en milieu océanique. En particulier, Moore et al. [2019] évoque les difficultés d'estimation des matrices de covariance quelle que soit la méthode ensembliste ou va-

riationnelle utilisée. Dans la plupart des cas, l'AD contrevient à l'équation de conservation du moment et la correction de l'état peut introduire, de manière irréaliste et par intermittence, des ondes d'états de mer ou de surcotes qui constituent un verrou scientifique en milieu océanique et côtier.
Les défis concernent principalement la mise en œuvre des méthodes variationnelles d'ensemble (cf. section 3.5) pour tirer bénéfice du meilleur de chacune des méthodes ensembliste et variationnelle dans les centres opérationnels de prévision. La question de la disponibilité des observations et de leur qualité est évoquée par Moore et al. [2019], ainsi que la perspective de la future mission SWOT (Surface Water and Ocean Topography) qui devrait apporter des données satellitaires supplémentaires sur des résolutions plus fines (https://swot.cnes.fr/fr).

Barthélémy et al. [2017] présente les méthodes d'AD par filtrage récemment appliquées avec succès à des modèles 1D ou pseudo-2D (issus d'un couplage 1D/2D ou encore à casiers) d'hydrodynamique fluviale, du fait du coût du modèle numérique et de la taille du vecteur d'état qui restent limités dans un cadre 1D ou pseudo-2D.

### 3.6.2 Correction de l'état hydraulique (niveau d'eau, vagues ou surcotes)

#### 3.6.2.1 En milieu océanique et côtier

Dans le domaine côtier, la correction de l'état hydraulique, qui correspond en réalité à l'estimation de la condition initiale, est couramment utilisée et les références nombreuses.
L'incertitude est souvent portée sur les champs météorologiques (et les champs de vent) fournis par des prévisions d'ensemble de modèles météorologiques qui permettent de créer l'ensemble initial des états du système. Des EnKF fondés sur cette hypothèse sont ainsi couramment mis en œuvre pour l'amélioration de l'estimation des niveaux d'eau, des surcotes ou des états de mer en milieu côtier.
Almeida et al. [2015] corrige le modèle d'états de mer SWAN sur les côtes portugaises à l'aide d'un EnKF dans lequel l'observation est remplacée par l'état analysé du temps précédent. Zijl et al. [2015] cherche à optimiser la prévision des surcotes à l'aide d'un modèle opérationnel construit sur le plateau continental du nord-ouest de l'Europe. Les auteurs corrigent avec un EnKF stochastique l'état du système seul et montrent que des gains substantiels peuvent être obtenus dans les premières heures de prévision.
Etala et al. [2015] utilise un 4D-LETKF, EnKF déterministe de la famille des EnSRF, appliqué à un modèle océanique de prévision des surcotes sur la cote argentine (SMARA) au droit de Buenos Aires pour corriger les niveaux sur des fenêtres d'assimilation de 6 heures et améliorer la qualité des prévisions en assimilant des observations altimétriques réelles issues de l'événement de septembre 2011 et peu nombreuses. Les forçages météorologiques sont fournis par un système de prévision d'ensemble opérationnel.
Chen et al. [2011] utilise des mesures de vitesses pour corriger à l'aide d'un Quick-EnKF, qui n'utilise qu'une seule intégration de la moyenne du vecteur d'ébauche, les composantes de la

vitesse du système côtier situé près de l'estuaire de Changjiang dont l'hydrodynamique est gouvernée par 8 composantes de la marée, le débit du fleuve Changjiang et le vent.

De la même manière, Raboudi et al. [2019] présente un ETKF combinant deux techniques auxiliaires, une covariance hybride et une formulation de lissage "One step ahead" utilisant deux fois les observations, appliqué sur un modèle numérique bidimensionnel ADCIRC du golfe du Mexique. Les observations sont issues de simulations réalisées à l'aide d'un modèle à maillage plus raffiné que celui utilisé dans l'expérience d'AD. Les forçages de ce modèle sont les seuls champs de vent issus d'un modèle météorologique de résolution plus fine que ceux utilisés pour l'expérience d'AD. L'objectif de l'étude est de prédire l'état du système en une étape de lissage de l'état analysé du cycle précédent utilisant une matrice de covariance d'erreur d'ébauche hybride, puis d'analyse de l'état pseudo-prédit issu de l'intégration de l'état analysé de l'étape précédente. L'étape de lissage (resp. d'analyse) remplace la matrice de covariance classique par la combinaison linéaire de la matrice de covariance entre l'état analysé du pas de temps précédent (resp. le pseudo-état prédit) et les observations du cycle en cours et une matrice d'erreur de covariance d'ébauche dite "statique" issue de simulations en free-run réalisées sur de nombreux événements historiques. La double-intégration du modèle, nécessaire, est compensée par l'utilisation d'une matrice de covariance hybride qui permet de réduire par 2 le nombre de membres par rapport à l'ETKF classique.

#### 3.6.2.2 En milieu estuarien

En domaine estuarien, Bertino et al. [2002] utilise un KF déterministe réduit RRSQRT (Reduced Rank Square Root filter) développé par Canižares Torres [1999] pour améliorer la simulation des niveaux d'eau de la lagune de l'Oder en corrigeant l'état du système obtenu à l'aide du modèle hydrodynamique tridimensionnel TRIM3D à l'aide de mesures ***in situ***. La source des incertitudes permettant de construire la matrice de covariance d'ébauche provient des forçages aux frontières du modèle et des forçages météorologiques. Bien que la lagune semble presque fermée, la principale source d'erreur de modélisation identifiée réside dans la connaissance imparfaite des trois conditions aux limites reliées aux niveaux d'eau de la mer Baltique. L'état est ici corrigé, mais l'ensemble est généré en perturbant les conditions aux limites.

Gu et al. [2019] améliore, à l'aide d'une méthode par interpolation optimale (Beckers [2008]), l'état du système en assimilant des données de stratification de la salinité en milieu estuarien dans la baie de Gyeonggi (Corée du sud). Bertino et al. [2002] estime des concentrations en phytoplancton, nutriments et herbivores d'un modèle écologique 1D à l'aide d'un EnKF et d'un RRSQRT KF. Dans les deux cas, la condition initiale, c'est-à-dire l'état du système en tant qu'ébauche de l'intégration, est corrigée.

Pour corriger et estimer l'état hydrodynamique et la salinité dans l'estuaire du fleuve Columbia, au sud de la frontière entre le Canada et les États-Unis,compte-tenu des coûts de calcul induits par la simulation de l'état de tels systèmes, Frolov et al. [2009] propose un méta-modèle à 50 degrés

de liberté et un KF. La matrice de covariance du KF est fondée sur une méthode ensembliste utilisant des bases de données constituées offline pour la construction du méta-modèle.

D'autres KF déterministes sont mis en œuvre en hydrodynamique estuarienne bidimensionnelle. Canas et al. expose brièvement l'implémentation d'un SFEK appliqué au modèle hydrodynamique de l'estuaire du Tage au Portugal dans le " Mohid Water Modelling System" pour évaluer les niveaux d'eau et les vitesses d'un modèle hydrodynamique 2D dans un cadre synthétique. De bons résultats sont obtenus pour la réanalyse des niveaux d'eau sur une période de 15 jours, mais les vitesses dégradées rendent nécessaire l'utilisation d'un SEEK.

#### 3.6.2.3 Prise en compte de l'incertitude sur les conditions aux limites dans l'ébauche de l'ensemble

L'incertitude sur les forçages à la frontière du modèle, qui permet de créer l'ensemble d'ébauche dans le cas de l'EnKF et la matrice de covariance de l'état du système dans le cas de l'EKF et de ses dérivés, peut être prise en compte par perturbation des conditions aux limites liées à la marée et/ou des paramètres, tout en ne corrigeant que l'état du système.

Dans le domaine côtier, Karri et al. [2014] corrige l'état du modèle SRM (Singapore Regional Model) à l'aide d'un Ensemble Steady State Kalman Filter (EnSSKF). Elle compare ses performances avec un EnKF dans le cadre d'une expérience jumelle dont les observations synthétiques sont issues de la réanalyse de la tempête de 2004 pour un modèle de plus grande emprise. Les champs de vent et de pression ne sont pas appliqués sur les noeuds du modèle, dont l'emprise est insuffisante pour permettre la génération en son sein de surcotes météorologiques locales. Les surcotes météorologiques sont ainsi incluses dans les conditions aux limites. À l'instar de Tamura et al. [2014], qui corrige l'état du système à l'aide d'un EnKF en perturbant les coefficients de Strickler dans l'estuaire de St John en Floride, l'erreur sur la condition aux limites de type marée seule est fixée à 5 cm. Les états corrigés sont utilisés comme condition initiale pour une période de prévision de 24 h en ne modifiant pas les autres paramètres du modèle. Karri et al. [2014] montre une bonne amélioration des niveaux d'eau jusqu'à une échéance de 6 h, voire 12 h pour certaines stations, puis une dégradation de telle sorte que les résultats du run de contrôle sont meilleurs que ceux de l'EnKF ou de l'ESSKF.
De la même manière, Karri and Babovic [2017] utilise la plate-forme OPEN-DA (`http://www.openda.org/`) et un modèle 1D simplifié du détroit de Singapour forcé par la marée et évalue les performances d'un EnKF stochastique. En matière de forçages aux frontières du modèle hydrodynamique de la lagune de l'Oder, Bertino et al. [2002] met en lumière l'incertitude sur les conditions aux limites maritimes de la mer Baltique en contrôlant l'état du système.

Enfin, Barthélémy et al. [2017], sur le fleuve Adour, perturbe les forçages fluviaux d'un modèle 1D fondé sur mascaret pour corriger l'état du système, c'est-à-dire les niveaux d'eau et les débits le long du fleuve.

### 3.6.3 Correction des conditions aux limites et de l'état du système

L'incertitude sur les conditions aux limites du modèle ou les paramètres d'un modèle peut être mise en évidence par l'estimation de l'état du système ou faire l'objet d'une estimation conjointe ou duale avec l'état du système. Comme indiqué dans la section 3.4.3.2, plusieurs méthodes existent dans la littérature. L'une d'entre elles, décrite dans Evensen [2009b], consiste à "augmenter" le vecteur d'état à l'aide des forçages ou des paramètres. D'autres consistent à découpler l'estimation de l'état de celle du forçage ou du paramètre.

Dans le domaine de la modélisation micrométéorologique et en régime stationnaire, Defforge et al. [2019] adapte l'IEnKS (cf. section 3.5) à la correction des conditions aux limites (BC) qu'elle applique au modèle CFD Code_Saturne en 2D et 3D. Les auteurs utilisent à la fois des expériences jumelles et des observations de terrain, dans l'objectif d'améliorer les simulations de vent sur une topographie très complexe et en n'assimilant que quelques observations in situ. L'ébauche et la matrice de covariance d'ébauche sont construites à l'aide de l'analyse statistique de trois années de simulations WRF (Done et al. [2004]). Les auteurs démontrent que l'IEnKS réduit considérablement l'erreur et l'incertitude des BC et donc du champ de vent simulé et qu'il permet une meilleure précision pour l'évaluation du potentiel éolien.

Jean-Baptiste et al. [2011] utilise un vecteur d'état augmenté par les forçages inconnus et estime les forçages latéraux non jaugés sur le Rhône, dans le sud-est de la France, de manière conjointe avec l'état du système à l'aide d'un KF et d'un filtre à particules. Hartnack et al. [2006] utilise un EnKF pour estimer le niveau d'eau et le débit d'un modèle d'inondation 1D-2D, améliorant ainsi l'état hydraulique et le forçage.

Ricci et al. [2011], de même que Madsen and Skotner [2005] et Shiiba et al. [2000], procèdent à une estimation duale (en deux étapes) des forçages en amont de l'Adour, fleuve côtier du sud-ouest de la France. Les auteurs utilisent un EKF qui tient compte des non-linéarités du modèle et utilisent une matrice de covariance invariante (IKF), ce qui permet de réduire le coût de calcul de l'algorithme. L'IKF permet également de corriger l'état hydraulique (niveau d'eau et débit) et d'améliorer significativement la prévision jusqu'à 12 h de délai.

### 3.6.4 Estimation de l'état et des paramètres pour la correction des niveaux d'eau

#### 3.6.4.1 En rivière

D'autres études (Bonan et al. [2014], Bocquet and Sakov [2013]), fondées sur des méthodes ensemblistes ou variationnelles, estiment l'état du système et les paramètres (hors BC) des modèles par estimation conjointe ("état augmenté") ou duale. Par ailleurs, des travaux récents (Stroud et al. [2018]) proposent une approche bayésienne complète pour l'estimation des états du système et des paramètres au sein d'un EnKF.

Dans le domaine de la prévision hydrologique, Vrugt et al. [2006] réalise une estimation duale des débits et des paramètres du modèle hydrologique conceptuel Sacramento à l'aide d'un EnKF pour l'estimation de l'état, ce qui permet de traiter les erreurs de données sur le débit, les erreurs structurelles du modèle et l'incertitude des paramètres. L'estimation des paramètres est ici découplée à l'aide de l'approche SCEM-UA (Shuffled complex evolution metropolis stochastic ensemble optimization approach).
En amont de la rivière Marne, Habert et al. [2016] étend l'algorithme IKF de Ricci et al. [2011] à l'estimation des coefficients de frottement à l'aide de mesures ***in situ*** du niveau d'eau, améliorant ainsi la loi de tarage en entrée du modèle et, notamment, pendant la crue.

Dans un cadre synthétique et dans la perspective d'utiliser cette méthode avec les données du simulateur SWOT, Oubanas et al. [2018] utilise une méthode d'AD variationnelle appliquée à un modèle 1D SIC2 (`sic.g-eau.net`) de la rivière Garonne pour déterminer son débit conjointement avec les coefficients de frottement appliqués sous forme de Strickler sur les biefs du modèle. Les auteurs démontrent que le débit peut être reconstitué de manière fiable sans connaissance précise de la bathymétrie et des paramètres liés au frottement.
Cooper et al. [2018] utilise un ETKF pour corriger de manière conjointe les débits, niveaux d'eau et coefficients de frottement d'un domaine schématisant une rivière et son champ d'expansion de crue en assimilant des niveaux d'eau. La démarche est identique pour Ziliani et al. [2019] dans un cadre non-linéaire de rupture de barrage sur le Toce (Piémont, Italie). Dans les deux cas, les auteurs concluent à l'efficience d'un système de prévision intégrant un EnKF couplant l'estimation de l'état du système et des coefficients de frottement.

L'estimation des paramètres seuls, sans correction simultanée de la condition initiale, est courante dans le domaine de l'hydrologie et des écoulements souterrains (Nan and Wu [2011], Devegowda et al. [2010]). D'autres études explorent la capacité de l'EnKF (Evensen [1994]) à estimer les fonctions de covariance d'erreur d'ébauche et leur évolution temporelle à l'aide de l'ensemble.

#### 3.6.4.2 En milieu estuarien

En milieu estuarien, Tamura et al. [2014] corrige de manière conjointe l'état du système et les paramètres sur la rivière St John dans le Nord de la Floride pendant une période anticyclonique de 10 jours à l'issue de 20 jours de simulation. L'événement est compris entre le 21/09 et le 21/10/1999. Les membres du système sont créés en perturbant le coefficient de Manning et en appliquant une erreur de 5 cm sous la forme d'un bruit blanc aux forçages maritimes constitués exclusivement de la marée. Parmi les 4 stations d'observation disponibles, 3 sont utilisées par l'EnKF, la dernière sert à évaluer les performances de l'EnKF en réanalyse de l'événement. Les erreurs sur le run de contrôle sont inférieures à 5 cm pendant la période étudiée et descendent à 2 cm en utilisant l'EnKF. L'erreur de 5 cm appliquée au forçage lié à la marée dans Tamura et al. [2014] et Karri et al. [2014] correspond à la précision indiquée par le SHOM sur le signal de marée. Elle est d'un ordre de grandeur inférieur à celui de l'incertitude sur la surcote pour l'estuaire de

la Gironde, comme indiqué dans l'étude de quantification d'incertitudes (cf. chapitre 2).

En milieu schématique côtier et dans un cadre synthétique (c'est-à-dire où les observations assimilées sont issues de simulations dont les entrées sont perturbées par rapport à celles du run de contrôle), Siripatana et al. [2018] utilise un EnKF (SEIK) avec un vecteur de contrôle augmenté pour l'estimation conjointe de l'état du système et des paramètres. Celui-ci est appliqué à un modèle côtier simplifié pour estimer le coefficient de Manning, dont la valeur en chaque nœud est intégrée dans le vecteur de contrôle, et corriger l'état du système simultanément. Dans un second temps, les composantes modales issues de la décomposition de Karhunen-Loève (KL) d'un champ spatial de frottement sont contrôlées à la place des coefficients de frottement appliqués en chaque nœud du maillage. Les auteurs montrent que les résultats sont meilleurs, mais dépendent du nombre de modes retenu pour la décomposition de KL.

### 3.6.5 Nature des observations

Concernant les données assimilées autres que celles mesurées ***in situ***, plusieurs études ont démontré les bénéfices apportés par assimilation des observations satellitaires à l'aide d'un EKF (Pedinotti et al. [2014], Oubanas et al. [2018]). Biancamaria et al. [2016] dresse l'inventaire des travaux intégrant les données synthétiques de la future mission SWOT dans des systèmes couplant des modèles hydro-météo-environnementaux et les données issues des simulateurs SWOT. Il décline, en particulier, les futurs usages possibles de la mission SWOT dans le domaine de la caractérisation des surfaces inondées ou des niveaux d'eau pendant les crues.

### 3.6.6 Prévision dans un cadre opérationnel

Les travaux cités précédemment annoncent souvent une amélioration des modèles de prévision. En réalité, les auteurs testent des méthodes d'AD en réanalyse (ou "hindcast") d'événements réels, mais encore plus souvent à l'aide d'expériences jumelles rejouant des événements de tempêtes. C'est ainsi le cas des travaux de Karri and Babovic [2017] ou de Chen et al. [2011]. Quelques travaux cités précédemment étudient dans un contexte opérationnel le bénéfice de la mise en place d'une méthode ensembliste appliquée à un modèle de prévision.

Dans le domaine de la prévision des surcotes en domaine côtier, Zijl et al. [2015] utilise un SSKF (Steady State Kalman Filter) pour corriger l'état sur le maillage du modèle DCSM-v6 de surcotes construit sur le plateau continental de l'Europe de l'Ouest et en mer du Nord. Les observations des stations les plus précises en termes de modélisation de la marée sont assimilées pendant une année. Les scores en termes de RMSE, de RMSE sur les pleines mers pour les stations assimilées ou non, ainsi que la capacité du modèle à reproduire la surcote survenue lors de la tempête de novembre 2007 à l'Ouest de l'embouchure de l'Escaut, sont évalués. Chaque cycle d'assimilation dure 6 heures et l'étude opérationnelle est fondée sur l'amélioration de l'état du système issue de la phase d'analyse à l'issue du cycle et l'utilisation des prévisions d'ensemble météorologiques à

horizon de 48 h.

Sur l'estuaire de l'Adour, Barthélémy et al. [2017] met en place un prototype incluant le modèle de prévision des crues de l'Adour maritime fondé sur le code de calcul 1D Mascaret (http://www.opentelemac.org/) et un filtre de Kalman d'Ensemble (EnKF), dont le vecteur de contrôle est composé de l'état du système hydraulique, associé (niveaux d'eau et débits) à une méthode de spatialisation de l'inflation fondée sur l'utilisation du critère de Desroziers (Desroziers et al. [2005]). Les membres de l'ensemble sont générés en ajoutant un bruit corrélé temporellement aux forçages amont et aval. Après avoir évalué les performances du prototype à l'aide d'expériences synthétiques puis en réanalyse sur des événements réels, Barthélémy [2015] évalue le prototype en mode prévision à l'aide d'une expérience synthétique issue de la réanalyse de l'événement de 2004. Chaque cycle d'assimilation est suivi de 12 h de prévision. Cette durée correspond ici au temps de transfert du forçage amont vers la condition aux limites maritimes à l'aval du modèle Mascaret. Par ailleurs, les forçages amont sont supposés constants et égaux aux derniers forçages disponibles avant la phase de prévision. Barthélémy [2015] conclut que la qualité de la prévision au vu des différents critères de performance calculés aux échéances inférieures aux temps de transfert de la condition à la limite amont vers les stations aval dépend fortement de la connaissance des forçages avant et pendant la prévision. L'auteur souligne que des forçages constants en mode prévision ne permettent pas d'établir des prévisions ensemblistes valides à moyen et long terme.

## 3.7 Conclusions

L'état de l'art de la section 3.6 décrit les pratiques actuelles de mise en œuvre des méthodes d'AD pour l'amélioration des états des systèmes dans les domaines océaniques, côtiers et fluviaux. Il permet de mettre en lumière les différences et les complémentarités entre l'approche proposée dans le cadre de ce travail de thèse et celles utilisées dans la littérature.

Il apparaît que, si la correction des conditions aux limites apparaît primordiale en domaine estuarien ou fluvial, son incertitude peut être prise en compte dans la génération de l'ensemble en corrigeant l'état du système (Bertino et al. [2002]) ou directement dans le vecteur de contrôle (Ricci et al. [2011]). Des travaux (Canižares Torres [1999] ; Zijl et al. [2015]) cités dans la section 3.6 s'appuient sur la matrice de covariance d'erreur de l'état du système aux abords de la frontière liquide pour déterminer l'évolution temporelle de la structuration de l'anomalie. La recherche d'une structuration temporelle de l'erreur sur les forçages aux frontières du modèle en lien avec la caractérisation physique du processus (surcote météorologique ou débit fluvial) paraît peu abordée dans la littérature. Or, les conditions aux limites sont souvent fournies par des modèles de plus grande emprise dans le cas des forçages maritimes ou par des modèles hydrologiques ou hydrauliques sur le bassin versant amont de la zone étudiée dans le cas des forçages fluviaux (cf. section 1.2.2.3 dans le cas du modèle de prévision des hautes eaux de l'estuaire de la

Gironde). Alors que l'usage des méthodes d'AD consiste à corriger l'état du système ***prior*** (donc dans le passé), la correction des conditions aux limites a pour objet de modifier spatialement l'état du système pour diminuer l'erreur qui y est propagée dans le temps.

Les coefficients de frottement sont estimés dans plusieurs études, que ce soit dans le domaine de l'hydrodynamique fluviale 1D (Habert et al. [2016] ; Cooper et al. [2018]) ou 2D (Ziliani et al. [2019]), ainsi qu'en domaine estuarien schématique (Siripatana et al. [2018]). Cependant, la littérature montre qu'ils sont estimés seuls (Habert et al. [2016] ; Siripatana et al. [2018]), conjointement avec l'état du système ou de manière duale avec celui-ci (Tamura et al. [2014], Moradkhani et al. [2005]). L'estimation de la structuration spatiale bidimensionnelle des coefficients de frottement au travers d'une décomposition de KL dans un cadre stationnaire est investie par Siripatana et al. [2018]. Cependant, l'évolution temporelle des coefficients de frottement dans son interaction avec la marée et les forçages maritimes ou fluviaux n'est pas abordée dans la littérature.

Dans le cadre de ces travaux de thèse, il est proposé de caractériser, dans un cadre de modélisation hydrodynamique bidimensionnelle en milieu estuarien, donc en régime instationnaire, à l'aide d'un EnKF stochastique, la structuration temporelle de l'erreur des forçages aux limites (BC) du modèle dans un espace réduit de KL conjointement avec les paramètres du modèle. La correction de l'état est une conséquence de celle des conditions aux limites et des paramètres. La méthode proposée est notée EnKF-$\gamma$-KLBC dans la suite du manuscrit.
Par ailleurs, alors que la section 3.4.3.2 présente les équations permettant la mise en œuvre d'un EnKF stochastique pour l'estimation des paramètres $\gamma$ (EnKF-$\gamma$) d'un modèle d'évolution dynamique pour une observation disponible à un temps donné, l'EnKF-$\gamma$-KLBC est adapté ici pour utiliser plusieurs observations par cycle d'assimilation dans le cadre d'un filtrage séquentiel.
Enfin, il est proposé de traiter le problème de sub-optimalité de l'EnKF en utilisant une méthode de rééchantillonnage fondée sur la persistance de la variance initiale des variables de contrôle au fur à mesure des cycles d'assimilation.

Le chapitre 4 décrit la méthodologie de l'EnKF-$\gamma$-KLBC proposée et appliquée au modèle hydrodynamique bidimensionnel de l'estuaire de la Gironde sous le nom de EnKF-$\gamma$-KLBC-MEG. L'EnKF-$\gamma$-KLBC-MEG est validé et analysé dans un cadre synthétique dans le chapitre 5 puis évalué dans un cadre d'expériences réelles dans le chapitre 6.

# Chapitre 4

# Assimilation de données ensembliste pour l'estimation conjointe des paramètres et des forçages dans un espace réduit de Karhunen-Loève

**Sommaire**


**4.1 EnKF-$\gamma$-KLBC . . . . . 104**
4.1.1 Cyclage des fenêtres d'assimilation de données . . . . . 104
4.1.2 Équations de l'EnKF-$\gamma$-KLBC . . . . . 105
4.1.2.1 Étape de prévision . . . . . 106
4.1.2.2 Étape d'analyse . . . . . 106
4.1.3 Initialisation du vecteur de contrôle de l'EnKF-$\gamma$-KLBC . . . . . 107
4.1.4 Rééchantillonnage de l'ensemble de contrôle . . . . . 107
4.1.5 Synthèse de l'EnKF-$\gamma$-KLBC . . . . . 108
**4.2 Application de l'EnKF-$\gamma$-KLBC à l'estuaire de la Gironde . . . . . 109**
4.2.1 Choix du vecteur de contrôle . . . . . 109
4.2.2 Initialisation du vecteur des variables perturbées du système . . . . . 111
4.2.3 Conséquences d'un rééchantillonnage sur la matrice de covariance croisée entre le vecteur de contrôle et les observations simulées . . . . . 111
4.2.4 Erreurs sur l'observation . . . . . 112
**4.3 Bilan et perspectives générales . . . . . 113**

Le présent chapitre décrit la méthode d'assimilation de données (AD) utilisée dans le cadre de ces travaux de thèse et notée EnKF-$\gamma$-KLBC. Fondée sur l'EnKF stochastique présenté en section 3.4.3 (EnKF-$\gamma$), elle est adaptée dans la section 4.1 pour l'estimation conjointe des paramètres de frottement et d'influence du vent, notés $\boldsymbol{\theta}$, et des conditions aux limites (BC) décomposées dans un espace réduit de Karhunen-Loève (KL) dont les coefficients sont notés $\boldsymbol{\alpha}$. $\boldsymbol{\theta}$ et $\boldsymbol{\alpha}$ sont regroupés dans le vecteur $\boldsymbol{\gamma}$ des paramètres du modèle. La méthode décrite est notée EnKF-$\gamma$-KLBC.

L'initialisation de l'ensemble de contrôle, issue d'une séquence de Sobol' selon la méthodologie décrite dans le chapitre 2, est présentée dans la section 4.2.2. Une méthode de rééchantillonnage fondée sur la persistance de la variance initiale des variables de contrôle est proposée en section 4.1.4 pour traiter le problème de sub-optimalité de l'EnKF.

La section 4.2 présente enfin l'EnKF-$\gamma$-KLBC-MEG qui consiste à appliquer l'EnKF-$\gamma$-KLBC au modèle bidimensionnel de prévision des hautes eaux de l'estuaire de la Gironde, fondé sur le code Telemac2D. Les conséquences de la méthode de rééchantillonnage sur l'évolution temporelle des composantes de la matrice de covariance croisée entre les variables de contrôle modales et les prévisions simulées sont décrites et analysées dans la section 4.2.3.

## 4.1 EnKF stochastique adapté à l'estimation conjointe des paramètres et des conditions aux limites dans un espace réduit de Karhunen-Loève

La section 3.4.3.2 a présenté les équations permettant la mise en œuvre d'un EnKF stochastique pour l'estimation des paramètres $\boldsymbol{\gamma}$ d'un modèle d'évolution dynamique pour une observation disponible au temps $t_k$.

La figure 4.2 récapitule de manière synthétique les étapes de l'EnKF-$\gamma$-KLBC qui sont détaillées dans la section 4.1.5.

### 4.1.1 Cyclage des fenêtres d'assimilation de données

Considérons une fenêtre (ou cycle) d'assimilation de données $\mathcal{F}_k$ définie par un vecteur temporel $[t_{k,deb}\,;t_{k,fin}]$ où $t_{k,deb}$ (resp. $t_{k,fin}$) est le temps initial (resp. final) de la fenêtre d'assimilation $\mathcal{F}_k$ de durée $L\Delta t$, telle que $[t_{k,deb}\,;t_{k,fin}] = [T_0\,;T_L]$. $\Delta t$ (en heures (h)) est la période d'assimilation des observations disponibles. $L$ est le nombre de pas de temps $\Delta t$ dans une fenêtre d'assimilation.

L'EnKF stochastique permet d'enchaîner les fenêtres d'assimilation de données $\{\mathcal{F}_k\}_{k\in\{1,..,nb_{cycles}\}}$ où $nb_{cycles}$ est le nombre de cycles d'assimilation de données d'une expérience.
Les cycles d'assimilation peuvent se superposer. $nb_{cycles}$ dépend de la durée totale de l'expérience, mais aussi de la durée $L\Delta t = (t_{k,fin} - t_{k,deb})$ de chaque cycle et de la durée $S\Delta t = (t_{k,deb} - t_{k-1,deb})$ qui sépare deux cycles d'assimilation successifs, conformément à la figure 4.1.

$L$ et $S$ sont supposés constants dans le cadre de ce travail de thèse.

$n_{obs}$ est défini comme le nombre total d'observations disponibles sur $\mathcal{F}_k$. Si $N_e$ désigne le nombre de membres de l'ensemble, ces observations multi-spatialisées et multi-temporelles sont perturbées selon l'équation (3.20) et regroupées dans le vecteur d'observation $\mathbf{y}_k^0$ ($\in \mathbb{R}^{n_{obs} \times N_e}$).

La figure 4.1, adaptée de Bocquet and Sakov [2014], représente l'enchaînement des cycles d'assimilation $\mathcal{F}_{k-1}$ et $\mathcal{F}_k$ dans le cas particulier où $L = 6$ et $S = 2$. Si les observations $\mathbf{y}^0$ sont disponibles toutes les $\Delta t = 1\ h$, alors chaque cycle d'assimilation dure $L\Delta t = 6\ h$ et le cycle $\mathcal{F}_{k-1}$ est séparé de $S\Delta t = 2\ h$ du cycle $\mathcal{F}_k$. Les observations assimilées au cours de chaque cycle sont représentées par des points colorés en noir. Dans ce cas particulier, les $(L-S) = 4$ dernières observations du cycle $\mathcal{F}_{k-1}$ sont également assimilées au cours de $\mathcal{F}_k$. Dans le cas particulier où $L = S$, les cycles d'assimilation sont disjoints et chaque observation est assimilée une seule fois.

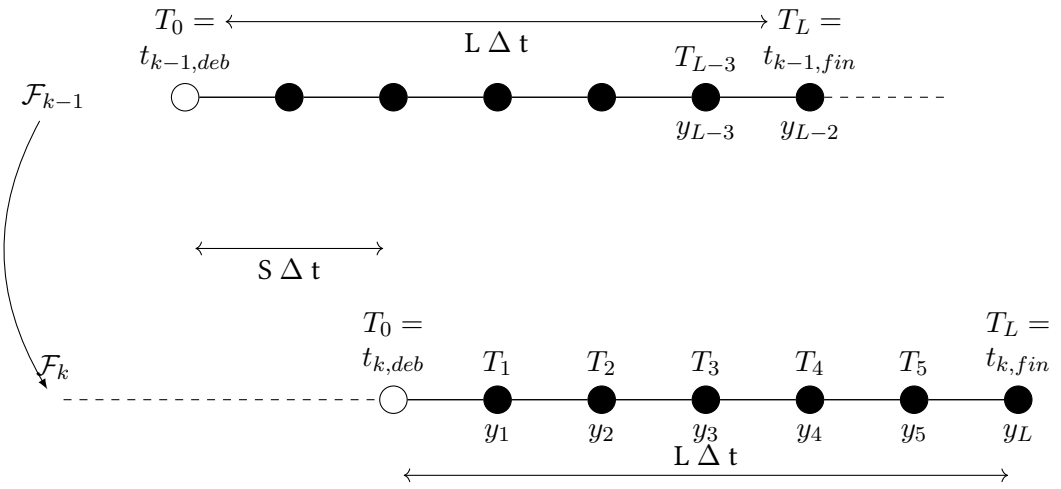



FIGURE 4.1 : Exemple d'enchaînement de deux cycles d'assimilation de données $\mathcal{F}_{k-1}$ et $\mathcal{F}_k$ pour l'EnKF-$\gamma$-KLBC pour $L = 6$ et $S = 2$ (adapté de Bocquet and Sakov [2014]).

### 4.1.2 Équations de l'EnKF-$\gamma$-KLBC

L'équation 4.1 permet de perturber chaque condition aux limites (BC) à l'aide d'une distribution gaussienne $\{\alpha_i\}_{i\in(1,\cdots,p)}$ de coefficients appliqués à chacun des $p$ modes issus de la décomposition de KL de la fonction d'autocorrélation d'un forçage fonctionnel (dépendant du temps), tel que la condition aux frontières maritime et fluviales, par exemple (cf. 2). $\boldsymbol{\alpha}$ est le vecteur des coefficients modaux.

$$q_p = (q_p(t_1), \ldots, q_p(t_N))^T = \sum_{i=1}^{p} \sqrt{\lambda_i}\Phi_i\alpha_i. \tag{4.1}$$

Soit $\theta$ l'espace des paramètres supposés constants lors d'une phase préalable de calage (coefficients de frottement et coefficient d'influence du vent d'un modèle hydrodynamique, par

exemple).
Si $\boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\theta} \end{pmatrix}$, les équations de l'EnKF-$\gamma$-KLBC sont identiques à celles présentées dans la section 3.4.3.2 du chapitre 3 en considérant que l'indice $k$ est désormais relatif à un cycle d'assimilation $\mathcal{F}_k$ et non plus au temps $t_k$. Les étapes de l'EnKF-$\gamma$-KLBC sont rappelées ci-dessous. Elles permettent la mise œuvre de l'EnKF-$\gamma$-KLBC, EnKF stochastique pour l'estimation du vecteur de contrôle $\boldsymbol{\gamma}$. L'état initial du système, au début de chaque fenêtre d'assimilation $\mathcal{F}_k$, correspond à l'intégration du modèle $\mathbf{M}_{k:k-1}$ en utilisant l'état analysé $\mathbf{x}^a_{k-1}$ et les paramètres $\boldsymbol{\gamma}^a_{k-1}$ issus du cycle d'assimilation $\mathcal{F}_{k-1}$.

#### 4.1.2.1 Étape de prévision

L'étape classique de prévision consiste en la propagation du vecteur de contrôle $\boldsymbol{\gamma}$ de $\mathcal{F}_{k-1}$ à $\mathcal{F}_k$ selon le modèle de persistance appliqué à chacune de ses composantes $\boldsymbol{\theta}$ et $\boldsymbol{\alpha}$, même si les conditions aux limites sont variables dans le temps :

$$\boldsymbol{\gamma}^f_k = \boldsymbol{\gamma}^a_{k-1} + \boldsymbol{\tau}_k \quad avec\ \boldsymbol{\tau} \sim \mathcal{N}(0, \mathbf{P}^{\gamma}_k). \tag{4.2}$$

$\mathbf{P}^{\gamma}_k$ est la matrice de covariance d'erreur d'ébauche des paramètres contrôlés sur $\mathcal{F}_k$, calculée à l'aide des membres de l'ensemble.

La trajectoire d'ébauche $\mathbf{x}^f_k$, qui représente l'état du système intermédiaire, est obtenue en utilisant les équations suivantes :

$$\mathbf{x}^f_k = \mathbf{M}_{k:k-1}(\mathbf{x}^a_{k-1}, \boldsymbol{\gamma}^f_k), \tag{4.3}$$

$$\mathbf{y}^f_k = \mathbf{H}_k(\mathbf{x}^f_k, \boldsymbol{\gamma}^f_k). \tag{4.4}$$

$\mathbf{H}_k : \mathbb{R}^m \longrightarrow \mathbb{R}^{n_{obs}}$ est l'opérateur d'observation linéaire de l'espace de l'état du modèle vers l'espace des observations où $m$ est la dimension du vecteur d'état du système et $n_{obs}$ la dimension du vecteur d'observation $\mathbf{y}^0_k$. Il s'agit d'une matrice de sélection, d'extraction et d'interpolation des résultats du modèle d'évolution dans l'espace d'observation.

$\mathbf{M}_{k:k-1} : \mathbb{R}^m \longrightarrow \mathbb{R}^m$ est le modèle d'évolution de l'état du système de $t_{k-1}$ à $t_k$ linéarisé.

#### 4.1.2.2 Étape d'analyse

L'actualisation, ou analyse, des paramètres du modèle est réalisée à l'aide de l'équation classique de l'EnKF (équation (3.27)) :

$$\boldsymbol{\gamma}^a_k = \boldsymbol{\gamma}^f_k + \mathbf{K}^{\gamma}_k\ (\mathbf{y}^0_k - \mathbf{y}^f_k). \tag{4.5}$$

$\mathbf{K}_k^{\boldsymbol{\gamma}}$ est la matrice gain de Kalman qui permet de corriger la trajectoire des paramètres $\boldsymbol{\gamma}$ et est obtenue à l'aide de l'équation suivante :

$$\mathbf{K}_k^{\boldsymbol{\gamma}} = \mathbf{P}_{k-1}^{\boldsymbol{\gamma},y}\left[\mathbf{P}_{k-1}^{\boldsymbol{\gamma}} + \mathbf{R}_k\right]^{-1}. \tag{4.6}$$

$\mathbf{P}_k^{\boldsymbol{\gamma}}$ et $\mathbf{P}_k^{\boldsymbol{\gamma},y}$, estimés stochastiquement, sont respectivement la matrice de covariance d'erreur de l'ensemble des paramètres prédit et la matrice de covariance croisée entre l'ensemble des paramètres et l'ensemble prédit des observations simulées.

En utilisant les paramètres analysés $\boldsymbol{\gamma}_k^a$, il est alors possible de calculer le nouvel état prédit du système pour chaque temps du cycle d'assimilation $\mathcal{F}_k$ :

$$\mathbf{x}_k^a = \mathbf{M}_{k:k-1}(\mathbf{x}_{k-1}^a, \boldsymbol{\gamma}_k^a), \tag{4.7}$$

$$\mathbf{y}_k^a = \mathbf{H}_k(\mathbf{x}_k^f, \boldsymbol{\gamma}_k^a). \tag{4.8}$$

### 4.1.3 Initialisation du vecteur de contrôle de l'EnKF-$\gamma$-KLBC

Soit $n_{var}$ la dimension du vecteur de contrôle $\boldsymbol{\gamma}$ et $N_e$ le nombre de membres de l'ensemble. En s'appuyant sur la méthodologie de création des ensembles de paramètres et de forçages fonctionnels perturbés développée pour l'étude de quantification d'incertitudes présentée dans le chapitre 2, la phase d'initialisation de la méthode consiste à créer un ensemble d'ébauche $\boldsymbol{\gamma}_0$ au temps initial $t_0$ de dimension ($n_{var}$, $N_e$) en utilisant une séquence de Sobol'.

Chaque membre $i$ de l'ensemble de taille $N_e$ à $t_0$ est noté $\boldsymbol{\gamma}_{0,i}$. Les membres de l'ensemble sont caractérisés par un vecteur moyen $\overline{\boldsymbol{\gamma}_0}$ et une matrice de covariance $\mathbf{P}_0^{\boldsymbol{\gamma}}$.

En posant :

$$\mathbf{X}_0^{\boldsymbol{\gamma}} = \frac{1}{\sqrt{N_e - 1}}\left[\boldsymbol{\gamma}_{0,1} - \overline{\boldsymbol{\gamma}_0}, \cdots, \boldsymbol{\gamma}_{0,N_e} - \overline{\boldsymbol{\gamma}_0}\right] \quad \in \mathbb{R}^{n_{var} \times N_e} \tag{4.9}$$

avec

$$\overline{\boldsymbol{\gamma}_0} = \frac{1}{N_e}\sum_{n=0}^{N_e} \boldsymbol{\gamma}_{0,n} \quad \in \mathbb{R}^{n_{var} \times N_e}, \tag{4.10}$$

la matrice $\mathbf{P}_0^{\boldsymbol{\gamma}}$ de covariance d'erreur d'ébauche des paramètres du système à $t_0$ s'écrit :

$$\mathbf{P}_0^{\boldsymbol{\gamma}} = (\mathbf{X}_0^{\boldsymbol{\gamma}})(\mathbf{X}_0^{\boldsymbol{\gamma}})^T \quad \in \mathbb{R}^{n_{var} \times n_{var}}. \tag{4.11}$$

### 4.1.4 Rééchantillonnage de l'ensemble de contrôle

La phase de prévision par persistance des paramètres $\boldsymbol{\gamma}$ est décrite par l'équation (4.2) dans le cas général. Or, le nombre limité de membres de l'ensemble, $N_e$, produit des erreurs d'échan-

tillonnage lors du calcul des matrices de covariances qui peuvent conduire à l'effondrement de l'ensemble du fait du manque de dispersion en son sein. L'EnKF ignore alors les observations. Les méthodes d'inflation multiplicative et additive exposées dans la section 3.4.3.1.c sont couramment utilisées pour la correction de l'état du système. Compte-tenu de la nature du vecteur de contrôle qui ne contient ici "que" les paramètres du modèle et dans l'objectif de redisperser "artificiellement" l'ensemble à l'issue de la phase d'analyse, il est proposé de mettre en œuvre et de tester un rééchantillonnage de l'ensemble $\boldsymbol{\gamma}^a$ qui consiste à contrebalancer la sous-estimation des erreurs en appliquant à chaque membre de l'ensemble l'anomalie initiale $\sqrt{N_e-1}\ \mathbf{X}_0^{\boldsymbol{\gamma}}$ gouvernée par la séquence de Sobol' ayant permis de créer l'ensemble $\boldsymbol{\gamma}_0$ initial selon l'équation :

$$\boldsymbol{\gamma}_k^f = \overline{\boldsymbol{\gamma}_{k-1}^a} + \sqrt{N_e-1}\mathbf{X}_0^{\boldsymbol{\gamma}}, \tag{4.12}$$

ainsi :

$$\boldsymbol{\gamma}_k^f = \overline{\boldsymbol{\gamma}_{k-1}^a} + \boldsymbol{\gamma}_0 - \overline{\boldsymbol{\gamma}_0}. \tag{4.13}$$

Il s'ensuit que le rééchantillonnage utilise la moyenne de l'ensemble issue de la phase d'analyse précédente ($\overline{\boldsymbol{\gamma}}_k^f = \overline{\boldsymbol{\gamma}}_{k-1}^a$) et conserve la matrice de covariance d'erreur d'ébauche de l'ensemble initial ($\mathbf{P}_k^{\boldsymbol{\gamma}} = \mathbf{P}_0^{\boldsymbol{\gamma}}$).

### 4.1.5 Synthèse de l'EnKF-$\gamma$-KLBC

Conformément à la figure 4.2, les phases successives de l'EnKF-$\gamma$-KLBC sont ainsi :

- la prévision : au début du cycle $\mathcal{F}_k$=$[t_{k,deb}\,;t_{k,fin}]$, le vecteur de contrôle $\boldsymbol{\gamma}_k^f$, de dimension $(n_{var},\ N_e)$, est obtenu en utilisant soit la persistance de $\boldsymbol{\gamma}_{k-1}^a$ (cf. section 4.1.3) soit une méthode de rééchantillonnage (cf. section 4.1.4). La composante $\boldsymbol{\alpha}_k^f$ du vecteur de contrôle $\boldsymbol{\gamma}_k^f$ est transcrite dans l'espace des BC fonctionnelles du modèle sur $\mathcal{F}_k$.
- la simulation : pour chaque membre de l'ensemble, l'état du système est obtenu par intégration du modèle $\mathbf{M}_{k:k-1}$ sur $\mathcal{F}_k$ en utilisant les paramètres et les forçages $\boldsymbol{\gamma}_k^f$ et l'état initial $\mathbf{x}_{k-1}^a(t_{k,deb})$ issu de la fenêtre d'assimilation précédente $\mathcal{F}_{k-1}$.
- l'analyse : les résultats du modèle sont transférés dans l'espace des observations en chaque nœud et à chaque pas de temps de $\mathcal{F}_k$ pour lesquels une mesure est disponible. L'ensemble des mesures (la surface libre ou les composantes horizontales moyennées sur la verticale des vitesses de l'écoulement, par exemple) disponibles sur $\mathcal{F}_k$ est confronté aux résultats du modèle sur cette fenêtre d'assimilation obtenus avec les forçages $\boldsymbol{\alpha}_k^f$ et les paramètres $\boldsymbol{\theta}_k^f$. Le vecteur de contrôle "analysé" $\boldsymbol{\gamma}_k^a$ sur la fenêtre $\mathcal{F}_k$, prenant en compte les observations "assimilées", est alors calculé. La composante $\boldsymbol{\alpha}_k^a$ de $\boldsymbol{\gamma}_k^a$ est alors transcrite en termes de BC sur la fenêtre $\mathcal{F}_k$.
- l'optimisation de l'état du système consiste alors à intégrer l'état sur $\mathcal{F}_k$ en utilisant comme condition initiale $\mathbf{x}_{k-1}^a(t_{k,deb})$, ainsi que $\boldsymbol{\gamma}_k^a$, c'est-à-dire les BC correspondant à $\boldsymbol{\alpha}_k^a$ et les

paramètres $\boldsymbol{\theta}_k^a$.

## 4.2 Application de l'EnKF-$\gamma$-KLBC à l'estuaire de la Gironde

### 4.2.1 Choix du vecteur de contrôle

Les variables de contrôle, regroupées au sein du vecteur de contrôle $\boldsymbol{\gamma}$, sont corrigées à l'aide de l'EnKF-$\boldsymbol{\gamma}$-KLBC décrit dans la section 4.1.2 au cours des cycles d'assimilation successifs.

La composante $\boldsymbol{\theta}$ de $\boldsymbol{\gamma}$ intègre les paramètres usuels des équations de Saint-Venant décrits dans la section 1.2.2.1 sur lesquelles sont fondés les codes de calcul bidimensionnels à surface libre tels que Telemac2D. Ce sont en particulier les coefficients de frottement $K_s$ ou le coefficient d'influence du vent $C_d$ (constant en espace et en temps) qui s'appliquent à chaque pas de temps de calcul sur chaque nœud du maillage.

Comme décrit de manière pratique pour le modèle de prévision des hautes eaux dans l'estuaire de la Gironde en section 1.2.3, les coefficients de frottement du modèle d'évolution $(\mathcal{M})$ (Telemac2D ici) sont choisis constants au cours du temps et distribués en zones homogènes. A l'issue de la phase de calage, réalisée "offline" et une fois pour toutes, le modélisateur choisit un jeu de paramètres qui permet d'optimiser la valeur du score retenu, selon des critères définis au préalable (Nash de Pleine Mer, RMSE, etc.) pour un ou plusieurs événements dont il pourra pondérer l'importance (Hissel [2010]). La distribution spatiale de ces zones est "censée" refléter les variations physiques de la valeur du coefficient de frottement liée à la nature de la granulométrie dans le lit mineur ou majeur du cours d'eau, mais aussi à la variabilité associée au mouvement des sédiments sur des échelles de temps plus longues. D'un point de vue numérique et en pratique, il "absorbe" les hypothèses (turbulence, dimension etc.) liées au code. En réalité, la hauteur d'eau varie au cours du temps dans un cours d'eau du fait des apports fluviaux, mais aussi, en particulier dans un estuaire, de manière périodique en fonction de la marée à laquelle s'ajoutent les surcotes météorologiques. De fait, la réalité physique d'une distribution spatio-temporelle des coefficients de frottement sur le fond, telle que mise en œuvre dans un environnement d'assimilation de données, devient pertinente. L'optimisation du jeu de coefficients de Strickler a fait l'objet de travaux de calibration automatique à l'aide d'algorithmes génétiques dans le cadre du projet GirondeXL 3D (Huybrechts et al. [2018b]). Lors du processus de calage du modèle, mais aussi lors de son utilisation, les autres données d'entrée (données bathymétriques/topographiques et les forçages météorologiques, de même que les forçages temporels aux frontières liquides) sont considérées comme "parfaites".

En milieu côtier et estuarien, les forçages au droit de la frontière maritime liquide, au large, sont des chroniques temporelles de niveaux d'eau et, quand elles sont disponibles, de vitesses de courant. À l'amont des parties fluviales, des chroniques, temporelles ou stationnaires, de débits ou de niveaux d'eau sont imposées aux frontières du modèle.

L'EnKF-$\gamma$-KLBC développé dans le cadre de cette thèse permet de corriger les forçages aux limites du modèle en intégrant dans le vecteur de contrôle soit un incrément sur la variable physique, soit les coefficients modaux de la fonction d'autocorrélation d'une perturbation du signal. Ainsi, deux options existent pour la correction des conditions aux limites maritimes ou fluviales. Les variables de contrôle correspondantes sont notées $CLMAR$, $QGAR$ et $QDOR$.
La première option consiste à considérer la variable de contrôle $dsm$ qui correspond à une translation verticale homogène des niveaux d'eau imposés sur les nœuds de calcul de la frontière maritime. Sur la Garonne (resp. la Dordogne), un incrément $dQGAR$ (resp. $dQDOR$) peut être intégré dans le vecteur de contrôle et consiste en un incrément sur le débit de la Garonne (resp. la Dordogne). La translation verticale des conditions aux limites a déjà fait l'objet de travaux dans des cadres variationnels (Poncot et al. [2017] sur le Raz-Blanchard, Ricci et al. [2011]) et l'estuaire de la Gironde dans le cadre du projet GirondeXL 3D (Barthélémy [2018]). Elle équivaut à corriger le niveau moyen à la frontière du modèle, mais ne recherche pas la compréhension de la structure temporelle de l'erreur vis-à-vis de la physique du processus.
La deuxième option permet la prise en compte de la structure de l'erreur sur les forçages temporels en utilisant les propriétés physiques du signal à corriger (surcote/niveau d'eau ou débit). C'est pourquoi les coefficients modaux $\boldsymbol{\alpha}$ décrits dans le chapitre 2 de la fonction de corrélation de chaque forçage temporel associé à une frontière liquide sont inclus dans le vecteur de contrôle.

En résumé, les variables susceptibles d'être intégrées dans le vecteur de contrôle dans le cadre de l'utilisation de l'EnKF-$\gamma$-KLBC pour le modèle Telemac2D de l'estuaire de la Gironde sont les suivantes :

- les coefficients de Strickler des 4 zones $A_1$, $A_2$, $A_3$ et $A_4$ ($Ks1$, $Ks2$, $Ks3$, $Ks4$) (selon les notations du chapitre 2),
- le coefficient d'influence du vent $Cd$,
- la condition à la limite maritime ($CLMAR$) soit par un incrément global appliqué sur la frontière maritime (noté $dsm$ dans la suite du manuscrit), soit par les coefficients modaux gouvernant la condition à la limite maritime dans l'espace réduit de KL, de dimension $p_{CLMAR}$ : $\{\boldsymbol{\alpha}_{CLMAR,j}\}_{j\in\{1,..,p_{CLMAR}\}}$ (notés $(alpha\,\{j\})_{j\in\{1,..,p_{CLMAR}\}}$ sur les figures des chapitres 5 et 6),
- les conditions aux limites fluviales sur la Dordogne ($QDOR$) et la Garonne ($QGAR$), soit par un incrément global (notés respectivement $dQDOR$ et $dQGAR$), soit par les coefficients modaux appliqués aux composantes principales de la fonction d'autocorrélation des chroniques temporelles de débits obtenues dans l'espace réduit de KL de dimensions $p_{DOR}$ et $p_{GAR}$ : $\{\boldsymbol{\alpha}_{i,j}\}_{(i,j)\in(\{DOR,GAR\},\{1,..,p_j\})}$ avec $p_j \in \{p_{DOR}, p_{GAR}\}$ (notés $(Coeff\,\{i\}\,\{j\})_{(i,j)\in(\{DOR,GAR\},\{1,..,p_j\})}$ sur les figures des chapitres 5 et 6).

### 4.2.2 Initialisation du vecteur des variables perturbées du système

De manière pratique, au temps initial $t_0$ de début d'une expérience d'assimilation de données, la séquence de Sobol' $\mathcal{S}$ de dimension ($n_{var}$, $N_e$) permet de créer, pour chaque composante de $\mathcal{S}$, un ensemble de $N_e$ membres dont la distribution est uniforme.

Les coefficients de Strickler (resp. d'influence du vent $Cd$) suivent une loi uniforme sur l'intervalle $[Ksinit - \frac{std_Ks}{2} ; Ksinit + \frac{std_Ks}{2}]$ (resp. $[0.678E-6 ; 3.016E-6]$).

Pour chaque mode $i$ avec $i \in \{1, .., p_j\}$ et $p_j \in \{p_{CLMAR}, p_{DOR}, p_{GAR}\}$ de chaque forçage temporel, comme décrit dans le chapitre 2, une distribution gaussienne $\{\boldsymbol{\alpha}_{i,j}\}$ avec $j \in \{1, .., N_e\}$ est obtenue à partir de la distribution uniforme des coefficients correspondants de la séquence de Sobol' en utilisant la méthode de la transformée inverse (Devroye [2006]). Un vecteur temporel de perturbation $\mathbf{q}_{p,j}$ pour le membre $j$ obtenu à l'aide de l'équation (4.1) de l'ensemble est alors appliqué au forçage initial sur la fenêtre temporelle $\mathcal{F}_k$.

Afin de suffisamment disperser l'état du système correspondant à chaque membre de $\boldsymbol{\gamma}$ en début d'expérience, une simulation libre (sans assimilation de données) ou "spin-up" de quelques heures est mise en œuvre. L'état initial de ce "spin-up" est commun à tous les membres.

### 4.2.3 Conséquences d'un rééchantillonnage sur la matrice de covariance croisée entre le vecteur de contrôle et les observations simulées

La technique de rééchantillonnage proposée dans la section 4.1.4 correspond à une redispersion de l'ensemble de contrôle à l'issue de la phase d'analyse du cycle d'assimilation $\mathcal{F}_k$ selon : $P_k^{\boldsymbol{\gamma}} = P_0^{\boldsymbol{\gamma}}$.

L'annexe A, dont les principaux résultats sont résumés ci-dessous, démontre de manière détaillée que le choix de cette méthode implique que l'évolution temporelle de la covariance $\mathbf{P}_k^{\boldsymbol{\gamma},\mathbf{y}} = (p_{i_1,i_{obs}})_{1\leq i_1\leq nvar;1\leq i_{obs}\leq n_{obs}}$ entre les niveaux d'eau calculés par Telemac2D et un coefficient modal choisi comme variable de contrôle est proportionnel au mode propre correspondant. Dans le cas d'un faible nombre de membres $N_e$, comme c'est le cas ici, l'erreur sur la covariance entre un coefficient modal et une station d'observation peut alors être calculée à chaque pas de temps.

La démonstration suppose l'existence d'une fonction de transfert $\beta_{i_{obs}}$ de toute perturbation appliquée à la condition aux limites à la frontière liquide vers chaque station $i_{obs}$ de l'estuaire de la Gironde sous l'influence de cette condition aux limites. Il a ainsi été vérifié que, pour la condition aux limites maritimes, $\beta_{i_{obs}} = 1 + O(1)$ pour les stations les plus proches de l'embouchure de l'estuaire de la Gironde, comme le montre la figure 4.3. Sur les frontières fluviales, pour les stations de Pessac et de La Réole, il vient immédiatement $\beta_{i_{obs}} = 1$. $\beta_{i_{obs}} \neq 1$ correspond physiquement aux non-linéarités du système lors de la propagation de la perturbation vers l'intérieur de l'estuaire.

Si, au temps $t_i$, $inp$ est la ligne de $\mathbf{P}_k^{\boldsymbol{\gamma},\mathbf{y}}(t_i)$ correspondant au coefficient modal $\boldsymbol{\alpha}^{np}$ et $iobs$ la

colonne correspondant à la station d'observation de la matrice $\mathbf{P}_k^{\boldsymbol{\gamma},\mathbf{y}}(t_i)$, alors :

$$\forall k \in \{1, \cdots, nb_{cycles}\}\,,\ p_{inp,iobs}^{(k)}(t_i) \underset{N_e \to +\infty}{\longrightarrow} \beta_{iobs}(t_i)\sqrt{\lambda_{np}}\ \phi_{np}(t_i). \tag{4.14}$$

($\lambda_{np}$, $\phi_{np}$) sont les $np^e$ valeur propre et vecteur propre de $\mathcal{R}^{(0)}$, solution de l'équation de Fredholm, comme indiqué dans le chapitre 2, où :

$$\mathcal{R}^{(0)}(t_i, t_j) = \sum_{np=1}^{n_{modes}} \lambda_{np}\ \phi_{np}(t_i)\ \phi_{np}(t_j) \tag{4.15}$$

et :

$$\begin{aligned}
\mathbf{R}_{dsm}^{(0)}\Phi_{np} &= \lambda_{np}\Phi_{np},\\
\Phi_{np} &= (\phi_{np}(t_1)...\phi_{np}(t_{nt})),\\
\mathbf{R}_{dsm}^{(0)} &= (\mathcal{R}^{(0)}(t_i, t_j)_{1\leq i,j\leq n_p}).
\end{aligned}$$

Au temps $t_i$, si $n_{modes} \in \{p_{CLMAR}, p_{DOR}, p_{GAR}\}$ est le nombre de modes retenu pour la frontière liquide considérée, l'erreur introduite par l'échantillonnage est alors :

$$e_{inp,iobs}^{(0)}(t_i) = \frac{N_e}{N_e - 1}\ \beta_{iobs}(t_i) \left( \sum_{\substack{np'=1 \\ np' \neq np}}^{n_{modes}} \sqrt{\lambda_{np'}}\ D_{np,np'}\ \phi_{np'}(t_i) \right), \tag{4.16}$$

où :

$$D_{np,np'} = \frac{1}{N_e} \sum_{l=1}^{N_e} \left( \boldsymbol{\alpha}_0^{np,l} \boldsymbol{\alpha}_0^{np',l} \right) \tag{4.17}$$

Du fait de la redispersion selon la séquence de Sobol' initiale des coefficients modaux appliqués à chaque mode propre de la fonction d'autocorrélation temporelle, l'évolution temporelle de la covariance entre les niveaux d'eau calculés par Telemac2D et un coefficient modal donné est donc proportionnelle au mode propre correspondant. Dans le cas d'un faible nombre de membres $N_e$, comme c'est le cas ici, il est également possible de calculer l'erreur sur la covariance entre un coefficient modal et une station d'observation. Cette erreur peut être évaluée à l'aide de l'équation (4.16). Les écarts au mode portant le signal proviennent, d'une part, des non-linéarités du système ($\beta_{i_{obs}} \neq 1$) et des erreurs d'échantillonnage liées à la valeur de $N_e$.

### 4.2.4 Erreurs sur l'observation

Les observations à la disposition du modélisateur et du prévisionniste sont de natures différentes (***in situ***, satellitaires, numériques) selon le cadre opérationnel ou de réanalyse d'un événement

dans lequel il se trouve. Dans un cadre opérationnel, pour anticiper le fonctionnement hydrodynamique du système, le prévisionniste s'appuie sur les observations des niveaux d'eau issus des marégraphes et limnigraphes installés sur le domaine d'étude, ainsi que sur les prévisions issues de modèles numériques météorologiques, hydrologiques et de propagation.
Dans un cadre de réanalyse d'un événement, l'hydraulicien peut avoir à sa disposition des mesures de champs hydrodynamiques (niveaux d'eau, vitesses des courants, états de mer etc.) issues soit de campagnes satellitaires, souvent non disponibles en temps réel, soit ***in situ***.

La section 3.2.2 évoque les deux types d'erreur associés à chaque observation :

- l'erreur de mesure instrumentale, faible, dépend de l'appareil de mesure utilisé,
- l'erreur de représentativité, quant à elle, exprime dans quelle mesure une observation est représentative du phénomène modélisé. Elle dépend du modèle utilisé et, en particulier, de sa résolution, ainsi que de la qualité de la mesure.

Dans le cadre de l'utilisation de l'EnKF-$\gamma$-KLBC pour la correction des forçages et des paramètres du modèle Telemac2D de l'estuaire de la Gironde, les données d'observation utilisées sont des données de niveaux d'eau issues des marégraphes et limnigraphes installés en 13 points d'observation de l'estuaire, comme indiqué dans le chapitre 2. Les erreurs d'observation sont supposées indépendantes, gaussiennes et d'écart-type $\sigma_{obs} = 1\ cm$ constant dans le temps. L'ordre de grandeur est le même que celui indiqué par Barthélémy [2015] qui évalue l'écart-type des erreurs de mesure des stations d'observation ***in situ*** sur le fleuve Adour à 2.5 cm. De fait, la matrice de covariance d'erreur d'observation utilisée est : $\mathbf{R} = {\sigma_{obs}}^2 \mathbf{I}_{n_{obs},n_{obs}}$ où $I_{n_{obs},n_{obs}}$ est la matrice identité dans $\mathbb{R}^{n_{obs} \times n_{obs}}$.

Aucune étude de sensibilité sur le paramétrage ou d'éventuelles méthodes d'estimation de $\mathbf{R}$ n'a été réalisée dans le cadre de ces travaux de thèse. Des travaux ultérieurs de recherche devront être menés sur le sujet en s'appuyant éventuellement sur les résultats de recherche de Tandeo et al. [2018].

## 4.3 Bilan et perspectives générales

Ce chapitre a présenté, dans un premier temps, la méthode d'assimilation de données ensembliste EnKF-$\gamma$-KLBC qui permet d'estimer de manière conjointe les paramètres d'un modèle d'évolution dynamique et de ses conditions aux limites dans un espace réduit de Karhunen-Loève. Le cyclage de ses fenêtres d'assimilation de données a été détaillé, de même que les équations de son algorithme en phase de prévision, d'analyse et d'optimisation de l'état du système. Une méthode de rééchantillonnage fondée sur la séquence de Sobol' ayant servi à créer le vecteur de contrôle initial est également présentée.

La seconde partie du chapitre a été consacrée à l'application de l'EnKF-$\gamma$-KLBC au modèle de prévision des hautes eaux de l'estuaire de la Gironde fondé sur le code numérique bidimensionnel Telemac2D. Les variables de contrôle dans ce cadre précis ont été détaillées, mettant en lumière

la possibilité d'utiliser l'EnKF-$\gamma$-KLBC-MEG pour la correction directe des frontières liquides via un incrément, mais aussi l'opportunité de corriger les conditions aux limites temporelles dans un espace réduit de KL afin d'estimer la structure temporelle de l'erreur aux frontières du modèle. L'estimation des coefficients modaux et celle des paramètres du modèle peut être réalisée de manière conjointe ou indépendante. L'initialisation du vecteur de contrôle et de l'état perturbé du système, ainsi que le paramétrage de la matrice de covariance d'erreur d'observation, sont également abordés.

Les conséquences de la méthode de rééchantillonnage proposée dans la section 4.1.4 sur l'évolution temporelle de la matrice de covariance entre les variables de contrôle modales et les niveaux d'eau prédits sont détaillées dans la section 4.2.3. Il y est montré que $\mathbf{P}^{\boldsymbol{\gamma},y}$ reste alors quasi-invariante au fil des cycles d'assimilation et que la composante de $\mathbf{P}^{\boldsymbol{\gamma},y}$ pour chaque mode corrigé est proportionnelle au mode lui-même (aux non-linéarités du système estuarien et aux erreurs d'échantillonnage près). En introduisant une fonction de transfert entre une perturbation appliquée à la frontière liquide et celle propagée à chaque station d'observation, l'erreur d'échantillonnage peut être évaluée. Il peut alors être proposé de substituer à $\mathbf{P}_k^{\boldsymbol{\gamma},\mathbf{y}}$ une matrice de covariance issue des simulations réalisées dans le cadre de l'étude de quantification d'incertitudes réalisée avec 10000 membres afin de s'affranchir des erreurs liées à un nombre de membres $N_e$ trop faible, selon la méthodologie proposée par Raboudi et al. [2019]. L'étude de quantification d'incertitudes fondée sur $N_e = 10000$ membres pourrait également permettre d'évaluer la fonction de transfert du large vers chaque station d'observation. Cette piste de recherche n'a pas été mise en œuvre ici et devra faire l'objet de travaux ultérieurs.

Par ailleurs, les conditions météorologiques (champ de pression et composantes horizontales du champ de vent à 10 m) en chaque nœud du maillage sont des forçages spatio-temporels bidimensionnels du modèle. Du fait de la difficulté de créer par krigeage conditionnel un ensemble aux caractéristiques physiques acceptables et de taille supérieure à celle des 35 membres de la Prévision d'Ensemble ARPEGE (PEARP) déjà disponibles, les forçages météorologiques ne peuvent pas être corrigés actuellement et devront être intégrés dans l'EnKF-$\gamma$-KLBC ultérieurement.

Enfin, il peut paraître étonnant, compte-tenu des résultats obtenus dans le cadre de l'analyse de sensibilité globale menée sur l'estuaire de la Gironde décrite dans le chapitre 2, d'avoir intégré dans les variables de contrôle possibles les débits fluviaux $QGAR$ et $QDOR$, ainsi que le coefficient d'influence du vent $Cd$. Les raisons de ce choix sont les suivantes.
D'une part, l'EnKF-$\boldsymbol{\gamma}$-KLBC n'a pas vocation à être utilisé uniquement pour le modèle de prévision des crues de l'estuaire de la Gironde. Ainsi, le modèle de surcotes développé dans le cadre du projet de recherche "Gestion et impact du changement climatique sur le littoral" (Laborie et al. [2015]) est sensible au coefficient d'influence du vent. Par ailleurs, sous réserve de l'utilisation d'un maillage suffisamment raffiné en amont d'un modèle hydrodynamique, l'EnKF-$\boldsymbol{\gamma}$-KLBC peut permettre de reconstituer la chronique de mesures de débits en amont d'une rivière. Cette fonctionnalité peut être utile pour la construction des courbes de tarage. Le choix le plus complet possible dans les variables de contrôle implémenté dans l'EnKF-$\gamma$-KLBC doit permettre son

utilisation dans d'autres cadres d'étude et de recherche.
D'autre part, la possibilité de corriger toutes les variables présentes dans le prototype de quantification d'incertitudes permet de confirmer, ou non, les analyses apportées dans le chapitre 2 du manuscrit.

Le chapitre 5 suivant décrit les expériences réalisées dans un cadre synthétique à l'aide du modèle Telemac2D de l'estuaire de la Gironde décrit dans le chapitre 1 et l'EnKF-$\gamma$-KLBC décrit dans ce chapitre. L'objectif est à la fois de valider la méthodologie, d'optimiser le paramétrage de l'EnKF-$\gamma$-KLBC, mais aussi d'en retirer les enseignements pour une utilisation avec des données observées réelles ou un cadre de prévision opérationnelle abordé dans le chapitre 6.

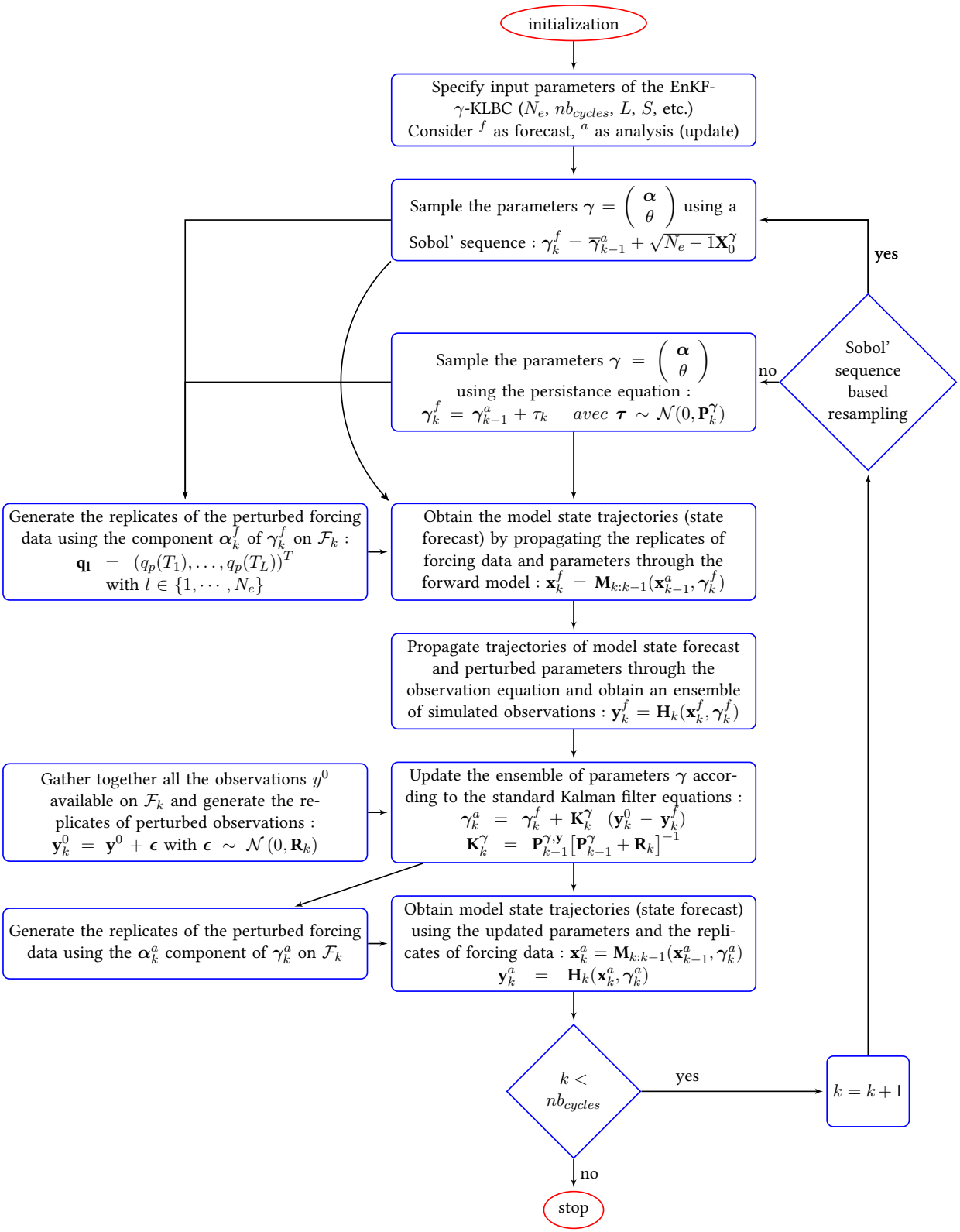



FIGURE 4.2 : Dual parameter and boundary condition estimation flowchart using the stochastic EnKF with reduced boundary conditions in KL space (EnKF-$\gamma$-KLBC) (adapté de Moradkhani et al. [2005]).

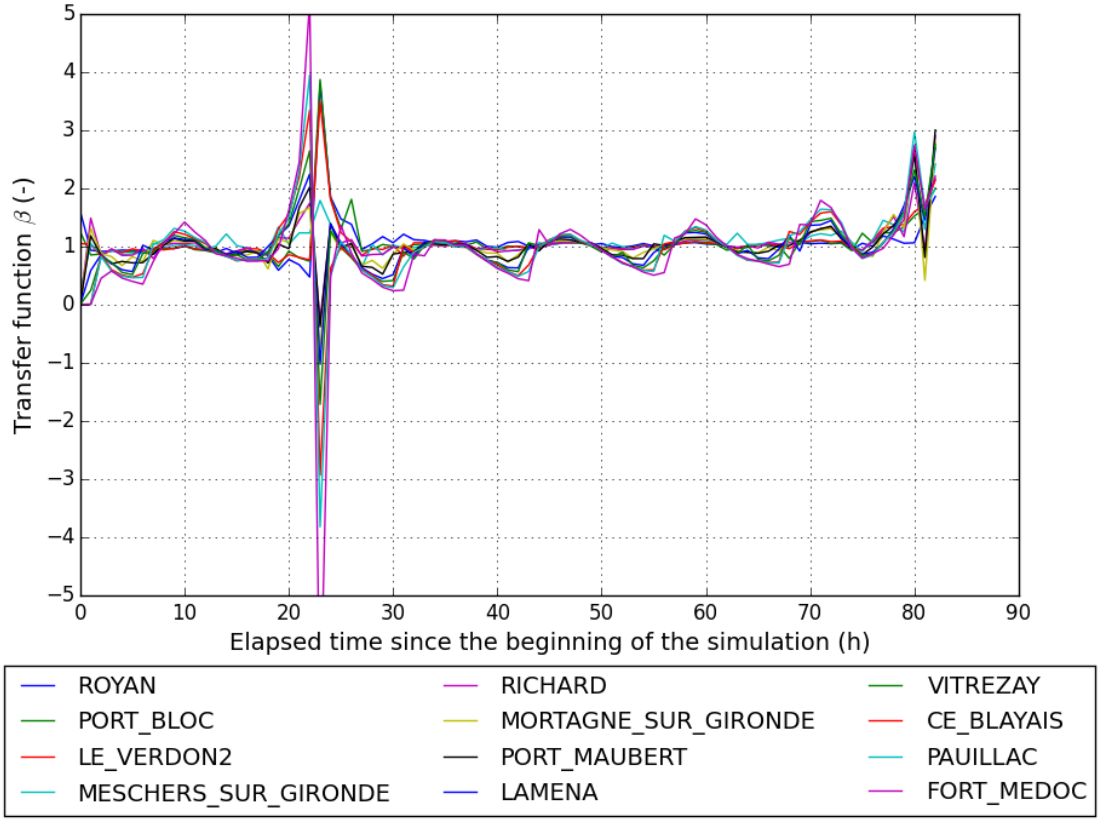



FIGURE 4.3 : Fonction de transfert $\beta$ pour les stations de l'estuaire de la Gironde.

# Chapitre 5

# Validation de l'EnKF-$\gamma$-KLBC-MEG dans un cadre synthétique

**Sommaire**


**5.1 Introduction . . . 121**

**5.2 Dispositif Expérimental Synthétique . . . 122**

5.2.1 Description et objectifs du dispositif d'expériences synthétiques . . . 128

5.2.2 Runs vrais . . . 129

5.2.2.1 Run "vrai" $True1$ : perturbations permanentes appliquées aux coefficients de frottement et aux forçages temporels . . . 130

5.2.2.2 Run "vrai" $True2$ : perturbation variable en temps des forçages et coefficients de Strickler constants ou variables en temps . . . 130

5.2.3 Caractéristiques des expériences jumelles . . . 134

5.2.3.1 Observations synthétiques, vecteur de contrôle et ébauche . . . 134

5.2.3.2 Run de contrôle . . . 135

5.2.3.3 Autres paramètres . . . 137

**5.3 Résultats : estimation des coefficients de Strickler et des forçages dans l'espace réduit de Karhunen-Loève . . . 137**

5.3.1 Estimation de coefficients de Strickler permanents (Exp. 1) . . . 137

5.3.2 Coefficients de Strickler périodiques et redispersion de l'ensemble (Exp. 2) 138

5.3.2.1 Évolution de l'ensemble . . . 138

5.3.2.2 Matrice de covariance d'erreur d'ébauche de l'ensemble . . . 141

5.3.3 Évolution des ensembles dans les espaces de contrôle et d'observation pour l'estimation conjointe de paramètres et de forçages dans l'espace réduit de Karhunen-Loève (Exp. 3) . . . 141

5.3.3.1 Mise en évidence de l'équifinalité . . . 141

5.3.3.2 Reconstruction des conditions aux limites maritime et fluviales 142

5.3.3.3 Matrice de covariance d'erreur d'ébauche de l'ensemble de contrôle pour l'estimation simultanée des paramètres de frottement et des forçages dans l'espace réduit de Karhunen-Loève 145
5.3.4 Conclusions . . . 145
**5.4 Résultats : amélioration des prévisions de niveaux d'eau par l'EnKF-$\gamma$-KLBC-MEG . . . 147**
5.4.1 Prévision des niveaux d'eau par correction de coefficients de Strickler avec perturbation vraie permanente (Exp. 1) et confirmation de l'étude de quantification d'incertitudes . . . 147
5.4.1.1 Prévision des niveaux d'eau sur $A1$ . . . 147
5.4.1.2 Prévision des niveaux d'eau sur $A2$ . . . 148
5.4.1.3 Prévision des niveaux d'eau sur $A3$ . . . 149
5.4.1.4 Prévision des niveaux d'eau sur $A4$ . . . 149
5.4.1.5 Conclusions . . . 150
5.4.2 Niveaux d'eau et scores le long de l'estuaire par correction des coefficients de Strickler avec valeurs vraies périodiques (Exp. 2) . . . 150
5.4.3 Niveaux d'eau et scores le long de l'estuaire par correction des paramètres de frottement (valeurs vraies constantes par morceaux) et des forçages temporels dans l'espace réduit de Karhunen-Loève (Exp. 3-b) 151
5.4.4 Conclusions . . . 151
**5.5 Optimisation des performances de l'EnKF-$\gamma$-KLBC-MEG . . . 152**
5.5.1 Localisation des stations de mesure . . . 152
5.5.2 Nombre de membres de l'ensemble . . . 153
5.5.2.1 Description des expériences . . . 153
5.5.2.2 Performances de l'EnKF-$\gamma$-KLBC-MEG pour la simulation des niveaux d'eau . . . 154
5.5.2.3 Matrices de covariance croisée $\mathbf{P}^{\gamma,y}$ . . . 154
5.5.2.4 Conclusion . . . 157
5.5.3 Influence du maillage . . . 157
5.5.4 Configuration du cycle d'assimilation . . . 158
**5.6 Caractérisation des covariances croisées de l'EnKF-$\gamma$-KLBC-MEG . . . 158**
5.6.1 Influence du rééchantillonnage sur les fonctions de covariance entre les coefficients de frottement et les observations simulées . . . 160
5.6.2 Influence du rééchantillonnage sur les fonctions de covariance croisée pour un vecteur de contrôle intégrant les coefficients de frottement et les coefficients modaux associés aux forçages (Exp. 3-b) . . . 160
**5.7 Conclusions sur les performances de l'EnKF-$\gamma$-KLBC-MEG dans un cadre synthétique et perspectives . . . 162**

## 5.1 Introduction

Les observations utilisées dans le cadre d'expériences d'assimilation de données peuvent être réelles ou synthétiques.

En milieu estuarien, les observations réelles de niveaux d'eau sont obtenues à l'aide de mesures marégraphiques, de campagnes de mesures ponctuelles ou continues (réseau de bouées Candhis pour les états de mer sur le proche côtier, campagnes de mesures réalisées dans le cadre de GirondeXL 3D pour l'estuaire de la Gironde) ou par des mesures issues de campagnes satellitaires. Les observations satellitaires ne sont pas utilisées ici, mais leur utilisation est en cours d'étude sur l'estuaire de la Gironde dans le cadre de travaux menés en collaboration avec l'Université de Rouen, EDF R&D et le CERFACS (projet SWOT-Estuaires du programme TOSCA du CNES). L'assimilation de données "in situ" réelles pour l'estuaire de la Gironde fait l'objet du chapitre 6 du présent manuscrit.

Les observations assimilées peuvent également être synthétiques, c'est-à-dire obtenues en intégrant le modèle sur un autre jeu de forçages et de paramètres, dit "vrai", que la méthode d'assimilation de données cherche alors à atteindre en partant d'un autre jeu dit d'"ébauche" (ou "background"). L'intérêt du cadre synthétique est de maîtriser l'ensemble des erreurs prescrites sur les paramètres et les forçages et de pouvoir tester la sensibilité des performances du modèle à celles-ci. Par ailleurs, contrairement aux mesures réelles, les observations synthétiques ne sont pas lacunaires et leur disponibilité est choisie par le modélisateur. L'inconvénient réside dans le fait que la structure de l'erreur du modèle se retrouve dans celle de l'observation, mettant de côté les erreurs de représentativité (processus et résolution non prises en compte par le modèle). L'utilisation d'un autre modèle (soit plus raffiné soit prenant en compte d'autres processus physiques) que celui auquel la méthode d'assimilation de données est appliquée peut permettre d'atténuer la problématique de la représentativité. Raboudi et al. [2019] et Karri et al. [2014] utilisent ce cadre synthétique particulier dans le cas où l'incertitude sur l'état du système, lui-même corrigé, n'est prise en compte que par les forçages météorologiques.

L'EnKF-$\gamma$-KLBC, décrit dans le chapitre 4, est ici appliqué au modèle de prévision des hautes eaux de l'estuaire de la Gironde de la section 1.2.2 sous la forme d'un EnKF-$\gamma$-KLBC-MEG. Ce dernier est validé dans ce chapitre dans un cadre synthétique en utilisant le dispositif d'expériences synthétiques décrit dans la section 5.2.
Les performances de l'EnKF-$\gamma$-KLBC-MEG sont évaluées, d'une part, dans la section 5.3 au regard de sa capacité à retrouver les paramètres et forçages $\boldsymbol{\gamma}^{true}$ d'expériences de référence. D'autre part, elles se déclinent dans la section 5.4 au travers de l'amélioration de la prévision des niveaux d'eau le long de l'estuaire. Au regard des constats de sub-optimalité et d'équifinalité mis en évidence dans ces deux sections, le paramétrage de l'EnKF-$\gamma$-KLBC-MEG est optimisé dans la section 5.5. La matrice de covariance croisée entre les variables de contrôle et les observations simulées est analysée dans la section 5.6.

## 5.2 Dispositif Expérimental Synthétique

Une expérience jumelle (ou synthétique) est caractérisée par le choix des observations synthétiques $\mathbf{y}^0$ assimilées au cours des cycles d'assimilation et du vecteur de contrôle $\boldsymbol{\gamma}$. Ce dernier, par sa composition, sa taille $N_e$ et son ébauche, détermine à la fois le run de contrôle par rapport auquel sont comparées les performances de la méthode, mais aussi la valeur moyenne initiale $\overline{\boldsymbol{\gamma}}_0$ et la matrice de covariance d'ébauche $\mathbf{P}_0^{\boldsymbol{\gamma}}$ à partir de laquelle les ensembles évolueront.

Le Dispositif Expérimental Synthétique, construit et mis en œuvre dans le but de valider l'EnKF-$\gamma$-KLBC-MEG, est décrit dans la section 5.2.1. Les expériences de référence (runs vrais) et les expériences synthétiques, décrites en détail dans les sections 5.2.2 et 5.2.3, sont respectivement récapitulées dans le tableau 5.1 et dans les tableaux 5.2 à 5.4. Le tableau 5.5 décrit les autres paramètres (durée des fenêtres d'assimilation et décalage entre elles, nombre de membres $N_e$, rééchantillonnage, etc.). Les détails sont donnés par la suite.

TABLE 5.1 : Synthèse des expériences de référence $True1$, $True2a$, $True2b$ et $True2c$ dont les observations synthétiques sont extraites et assimilées au cours des expériences jumelles. Pour chaque expérience de référence ($True1$ ou $True2$), les coefficients de frottement utilisés sont indiqués dans la colonne $Ks^{true}$ (selon le zonage présenté dans le chapitre 1). Concernant les forçages temporels, une perturbation constante appliquée aux forçages réanalysés de l'événement de Février 2003 est indiquée dans la colonne "$d\{q\}^{true}$" ($q \in (sm, QGAR, QDOR)$). Une perturbation variable dans le temps dans l'espace réduit de KL est indiquée dans la colonne "$\alpha_{i,j}^{true}$" (pour $i \in (CLMAR, QGAR, QDOR)$ et $j \in (1, \cdots, p_i)$ où $p_i$ est le nombre de modes retenus pour le forçage $i$). Pour toutes les expériences de référence, $Cd$ est constant et est égal à $2.57 \times 10^{-6}$.

| Exp. | $Ks^{true}$ $[Ks1, Ks2, Ks3, Ks4]^{true}$ | $d\{q\}^{true}$ pour q $\in (sm, QGAR, QDOR)$ | $\alpha_{i,j}^{true}$ pour $i \in (CLMAR, QGAR, QDOR)$ et $j \in (1, \cdots, p_i)$ |
|---|---|---|---|
| $True1$ | [66,34,44,35] | $dsm$=19 cm<br>$dQGAR$=625 $m^3/1$<br>$dQDOR$=89 $m^3/1$ | |
| $True2a$ | [60,60,50,40] | | |
| $True2b$ | Ks constants par morceaux (cf. Fig. 5.2) | | constants (cf. tableau 5.6) |
| $True2c$ | périodiques (cf. Fig. 5.3) | | |

TABLE 5.2 : Observations synthétiques (à gauche), vecteur de contrôle (en haut à droite) et ébauche (en bas à droite) pour l'Exp. 1. Les observations synthétiques sont des niveaux d'eau à un pas de temps horaire. $CLMAR$, $Cd$, $QGAR$ et $QDOR$ ne sont ni perturbés ni contrôlés. "Exp." désigne le nom de l'expérience synthétique. La partie inférieure "Ébauche" des tableaux fournit la moyenne initiale de l'ensemble (ligne "Moy.") et sa distribution initiale (ligne "Loi") pour chaque variable de contrôle. $\mathcal{U}(x_1; x_2)$ désigne la loi uniforme sur l'ensemble support $[x_1 ; x_2]$.

| Exp. | Observations | | $K_s$ | | | |
|---|---|---|---|---|---|---|
| | | | $Ks1$ | $Ks2$ | $Ks3$ | $Ks4$ |
| | Run "vrai" | Localisation | **Vecteur de contrôle** | | | |
| 1a | $True1$ | Le Verdon | ✓ | | | |
| 1b | $True1$ | Pauillac | | ✓ | | |
| 1c | $True1$ | Bordeaux PABx | | | ✓ | |
| 1d | $True1$ | La Réole | | | | ✓ |
| | | | **Ébauche** | | | |
| Exp. | | | 1a | 1b | 1c | 1d |
| Moy. | | | 35 | 60 | 60 | 60 |
| Loi | | | $\mathcal{U}(5; 65)$ | $\mathcal{U}(20; 100)$ | $\mathcal{U}(20; 100)$ | $\mathcal{U}(20; 100)$ |

TABLE 5.3 : Observations synthétiques (à gauche), vecteur de contrôle (en haut à droite) et ébauche (en bas à droite) pour l'Exp. 2. Les observations synthétiques sont des niveaux d'eau à un pas de temps horaire. $Cd$, $Ks4$, $QGAR$ et $QDOR$ ne sont ni perturbés ni contrôlés. "Exp." désigne le nom de l'expérience synthétique. La partie inférieure "Ébauche" des tableaux fournit la moyenne initiale de l'ensemble (ligne "Moy.") et sa distribution initiale (ligne "Loi") pour chaque variable de contrôle. $\mathcal{U}(x_1; x_2)$ désigne la loi uniforme sur l'ensemble support $[x_1 ; x_2]$.

| Exp. | Observations | | $K_s$ | | | $CLMAR$ | |
|---|---|---|---|---|---|---|---|
| | Run "vrai" | Localisation | $Ks1$ | $Ks2$ | $Ks3$ | $dsm$ | $\alpha$ |
| | | | **Vecteur de contrôle** | | | | |
| 2-a | $True2c$ | Le Verdon, Richard, Laména, Pauillac, Fort-Médoc, Bec d'Ambès, Le Marquis, Bassens, Bordeaux Pabx, Bordeaux2. | ✓ | ✓ | ✓ | | |
| 2-b | $True2c$ | Le Verdon, Pauillac, Bordeaux Pabx. | ✓ | ✓ | ✓ | | |
| 2-c<br>2-d | $True2c$ | Le Verdon, Richard, Laména, Pauillac, Fort-Médoc, Bec d'Ambès, Le Marquis, Bassens, Bordeaux Pabx, Bordeaux2. | ✓ | ✓ | ✓ | | ✓ |
| | | | **Ébauche** | | | | |
| Moy. | | | | 35 | | (2-c,2-d) | 0 |
| Loi | | | | 2-a : $\mathcal{U}(0; 70)$<br>2-b : $\mathcal{U}(25; 45)$<br>2-c et d : $\mathcal{U}(25; 45)$ | | ± 30 cm | $\mathcal{U}^{-1}(0, 1)$ |

TABLE 5.4 : Observations synthétiques (à gauche), vecteur de contrôle (en haut à droite) et ébauche (en bas à droite) pour l'Exp. 3. Les observations synthétiques sont des niveaux d'eau à un pas de temps horaire. $Ks4$ n'est ni perturbé ni contrôlé et est égal à 40. "Exp." désigne le nom de l'expérience synthétique. La partie inférieure "Ébauche" des tableaux fournit la moyenne initiale de l'ensemble (ligne "Moy.") et sa distribution initiale (ligne "Loi") pour chaque variable de contrôle. $\mathcal{U}(x_1; x_2)$ désigne la loi uniforme sur l'ensemble support $[x_1 ; x_2]$.

| | **Observations** | | | $K_s$ | | | $CLMAR$ | | $QGAR$ | | $QDOR$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Exp. | Run "vrai" | Localisation | $Cd$ | $Ks1$ | $Ks2$ | $Ks3$ | $dsm$ | $\alpha$ | $dQ$ | $\alpha$ | $dQ$ | $\alpha$ |
| | | | **Vecteur de contrôle** | | | | | | | | | |
| 3-a<br>3-e<br>3-f<br>3-i<br>3-j | $True2b$ | Royan, Richard,<br>Pauillac,<br>Bec d'Ambès,<br>Bordeaux2,<br>La Réole, Pessac. | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | | ✓ |
| 3-b<br>3-g<br>3-h | $True2b$ | Le Verdon, Richard,<br>Laména, Pauillac,<br>Fort-Médoc, Bec d'Ambès,<br>Le Marquis, Bassens,<br>Bordeaux Pabx, Bordeaux2,<br>La Réole, Pessac. | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | | ✓ |
| 3-c | $True2b$ | Lamena,<br>La Réole,<br>Pessac. | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | | ✓ |
| 3-d | $True2b$ | Le Verdon,<br>Pauillac,<br>La Réole,<br>Pessac. | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | | ✓ |
| | | | **Ébauche** | | | | | | | | | |
| Moy. | | | $1.80\ 10^{-6}$ | 35 | 35 | 35 | - | 0 | - | 0 | - | 0 |
| Loi | | | $\mathcal{U}(0.678E-6\ ,$<br>$3.016E-6)$ | $\mathcal{U}(25; 45)$<br>sauf 3-h : $\mathcal{U}(30; 40)$ | | | ± 30 cm | $\mathcal{U}^{-1}(0, 1)$ | ± 20 % | $\mathcal{U}^{-1}(0, 1)$ | ± 20 % | $\mathcal{U}^{-1}(0, 1)$ |

TABLE 5.5 : Autres paramètres d'entrée de la maquette pour les Exp. 1, Exp 2 et Exp. 3.

| Exp. | Durée totale (h) | cycle d'assimilation | | Durée des runs libres (h) | | | $N_e$ | $P_k^\gamma = P_0^\gamma$ |
|---|---|---|---|---|---|---|---|---|
| | | L | S | avant "spin-up" | pendant "forecast" | après "forecast" | | |
| 1a | 24 | 1 | 1 | 6 | 0 | 0 | 50 | NON |
| 1b | 24 | 1 | 1 | 6 | 0 | 0 | 50 | NON |
| 1c | 24 | 1 | 1 | 6 | 0 | 0 | 50 | NON |
| 1d | 24 | 1 | 1 | 6 | 0 | 0 | 50 | NON |
| Exp. 2-a | 96 | 1 | 1 | 6 | 0 | 0 | 100 | NON |
| Exp. 2-b | 96 | 1 | 1 | 6 | 0 | 0 | 100 | OUI |
| Exp. 2-c | 96 | 1 | 1 | 6 | 0 | 0 | 100 | OUI |
| Exp. 2-d | 96 | 1 | 1 | 6 | 0 | 0 | 200 | OUI |
| Exp. 3-a, -b, -c, -d | 96 | 1 | 1 | 6 | 0 | 0 | 100 | OUI |
| Exp. 3-e | 96 | 1 | 1 | 6 | 0 | 0 | 50 | OUI |
| Exp. 3-f | 96 | 1 | 1 | 6 | 0 | 0 | 125 | OUI |
| Exp. 3-g | 96 | 6 | 2 | 6 | 0 | 0 | 100 | OUI |
| Exp. 3-h, -i, -j | 96 | 1 | 1 | 6 | 0 | 0 | 100 | OUI |

### 5.2.1 Description et objectifs du dispositif d'expériences synthétiques

Trois séries d'expériences jumelles, notées Exp. 1, Exp. 2 et Exp. 3, ont permis d'étudier et de valider l'EnKF-$\gamma$-KLBC-MEG afin d'en tirer des enseignements utiles à l'assimilation de données réelles.

L'Exp. 1 est déclinée sous la forme de 4 sous-expériences de validation 1-a, -b, -c et -d, consistant à estimer respectivement et indépendamment le coefficient de Strickler $Ks_i$ de chaque zone $A_i$ (pour $i \in \{1, .., 4\}$), conformément aux notations introduites dans le chapitre 2. Les observations synthétiques utilisées sont issues de l'expérience de référence $True1$ décrite dans la section 5.2.2.1. Elles sont extraites en une seule station située dans la zone d'influence du coefficient considéré. La valeur "vraie"', $Ks_i^{true}$, cible de l'expérience, est constante au cours du temps. La zone d'influence est supposée connue grâce à l'étude de quantification d'incertitudes décrite dans le chapitre 2 de ce manuscrit. L'un des objectifs de l'Exp. 1 est de montrer que l'EnKF-$\gamma$-KLBC-MEG estime de manière satisfaisante des coefficients "vrais" de Strickler permanents pour chaque zone à l'aide d'observations en une seule station et d'étudier l'évolution de la dispersion de l'ensemble de contrôle au cours du temps. Cette expérience met en évidence la sub-optimalité de la méthode et la nécessité, pour les paramètres de frottement du modèle, de mettre en œuvre une méthode de rééchantillonnage du vecteur de contrôle pour redisperser l'ensemble au début de chaque cycle d'assimilation. Elle permet également de confirmer, par le choix d'une unique variable de contrôle pour chaque sous-expérience au sein de l'EnKF-$\gamma$-KLBC-MEG, les zones d'influence de chacune d'entre elles déterminées dans l'étude de quantification d'incertitudes du chapitre 2.

La seconde expérience jumelle est notée Exp. 2. Elle s'intéresse à l'estimation de coefficients de Strickler dont les valeurs vraies sont variables en temps de manière périodique. Les données assimilées sont les données synthétiques $True2c$, décrites dans la section 5.2.2.2, localisées en différents points d'intérêt de l'estuaire. Il s'agit d'évaluer la pertinence de la technique de rééchantillonnage exposée en section 4.1.4 pour l'estimation de coefficients de frottement périodiques dans le temps. Ses conséquences vis-à-vis des corrélations entre les composantes du vecteur de contrôle $\boldsymbol{\gamma}$ et leurs covariances croisées avec les observations simulées par Telemac2D sont étudiées dans le cas où $\boldsymbol{\gamma}$ est composé des seuls coefficients de Strickler $Ks1$, $Ks2$ et $Ks3$. Les Exp. 2-c et 2-d permettent de compléter l'analyse précédente en corrigeant également le forçage de la frontière maritime dans l'espace réduit de KL. L'une des différences entre les sous-expériences Exp 2-a et Exp 2-b réside dans la localisation des "mesures" synthétiques. Ainsi, le vecteur d'observations de l'Exp. 2-b ne contient que trois des stations de mesures utilisées pour les Exp. 2-a, 2-c et 2-d. L'Exp. 2 permet ainsi d'étudier l'influence de la localisation des observations synthétiques sur les performances de l'EnKF-$\gamma$-KLBC-MEG pour une meilleure prévision des niveaux d'eau sur les points d'intérêt de l'estuaire présentés dans la section 1.1.3.2.

L'Exp. 3, enfin, a pour objectif d'évaluer les performances en termes de prévision des niveaux d'eau de l'EnKF-$\gamma$-KLBC-MEG en ces mêmes points de sortie, lorsque les coefficients de Strickler

et les forçages fonctionnels maritimes ou fluviaux au travers de l'estimation des coefficients modaux de l'espace réduit de KL sont corrigés conjointement. Les données synthétiques assimilées sont celles du run vrai $True2b$, décrit dans la section 5.2.2.2 et obtenues avec des Strickler vrais variables (constants par morceaux). La problématique de l'équifinalité du vecteur de contrôle vis-à-vis de la reconstruction des conditions aux limites est mise en évidence. A l'issue de l'expérience, l'erreur entre les conditions aux limites reconstruites par l'EnKF-$\gamma$-KLBC $CLMAR^a$ et les conditions aux limites "vraies" de $True2b$ ($CLMAR^{true}$) est évaluée. L'influence de la longueur de l'ensemble support de la loi uniforme pour l'ébauche des coefficients de Strickler est également étudiée à l'aide de l'Exp. 3-h.

Ce dispositif d'expériences synthétiques a également permis d'étudier l'influence du paramétrage de l'EnKF-$\gamma$-KLBC-MEG dans le but, d'une part, d'en optimiser les performances en matière de prévision des niveaux d'eau dans l'estuaire. Il permet, d'autre part, d'analyser l'inertie du système vis-à-vis des corrections apportées par les paramètres et les forçages dans l'espace réduit de KL, en particulier, ainsi que l'évolution des ensembles de contrôle, des ensembles d'observation simulés et de leurs covariances. L'influence du nombre de membres de l'ensemble ($N_e$) de contrôle, du choix de la localisation des stations d'observation synthétique et de la configuration des cycles d'assimilation a ainsi été étudiée à l'aide des expériences jumelles. Les résultats sont présentés dans la section 5.5.1.

L'influence du rééchantillonnage sur l'évolution temporelle des matrices de covariance lorsque les coefficients modaux associés aux forçages fonctionnels sont intégrés dans le vecteur de contrôle est analysée dans la section 5.6.

### 5.2.2 Runs vrais

Les observations utilisées dans ce chapitre, dites "synthétiques", sont issues de simulations réalisées à l'aide du modèle numérique de l'estuaire de la Gironde fondé sur Telemac2D alimenté par des jeux de forçages et/ou de paramètres supposés "vrais", différents de ceux obtenus à l'issue du calage du modèle sur l'événement de Février 2003 (cf. section 1.2.3). Les résultats numériques (niveaux d'eau) de ces simulations sont extraits en certains points de sortie du maillage le long de l'estuaire de la Gironde, choisis parmi les points de mesures représentés sur la figure 1.8. Ils sont bruités selon l'équation 3.2, puis assimilés en tant qu'observations au cours des différentes expériences.

Trois jeux d'observations synthétiques ont été utilisés. Les paramètres et les forçages du modèle pour les 3 simulations ayant permis de les obtenir sont détaillés dans le tableau 5.1.

#### 5.2.2.1 Run "vrai" $True1$ : perturbations permanentes appliquées aux coefficients de frottement et aux forçages temporels

Les observations synthétiques $True1$ sont issues de simulations utilisant un incrément constant en temps appliqué aux coefficients de Strickler et aux forçages temporels du modèle. Les incréments appliqués aux forçages ($dsm^{true} = 19\ cm$, $dQGAR^{true} = 625$ m$^3$/s et $dQGAR^{true} = 89$ m$^3$/s) correspondent physiquement à une incertitude, sous forme de biais constant en temps, pertinente sur la condition à la limite maritime et égale à 8 % au pic de crue de la Garonne pendant l'événement de 2003.

#### 5.2.2.2 Run "vrai" $True2$ : perturbation variable en temps des forçages et coefficients de Strickler constants ou variables en temps

Les observations synthétiques $True2$ sont issues de simulations pour lesquelles les niveaux d'eau imposés à la frontière maritime et le débit à l'amont de la Garonne (resp. de la Dordogne) sont choisis parmi un ensemble de 1000 membres perturbés à l'aide de la méthode décrite dans le chapitre 2. Cette perturbation est issue d'un unique jeu de coefficients modaux constant dans le temps dont les valeurs sont récapitulées pour chaque mode dans le tableau 5.6.

TABLE 5.6 : Coefficients "vrais" associés aux modes de $CLMAR$, $QGAR$ et $QDOR$ pour l'expérience de référence $True2$

| | $CLMAR$ | $QGAR$ | $QDOR$ |
|---|---|---|---|
| Mode 1 | -4.71×10$^{-01}$ | 5.56×10$^{-03}$ | 1.01 |
| Mode 2 | -9.27×10$^{-01}$ | -1.17 | 9.44×10$^{-01}$ |
| Mode 3 | -7.26×10$^{-01}$ | -1.13 | -1.41 |
| Mode 4 | 4.14×10$^{-01}$ | -4.26×10$^{-01}$ | 4.82×10$^{-01}$ |
| Mode 5 | 1.34 | - | - |
| Mode 6 | 9.62×10$^{-01}$ | - | - |
| Mode 7 | -1.32 | - | - |

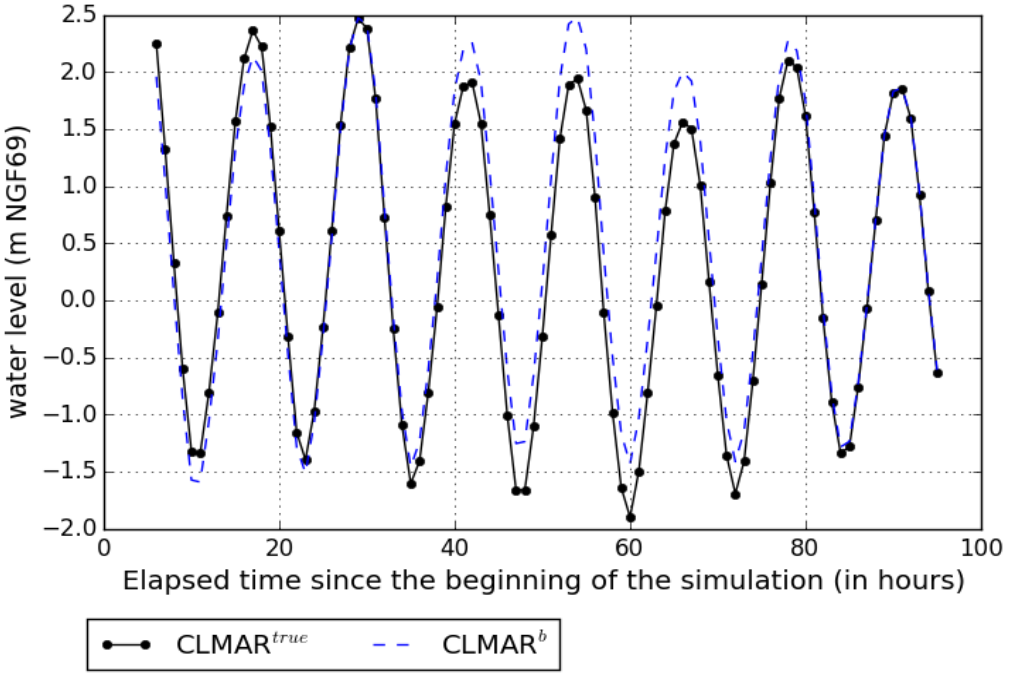
Les niveaux d'eau (resp. le débit de la Garonne) "vrais" ("$CLMAR^{true}$") (resp. "$QGAR^{true}$") résultant de la perturbation appliquée aux forçages de la tempête de 2003 pour la condition à la limite maritime (resp. fluviale à l'amont de la Garonne) sont représentés en trait continu noir sur la figure 5.15.

L'expérience de référence $True2$ se décline en trois jeux de simulations synthétiques $True2_a$, $True2_b$ et $True2_c$ correspondant à trois hypothèses d'évolution temporelle des coefficients de Strickler pour chaque zone.
Pour l'expérience $True2_a$, les coefficients de Strickler utilisés sont les valeurs constantes issues du calage du modèle sur la tempête de Février 2003, c'est-à-dire $Ks1^{true}$ = 60, $Ks2^{true}$ = 60,

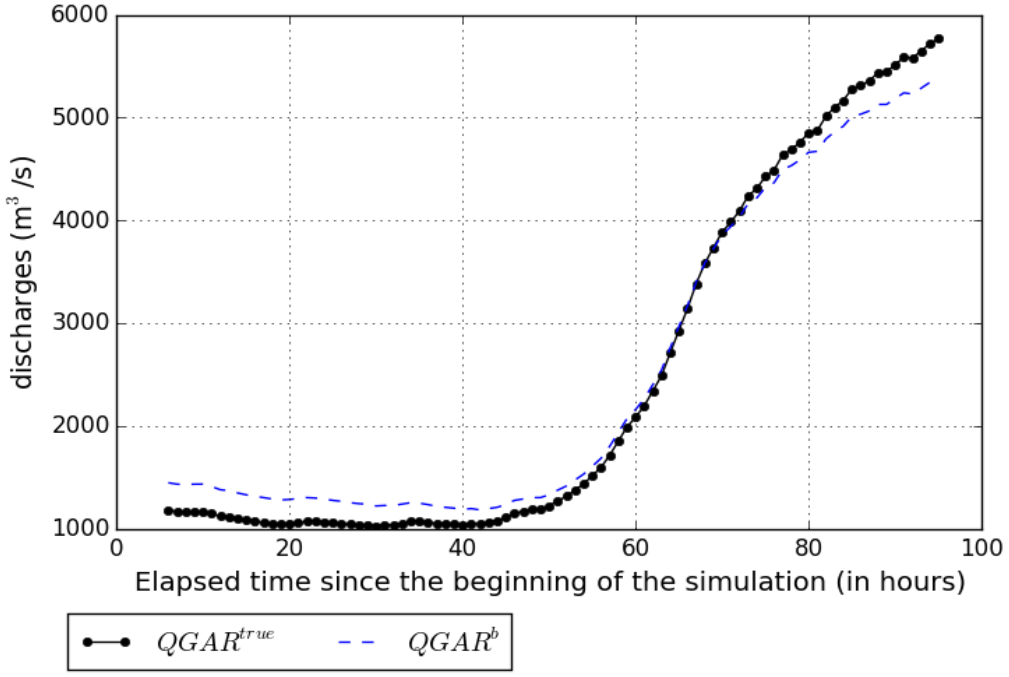
$Ks3^{true}$ = 50, $Ks4^{true}$ = 40.

Pour l'expérience $True2_b$, les coefficients de Strickler utilisés sont variables en temps et sont constants par morceaux (paliers). Ils sont représentés sur la figure 5.2. Chaque palier dure environ 12 h, soit un cycle complet de marée.

Les coefficients de Strickler de l'expérience de référence $True2_c$ suivent une fonction temporelle sinusoïdale représentée sur la figure 5.3 de période $T = 6$ heures, soit à peu près $\frac{1}{2}$ marée centrée autour des valeurs de l'expérience $True2_a$.




(a) CLMAR




(b) QGAR

FIGURE 5.1 : Évolution temporelle (depuis le début de l'événement de 2003) pour l'expérience de référence $True2$ : a) de la condition à la limite maritime $CLMAR$ ; b) de la condition à la limite fluviale sur la Garonne. $CLMAR^{true}$ et $QGAR^{true}$ sont les valeurs "vraies" dont sont issues les observations synthétiques $\mathbf{y}^0$. $CLMAR^b$ et $QGAR^b$ sont les forçages d'ébauche ("background") du run de contrôle.

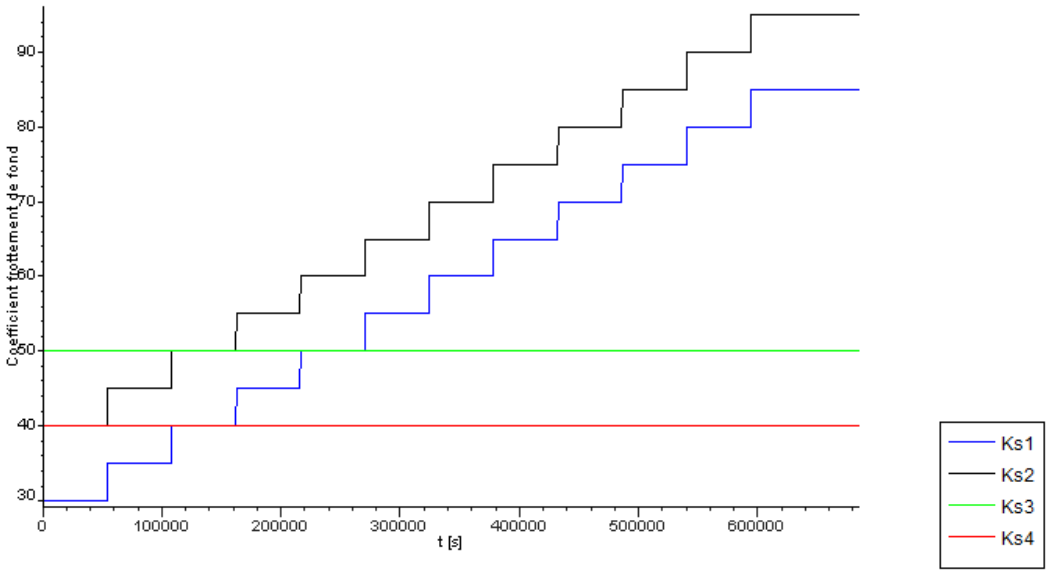



FIGURE 5.2 : Évolution temporelle des coefficients de Strickler pour l'expérience $True2_b$

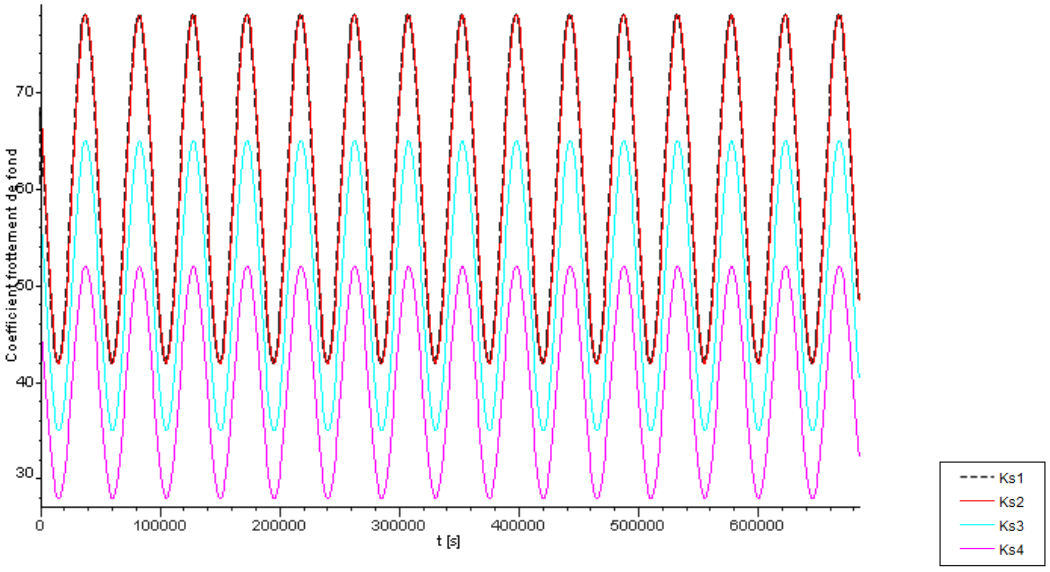



FIGURE 5.3 : Évolution temporelle des coefficients de Strickler pour l'expérience $True2_c$

Les résultats de sortie du run "vrai" peuvent être enregistrés à une fréquence définie par le modélisateur (ici, toutes les 10 mn). Cependant, dans le cadre de ces travaux, une observation sur 6, c'est-à-dire une donnée horaire, cohérente avec la dynamique du processus physique et conforme aux pratiques décrites dans l'état de l'art de la section 3.6, est utilisée. Dans le cadre de l'assimilation de données réelles comme dans le cadre synthétique, le choix de la fréquence et de l'origine des temps de la période de simulation est un point à débattre, car il peut avoir une grande influence sur la qualité des résultats, mais également sur les coûts de calcul.

Concernant la nature des observations synthétiques, toute variable hydrodynamique de type niveau d'eau total (cote de la surface libre H) ou composante horizontale moyennée sur la verticale (U et V) du champ de vitesse pourrait être assimilée par l'EnKF-$\gamma$-KLBC développé et mis en œuvre ici. Dans le cadre de ces travaux, seules des observations de niveaux d'eau sont utilisées. L'objectif de ce travail étant l'examen de l'apport d'une méthode d'assimilation de données pour la simulation et la prévision des niveaux d'eau sur l'estuaire de la Gironde, l'utilisation dans un premier temps de niveaux d'eau seuls est cohérente. Compte-tenu de la disponibilité de campagnes de mesures sur l'estuaire de la Gironde, réalisées dans le cadre de GirondeXL 3D, l'une des suites possibles de ces travaux sera de déterminer si l'assimilation de données de champs de vitesse permettrait d'améliorer les résultats obtenus sur expérience réelle dans le chapitre 6.

### 5.2.3 Caractéristiques des expériences jumelles

L'objectif est de valider et d'étudier les performances de l'EnKF-$\gamma$-KLBC-MEG en assimilant les observations synthétiques extraites des runs "vrais" décrits en section 5.2.2. Tous les paramètres et/ou forçages temporels du modèle Telemac2D de l'estuaire de la Gironde indiqués dans la section 4.2.1 peuvent être intégrés séparément ou conjointement dans le vecteur de contrôle.

#### 5.2.3.1 Observations synthétiques, vecteur de contrôle et ébauche

Les tableaux 5.2, 5.3 et 5.4 récapitulent, d'une part, les caractéristiques des observations synthétiques (qualité, nature, localisation, fréquence et run "vrai") utilisées pour les Exp. 1, Exp 2 et Exp. 3. Pour chaque expérience de la colonne "Exp.", des observations synthétiques horaires ($\Delta t = 1\ h$) de niveaux d'eau sont extraites des simulations du run "vrai" pour chaque station de la colonne "Localisation".
D'autre part, ils synthétisent, pour chaque expérience jumelle, la composition des vecteurs de contrôle et l'ébauche du vecteur de contrôle (moyenne et loi suivie par l'ensemble) au début de l'expérience.
La partie supérieure de chaque tableau intitulée "Vecteur de contrôle" désigne les variables de contrôle incluses dans le vecteur de contrôle $\boldsymbol{\gamma}$ pour chaque expérience de la colonne "Exp.". Le symbole ✓ indique que la variable correspondante est contrôlée. $dsm$ (resp. $dQ$) désigne l'incrément (contrôlé ou résultant du contrôle des forçages dans l'espace réduit de KL) sur la condition

à la limite maritime (resp. fluviale associée). $\alpha$ qualifie la présence des coefficients modaux du forçage correspondant ($CLMAR$, $QGAR$ ou $QDOR$) dans le vecteur de contrôle.
Lorsque la condition à la limite maritime est perturbée dans l'espace réduit de KL, l'évolution de l'écart-type de la perturbation appliquée initialement correspondant à l'ensemble $\alpha^0_{CLMAR}$ est représentée sur la figure 5.4. L'incertitude sur le forçage maritime est alors égal à environ $\pm$ 30 cm.

#### 5.2.3.2 Run de contrôle

Le run de contrôle est la simulation sans assimilation de données, c'est-à-dire dont les paramètres et forçages, fixés au vecteur de contrôle d'ébauche $\boldsymbol{\gamma}^b$ indiqués dans les tableaux 5.2 à 5.4, ne sont pas corrigés au cours du temps. Sur la figure 5.15, la chronique initiale (ou ébauche) ("$CLMAR^b$") (resp. "$QGAR^b$") est représentée à l'aide d'une ligne en pointillés. La moyenne de l'ébauche des forçages correspond aux forçages réanalysés de la tempête de Février 2003. Les observations simulées issues de cette simulation, notées "$\mathbf{y}^b$", sont comparées aux observations simulées "$\mathbf{y}^a$" à l'issue des cycles d'assimilation de données pour déterminer les gains obtenus avec l'EnKF-$\gamma$-KLBC-MEG par rapport à cette simulation de contrôle.

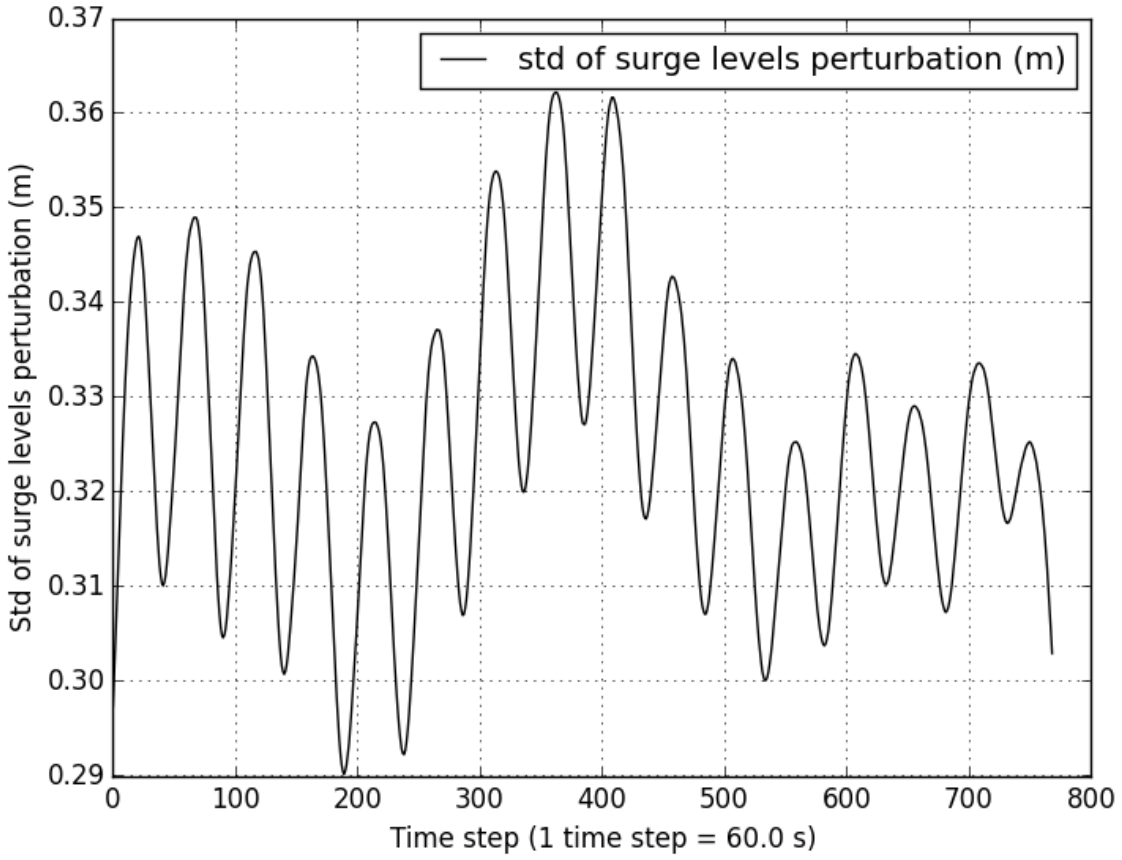



FIGURE 5.4 : Évolution de l'écart-type de la distribution initiale (d'ébauche) des perturbations appliquées à la surcote sur la frontière maritime du modèle $CLMAR^b$.

#### 5.2.3.3 Autres paramètres

Le tableau 5.5 synthétise les autres paramètres d'entrée de l'EnKF-$\gamma$-KLBC pour chaque expérience jumelle. L'Exp. 1 dure 24 h. Les Exp. 2 et Exp. 3 durent 96 h. Toutes débutent pas un spin-up de 6 h.

Le nombre de pas de temps de chaque cycle d'assimilation et entre deux cycles d'assimilation de données successifs valent respectivement $L$ et $S$, conformément au cyclage des fenêtres d'assimilation de données présenté dans la section 4.1.1. $L = 1$ et $S = 1$ pour toutes les Exp. 1, Exp 2 et Exp. 3, sauf pour l'Exp 3-g pour laquelle $L = 6$ et $S = 2$ et qui permet d'étudier l'influence sur les résultats de $L$ et de $S$. Ainsi, sauf pour l'Exp 3-g, il n'y a pas de recouvrement entre deux cycles d'assimilation successifs et chaque observation horaire n'est assimilée qu'une seule fois. L'utilisation (ou non) de la technique de rééchantillonnage décrite dans la section 4.1.4 est indiquée dans la colonne "$\mathbf{P}_k^\gamma = \mathbf{P}_0^\gamma$".

Pour l'Exp. 1, la méthode de rééchantillonnage présentée dans la section 4.1.4 n'est pas appliquée. De fait, la redispersion de l'ensemble au début de chaque cycle d'assimilation de données est assurée par l'injection de bruits gaussiens dans le vecteur de contrôle analysé conformément à l'équation (4.2) du chapitre 4. Pour la variable de contrôle $Ks1$, par exemple, l'ensemble initial $Ks1^b$ est composé de $N_e = 50$ membres tirés dans une loi uniforme $\mathcal{U}(20, 50)$ ($\overline{Ks1^b} = 35$). Cette valeur relativement petite est justifiée par le coût en simulations réalisées avec le code de calcul Telemac2D.

Comme l'indique le tableau 5.5, $N_e$ vaut 100 membres pour les sous-expériences 2-a, 2-b et 2-c et $N_e$=200 pour l'Exp. 2-d. Pour l'Exp. 3, $N_e$ est égal à 100 membres pour les sous expériences 3-a à 3-d et 3-g à 3-i. Une analyse de l'influence du nombre de membres, fondée sur les résultats des Exp. 3-e et 3-f et l'Exp. 2, est réalisée dans la section 5.5.2.
L'Exp. 2 permet également de tester la méthode de rééchantillonnage proposée dans le chapitre 4, dans la mesure où celle-ci n'est pas utilisée pour l'Exp. 2-a.

L'influence du raffinement du maillage utilisé pour l'assimilation des données ou pour les observations synthétiques est analysé à l'aide des Exp. 3-i et Exp. 3-j dans la section 5.5.3.

Les figures représentant les résultats des sections suivantes sont regroupées à la fin du chapitre.

## 5.3 Résultats : estimation des coefficients de Strickler et des forçages dans l'espace réduit de Karhunen-Loève

### 5.3.1 Estimation de coefficients de Strickler permanents (Exp. 1)

L'Exp. 1 permet d'évaluer les performances de l'EnKF-$\gamma$-KLBC-MEG pour l'estimation d'une perturbation permanente appliquée à chaque coefficient de Strickler sur chaque zone séparément en n'assimilant qu'une seule observation. La capacité de l'EnKF-$\gamma$-KLBC-MEG à retrouver des biais

sur les chroniques de débits et les conditions à la limite maritime constants en temps a été validée, mais les résultats ne sont pas montrés ici.

Pour $i \in \{1, \cdots, 4\}$, l'évolution de l'erreur entre $\overline{Ks_i^a}$ (resp. $\overline{Ks_i^b}$) et $Ks_i^{true}$ pour les Exp. 1-a à 1-d est représentée sur les figures 5.6-a à 5.6-d par un trait continu rouge (resp. bleu). La répartition des $N_e$ composantes de l'ensemble $Ks_i^a$ est représentée à l'aide de "boîtes à moustaches" (Jakobowicz [2018]).
Elles montrent que la moyenne des ensembles $\overline{Ks_i^a}$ converge rapidement vers la valeur "vraie" $Ks_i^{true}$. Ainsi, l'ensemble met 3 (resp. 1 et 2) cycles d'assimilation pour atteindre les valeurs "vraies" de $Ks1^{true}$ et $Ks3^{true}$ (resp. $Ks2^{true}$ et $Ks4^{true}$).

Cependant, l'examen de l'évolution des variables sur la figure 5.6 montre une diminution de la dispersion de l'ensemble, conduisant à la mise en place nécessaire d'une stratégie de rééchantillonnage autour de la nouvelle moyenne de l'ensemble. Ce problème de sous-dispersion de l'ensemble au cours des cycles d'assimilation est bien connu sous le nom de sub-optimalité de l'EnKF.

### 5.3.2 Coefficients de Strickler périodiques et redispersion de l'ensemble (Exp. 2)

#### 5.3.2.1 Évolution de l'ensemble

La section 5.3.1 a validé l'utilisation de l'EnKF-$\gamma$-KLBC pour l'estimation de coefficients de Strickler constants indépendamment les uns des autres. La seconde expérience jumelle, notée Exp. 2, s'intéresse à l'estimation simultanée de coefficients de Strickler variables dans le temps en assimilant des données synthétiques localisées en différents points d'intérêt de l'estuaire. Le coefficient de Strickler pondère la force de frottement de l'écoulement sur le fond du cours d'eau. Il dépend en réalité de la hauteur d'eau. Dans le cas d'un cours d'eau soumis à l'influence maritime, le niveau d'eau, somme du signal de marée et du signal de surcote météorologique instantanée, varie de manière périodique. Il paraît donc pertinent d'appréhender la capacité de l'EnKF-$\gamma$-KLBC-MEG à corriger des coefficients de Strickler variables dans le temps à l'aide des mesures relevées le long de l'estuaire.

Les deux différences entre les sous-expériences Exp 2-a et Exp 2-b résident, d'une part, dans la localisation des "mesures" synthétiques. Ainsi, le vecteur d'observations de l'Exp. 2-b ne contient que trois des stations de mesures utilisées pour l'Exp. 2-a. D'autre part, à l'issue de la phase d'analyse de chaque cycle d'assimilation $\mathcal{F}_k$ de l'Exp. 2-b, le vecteur de contrôle analysé est redispersé à l'aide de la technique de rééchantillonnage décrite ci-dessus. Par conséquent, l'objectif de l'Exp. 2 est également d'évaluer la pertinence de techniques de rééchantillonnage vis-à-vis des corrélations existant naturellement entre les composantes du vecteur de contrôle, lorsque celui-ci est uniquement composé des coefficients de Strickler $Ks1$, $Ks2$ et $Ks3$. Dans ce cadre, les forçages temporels ne sont pas corrigés.

Les évolutions de chaque variable de contrôle sans (Exp. 2-a) et avec rééchantillonnage (Exp. 2-b)

sont respectivement représentées sur les figures 5.7 et 5.8.

Pour $i \in \{1, \cdots, 4\}$, la valeur moyenne de l'ébauche (resp. "vraie") de la variable de contrôle $Ks_i$ est représentée par la courbe "$Ks_i^b$" (resp. "$Ks_i^{true}$") en trait continu bleu (resp. noir). La trajectoire de la moyenne de la variable de contrôle au fur et à mesure des cycles d'assimilation est représentée par la courbe "$Ks_i^a$" en trait continu rouge. Chaque marqueur correspond à la valeur moyenne de la variable de contrôle après la phase d'analyse. Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps.

Les résultats obtenus illustrent la capacité de la maquette d'assimilation de données à estimer les coefficients de Strickler variables de manière périodique en temps.

Dans le cadre de l'Exp. 2-a, l'EnKF-$\gamma$-KLBC-MEG parvient à corriger de manière pertinente et dans une gamme de valeurs physiques et proches de la valeur cible les coefficients de Strickler des zones $A_1$, $A_2$ et $A_3$ pendant les 18 premiers cycles d'assimilation. La figure 5.7 montre cependant, comme pour l'Exp. 1, que, pour chaque variable de contrôle, la dispersion de l'ensemble décroît au bout de deux cycles de marée, de telle sorte que chaque membre de l'ensemble est confondu avec sa moyenne après 24 h d'expérience. Les observations sont alors ignorées, menant à une analyse non satisfaisante. Aucun signal haute fréquence n'apparaît sur les variables de contrôle à l'issue de la phase d'analyse, malgré l'utilisation de 12 stations d'observations synthétiques le long de l'estuaire.

Pour $i \in \{1, \cdots, 4\}$, l'évolution des ensembles $Ks_i^a$ au cours de l'Exp. 2-b, représentée sur la figure 5.8, montre la bonne capacité de l'EnKF-$\gamma$-KLBC-MEG à reproduire simultanément l'évolution temporelle périodique des trois coefficients de frottement tout au long des 7 cycles de marée de l'expérience. Cependant, contrairement au vecteur de contrôle obtenu pour l'Exp. 2-a, des perturbations haute fréquence apparaissent sur les signaux des variables de contrôle. Ces perturbations sont liées à la mise en œuvre de la technique de rééchantillonnage. L'une des solutions ***a priori*** pour lisser ce signal peut être de choisir une durée $L$ pour chaque cycle d'assimilation supérieure à 1 h, comme c'est le cas pour l'Exp. 5-g dont les résultats sont exposés dans la section 5.5.4. L'autre solution pourrait consister en un ensemble composé d'un plus grand nombre $N_e$ de membres.

La figure 5.8-a montre que l'amplitude de la variation, ainsi que la périodicité du signal, sont bien reproduites pour $Ks1$. L'amplitude de $Ks2$ est amplifiée par l'EnKF-$\gamma$-KLBC-MEG tout en restant dans une gamme de valeurs physiquement acceptables, bien qu'un peu élevée ($> 80$) au moment des pleines mers (PM). Cependant, la périodicité du signal est également bien reproduite. L'estimation du signal périodique $Ks3$ est plus difficile. Si la périodicité et les ***minima*** de $Ks3$ sont bien représentés, il reste que l'amplitude sur les 60 premiers cycles est trop grande, conduisant à des valeurs trop élevées de ce paramètre, qui peut même atteindre 120. Ce phénomène peut être dû au rééchantillonnage ou à l'équifinalité de $Ks2$ et $Ks3$.
Le rapport de Barthélémy [2018] consacré au calage automatique de paramètres dans le cadre

du projet GirondeXL 3D utilise une méthode d'assimilation de données 3D-Var pour caler le niveau moyen de la marée, le coefficient d'amplitude de la marée, ainsi que les coefficients de frottement $Ks1$, $Ks2$ et $Ks3$. $Ks3$ est uniquement appliqué sur le tronçon fluvial de la Garonne. Cette expérience est menée sur la période estivale anticyclonique (sans surcote météorologique) du 12/08/2015 au 20/08/2015 en assimilant les données de toutes les stations de l'Exp. 2-b à une fréquence horaire. Il ressort de cette étude, d'une part, la difficulté d'estimer simultanément par cette méthode les coefficients liés à la condition à la limite maritime (amplitude et niveau moyen) et les coefficients de frottement et, d'autre part, la difficulté d'estimation du paramètre $Ks3$ dont l'influence est pourtant limitée à la zone $A_3$ autour de Bordeaux. Les valeurs de $Ks3$ estimées sur cette période anticyclonique atteignent 100 pour des niveaux d'eau à Bordeaux similaires à ceux de l'Exp. 2.
Ainsi, l'estimation du paramètre $Ks3$ qui conditionne les niveaux d'eau à Bordeaux constitue un point délicat, indépendamment du raffinement du maillage et de la méthode d'assimilation (ensembliste ou variationnelle) mise en œuvre. Force est de constater, par ailleurs, que la variation de $Ks3$ simulée lors de l'expérience de référence $True2c$ est de 30 unités en 6 heures, ce qui est un peu "raide" et assez peu réaliste de l'évolution réelle de la variable. Les difficultés d'estimation de $Ks3$ peuvent ainsi être liées, dans notre cas, au raffinement du maillage "grossier" dans cette zone sensible située à la confluence des deux rivières, aussi bien qu'à l'amplitude de la cible choisie pour l'expérience jumelle. Cette dernière explication est, cependant, à considérer avec précaution, dans la mesure où la différence entre $Ks\{i\}^{true}$ et $Ks\{i\}^{b}$ pour $i \in \{1,2\}$ des zones $A_1$ et $A_2$ est encore plus grande que pour la zone $A_3$, comme le montre la figure 5.3. Cependant les processus non-linéaires en jeu dans la zone de confluence sur laquelle s'applique $Ks3$ et, en particulier, la relation fortement non-linéaire entre les niveaux d'eau et $Ks3$ peuvent aboutir à des solutions apportées par l'EnKF-$\gamma$-KLBC (et le 3DVar) sous-optimales et conduire aux difficultés rencontrées.

En conclusion, les Exp. 2-a et 2-b montrent que la redispersion de l'ensemble autour de la moyenne de l'ensemble après la phase d'analyse selon le plan de perturbation initial peut permettre de résoudre le problème de sous-dispersion au fur et à mesure de l'expérience pour une meilleure estimation des niveaux d'eau le long de l'estuaire de la Gironde et de reproduire l'évolution temporelle périodique des coefficients de frottement. Il convient toutefois de souligner que, pour des observations synthétiques issues d'une évolution périodique des coefficients de frottement, la redispersion de l'ensemble selon cette méthode induit, d'une part, des perturbations haute fréquence de la variable de contrôle et, d'autre part, une augmentation de l'amplitude du signal corrigé pouvant aboutir à des valeurs trop "élevées" des coefficients de Strickler, notamment dans la zone sensible de la confluence entre la Dordogne et la Garonne. Une vigilance particulière semble ainsi devoir être portée lors de l'intégration du paramètre $Ks3$ dans le vecteur de contrôle.

#### 5.3.2.2 Matrice de covariance d'erreur d'ébauche de l'ensemble

Le vecteur de contrôle d'ébauche initial $\boldsymbol{\gamma}_0$ est construit de telle sorte que $\mathbf{P}_0^{\boldsymbol{\gamma}}$ soit diagonale. Il est donc supposé qu'il n'y a pas de corrélation entre les erreurs d'ébauche des différentes zones de frottement. L'Exp. 2-a permet de suivre au cours d'un cycle de marée (du cycle 0 au cycle 18), c'est-à-dire avant que l'ensemble ne soit sous-dispersé, l'évolution de la corrélation entre les variables $Ks1$, $Ks2$ et $Ks3$.
La figure 5.9 représente l'évolution de la matrice de covariance d'erreur d'ébauche $\mathbf{P}_k^{\boldsymbol{\gamma}}$ pour l'Exp 2-a pour chaque variable de contrôle avec les autres variables (en abscisse) au cours des cycles d'assimilation $\mathcal{F}_k$ pour $k \in \{1, \cdots, nb_{cycles}\}$ (en ordonnée) sous la forme d'un colorbar. Les variables corrélées positivement (resp. négativement) sont représentées en nuances de rouge (resp. bleu).
Cette figure met en lumière le lien entre l'évolution des corrélations entre les coefficients de Strickler selon les phases (flot et jusant) de propagation de la marée. Elle montre que la corrélation entre $Ks1$ et $Ks2$ est prédominante pendant la phase de flot, c'est-à-dire lorsque la marée remonte dans l'estuaire, tandis qu'une corrélation entre $Ks2$ et $Ks3$ s'instaure au jusant, mettant en évidence une tendance à évoluer dans le même sens lorsque l'estuaire se vide.
Il est important de noter que la mise en place d'une redispersion fondée sur $\mathbf{P}_0^{\boldsymbol{\gamma}}$ fait perdre cette dynamique liée aux processus physiques entre les coefficients de Strickler, dans la mesure où, $\forall k \in \{1, .., nb_{cycles}\}$, $\mathbf{P}_k^{\boldsymbol{\gamma}} = \mathbf{P}_0^{\boldsymbol{\gamma}}$ .

### 5.3.3 Évolution des ensembles dans les espaces de contrôle et d'observation pour l'estimation conjointe de paramètres et de forçages dans l'espace réduit de Karhunen-Loève (Exp. 3)

L'objectif de l'Exp. 3 est d'étudier l'évolution du vecteur de contrôle et des gains apportés par assimilation de données à l'issue des phases d'analyse des différents cycles d'assimilation pour ce qui concerne les coefficients de Strickler $Ks1$, $Ks2$ et $Ks3$ et en corrigeant simultanément les conditions maritimes (resp. fluviales) à l'aide de 7 (resp. 4) coefficients modaux.

#### 5.3.3.1 Mise en évidence de l'équifinalité

Les figures 5.10, 5.11, 5.12, 5.13 et 5.14 montrent l'évolution de chacune des variables de contrôle au cours des cycles d'assimilation de données pour l'Exp. 3-b, en utilisant le même formalisme que pour l'Exp. 2. L'évolution temporelle des variables de contrôle pour les Exp. 3-a, Exp. 3-e et Exp. 3-f (non montrés ici) pour lesquelles $N_e$ est égal à 100, 50 et 125 membres montrent une absence d'influence de $N_e$ sur les constats d'équifinalité mis en évidence dans cette section.

Les figures 5.10-a, -b et -c montrent l'évolution de $Ks1$, $Ks2$ et $Ks3$ au cours du temps. Au cours des cycles d'assimilation, $Ks1$ et $Ks2$, représentés sur les figures 5.10-a et -b, suivent une pente moyenne égale à celle de la cible malgré la présence d'oscillations d'amplitude maximale

égale à 10 pour $Ks1$ et pouvant atteindre 35 lors du passage d'un palier à l'autre pour $Ks2$ sans compromettre la stabilité du système. Comme pour l'Exp. 2, ces oscillations haute fréquence sont décorrélées de la marée et la technique de rééchantillonnage utilisée peut en être responsable. Malgré ces oscillations, les valeurs prises par ces variables de contrôle sont physiques et acceptables. La tendance générale de leur évolution est satisfaisante.
La variable de contrôle $Ks3$ est représentée sur la figure 5.10-c. Sa valeur "vraie" $Ks3^{true}$ est constante dans le temps et vaut 50. $Ks3$ présente des oscillations dont l'amplitude maximale atteint 15 autour de valeurs comprises entre 40 et 60. A la fin de l'expérience, $Ks3$ montre des oscillations plus faibles autour de $Ks3^{true}$, ce qui permet de valider le bon comportement de l'EnKF-$\gamma$-KLBC-MEG.

Les figures 5.11 et 5.12 montrent l'évolution des 7 coefficients modaux de la condition à la limite maritime au cours du temps. Les coefficients modaux obtenus à l'issue des phases d'analyse successives sont variables dans le temps et ne convergent pas vers leurs valeurs cibles respectives, qui sont, elles, constantes dans le temps (cf. tableau 5.6).

Le même constat peut être fait pour les 4 coefficients modaux des chroniques de débits appliquées en entrée de la Garonne et de la Dordogne, respectivement représentés sur les figures 5.13 et 5.14.

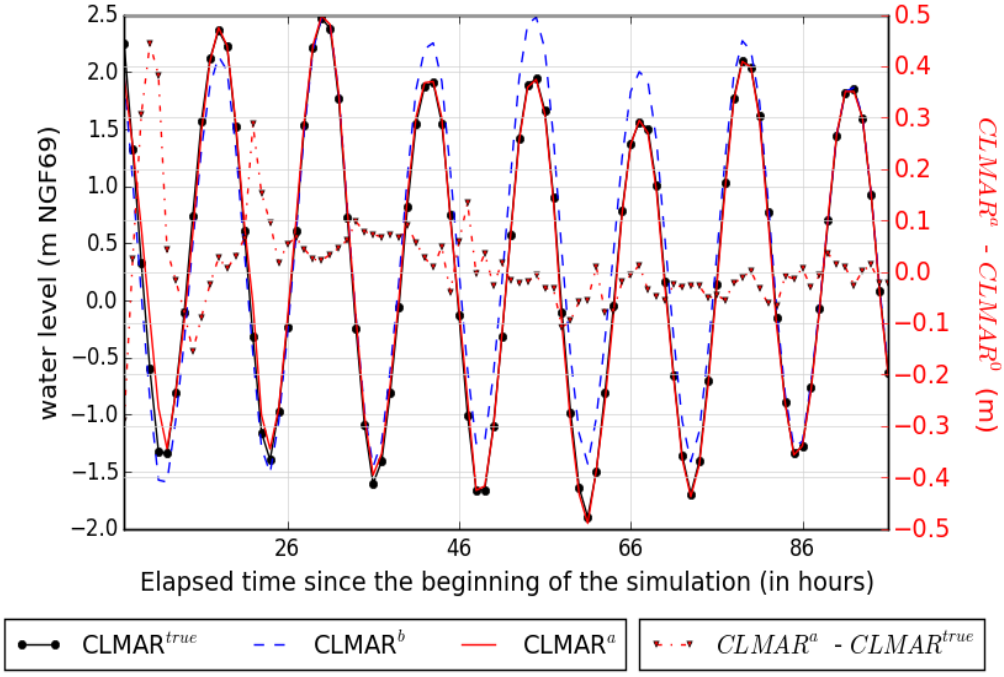
Il est, par conséquent, intéressant d'analyser si ces jeux de coefficients modaux différents du jeu "vrai" permettent de reconstituer $CLMAR$ à la frontière maritime et les débits $QGAR$ et $QDOR$ des rivières Garonne et Dordogne, afin de constater le phénomène d'équifinalité au travers de la correction des forçages.

#### 5.3.3.2 Reconstruction des conditions aux limites maritime et fluviales

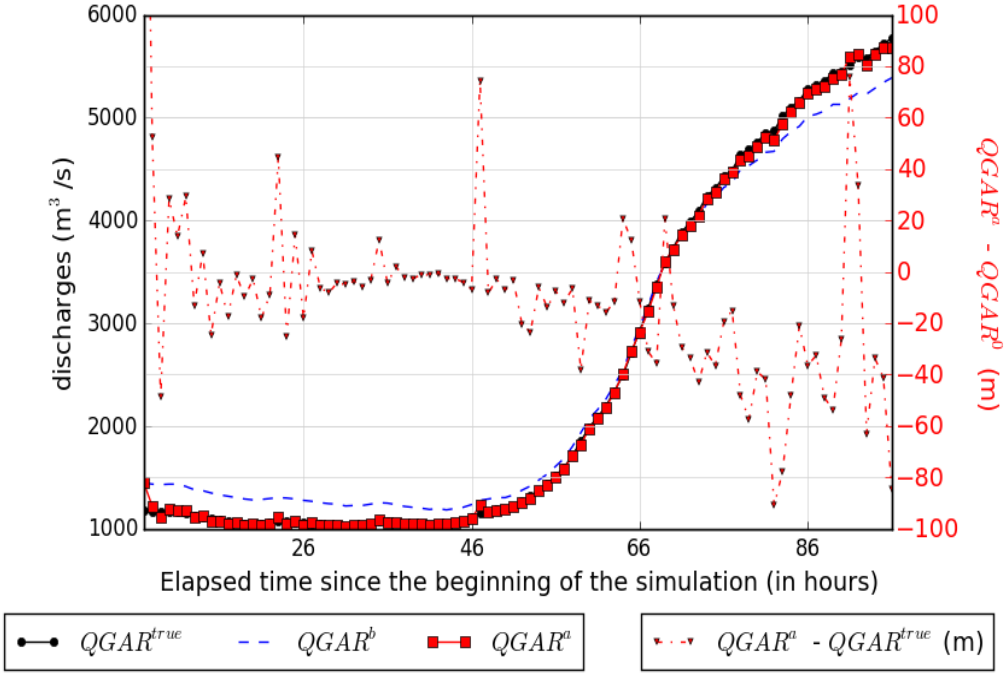
Les conditions aux limites maritime $CLMAR^a$ et fluviale $QGAR^a$ sur la Garonne toutes deux corrigées par l'EnKF-$\gamma$-KLBC-MEG sont représentées par une ligne rouge continue sur la figure 5.5. Les signaux d'ébauche et d'observation sont respectivement représentés par des traits continus bleu et noir. L'erreur de la condition aux limites analysée $CLMAR^a$ (resp. $QGAR^a$) par rapport au forçage vrai $CLMAR^{true}$ (resp. $QGAR^{true}$) est représentée en pointillés rouges à l'aide de l'axe des ordonnées de droite.

L'erreur instantanée entre le signal maritime corrigé et le forçage vrai est inférieure à 10 cm à la fin de l'expérience. De la même manière, le débit de chaque rivière est reconstitué à 2 % près à l'issue de l'Exp. 3-b.
Ces résultats montrent qu'un jeu de coefficients modaux associés à un forçage donné différent de ceux du forçage vrai permet une reconstruction satisfaisante des conditions aux limites du modèle, dans un contexte où les coefficients de Strickler de la confluence à l'embouchure sont eux-aussi corrigés. L'Exp. 3 met donc en évidence, comme l'Exp. 2, une problématique d'équifinalité entre les coefficients modaux et les forçages aux frontières liquides qu'il s'agira d'analyser au regard des performances de l'EnKF-$\gamma$-KLBC-MEG en matière de prévision des niveaux d'eau

dans la section 5.4.




(a) CLMAR




(b) QGAR

FIGURE 5.5 : Conditions aux limites du modèle corrigées à l'aide de l'EnKF-$\gamma$-KLBC-MEG pour l'Exp. $3-b$. Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite.

#### 5.3.3.3 Matrice de covariance d'erreur d'ébauche de l'ensemble de contrôle pour l'estimation simultanée des paramètres de frottement et des forçages dans l'espace réduit de Karhunen-Loève

De la même manière que pour l'Exp. 2-b, la matrice de covariance d'erreur d'ébauche de l'ensemble, représentée sur la figure 5.16, reste invariable en temps du fait du choix de la technique de rééchantillonnage mise en œuvre dans l'EnKF-$\gamma$-KLBC-MEG.
Du fait de la taille moyenne de l'ensemble ($N_e$=100), des corrélations apparaissent lors de sa génération. Des corrélations positives (resp. négatives) peuvent ainsi être notées entre les variables $Ks2$ et $alpha7$, entre $GARcoef1$ et $DORcoef1$, entre $GARcoef3$ et $DORcoef3$, entre $alpha3$ et $DORcoef2$, entre $alpha2$ et $DORcoef4$ (resp. entre $Ks1$ et $alpha6$, entre $alpha4$ et $GARcoef1$ et entre $alpha5$ et $DORcoef4$). Ces corrélations intempestives liées à la taille de l'ensemble sont importantes, car elles peuvent conduire à la correction d'une variable de contrôle à l'aide de stations de mesures ***a priori*** en-dehors de sa zone d'influence au travers de la matrice de covariance croisée entre le vecteur de contrôle et les observations simulées. Ce point est développé dans la section 5.6.

### 5.3.4 Conclusions

Les trois expériences jumelles Exp 1, Exp 2 et Exp 3 ont permis d'évaluer la capacité de l'EnKF-$\gamma$-KLBC-MEG à estimer les paramètres et forçages du modèle (incrément ou espace réduit de KL) dans différentes configurations. D'autres expériences ont été menées pour atteindre cet objectif dont les résultats ne sont pas montrés ici. Ils concernent l'estimation d'un biais constant sur les forçages sans rééchantillonnage, la correction du forçage maritime dans l'espace réduit de KL avec et sans rééchantillonnage et la correction du forçage maritime dans l'espace réduit de KL de manière conjointe avec des coefficients de Strickler constants dans le temps $Ks1$, $Ks2$ et $Ks3$ avec et sans rééchantillonnage.

L'estimation d'une perturbation constante sur les paramètres contrôlés séparément ou simultanément et/ou les forçages a été validée (Exp. 1), ainsi que l'estimation simultanée des coefficients de frottement variables en temps de manière périodique (Exp. 2-a et Exp. 2-b) ou constants par morceaux conjointement avec les forçages dans l'espace réduit de KL (Exp. 3).

Deux points d'attention ont été mis en évidence.

D'une part, une sous-dispersion de l'ensemble à l'issue des phases d'analyse apparaît lorsque le vecteur de contrôle est uniquement composé d'incréments sur le paramètre ou le forçage. Cette sous-dispersion a été résolue à l'aide d'une méthode de rééchantillonnage fondée sur la persistance de la matrice de covariance initiale d'ébauche $\mathbf{P}_0^{\gamma}$. Cependant, la méthode fait apparaître des signaux haute fréquence dans l'évolution de la moyenne de l'ensemble. Il peut être envisagé d'augmenter la durée $L$ de chaque fenêtre d'assimilation ou d'augmenter le nombre de membres $N_e$ de l'ensemble.

D'autre part, l'intégration du coefficient de frottement $Ks3$ appliqué sur la zone de la confluence entre les rivières Garonne et Dordogne dans le vecteur de contrôle simultanément avec les autres coefficients $Ks1$, $Ks2$ et les coefficients modaux des forçages induit des valeurs non-physiques de ce coefficient. Cela met en lumière la sub-optimalité de l'EnKF-$\gamma$-KLBC-MEG pour cette zone soumise à des processus physiques non-linéaires. La solution retenue ici est de contraindre l'intervalle des valeurs autorisées pour les $Ks$.
Ainsi, une attention particulière doit être portée sur deux points. Le premier point réside dans les bornes initiales de la loi uniforme, notamment si la longueur du support de celle-ci est maintenue au fur et à mesure des cycles, comme c'est le cas avec le rééchantillonnage utilisé. Le second point concerne le fait que les coefficients de Strickler sont des valeurs réelles positives. Dans l'EnKF-$\gamma$-KLBC développé ici, le choix a été fait d'utiliser une longueur de l'intervalle support de la loi uniforme faible (10 ou 20) et de contraindre les valeurs des moyennes des coefficients de Strickler $Ks^a$ à des valeurs supérieures à 10 pour 2 raisons. Il s'agit, d'une part, d'éviter l'instabilité des résultats de l'EnKF-$\gamma$-KLBC constatée pour une différence entres les bornes de la loi uniforme égale à 35. D'autre part, l'objectif est d'empêcher que le vecteur de contrôle ne contienne des valeurs peu physiques de coefficient de Strickler, c'est-à-dire inférieures à 10, voire négatives. Ce dernier point est généralement traité dans la littérature par l'utilisation de fonctions d'anamorphisme, telle que la fonction log (Bocquet and Sakov [2013]), par exemple. (Simon and Bertino [2009], Simon and Bertino [2012]) utilisent une fonction d'anamorphisme dans le cas d'un EnKF couplé à un modèle environnemental 3D dont les concentrations qui composent le vecteur de contrôle doivent rester positives à l'aide d'une fonction d'anamorphisme fondée sur la marginale supposée gaussienne du vecteur d'état. Leur méthode n'est cependant pas appliquée au cas d'une distribution multi-variée de variables de contrôle telles que c'est le cas ici, mais pourrait servir de base à des travaux futurs pour rendre plus "propre" l'EnKF-$\gamma$-KLBC.

Enfin, un phénomène d'équifinalité est mis en évidence dans plusieurs cas. L'équifinalité est systématique lorsque les forçages sont corrigés par les coefficients modaux de l'espace réduit de KL, seuls, ou conjointement avec des paramètres ($Ks$ ou $Cd$) et donc lorsque la perturbation sur le forçage est variable dans le temps. Elle apparaît également pour des $Ks$ variables en temps et corrigés simultanément (Exp. 2-a, Exp. 2-b) conjointement ou non avec les forçages dans l'espace réduit de KL (Exp. 3). Elle est moins sévère lors de l'estimation de valeurs de Strickler constantes par morceaux de l'Exp. 3 (cf. figure 5.10). Enfin, l'équifinalité n'apparaît pas lorsque la valeur "vraie" du paramètre ou de la perturbation du forçage (incrément) est constante dans le temps.

Malgré l'équifinalité, l'erreur sur la condition à la limite maritime $CLMAR^a$ pour les Exp. 2 et Exp. 3 est inférieure à 10 cm. Dans ce cadre synthétique où les coefficients modaux "vrais" associés à chaque condition aux limites sont constants, l'EnKF-$\gamma$-KLBC-MEG reconstruit donc correctement les forçages temporels. Par ailleurs, l'équifinalité se situe à un niveau inhabituel entre les coefficients modaux et les conditions aux limites physiques. Deux jeux de coefficients modaux différents permettent d'obtenir la même condition aux limites.

Au vu de ces conclusions, la section 5.4 évalue les performances de l'EnKF-$\gamma$-KLBC-MEG pour

l'amélioration de la prévision des niveaux d'eau pour les mêmes expériences synthétiques.

## 5.4 Résultats : amélioration des prévisions de niveaux d'eau par l'EnKF-$\gamma$-KLBC-MEG

### 5.4.1 Prévision des niveaux d'eau par correction de coefficients de Strickler avec perturbation vraie permanente (Exp. 1) et confirmation de l'étude de quantification d'incertitudes

Les sections suivantes présentent les scores en termes d'erreurs quadratiques moyennes (RMSE) obtenues en chaque station d'observation de l'estuaire de la Gironde sur la durée de l'Exp. 1, soit 24 h.

Les erreurs quadratiques moyennes entre l'observation horaire "vraie" "$\mathbf{y}^0$" et l'observation simulée "$\mathbf{y}^a$" (resp. "$\mathbf{y}^b$") obtenue avec (resp. sans) assimilation de données (en rouge "rmse avec DA") (resp. en bleu "rmse sans DA") et le gain (en jaune "gain (en m)") pour 26 stations d'observation situées le long de l'estuaire de la Gironde, ainsi que la moyenne sur les stations d'observation de ces quantités (colonne "Global"), sont représentées sous forme d'histogrammes sur la figure 5.17 pour chaque sous-expérience de l'Exp. 1.

Par ailleurs, les figures 5.18 à 5.21 représentent l'évolution des niveaux d'eau à (a) Le Verdon, (b) Pauillac, (c) Bec d'Ambès, (d) Bordeaux (marégraphe du Port Autonome de Bordeaux), (e) La Réole et (f) Pessac pour chacune des Exp. 1-a à 1-d. Sur ces figures, "$\mathbf{y}^b$" représente la ligne d'eau sans assimilation de données (run de contrôle) au cours du temps. Les observations synthétiques $\mathbf{y}^0$ sont représentées par des étoiles noires vers lesquelles le niveau d'eau $\mathbf{y}^a$, correspondant à la projection de la moyenne de l'ensemble de contrôle $\overline{\boldsymbol{\gamma}^a}$ dans l'espace des observations, en trait continu rouge, doit converger au cours de l'expérience.

#### 5.4.1.1 Prévision des niveaux d'eau sur $A1$

Comme le montre la figure 5.17-a, la perturbation de $Ks1$ seul et sa correction à l'aide d'observations synthétiques au Verdon pendant l'Exp. 1-a a une incidence en termes de niveaux d'eau sur l'ensemble de l'estuaire jusqu'aux limites de la frontière maritime. L'impact est plus sensible en termes de niveau d'eau de Royan à Fort-Médoc, conformément aux résultats de l'étude de quantification d'incertitudes. L'assimilation des données au Verdon pour corriger $Ks1$ permet d'obtenir une RMSE inférieure à 10 cm pour 24 heures d'expérience jumelle sur toute la partie de l'estuaire en aval de la confluence. Le gain obtenu par assimilation de données présente une forme dissymétrique en forme de cloche avec une valeur maximale à Richard (40 cm de gain pour une RMSE avant assimilation de 48 cm). Le gain est croissant du Verdon à Richard et décroît ensuite de manière linéaire depuis Richard vers l'amont des deux rivières. Ce dernier point

s'explique par l'influence décroissante de $Ks1$ depuis l'embouchure de l'estuaire vers les parties en amont de la confluence entre la Garonne et la Dordogne.
La figure 5.18 montre qu'au Verdon, dont les observations sont assimilées toutes les heures, la ligne d'eau après assimilation rejoint la mesure dès le premier cycle d'assimilation. Le niveau d'eau obtenu après assimilation et les observations sont alors confondus pour le reste de l'expérience. Le même constat un peu différé dans le temps du fait de la propagation du signal de marée peut être observé pour toutes les autres stations qui sont, quant à elles, utilisées pour validation. Ainsi, du fait de la propagation du signal de marée et de l'influence du Strickler sur la vitesse de propagation de l'onde de marée, le signal converge vers la ligne d'eau mesurée à Pauillac en 1 cycle d'assimilation, 6 à Bec d'Ambès, Bordeaux et La Réole et en 8 cycles pour Pessac. Il peut être étonnant de constater que l'étude de quantification d'incertitudes a montré que l'influence de $Ks1$ ne se faisait sentir que sur la partie aval de l'estuaire, alors que l'EnKF-$\gamma$-KLBC-MEG conduit à élargir cette zone d'influence. Deux explications peuvent justifier ce constat : d'une part le fait que l'état initial ne correspond pas aux conditions aux limites de l'expérience. Cependant, au bout de 6 heures de simulation, la condition initiale est ***a priori*** "évacuée" au profit de la condition à la limite maritime. La seconde réside dans le fait que $Ks1$ influence la vitesse de propagation de l'onde de marée et que, du fait de la faible taille $N_e$ de l'ensemble, cette influence s'étend numériquement jusqu'à Bordeaux.

#### 5.4.1.2 Prévision des niveaux d'eau sur $A2$

Comme le concluait également l'étude de quantification d'incertitudes, l'influence de $Ks2$, variable de contrôle de l'Exp 1-b, est particulièrement sensible sur la zone médiane de l'estuaire de la Gironde, de Richard à Cadillac (resp. Ambès) sur la Garonne (resp. la Dordogne). Comme le montre la figure 5.17-b, c'est bien sur cette zone que l'assimilation des données à Pauillac permet d'obtenir un gain sensible compris entre 10 cm et 35 cm en termes de RMSE, la divisant d'un facteur 2 à 3 par rapport au run de contrôle. La RMSE avec assimilation est ainsi inférieure à 15 cm à l'aval de Pauillac. Le gain obtenu par assimilation des données à Pauillac est croissant depuis l'embouchure jusqu'à Bec d'Ambès où il est maximal et vaut 37 cm pour une RMSE initiale de 68 cm. Le gain obtenu décroît ensuite du Bec d'Ambès vers l'amont de chacune des deux rivières. La figure 5.19-a montre, conformément à l'étude de quantification d'incertitudes, que la station du Verdon à l'embouchure de l'estuaire n'est que peu influencée par la correction du coefficient de Strickler $Ks2$ mis en œuvre sur la zone médiane. Au Verdon, la RMSE sans assimilation de données égale seulement 9 cm et est divisée par 3 à la fin des cycles d'assimilation. Pour l'expérience 1-b, les observations synthétiques extraites à Pauillac sont assimilées toutes les heures. En ce point de sortie, la ligne d'eau après assimilation se confond avec la mesure après 4 cycles d'assimilation de 1 h. La RMSE sur les 24 h d'expérience est divisée par 3 et décroît de 45 cm à 15 cm. Toutes les autres stations sont utilisées pour validation. Du fait de la propagation du signal de marée et de l'influence du Strickler sur la vitesse de propagation de l'onde de marée, le signal

met respectivement 5 et 6 heures pour rejoindre les lignes d'eau "observées" au Bec d'Ambès et à Bordeaux.

#### 5.4.1.3 Prévision des niveaux d'eau sur $A3$

L'Exp. 1-c consiste en la correction du paramètre $Ks3$ par assimilation des observations synthétiques localisées au niveau du marégraphe du Port Autonome de Bordeaux. La figure 5.17-c montre que la zone d'influence de la perturbation de $Ks3$ est située de Lamena à Langon (resp. Libourne) sur la Garonne (resp. sur la Dordogne). La RMSE obtenue avec assimilation de données sur cette zone est inférieure à 1 cm, alors qu'elle est supérieure à 5 cm sans assimilation de données de Bassens à Cadillac sur la Garonne. Le gain obtenu par assimilation de données est croissant depuis Vitrezay jusque Bègles sur la Garonne où il est maximal. Il décroît ensuite jusqu'à Langon. Sur la Dordogne, la correction de $Ks3$ par assimilation des données à Bordeaux se fait sentir jusqu'à Ambès.
La représentation des lignes d'eau sur 6 stations de l'estuaire (Bec d'Ambès, Bassens, Bordeaux, Bègles, Cadillac et Langon) sur la figure 5.20 montre que la ligne d'eau après assimilation de données est confondue avec l'observation au bout du 3$^{\text{ème}}$ cycle d'assimilation, soit 3 heures après le début de l'expérience d'assimilation. L'influence de $Ks3$ étant plutôt sensible sur le jusant, alors que celles de $Ks1$ et $Ks2$ se manifestent à la fois sur la célérité et l'amplitude de l'onde au flot et au jusant, c'est sur cette première partie du cycle de marée que la correction du coefficient de Strickler, notamment à Bègles où la RMSE sur l'événement décroît de 14 cm à 1 cm, permet d'améliorer les résultats en termes de niveaux d'eau.

#### 5.4.1.4 Prévision des niveaux d'eau sur $A4$

L'expérience 1-d consiste en la correction du paramètre $Ks4$ par assimilation des observations horaires à La Réole. Conformément aux résultats de l'étude de quantification d'incertitudes, l'influence du coefficient de Strickler $Ks4$ est localisée sur les parties amont de la Garonne (resp. la Dordogne) de Cadillac (resp. Libourne) à La Réole (resp. Pessac). Les gains obtenus avec assimilation de données sont substantiels. L'assimilation de données permet ainsi de diviser par 3 la RMSE sur ces zones, comme le montre la figure 5.17-d.
La figure 5.21 montre qu'en assimilant les "mesures" à La Réole, les lignes d'eau calculées avec assimilation de données rejoignent l'observation au bout de 5 et 2 (resp. 7) cycles d'assimilation de données à Cadillac et La Réole (resp. Pessac) sur la Garonne (resp. la Dordogne). Il convient toutefois de noter que l'amélioration des niveaux d'eau sur la Dordogne du fait de l'assimilation des mesures synthétiques à la Réole résulte d'un artefact numérique, à savoir l'application d'un même coefficient de frottement $Ks4$ "vrai" sur la Garonne et la Dordogne pour le run vrai $True1$.

#### 5.4.1.5 Conclusions

L'Exp. 1, très simple, permet ainsi de confirmer les résultats de l'étude de quantification d'incertitudes pour la correction univariée des coefficients de Strickler de chaque zone en assimilant une observation synthétique en un seul point de sortie de l'estuaire situé dans la zone d'influence de la variable de contrôle.

### 5.4.2 Niveaux d'eau et scores le long de l'estuaire par correction des coefficients de Strickler avec valeurs vraies périodiques (Exp. 2)

Pour rappel, l'Exp. 2-b consiste à corriger les coefficients de Strickler sur les zones $A_1$, $A_2$ et $A_3$ à l'aide des observations synthétiques $True2c$ obtenues à l'aide de coefficients de frottement périodiques. Une redispersion de l'ensemble par rééchantillonnage après chaque phase d'analyse est mise en œuvre.

Les niveaux d'eau pour les 3 stations dont les "observations" sont assimilées et 3 stations de validation le long de l'estuaire pour l'Exp. 2-b sont respectivement représentés sur les figures 5.22 et 5.23.
La valeur absolue de l'erreur instantanée sur les niveaux d'eau, représentée par une ligne pointillée rouge sur l'axe des ordonnées de droite, est généralement inférieure à 10 cm pour les stations d'observation et de validation au cours de l'expérience, sauf à certains pics de basse mer à Bordeaux et Pauillac et pendant le flot au Marquis. Ces erreurs instantanées supérieures à 10 cm semblent être liées à un déphasage entre les signaux observés et corrigés.

Les RMSE sur la durée de l'Exp. 2 (96 h) sur chaque station de l'estuaire de la Gironde sont représentées sur la figure 5.24. Les RMSE sont inférieures à 10 cm pour toutes les stations situées de l'embouchure de l'estuaire à Fort-Médoc. Les zones sous influence de $Ks3$ et $Ks4$ présentent alors des RMSE plus grandes liées principalement à la difficulté de bien estimer les pics de basse mer et à une surestimation légère (environ 10 cm) des pics de pleine mer. L'EnKF-$\gamma$-KLBC-MEG permet de diviser globalement par 3 l'erreur quadratique. Il améliore également l'estimation des basses mers et des pleines mers, malgré l'introduction d'un léger déphasage dans le signal, amplifié par l'utilisation d'une fréquence horaire pour le calcul des scores.

Malgré la mauvaise analyse de $Ks3$, l'erreur sur les hauteurs d'eau est ainsi améliorée sur l'ensemble de l'estuaire.

### 5.4.3 Niveaux d'eau et scores le long de l'estuaire par correction des paramètres de frottement (valeurs vraies constantes par morceaux) et des forçages temporels dans l'espace réduit de Karhunen-Loève (Exp. 3-b)

Cette section présente et analyse les gains obtenus grâce à l'EnKF-$\gamma$-KLBC-MEG en termes de niveaux d'eau globalement et le long de l'estuaire de la Gironde pour l'Exp. 3-b. La RMSE maximale atteinte après assimilation de données sur la durée totale de l'expérience (96 h, spin-up inclus), représentée sur la figure 5.25 est inférieure à 11 cm pour toutes les stations, sauf à Bègles et Cadillac où elle atteint respectivement 12 et 13 cm.

Les lignes d'eau pour 6 stations de l'estuaire sont représentées sur la figure 5.26. L'erreur instantanée représentée en pointillés rouges est inférieure à 10 cm au bout d'un cycle de marée. Un éventuel déphasage du signal ne peut être incriminé que sur certaines stations, comme Laména, car, sur la plupart d'entre elles, l'EnKF-$\gamma$-KLBC-MEG rephase le signal sur les pleines mers.Ainsi, malgré une mauvaise estimation des coefficients modaux associés aux forçages temporels et de $Ks3$, en particulier, les résultats en termes de niveaux d'eau, obtenus en corrigeant tous les forçages et paramètres du modèle de l'estuaire de la Gironde à l'aide de ses principales stations, sont très bons malgré l'équifinalité déjà évoquée dans la section 5.3.

### 5.4.4 Conclusions

Malgré l'utilisation d'une technique de rééchantillonnage conduisant à des perturbations haute fréquence sur l'estimation des paramètres et des coefficients modaux associés aux forçages temporels et la mise en évidence d'un phénomène d'équifinalité, l'EnKF-$\gamma$-KLBC-MEG reconstruit correctement les conditions aux limites et estime avec une grande précision les niveaux d'eau le long de l'estuaire.

Il apparaît ainsi que, lorsque les observations synthétiques sont obtenues avec des coefficients modaux constants et des coefficients de Strickler constants, périodiques ou constants par morceaux au cours du temps, l'EnKF-$\gamma$-KLBC-MEG met en lumière d'autres jeux de paramètres et de forçages qui permettent d'obtenir les mêmes niveaux d'eau en sortie. De toute évidence, en l'absence de contrainte sur les vitesses, la relation qui relie les niveaux d'eau (y.c la condition initiale) et les paramètres/forçages du modèle numérique est surjective.

Cette caractéristique des écoulements est susceptible d'induire des problèmes concernant le bassin d'attraction dans lequel le vecteur de contrôle évolue. Elle semble être la conséquence du caractère instationnaire des perturbations appliquées au forçages et aux paramètres du modèle, ainsi qu'aux non-linéarités des processus physiques modélisés, en particulier dans la zone de la confluence des rivières Dordogne et Garonne.

C'est pourquoi il est intéressant d'étudier les matrices de covariance croisée des variables de contrôle avec les observations simulées, avec et sans technique de rééchantillonnage, du fait de leur rôle dans le calcul de la matrice de gain de Kalman qui contrôle l'innovation appliquée au

vecteur de contrôle au fil des cycles d'assimilation. La section 5.6 présente cette analyse.

Par ailleurs, le choix des paramètres (localisation des observations, nombre de membres, raffinement du maillage etc.) conditionne les performances de l'EnKF-$\gamma$-KLBC-MEG. La section 5.5 détermine sa configuration optimale.

## 5.5 Optimisation de l'EnKF-$\gamma$-KLBC-MEG

### 5.5.1 Localisation des stations de mesure

L'Exp. 3 permet d'étudier l'influence de la localisation des stations de mesure sur les performances de l'EnKF-$\gamma$-KLBC-MEG le long de l'estuaire de la Gironde. Pour rappel, les stations de mesure utilisées pour chacune de ces expériences sont indiquées dans le tableau 5.4.

L'Exp. 3-b utilise toutes les stations d'observation actuellement disponibles le long de l'estuaire de la Gironde. Les autres sous-expériences tentent d'optimiser les performances en termes de RMSE en faisant varier la localisation des stations dont les données sont assimilées.
L'Exp. 3-a utilise les stations de mesures de Royan, Richard, Pauillac, Bec d'Ambès, Bordeaux (limnigraphe), La Réole et Pessac. Le vecteur de contrôle est alors corrigé avec des données réparties de manière homogène sur l'ensemble de l'estuaire. Royan se situe à l'embouchure de l'estuaire et Richard à l'aval. Pauillac se situe dans la partie centrale de l'estuaire et Bec d'Ambès au niveau de la confluence entre la Garonne et la Dordogne. Enfin, Bordeaux se situe sur la Garonne en amont de la confluence. L'assimilation des données à Pessac et La Réole permet de corriger, si nécessaire, les débits injectés dans le modèle sur la Dordogne et la Garonne.
Les Exp. 3-c et 3-d réduisent encore le nombre de marégraphes dont les données sont assimilées. L'Exp. 3-c utilise uniquement les données de Laména, La Réole et Pessac, tandis que l'Exp. 3-d utilise les données au Verdon, à Pauillac, La Réole et Pessac. Les scores obtenus en termes de RMSE sur 96 h pour chaque configuration de l'Exp. 3 avec et sans assimilation de données (AD) sont récapitulés dans le tableau 5.7.

Pour une expérience dont le cycle d'assimilation est composé de fenêtres d'assimilation de durée $L = 1\ h$ consécutives et disjointes ($S = L$), l'Exp. 3-a présente les meilleurs scores en termes de RMSE sur toutes les stations, y compris celles de validation.
Dans cette configuration, le réseau de mesures idéal comprend les 7 marégraphes situés à Royan, Richard, Pauillac, Bec d'Ambès, Bordeaux (limnigraphe), auxquels il faut ajouter La Réole et Pessac, si les débits des rivières doivent être corrigés.

TABLE 5.7 : RMSE sans et avec EnKF-$\gamma$-KLBC-MEG pour l'Exp. 3

| Station | RMSE sans AD | RMSE avec AD | | | |
|---|---|---|---|---|---|
| | | Exp. 3-a | Exp. 3-b | Exp. 3-c | Exp. 3-d |
| Royan | 0.31 | 0.06 | 0.06 | 0.21 | 0.26 |
| Port-Bloc | 0.32 | 0.05 | 0.05 | 0.13 | 0.23 |
| Le Verdon | 0.32 | 0.04 | 0.05 | 0.12 | 0.23 |
| Meshers s/ Gir. | 0.33 | 0.05 | 0.06 | 0.18 | 0.25 |
| Richard | 0.34 | 0.06 | 0.08 | 0.27 | 0.28 |
| Mortagne s/ Gir. | 0.35 | 0.06 | 0.07 | 0.21 | 0.22 |
| Port Maubert | 0.37 | 0.06 | 0.07 | 0.18 | 0.19 |
| Laména | 0.41 | 0.07 | 0.08 | 0.17 | 0.18 |
| Vitrezay | 0.42 | 0.07 | 0.08 | 0.17 | 0.18 |
| CE Blayais | 0.45 | 0.08 | 0.09 | 0.19 | 0.2 |
| Pauillac | 0.47 | 0.08 | 0.09 | 0.19 | 0.21 |
| Fort Médoc | 0.51 | 0.08 | 0.09 | 0.21 | 0.23 |
| Bec d'Ambès | 0.58 | 0.09 | 0.1 | 0.25 | 0.29 |
| Le Marquis | 0.6 | 0.09 | 0.1 | 0.26 | 0.29 |
| St Louis de M | 0.6 | 0.09 | 0.1 | 0.24 | 0.33 |
| Bassens | 0.63 | 0.09 | 0.1 | 0.24 | 0.38 |
| Bordeaux Pabx | 0.65 | 0.1 | 0.11 | 0.23 | 0.44 |
| Bordeaux 2 | 0.63 | 0.1 | 0.11 | 0.23 | 0.46 |
| Bègles | 0.64 | 0.1 | 0.11 | 0.23 | 0.51 |
| Cadillac | 0.34 | 0.12 | 0.12 | 0.19 | 0.2 |
| Langon | 0.36 | 0.13 | 0.13 | 0.19 | 0.19 |
| La Réole | 0.42 | 0.09 | 0.09 | 0.11 | 0.1 |
| Ambès | 0.51 | 0.09 | 0.1 | 0.22 | 0.29 |
| St Vincent de P. | 0.32 | 0.09 | 0.09 | 0.17 | 0.19 |
| Libourne | 0.33 | 0.1 | 0.1 | 0.14 | 0.13 |
| Pessac | 0.72 | 0.1 | 0.1 | 0.08 | 0.08 |
| Global | 0.46 | 0.08 | 0.09 | 0.19 | 0.25 |

### 5.5.2 Nombre de membres de l'ensemble

#### 5.5.2.1 Description des expériences

La qualité d'un filtre de Kalman d'Ensemble est fortement dépendante du nombre de membres utilisé pour créer le vecteur de contrôle. L'influence du nombre de membres $N_e$ sur les performances de l'EnKF-$\gamma$-KLBC-MEG a été étudiée dans le cadre de ce travail, d'une part, en utilisant les résultats des expériences 3-a, -e et -f, pour lesquelles $N_e$ vaut 100, 50 et 125, et d'autre part, à l'aide des sous-expériences 2-c et 2-d, décrites dans la section 5.2, et pour lesquelles les valeurs respectives de $N_e$ sont 100 et 200 membres. Pour rappel, les coefficients de Strickler "vrais" des

sous-expériences 3-a, -e et -f sont les valeurs constantes par morceaux utilisées pour l'expérience de référence $True2b$ et ceux des sous-expériences 2-c et 2-d ceux de l'expérience de référence $True2c$ pour laquelle l'évolution temporelle des coefficients de Strickler est périodique.

#### 5.5.2.2 Performances de l'EnKF-$\gamma$-KLBC-MEG pour la simulation des niveaux d'eau

Les RMSE obtenues en chaque station de l'estuaire pour chacune des expériences décrites ci-dessus sont calculées sur 96 h et récapitulées dans le tableau 5.8. Pour l' Exp. 2, les résultats sont dégradés en choisissant des valeurs différentes de $N_e = 100$ à l'amont de Laména jusqu'au milieu des tronçons fluviaux. Les RMSE sont sensiblement identiques, mais un peu moins bonnes, en choisissant une valeur $N_e = 200$ pour l'Exp. 3. De toute évidence, le faible nombre de membres $N_e$ n'est pas la solution pour augmenter la performance de l'EnKF-$\gamma$-KLBC-MEG.

Ces résultats conduisent à conclure qu'un nombre de membres égal à 100, indépendamment du fait que les coûts de calcul sont sensiblement inférieurs, permet d'optimiser la performance de l'EnKF-$\gamma$-KLBC-MEG pour le critère RMSE dans un cadre synthétique.

Le diagramme de rang, ou diagramme de Talagrand, des observations au sein des niveaux d'eau simulés permet d'établir un diagnostic sur la capacité du modèle numérique à fournir une distribution de probabilité prévue (ou un ensemble de quantiles de prévision) dont tous les intervalles de prévision sont fiables (Brochero et al. [2011]).
Les diagrammes de rang des stations du Verdon, de Pauillac et de Bordeaux PABX sont représentés sur la figure 5.27 pour les Exp 2-c ($Ne = 100$) (à gauche) et Exp. 2-d ($Ne = 200$) (à droite). La distribution uniforme parfaite est représentée par une ligne continue bleue. Elle montre pour chaque station un diagramme de rang uniforme, même si le deuxième rang reflète un léger biais. En conséquence, la distribution des niveaux d'eau simulés est fiable. Cela signifie que les membres de l'ensemble projeté dans l'espace des observations, c'est-à-dire l'ensemble des observations simulées, et les observations suivent des variables aléatoires similaires.
Par ailleurs, les diagrammes de rang obtenus pour $Ne = 100$ (Exp. 2-c) et $Ne = 200$ (Exp. 2-d) sont proches. Quelques rangs sont plus proches du quantile parfait pour $Ne = 100$. C'est ainsi le cas pour le $6^{ème}$ rang de Pauillac, ainsi que les 4 derniers rangs de Bordeaux. Au contraire, d'autres rangs, tels les derniers rangs du Verdon, sont plus proches de la distribution uniforme pour $N_e$=200. De fait, aucune différence flagrante ne peut être mise en évidence.

#### 5.5.2.3 Matrices de covariance croisée $\mathbf{P}^{\gamma,y}$

La figure 5.28 montrent les fonctions temporelles de covariance croisée entre les variables de contrôle $Ks3$, $Cd$ et $\alpha_{GAR,1}$ ($GARcoeff1$ sur la figure) et les observations simulées pour les expériences 2-c et 2-d. Les fonctions de covariance croisée avec $Ks3$ sont peu sensibles au choix de $N_e$, comme le montrent les figures 5.28-a et -b concernant la corrélation de $Ks3$ avec les stations de Bordeaux et de Bassens. Les fonctions de covariance correspondant aux coefficients de

Strickler et aux premiers modes de $CLMAR$ (non montrées ici) sont similaires à celle représentée pour $Ks3$ sur les figures 5.28-a et 5.28-b et peu sensibles à $N_e$.

En revanche, les composantes liées au coefficient d'influence du vent $Cd$ et aux coefficients modaux des conditions aux limites fluviales le sont, comme l'illustrent les figures 5.28-c à 5.28-d. Des corrélations entre variables de contrôle et des covariances croisées entre variables de contrôle et observations simulées artificielles, comme c'est le cas entre Pessac et le coefficient modal $\alpha_{GAR,1}$ par exemple, sont ainsi amplifiées lorsque la valeur de $N_e$ est plus faible.

Les deux solutions immédiates sont soit de supprimer les variables concernées du vecteur de contrôle, soit d'augmenter le nombre de membres. Cette dernière solution ne se justifie pas ici, compte-tenu des RMSE sur les niveaux d'eau obtenues dans la section 5.5.2.2. Le fait de supprimer, dans le cas d'application présent qui repose sur les conditions météorologiques de la tempête de 2003, le coefficient d'influence du vent $Cd$ n'est pas traumatisant compte-tenu des résultats d'analyse de sensibilité globale le concernant. Cependant, il est discutable en cas de tempêtes d'intensités supérieures. Par ailleurs, Il faut noter que, malgré des corrélations artificielles avec des stations qui ne devraient pas subir l'influence des débits considérés, la reconstruction des débits amont par l'EnKF-$\gamma$-KLBC-MEG est peu sensible au nombre de membres choisis. En réalité, elle l'est beaucoup plus à la qualité des observations en amont du modèle, comme le suggèrent les conclusions de la section 5.5.3, consacrée à l'influence du maillage.

La dernière solution est donc de trouver un compromis entre la qualité des niveaux d'eau simulés dans l'estuaire à l'aide de l'EnKF-$\gamma$-KLBC-MEG et le nombre de membres, tout en gardant à l'esprit l'existence de ces "bruits" qui peuvent altérer les résultats de l'EnKF-$\gamma$-KLBC-MEG.

TABLE 5.8 : Influence du nombre de membres $N_e$ sur la RMSE dans un cadre synthétique. $N_e = 100$ (resp. $N_e = 200$) pour les Exp. 2-c et Exp. 3-a (resp. 2-d). $N_e = 50$ (resp. $N_e = 125$) pour l'Exp. 3-e (resp. 3-f). Les colonnes "(2-c) avec DA" (resp."(2-c) sans DA") et "(3-a) avec DA" indiquent les RMSE pour les Exp. 2-c et 3-a avec (resp. sans) l'EnKF-$\gamma$-KLBC-MEG. Les colonnes "(2-c)-(2d)" (resp. "(3-a)-(3-e)" ou "(3-a)-(3-f)") indiquent la différence entre la RMSE de l'Exp. 2-c (resp. Exp. 3-a) et celle de l'Exp. 2-d (resp. Exp. 3-e ou Exp. 3-f). La différence est négative (resp. positive) si la RMSE de Exp. 2-c ou Exp. 3-a est plus petite (resp. grande).

| Station | RMSE - Exp. 2 | | | RMSE - Exp. 3 | | |
|---|---|---|---|---|---|---|
| | (2-c) sans AD | (2-c) avec AD | (2-c)-(2-d) | (3-a) avec AD | (3-a)-(3-e) | (3-a)-(3-f) |
| Royan | 0.34 | 0.1 | 0 | 0.06 | -0.02 | -0.01 |
| Port-Bloc | 0.36 | 0.08 | 0 | 0.05 | 0 | 0.01 |
| Le Verdon | 0.35 | 0.07 | 0 | 0.04 | 0 | 0.01 |
| Meshers s/ Gir. | 0.39 | 0.08 | -0.01 | 0.05 | -0.01 | -0.01 |
| Richard | 0.46 | 0.1 | -0.02 | 0.06 | -0.04 | -0.02 |
| Mortagne s/ Gir. | 0.49 | 0.1 | -0.01 | 0.06 | -0.04 | -0.01 |
| Port Maubert | 0.54 | 0.1 | -0.01 | 0.06 | -0.05 | -0.02 |
| Laména | 0.62 | 0.13 | -0.01 | 0.07 | -0.06 | -0.04 |
| Vitrezay | 0.62 | 0.13 | -0.01 | 0.07 | -0.07 | -0.04 |
| CE Blayais | 0.68 | 0.14 | -0.02 | 0.08 | -0.08 | -0.05 |
| Pauillac | 0.69 | 0.14 | -0.01 | 0.08 | -0.07 | -0.06 |
| Fort Médoc | 0.75 | 0.16 | -0.01 | 0.08 | -0.07 | -0.06 |
| Bec d'Ambès | 0.82 | 0.18 | -0.01 | 0.09 | -0.08 | -0.08 |
| Le Marquis | 0.84 | 0.18 | -0.02 | 0.09 | -0.09 | -0.09 |
| St Louis de M | 0.85 | 0.2 | -0.01 | 0.09 | -0.09 | -0.08 |
| Bassens | 0.89 | 0.19 | -0.02 | 0.09 | -0.09 | -0.08 |
| Bordeaux Pabx | 0.92 | 0.2 | -0.02 | 0.1 | -0.09 | -0.08 |
| Bordeaux 2 | 0.9 | 0.2 | -0.01 | 0.1 | -0.09 | -0.08 |
| Bègles | 0.91 | 0.2 | -0.02 | 0.1 | -0.09 | -0.1 |
| Cadillac | 0.67 | 0.22 | -0.01 | 0.12 | -0.08 | -0.06 |
| Langon | 0.71 | 0.21 | -0.01 | 0.13 | -0.09 | -0.07 |
| La Réole | 0.79 | 0.14 | -0.01 | 0.09 | -0.04 | -0.03 |
| Ambès | 0.69 | 0.19 | -0.01 | 0.09 | -0.09 | -0.07 |
| St Vincent de P. | 0.4 | 0.2 | -0.02 | 0.09 | -0.08 | -0.04 |
| Libourne | 0.31 | 0.2 | -0.01 | 0.1 | -0.08 | -0.04 |
| Pessac | 0.98 | 0.14 | -0.01 | 0.1 | -0.02 | -0.03 |
| Global | 0.65 | 0.15 | -0.01 | 0.08 | -0.06 | -0.05 |

#### 5.5.2.4 Conclusion

Le choix du nombre de membres de l'ensemble, qui conditionne l'une des dimensions du vecteur de contrôle, mais aussi à la fois la performance de l'EnKF-$\gamma$-KLBC-MEG et les ressources informatiques pour sa mise en œuvre, est très important. L'analyse des scores en RMSE pour les expériences jumelles 2 et 3 et différentes valeurs de $N_e$ ne montre pas d'apport pour $N_e = 200$ membres et tend à justifier que l'utilisation de 100 membres permet d'optimiser les résultats de l'EnKF-$\gamma$-KLBC-MEG dans le cadre synthétique.

### 5.5.3 Influence du maillage

L'influence du maillage, dans le cadre d'une expérience jumelle, peut être étudiée de deux manières.

D'une part, pour l'Exp. 3-i, le maillage le plus raffiné ($mesh3$ du MEG3V7P3 du chapitre 1) a été utilisé au sein de l'EnKF-$\gamma$-KLBC-MEG en tant que modèle d'évolution $\mathcal{M}$ tout en utilisant les observations synthétiques issues de $True2b$ et obtenues avec le maillage grossier $mesh0$, toutes conditions égales par ailleurs avec l'Exp. 3-a. Les erreurs entre les observations simulées et l'observation synthétique sont représentées sur la figure 5.29. Les RMSE en chaque point de sortie de l'estuaire sur la durée de l'expérience (96 h) sont récapitulées dans le tableau 5.9. Elles révèlent une zone d'influence du raffinement du maillage sur la RMSE en amont de Richard au jusant et très marquée tout au long de l'expérience autour des stations de Langon et de Cadillac. Les autres stations sont peu impactées par le raffinement du maillage utilisé dans l'EnKF-$\gamma$-KLBC-MEG dans le cadre de cette expérience. De fait, l'utilisation d'un maillage très raffiné se justifie principalement dans le cadre de la nécessité d'amélioration des niveaux d'eau tout en amont de l'estuaire.

D'autre part, il est pertinent de vérifier que l'utilisation d'un maillage plus grossier $mesh0$ que le maillage $mesh3$ utilisé pour créer les observations synthétiques, qui se rapprochent alors plus des observations réelles, donne des résultats cohérents avec nos précédentes conclusions. Il a ainsi été montré dans le chapitre 1 du présent manuscrit consacré à la convergence en maillage du modèle de prévision des hautes eaux sur l'estuaire de la Gironde que ce modèle, choisi pour des raisons opérationnelles, présentait des zones, et notamment celles en amont des rivières Garonne et Dordogne, dont les résultats dépendent fortement du raffinement du maillage.

Les observations synthétiques de l'Exp. 3-j, dont les conditions d'expérience sont identiques à l'Exp. 3-a, sont obtenues avec le maillage le plus raffiné ($mesh3$) décrit dans la section 1.2.4. Les niveaux d'eau sont donc mieux représentés et plus proches des observations réelles à l'amont des deux rivières. Les différences entre les niveaux d'eau $\mathbf{y}$ issus de l'EnKF-$\gamma$-KLBC-MEG et les observations synthétiques obtenues avec le maillage opérationnel ($mesh0$) (resp. le maillage raffiné ($mesh3$)) au fil des cycles d'assimilation sont représentées sur la figure 5.29-a (resp. 5.29-c). Les RMSE en chaque station sont récapitulées dans le tableau 5.9. Comme prévu, les différences

sont peu sensibles à l'aval de la confluence entre la Garonne et la Dordogne et maximales sur les tronçons fluviaux, sur lesquels les niveaux d'eau sont sensibles à la propagation de l'onde qui dépend de la taille des éléments finis. En particulier, la RMSE à Langon (resp. à Libourne) est augmentée de 25 cm (resp. 41 cm). Il peut donc être noté que des gains en amont de Cadillac et de Libourne sont sensibles au raffinement du maillage utilisé pour créer les observations synthétiques de référence.

### 5.5.4 Configuration du cycle d'assimilation

La comparaison des résultats pour les expériences Exp 3-b et Exp 3-g, qui durent toutes deux 96 h, permet de considérer le bénéfice d'une fenêtre d'assimilation de données de durée $L = 6h$ décalée de 2 heures avec la fenêtre suivante, conformément à la figure 4.1 qui représente le cyclage des fenêtres utilisé pour l'EnKF-$\gamma$-KLBC. Dans le cadre de cette dernière expérience, pour chaque station, 6 observations sont assimilées par cycle et les 4 dernières sont également utilisées pour le cycle d'assimilation suivant. Aucun coefficient de pondération n'a été introduit et toutes les observations synthétiques ont le même poids. La RMSE est meilleure lorsqu'une fenêtre d'assimilation plus longue que $L = 1\ h$ est utilisée. Par ailleurs, les perturbations haute fréquence, constatées à la mise en œuvre de la redispersion de l'ensemble après chaque phase d'analyse, sont, par construction, lissées, comme le montre la figure 5.30 pour la variable de contrôle $Ks1$.

## 5.6 Caractérisation des covariances croisées entre les erreurs d'analyse du vecteur de contrôle et celles des observations simulées avec l'EnKF-$\gamma$-KLBC-MEG

La corrélation est une forme normalisée et adimensionnelle de la covariance. Cette dernière permet de quantifier l'écart conjoint de deux variables aléatoires par rapport à leurs espérances respectives. Elle caractérise donc les variations simultanées de deux variables aléatoires et le degré de dépendance linéaire entre celles-ci. Elle est positive lorsque les écarts entre les variables et leurs moyennes ont tendance à être de même signe, négative dans le cas contraire.

Comme indiqué dans la section 4.1.2, la matrice gain de Kalman est calculée à chaque étape d'analyse selon :

$$\mathbf{K}_k^{\gamma} = \mathbf{P}_{k-1}^{\gamma,y}\left[\mathbf{P}_{k-1}^{\gamma} + \mathbf{R}_k\right]^{-1}, \tag{5.1}$$

où $\mathbf{P}_k^{\gamma}$ et $\mathbf{P}_k^{\gamma,y}$, estimés stochastiquement, sont respectivement la matrice de covariance d'erreur de l'ensemble des paramètres prédit et la matrice de covariance croisée entre l'ensemble des paramètres et l'ensemble prédit des observations simulées.

TABLE 5.9 : Gains sur la RMSE pour différentes configurations synthétiques

| Station | RMSE de l'Exp. 3-b | | Gain par rapport à (3-b) | | Gain par rapport à (3-a) | |
|---|---|---|---|---|---|---|
| | sans DA | avec DA | de (3-g) | de (3-h) | de (3-i) | de (3-j) |
| Royan | 0,31 | 0,06 | 0 | -0,02 | -0,05 | -0,01 |
| Port-Bloc | 0,32 | 0,05 | 0 | -0,01 | -0,03 | 0 |
| Le Verdon | 0,32 | 0,05 | 0 | -0,02 | -0,04 | -0,02 |
| Meshers s/ Gir, | 0,33 | 0,06 | 0,01 | -0,03 | -0,08 | -0,06 |
| Richard | 0,34 | 0,08 | 0,02 | -0,04 | -0,01 | -0,02 |
| Mortagne s/ Gir, | 0,35 | 0,07 | 0,01 | -0,04 | -0,02 | -0,02 |
| Port Maubert | 0,37 | 0,07 | 0,01 | -0,04 | -0,01 | -0,02 |
| Laména | 0,41 | 0,08 | 0,02 | -0,04 | -0,01 | -0,02 |
| Vitrezay | 0,42 | 0,08 | 0,02 | -0,05 | -0,01 | -0,03 |
| CE Blayais | 0,45 | 0,09 | 0,02 | -0,05 | 0 | -0,03 |
| Pauillac | 0,47 | 0,09 | 0,02 | -0,05 | 0 | -0,04 |
| Fort Médoc | 0,51 | 0,09 | 0,02 | -0,04 | 0 | -0,04 |
| Bec d'Ambès | 0,58 | 0,1 | 0,02 | -0,05 | 0 | -0,05 |
| Le Marquis | 0,6 | 0,1 | 0,02 | -0,06 | -0,02 | -0,07 |
| St Louis de M | 0,6 | 0,1 | 0,02 | -0,06 | -0,04 | -0,08 |
| Bassens | 0,63 | 0,1 | 0,01 | -0,05 | -0,04 | -0,07 |
| Bordeaux Pabx | 0,65 | 0,11 | 0,01 | -0,04 | -0,02 | -0,05 |
| Bordeaux 2 | 0,63 | 0,11 | 0,02 | -0,04 | 0,01 | -0,05 |
| Bègles | 0,64 | 0,11 | 0,01 | -0,05 | 0,01 | -0,05 |
| Cadillac | 0,34 | 0,12 | 0,03 | -0,02 | -0,04 | -0,1 |
| Langon | 0,36 | 0,13 | 0,02 | -0,02 | -0,17 | -0,25 |
| La Réole | 0,42 | 0,09 | -0,02 | 0 | 0 | 0 |
| Ambès | 0,51 | 0,1 | 0,02 | -0,05 | -0,02 | -0,07 |
| St Vincent de P, | 0,32 | 0,09 | 0,01 | -0,05 | -0,01 | -0,08 |
| Libourne | 0,33 | 0,1 | 0,01 | -0,04 | -0,37 | -0,41 |
| Pessac | 0,72 | 0,1 | 0,04 | -0,01 | -0,01 | 0,03 |
| Global | 0,46 | 0,09 | 0,01 | -0,04 | -0,04 | -0,06 |

$\mathbf{P}_k^{\gamma,y}$ intervient dans le calcul de la matrice "gain" $\mathbf{K}$ à chaque étape d'analyse de l'EnKF-$\gamma$-KLBC-MEG. Selon l'équation :

$$\boldsymbol{\gamma}_k^a = \boldsymbol{\gamma}_k^f + \mathbf{K}_k^{\gamma} \, (\mathbf{y}_k^0 - \mathbf{y}_k^f), \tag{5.2}$$

elle détermine, à la somme de l'erreur numérique $\mathbf{P}_k^{y,y}$ et de l'erreur d'observation $\mathbf{R_k}$ près, la manière dont chaque observation contribue à la correction de la variable de contrôle. Son rôle est donc déterminant pour l'évolution de l'ensemble et la performance de l'EnKF-$\gamma$-KLBC-MEG.

### 5.6.1 Influence du rééchantillonnage sur les fonctions de covariance entre les coefficients de frottement et les observations simulées

La figure 5.31 montre l'évolution temporelle de la fonction de covariance croisée entre $Ks2^a$ et l'observation simulée $\mathbf{y}$ à Pauillac lorsque seul $Ks2$ est corrigé à l'aide de l'observation synthétique de $True1$ extraite en cette même station d'observation (Exp. 1-b) pour les configurations suivantes : a) $N_e = 10$ membres et aucun rééchantillonnage n'est utilisé pendant la durée de l'expérience 24 h (dont 6 h de spin-up) - b) $N_e = 10$ membres sur 24 h et un rééchantillonnage est utilisé - c) $N_e = 100$ membres pour une expérience de 96 h en utilisant la technique de rééchantillonnage. La figure 5.31-d montre l'évolution temporelle de la fonction de covariance croisée entre $Ks2^a$ et $\mathbf{y}$ en plusieurs stations d'observation, dont Pauillac, pour $N_e$=100 membres sur 96 h sans rééchantillonnage et avec correction simultanée des autres $Ks$ et du forçage maritime dans l'espace réduit de KL pour des valeurs vraies extraites de l'expérience de référence $True2a$.

La figure 5.31 montre que $Ks2$ est fortement corrélé avec le signal de marée, représenté au Verdon en pointillés bleus. La fonction de covariance est peu influencée par $N_e$ lorsque $Ks2$ est la seule variable de contrôle. En revanche, dans ce cas, le rééchantillonnage influe sur la fonction de covariance en transformant la forme sinusoïdale de la fonction en un signal créneau de valeur 1 autour des PM et -1 sinon. L'intégration d'autres variables dans le vecteur de contrôle sans rééchantillonnage (et affaissement de l'ensemble autour de 24 h après le début de l'expérience) montre également une fonction de covariance corrélée avec le signal de marée, sauf à Pessac et à La Réole.

### 5.6.2 Influence du rééchantillonnage sur les fonctions de covariance croisée pour un vecteur de contrôle intégrant les coefficients de frottement et les coefficients modaux associés aux forçages (Exp. 3-b)

Les fonctions de corrélation entre chaque variable de contrôle et les niveaux d'eau calculés par Telemac2D en chaque point de l'estuaire dont les données sont assimilées à l'aide de l'EnKF-$\gamma$-KLBC-MEG sont représentées pour l'Exp. 3-b (donc avec rééchantillonnage) sur les figures 5.32, 5.33, 5.34 et 5.35. Le niveau d'eau au Verdon est représenté en fonction du temps en ligne pointillée bleue à l'aide de l'axe des ordonnées de droite.

L'analyse des fonctions de corrélation liées à chaque variable de contrôle $Ks1$, $Ks2$ et $Ks3$ représentées sur la figure 5.32 permet d'établir, en accord avec l'étude de quantification d'incertitudes, d'une part, que ces trois paramètres sont corrélés avec les niveaux d'eau calculés aux stations sous influence maritime et, d'autre part, qu'ils sont décorrélés des niveaux d'eau aux

stations de Pessac (ligne rose à marqueur circulaire) et de La Réole (ligne violette à marqueur triangulaire vers le bas). La corrélation entre les coefficients de Strickler et les stations sous influence maritime dépend du cycle de marée, comme vu dans la section 5.6.1. En revanche, leur fonction de corrélation avec Pessac et La Réole est nulle pendant la durée de l'expérience.

Pendant les hautes eaux (niveau d'eau supérieur à la mi-marée) au Verdon, la corrélation de $Ks1$ avec toutes les stations sauf Pessac et la Réole est périodique et culmine à 0.4 au pic de pleine mer (PM) au Verdon. De manière pratique, $Ks1$ évolue de la même manière que le niveau d'eau pendant la fin du flot et le début du jusant. Pendant les basses eaux (fin du jusant et début du flot), la corrélation avec les niveaux d'eau calculés à Richard et Laména (resp. pour les autres stations) est négative et comprise dans l'intervalle [-0.6 ; -0.4] (resp. [-0.2 ; 0]). La corrélation de $Ks1$ avec les stations est proche de 0 en basses eaux (au Verdon) pour les stations en amont de la confluence entre la Dordogne et la Garonne. A l'aval de la confluence, elle est de nouveau en phase avec le cycle de marée et atteint sa valeur minimale à Richard (-0.4) au moment de la basse mer. Ainsi, pendant les basses eaux, $Ks1$ est décorrélé des stations en amont de la confluence et évolue dans le sens contraire des niveaux d'eau des stations situées à l'embouchure de l'estuaire (principalement Richard et Laména).

De manière plus accentuée, l'amplitude de la fonction de corrélation avec $Ks2$ est maximale pour Richard et Bec d'Ambès et jusque Laména. Elle présente également un maximum (resp. un minimum) au moment de la pleine mer (resp. de la basse mer) au Verdon. Une décorrélation a lieu au flot et au jusant à mi-marée. Ainsi, pendant les hautes eaux (resp. basses eaux) au Verdon, $Ks2$ évolue dans le même sens que les niveaux d'eau situés dans la zone entre Pauillac et Bordeaux (resp. dans le sens contraire).

$Ks3$ présente une corrélation périodique négative forte avec les stations de Bassens et Bordeaux et faible avec les autres stations. Ainsi, $Ks3$ est presque décorrélé des niveaux d'eau de toutes les stations sauf aux basses mers où $Ks3$ évolue en sens contraire des niveaux d'eau en aval de la confluence.

La figure 5.36 montre la fonction de covariance de $\alpha 1$ avec les stations d'observation de l'estuaire sans rééchantillonnage. Aucune tendance particulière ne peut être détectée. La figure 5.33 représente la corrélation entre les coefficients modaux associés au signal maritime CLMAR et les niveaux d'eau simulés en 12 points d'observation de l'estuaire. Tous les coefficients modaux sont faiblement corrélés aux niveaux d'eau des stations de La Réole et de Pessac situées aux frontières fluviales du modèle.

En rééchantillonnant l'ensemble selon $\mathbf{P}_0^\gamma$, il est intéressant d'observer que, malgré des oscillations d'ordre de grandeur inférieur, la fonction de corrélation entre un coefficient modal donné et les stations est proportionnelle au mode lui-même. Il s'agit de la vérification pratique des conséquences du rééchantillonnage, comme la section 4.2.3 l'a démontré.
Par ailleurs, la faible taille de l'ensemble et les non-linéarités véhiculées par le modèle hydrodynamique apparaissent sous la forme d'oscillations secondaires de même période que la marée.

L'amplitude des oscillations secondaires dépend de l'événement considéré, ainsi que de la station de mesures. Ainsi, la fonction de corrélation du mode 3 a une amplitude plus grande à Bordeaux qu'à Richard.

Le même phénomène de portance principale, démontré dans la section 4.2.3, associée à la période de chaque mode propre de la matrice de covariance du vecteur de contrôle peut être observé pour les conditions aux limites fluviales sur les figures 5.34 et 5.35. Ainsi, de manière très nette, la fonction de corrélation des coefficients modaux pour le débit de la Garonne (resp. de la Dordogne) est proportionnelle au mode propre correspondant pour la station de La Réole (resp. Pessac). L'influence de la marée ne se faisant pas sentir aux frontières amont des deux rivières, aucune oscillation n'est présente sur le signal représentant la fonction de corrélation. Les corrélations avec les autres stations en aval du tronçon fluvial présentent des oscillations liées à la période de la marée et au faible nombre de membres.

La figure 5.37 permet de visualiser l'évolution sur un cycle de marée de la matrice de covariance normalisée entre les variables de contrôle et les niveaux d'eau calculés par Telemac2D dans l'espace des observations.
De manière prévisible, les niveaux d'eau des stations de la Réole et de Pessac présentent une corrélation exclusivement avec les coefficients modaux des forçages fluviaux de la Garonne et de la Dordogne.
Les coefficients de Strickler $Ks1$, $Ks2$ et $Ks3$ ont des corrélations (positives ou négatives) significatives avec les zones de l'emprise situées respectivement entre l'embouchure (Le Verdon) et Pauillac, Laména et Bordeaux et, enfin, la zone de la confluence située entre Bassens et Bordeaux. Ces constats confirment l'analyse des zones d'influence de chacune de ces variables de l'étude de quantification d'incertitudes.
Les coefficients modaux du signal maritime ont une corrélation (positive ou négative) avec les niveaux d'eau de toutes les stations à juste titre "sous influence maritime". Pour le cycle de marée représenté, de 56 h à 71 h après le début de l'expérience, la corrélation (en valeur absolue) de chaque mode avec toutes les stations est maximale lors de la basse mer (BM) au Verdon et plus faible sauf pour les stations situées à l'embouchure de l'estuaire de la Gironde autour de la pleine mer (PM) au Verdon. Il peut être noté que les coefficients modaux $alpha5$, $alpha6$ et $alpha7$ n'ont pas de corrélation significative avec les niveaux d'eau calculés par Telemac2D aux stations dont les données sont assimilées pendant l'Exp. 3-b.

## 5.7 Conclusions sur les performances de l'EnKF-$\gamma$-KLBC-MEG dans un cadre synthétique et perspectives

Les performances de l'EnKF-$\gamma$-KLBC-MEG dans un cadre synthétique sont évaluées et l'influence de sa configuration étudiée dans ce chapitre.

Concernant l'estimation des paramètres et forçages du modèle, les paramètres $Ks1$ et $Ks2$ et

les forçages "vrais" sont correctement reconstitués au cours des expériences, même dans des situations de variation temporelle "rapide" du paramètre, comme c'est le cas pour l'Exp. 2. Une méthode de rééchantillonnage fondée sur la persistance de $\mathbf{P}_0^{\gamma}$ est testée pour remédier à l'effondrement de l'ensemble causé par l'EnKF-$\gamma$-KLBC-MEG pour les paramètres de frottement $Ks$ du modèle. De possibles difficultés pour l'estimation du paramètre $Ks3$, situé dans la zone de la confluence où les écoulements sont issus de processus physiques non-linéaires, ont également été mises en évidence.

Le phénomène d'"équifinalité", c'est-à-dire de deux vecteurs de contrôle temporels, l'un vrai $\boldsymbol{\gamma}^{true}$ et l'autre issu de l'analyse $\boldsymbol{\gamma}^{a}$ conduisant sensiblement au même état du système $\mathbf{y}$, est constaté lors de l'estimation de paramètres et forçages "vrais" instationnaires et est susceptible de poser problème lors de l'assimilation de données réelles.

Les niveaux d'eau sont, cependant, correctement estimés. L'erreur est inférieure à 10 cm le long de l'estuaire pour toutes les expériences, en-dehors des tronçons amont des fleuves Garonne et Dordogne, très sensibles à la qualité de l'observation, mais également au raffinement du maillage du modèle d'évolution. Il faut noter que les résultats en termes de niveaux d'eau sont en particulier bons aux stations faisant l'objet d'une validation et dont les mesures ne sont pas assimilées au cours des expériences.

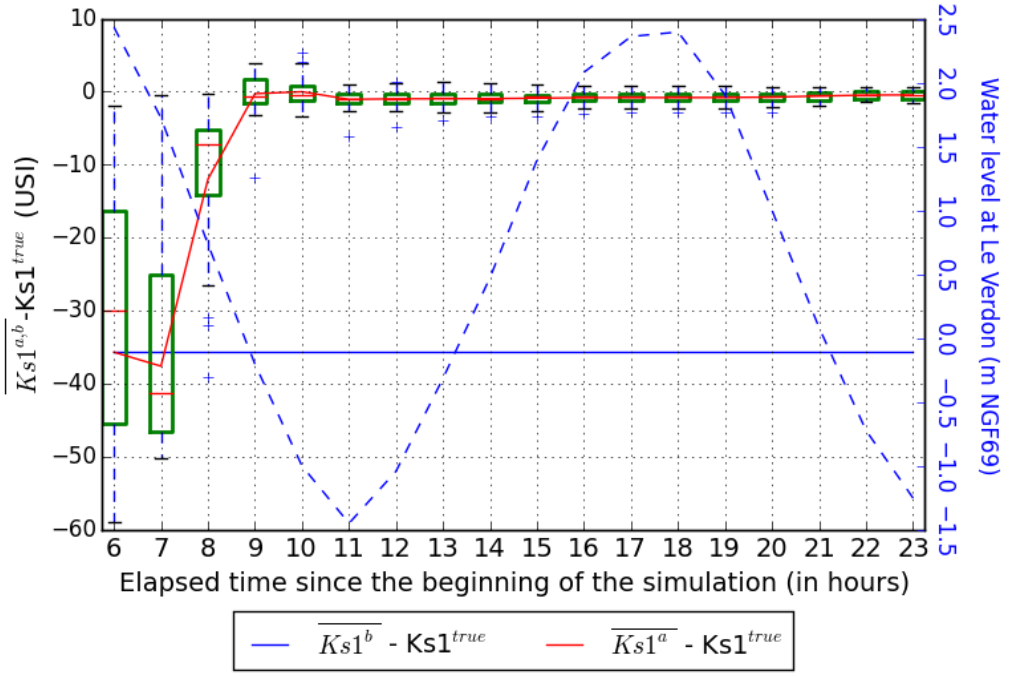
Les résultats des différentes expériences jumelles menées sont donc satisfaisants, mais mettent d'ores et déjà en lumière des difficultés susceptibles d'apparaître lors de la future utilisation de mesures réelles, du fait du nombre de membres $N_e$ faible, responsable d'erreurs d'échantillonnage, mais aussi de l'équifinalité qui peut impacter l'estimation des variables de contrôle par l'EnKF-$\gamma$-KLBC-MEG.

Des perspectives existent pour l'amélioration des performances de l'EnKF-$\gamma$-KLBC-MEG dans un cadre synthétique. D'autres expériences jumelles permettraient d'aller plus loin dans l'étude des spécifications optimales. Une expérience mettant en œuvre des coefficients modaux variables au cours du temps permettrait d'apprécier l'inertie de l'EnKF-$\gamma$-KLBC-MEG vis-à-vis de ces paramètres, comme c'est le cas pour les coefficients de Strickler. Par ailleurs, des durées de fenêtres d'assimilation de données plus longues associées à d'autres valeurs de décalage entre deux fenêtres successives, à l'aide des paramètres $L$ et $S$, pourraient également être testées.

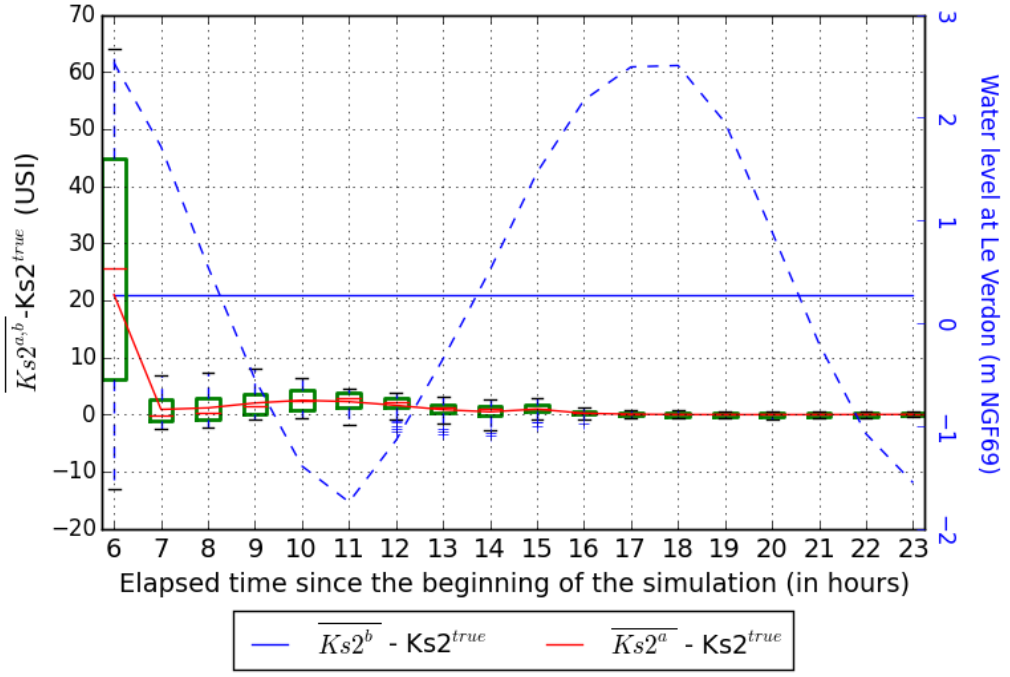
L'une des pistes de solution pour atténuer le problème de l'équifinalité, non explorée dans le cadre de ce travail, consiste éventuellement à découpler la correction des paramètres du système (qui influent sur la propagation de l'onde de marée principalement) de la correction des forçages aux frontières maritimes et fluviales dans l'espace réduit de KL.

Enfin, le filtre de Kalman itératif (IEnKF) décrit dans Bocquet and Sakov [2013] permet une meilleure prise en compte des non-linéarités du système, comme c'est le cas des écoulements de la zone de la confluence de l'estuaire de la Gironde. En revanche, le coût de calcul pour un système estuarien, instationnaire, n'est pas anodin.
En effet, le nombre total de simulations de l'expérience compte, en mode réanalyse (sans phase

de prédiction à la fin de chaque cycle d'assimilation), $N_e \times nb_{cycles}$ intégrations du modèle sur un cycle d'assimilation $\mathcal{F}_k$ défini par $[t_{k,deb}\ ;\ t_{k,fin}]$ en phase d'analyse et $N_e \times nb_{cycles}$ intégrations sur $[t_{k,deb}\ ;\ t_{k+1,deb}]$ en phase d'optimisation de l'état du système auxquelles s'ajoutent $N_e$ intégrations d'initialisation (spin-up) en tout début d'expérience, soit $N_e \times (2 \times nb_{cycles} + 1)$ intégrations du modèle. Il faudrait alors multiplier le nombre d'intégrations du système en phase d'analyse par le nombre d'itérations nécessaire pour atteindre la convergence de l'algorithme de l'IEnKF. Son implantation au sein d'un IEnKF-$\gamma$-KLBC doit être validée pour évaluer ses performances par rapport à l'EnKF-$\gamma$-KLBC-MEG pour un système instationnaire réel tel que celui de l'estuaire de la Gironde.

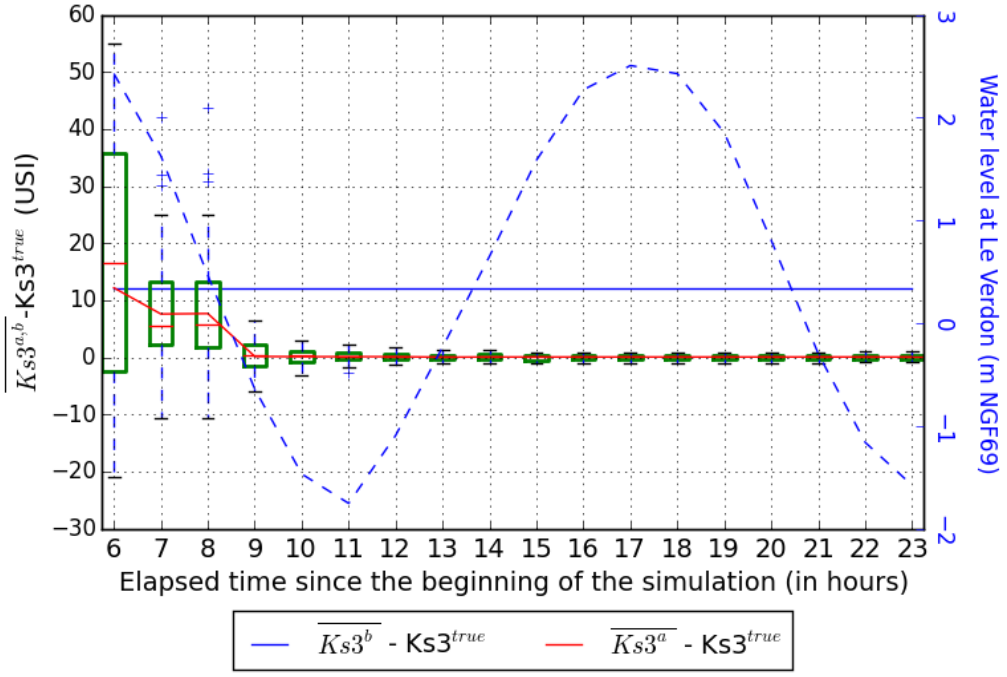



(a) 1-a

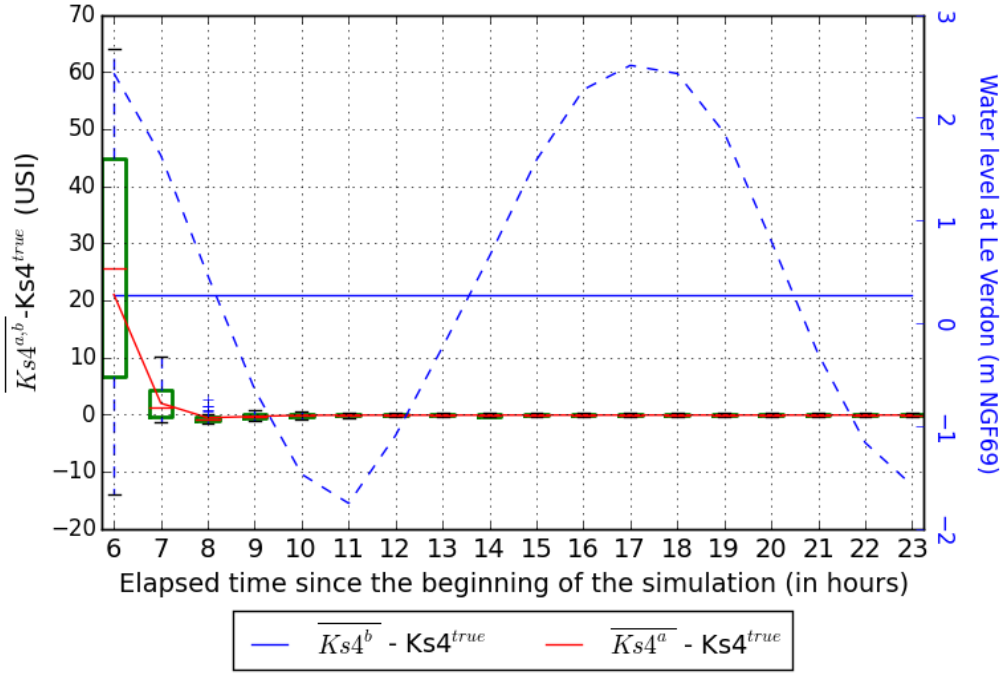



(b) 1-b

FIGURE 5.6 : Évolution de la différence entre la moyenne de l'ensemble après assimilation $Ks_i^a$ (resp. d'ébauche $Ks_i^b$) et la valeur "vraie" $Ks_i^{true}$ ($i \in \{1, 2, 3, 4\}$) pour les sous-expériences Exp 1-a à 1-d au début de chaque cycle d'assimilation (L=1h - S=1h) - 1 pas de temps (time step) = 1 h - durée de l'Exp. 1 : 24 h. Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. (début)

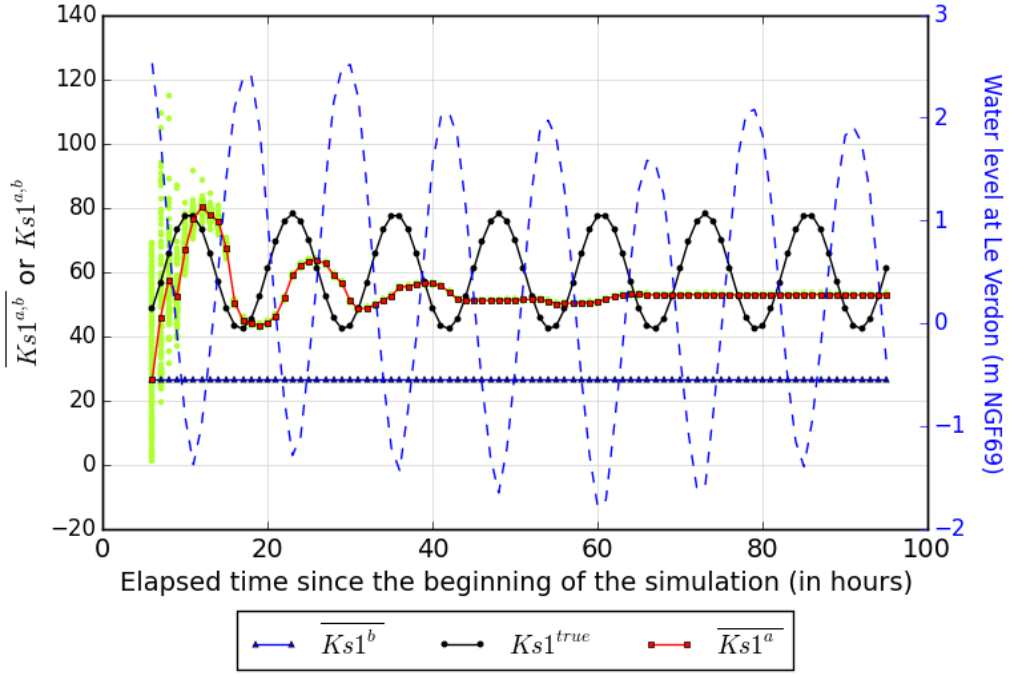



(c) 1-c

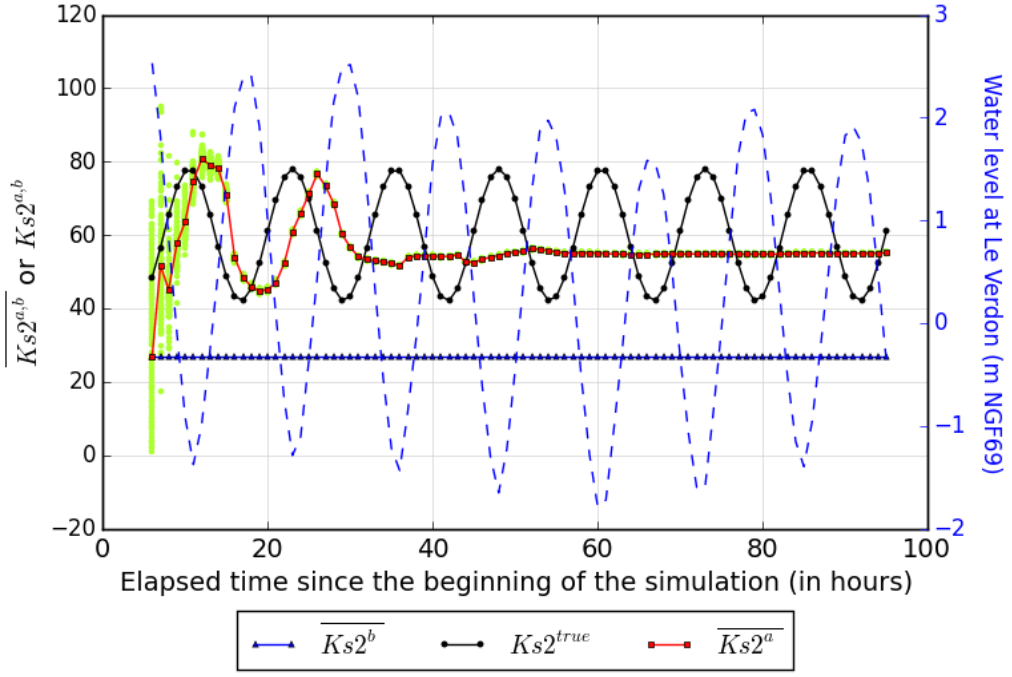



(d) 1-d

FIGURE 5.6 : Évolution de la différence entre la moyenne de l'ensemble après assimilation $Ks_i^a$ (resp. d'ébauche $Ks_i^b$) et la valeur "vraie" $Ks_i^{true}$ ($i \in \{1, 2, 3, 4\}$) pour les sous-expériences Exp 1-a à 1-d au début de chaque cycle d'assimilation (L=1h - S=1h) - 1 pas de temps (time step) = 1 h - durée de l'Exp. 1 : 24 h. Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. (fin)

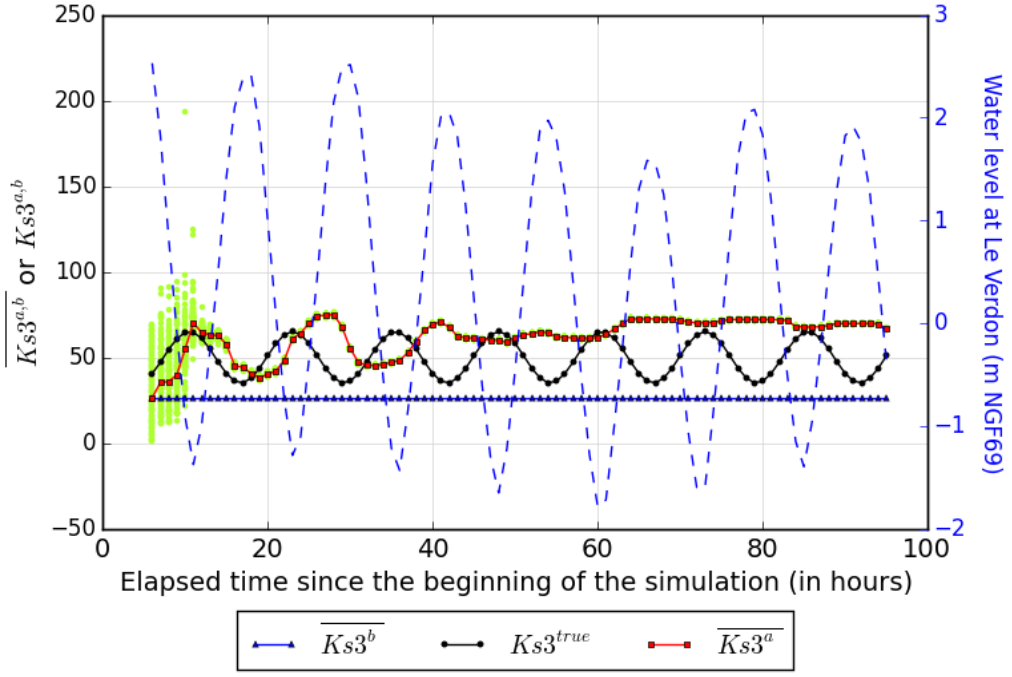



(a) $Ks1$



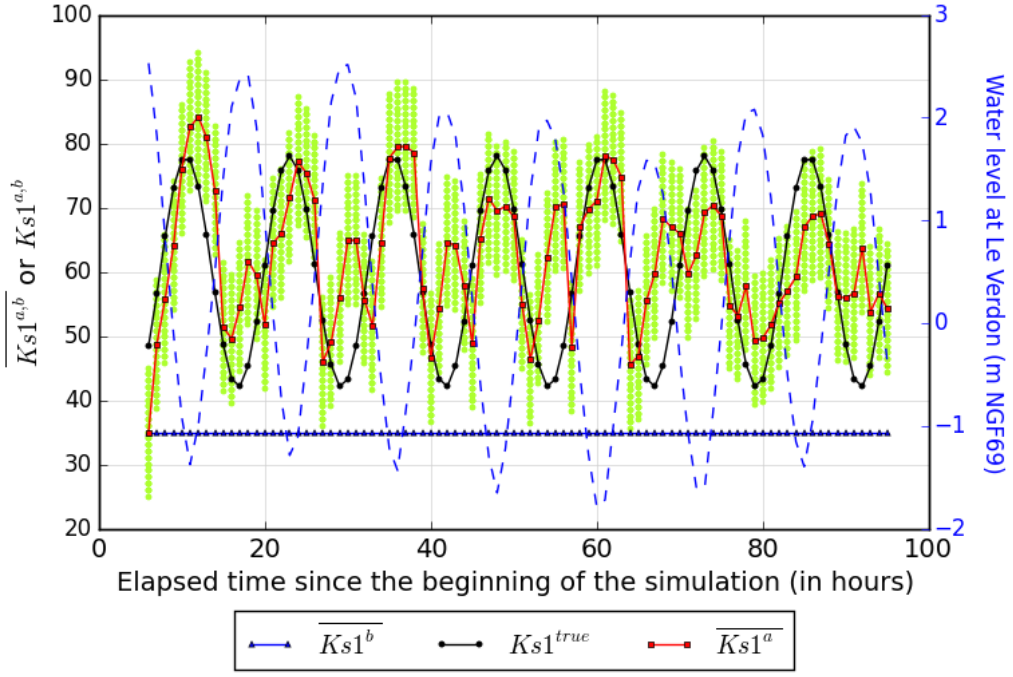

(b) $Ks2$

FIGURE 5.7 : Évolution de $Ks1$, $Ks2$ et $Ks3$ pour l'Exp 2-a. Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (début)

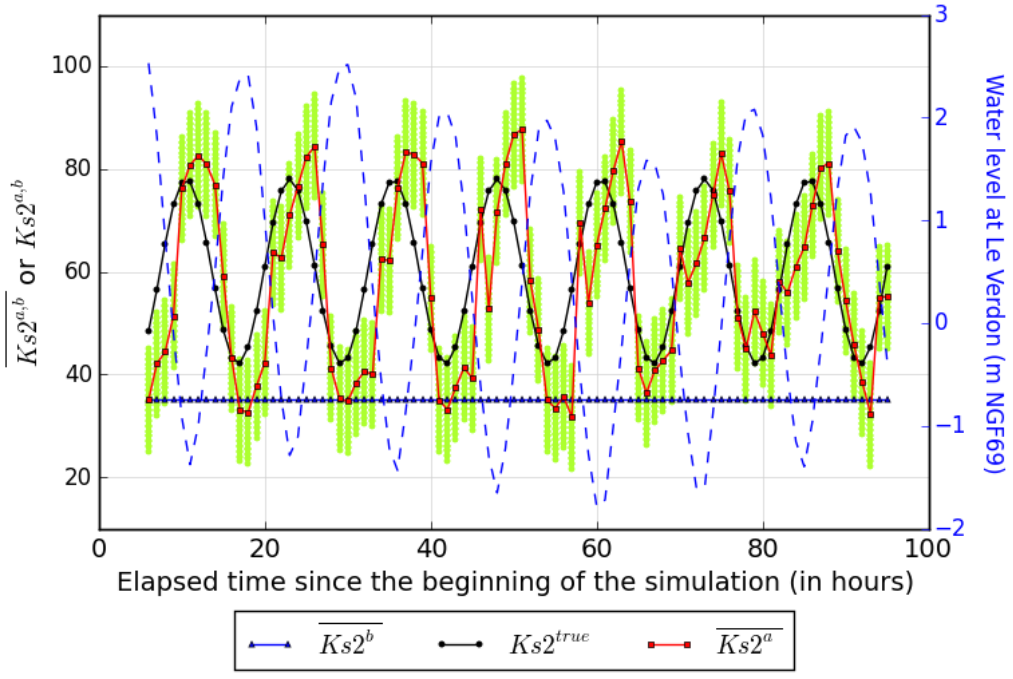



(c) $Ks3$

FIGURE 5.7 : Évolution de $Ks1$, $Ks2$ et $Ks3$ pour l'Exp 2-a. Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (fin)

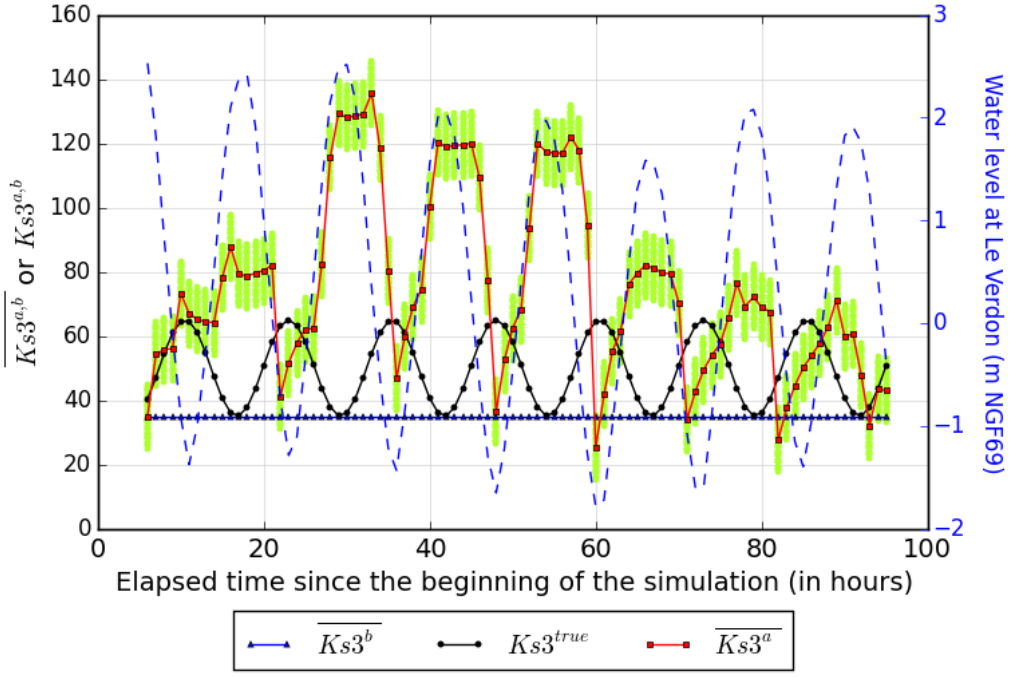



(a) $Ks1$



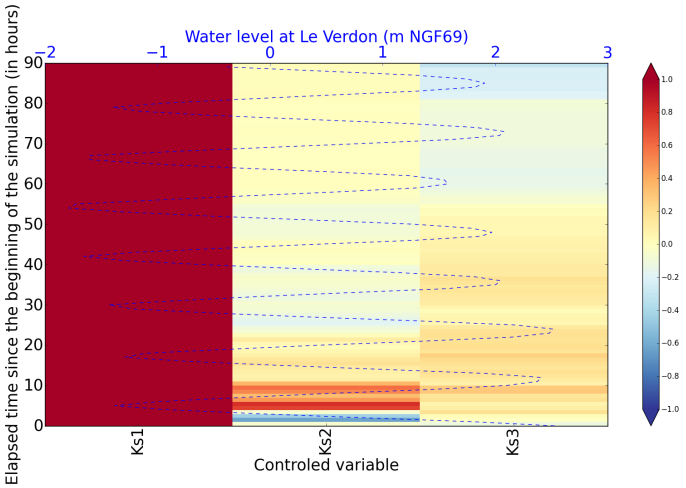

(b) $Ks2$

FIGURE 5.8 : Évolution de $Ks1$, $Ks2$ et $Ks3$ avec réensemencement pour l'Exp. 2-b. Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (début)

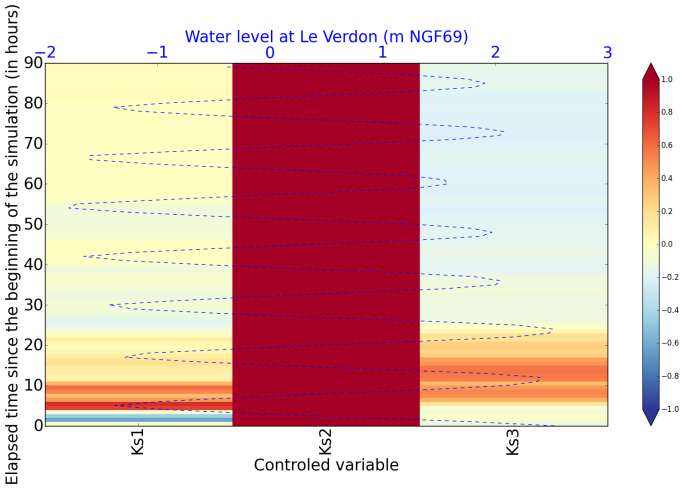



(c) $Ks3$

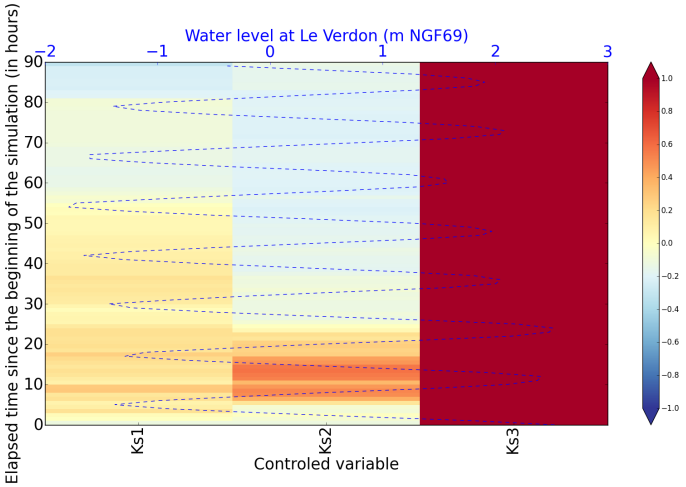
FIGURE 5.8 : Évolution de $Ks1$, $Ks2$ et $Ks3$ avec réensemencement pour l'Exp. 2-b. Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (fin)




(a) $Ks1$

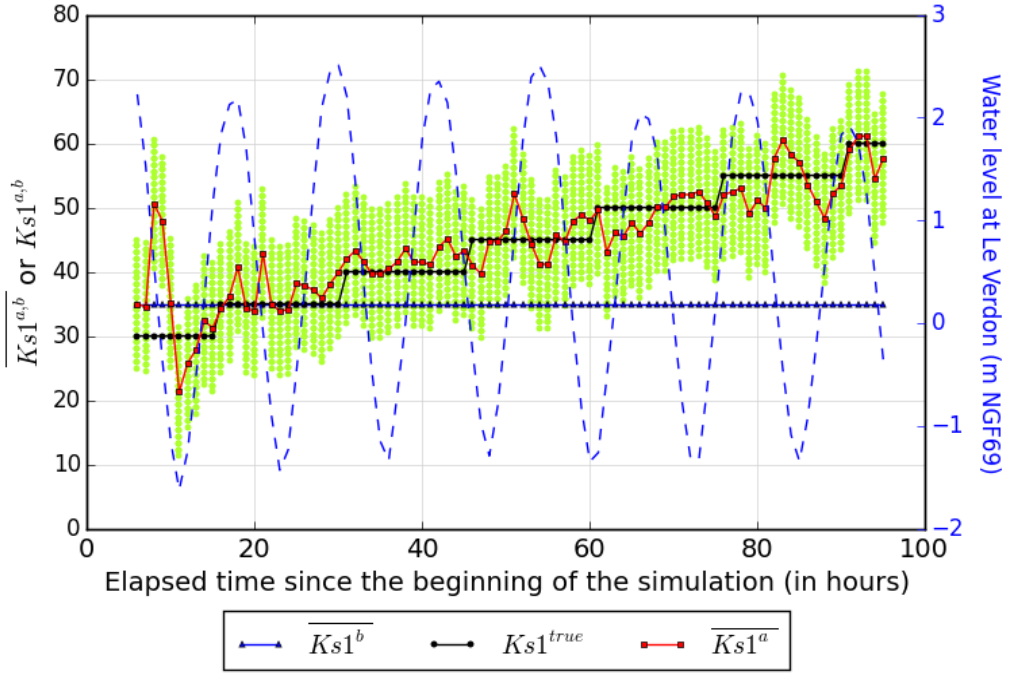



(b) $Ks2$

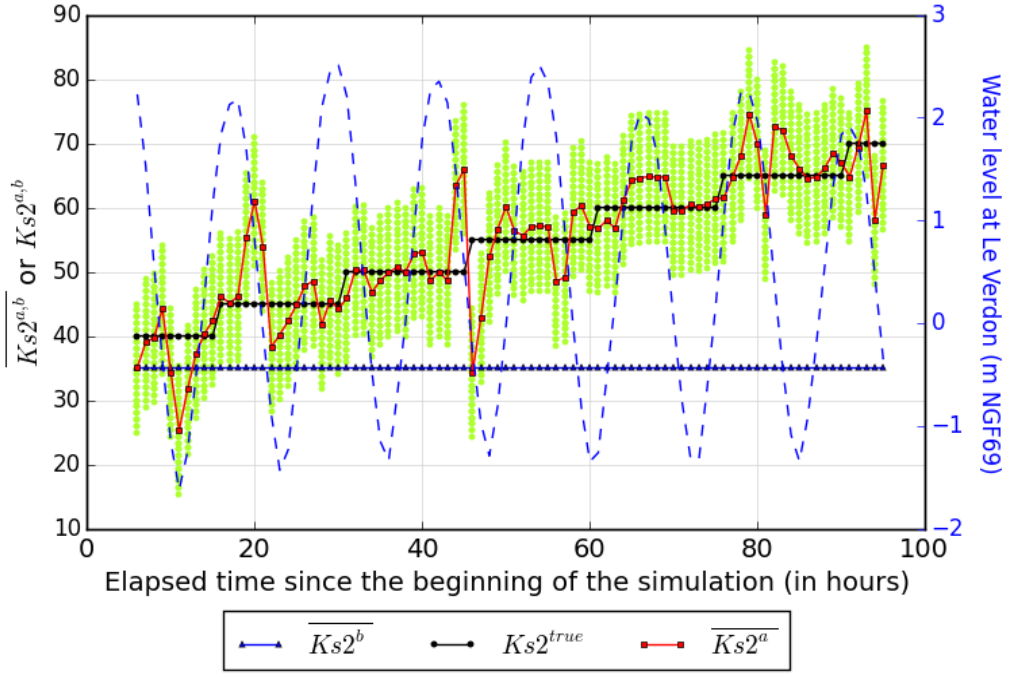



(c) $Ks3$

FIGURE 5.9 : Évolution de $\mathbf{P}^{\gamma}$ pour l'Exp. 2-a

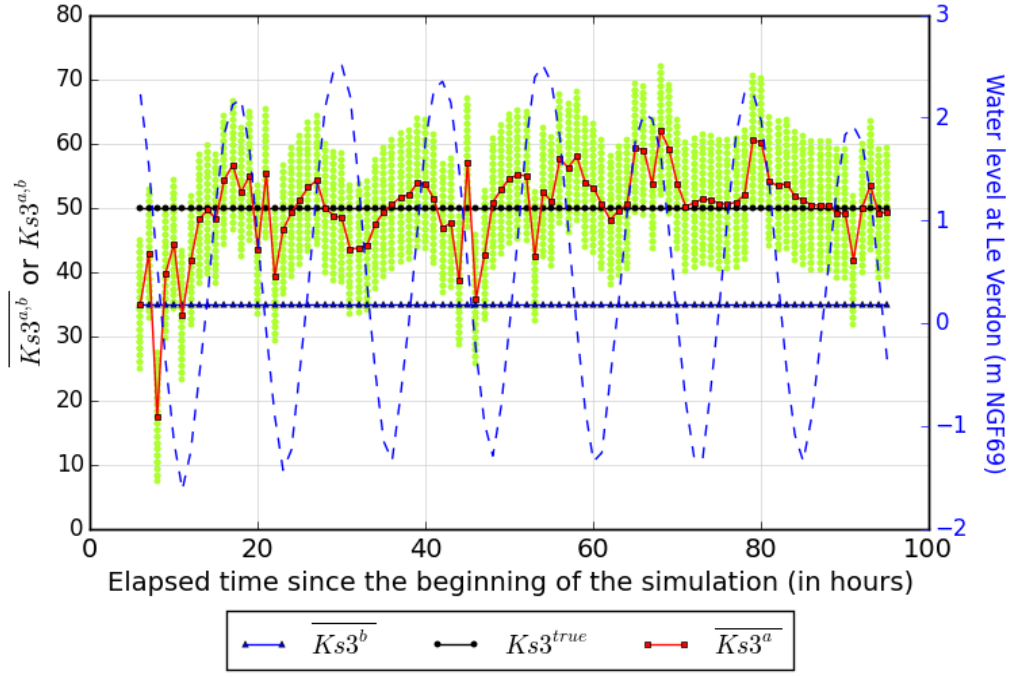



(a) $Ks1$



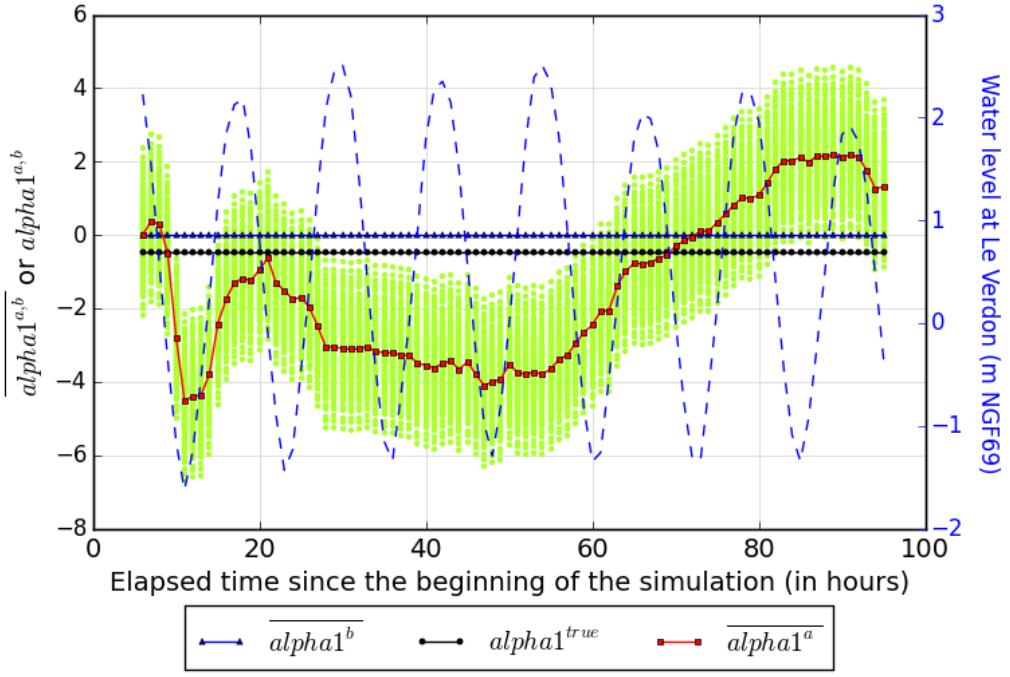

(b) $Ks2$

FIGURE 5.10 : Évolution de $Ks1$, $Ks2$ et $Ks3$ pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (début)

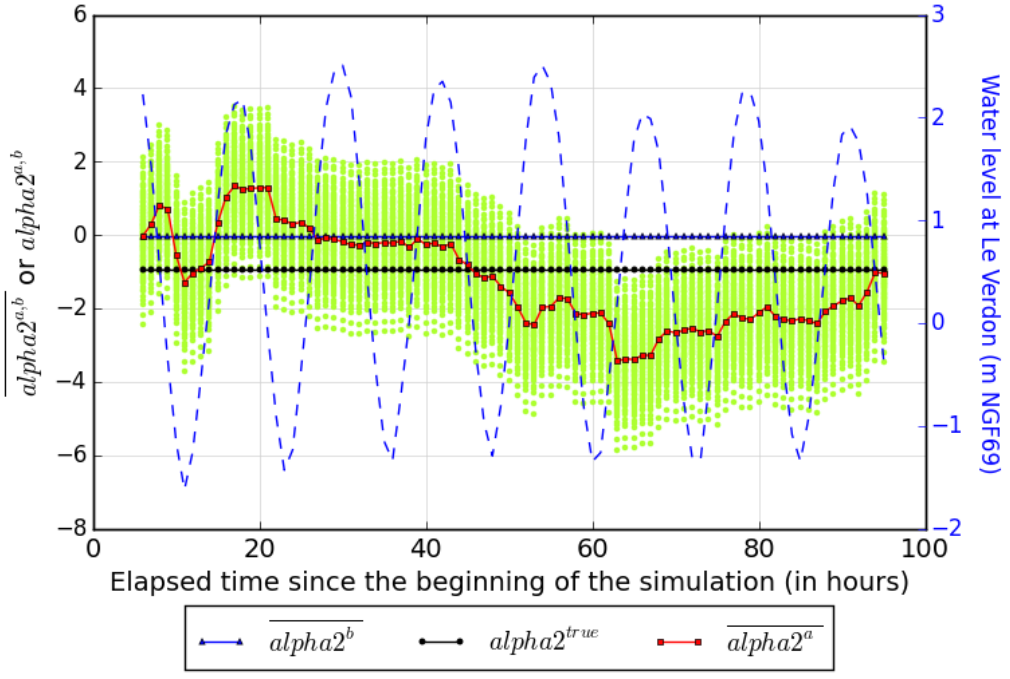



(c) $Ks3$

FIGURE 5.10 : Évolution de $Ks1$, $Ks2$ et $Ks3$ pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (fin)

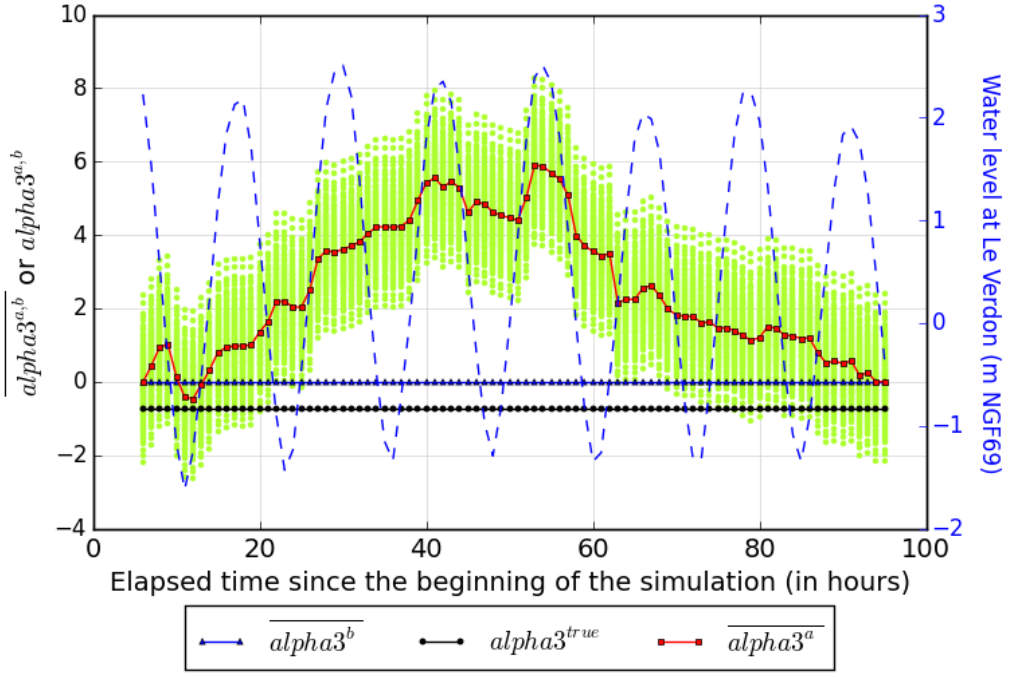



(a) CLMAR - mode 1

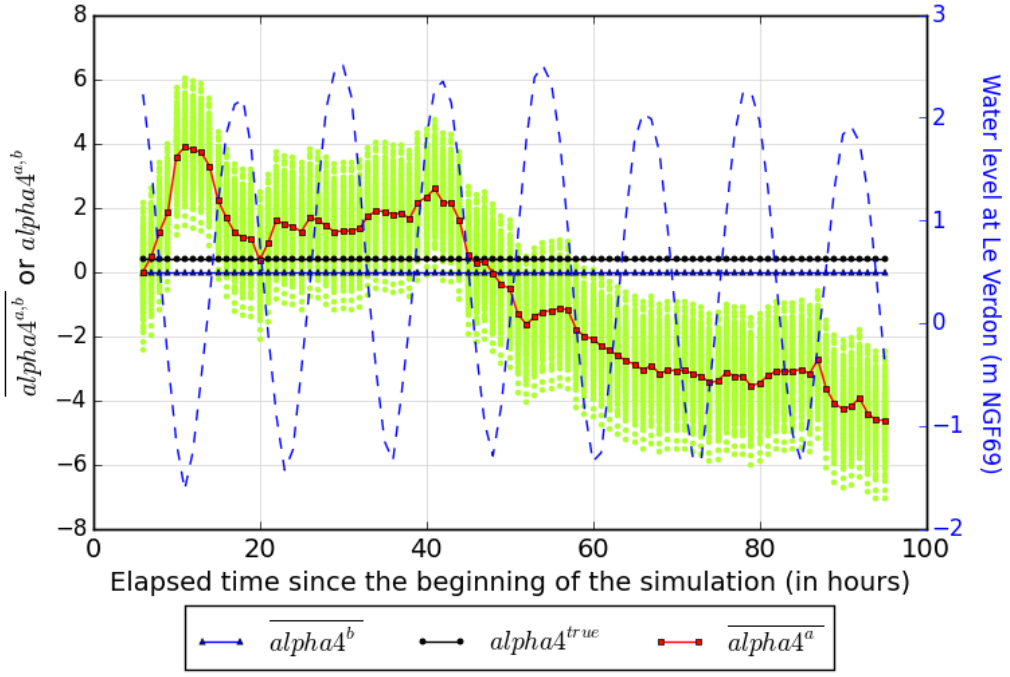



(b) CLMAR - mode 2

FIGURE 5.11 : Évolution des modes 1, 2, 3, 4 de CLMAR pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (début)

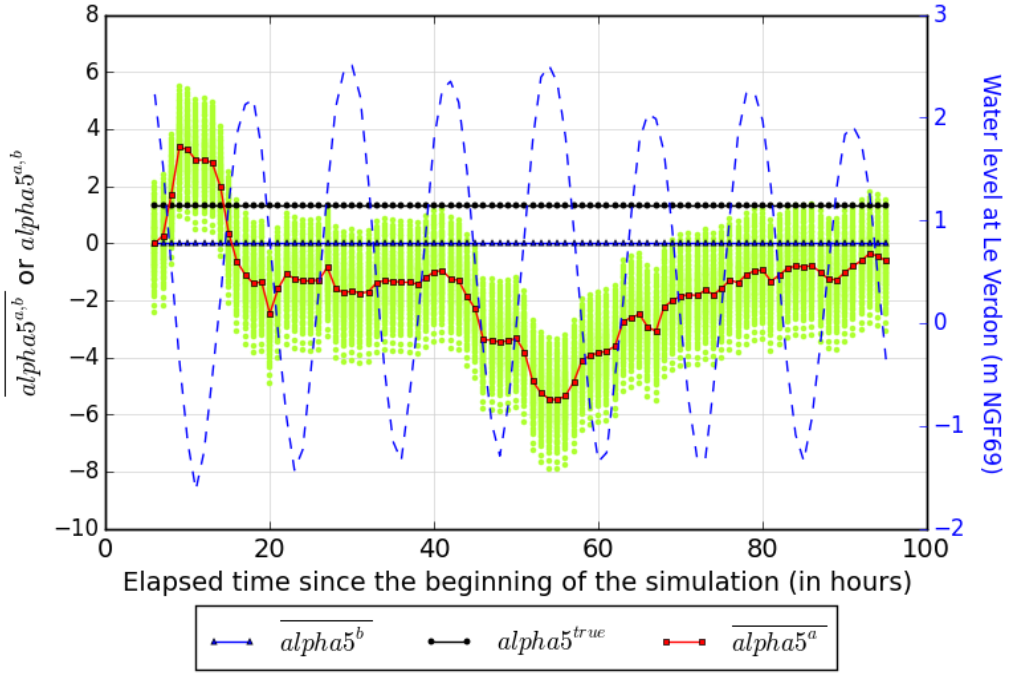



(c) CLMAR - mode 3

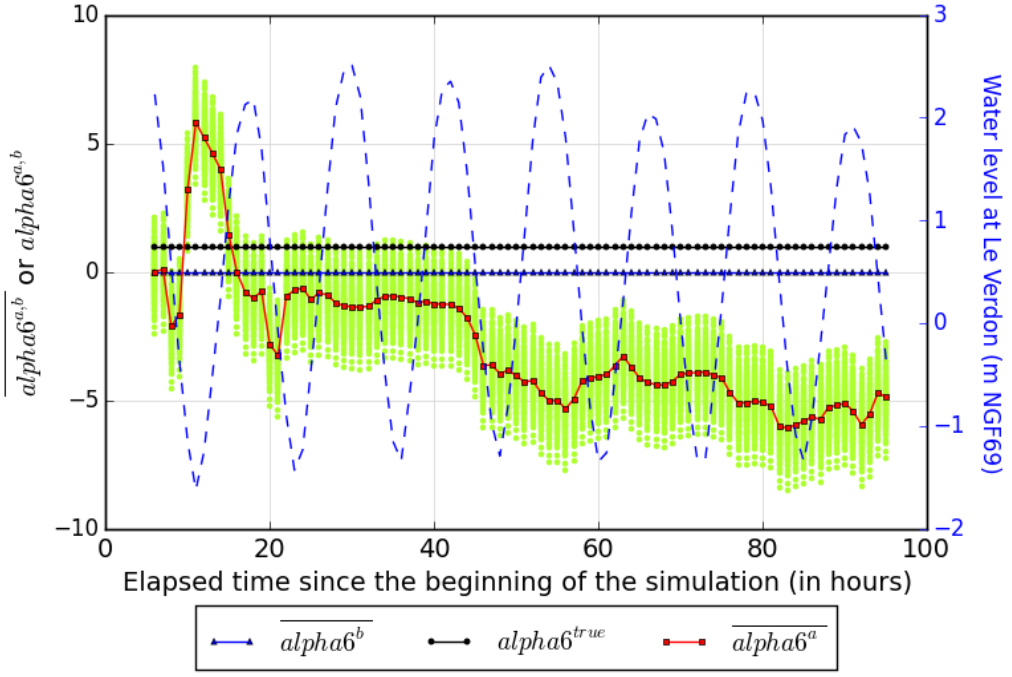



(d) CLMAR - mode 4

FIGURE 5.11 : Évolution des modes 1, 2, 3, 4 de CLMAR pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (fin)

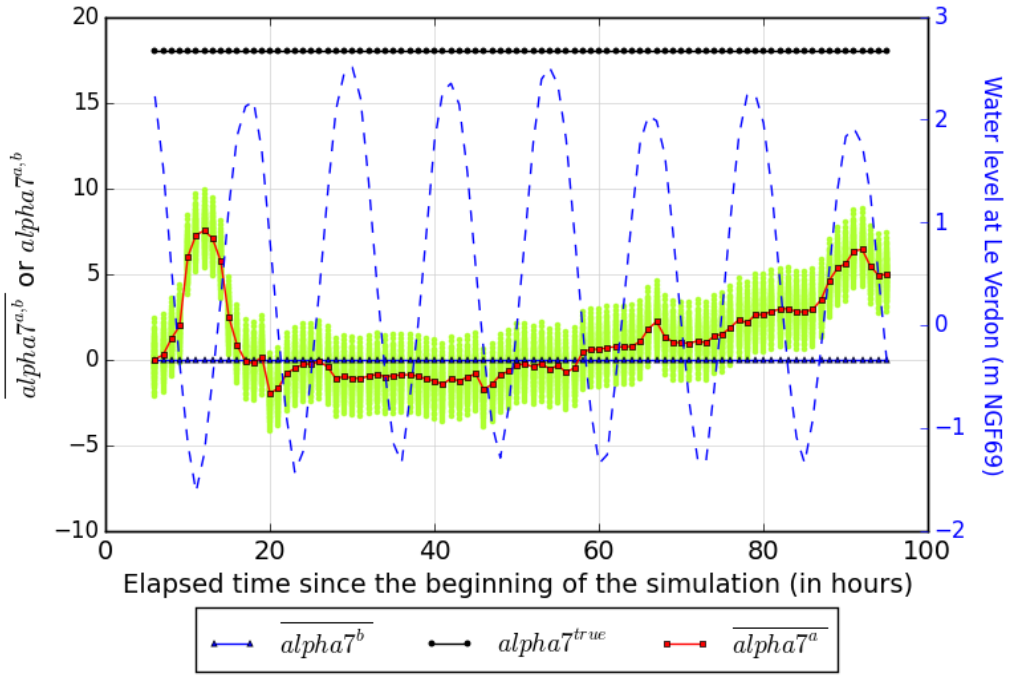



(a) CLMAR - mode 5



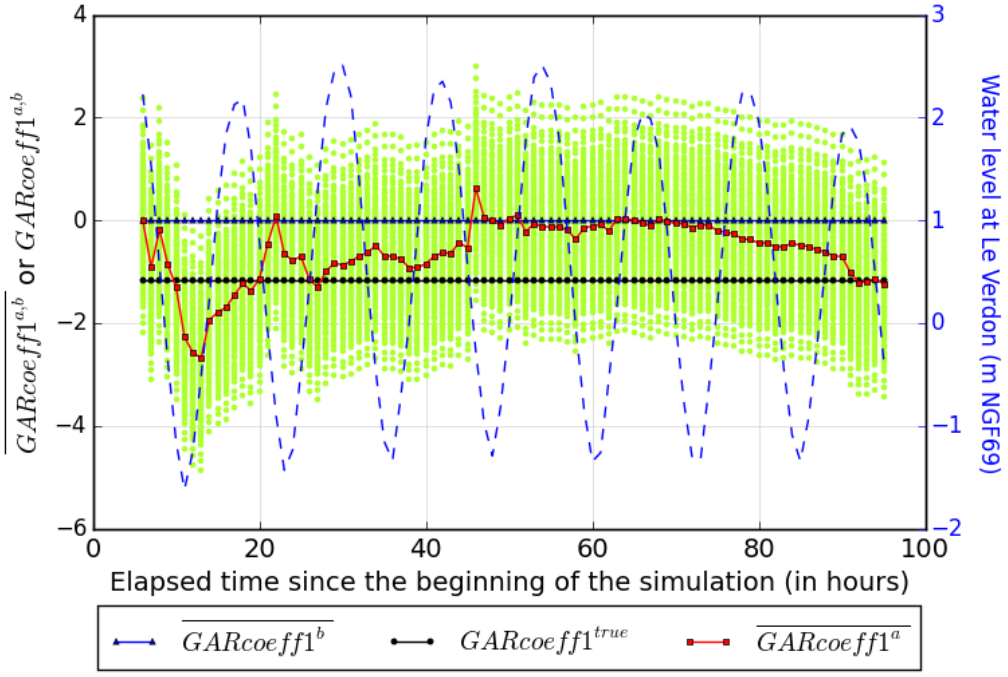

(b) CLMAR - mode 6

FIGURE 5.12 : Évolution des modes 5, 6, 7 de CLMAR pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (début)

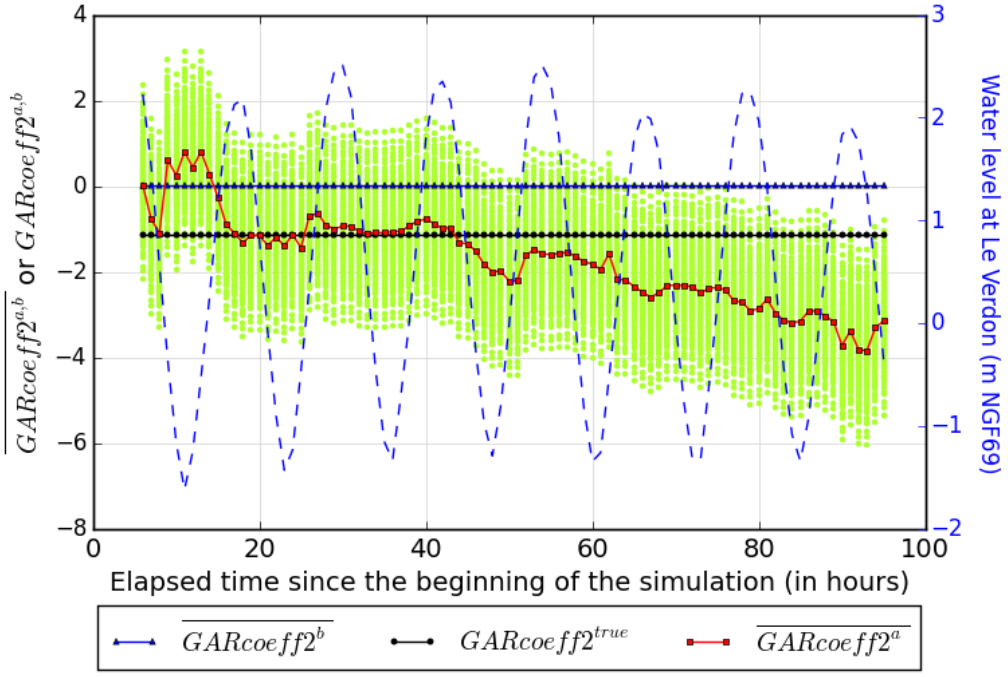



(c) CLMAR - mode 7

FIGURE 5.12 : Évolution des modes 5, 6, 7 de CLMAR pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (fin)

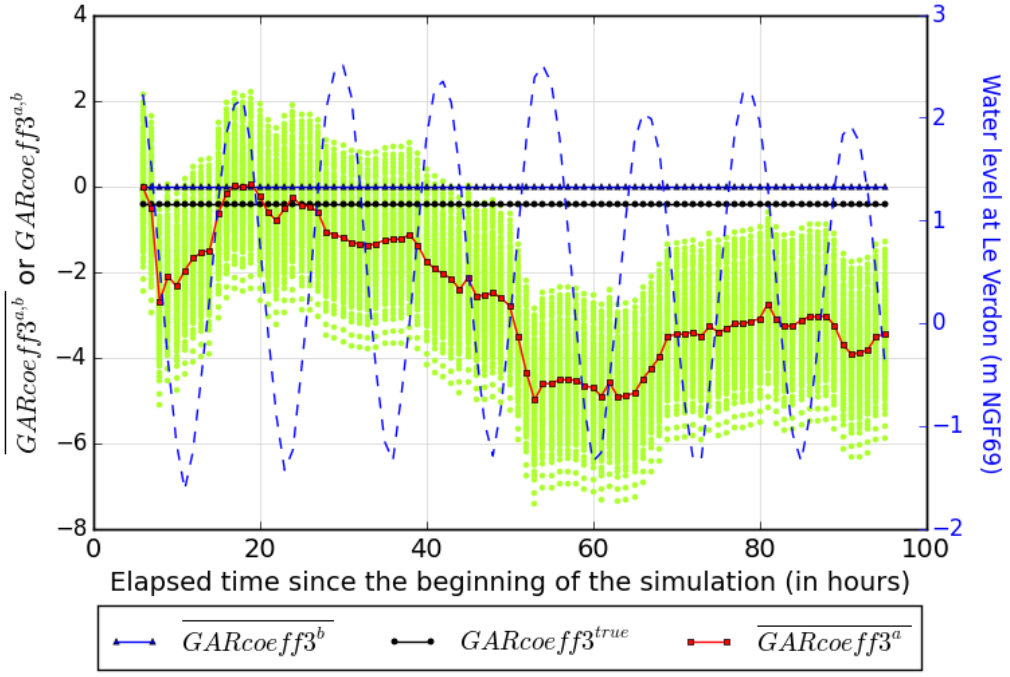



(a) QGAR - mode 1

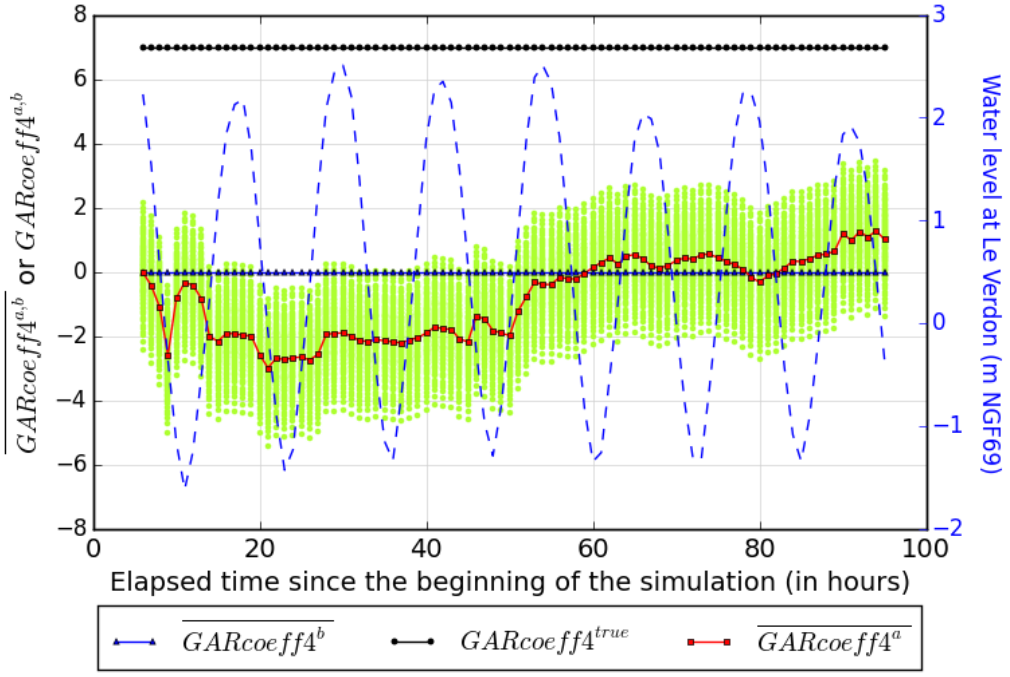



(b) QGAR - mode 2

FIGURE 5.13 : Évolution des modes de QGAR pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (début)

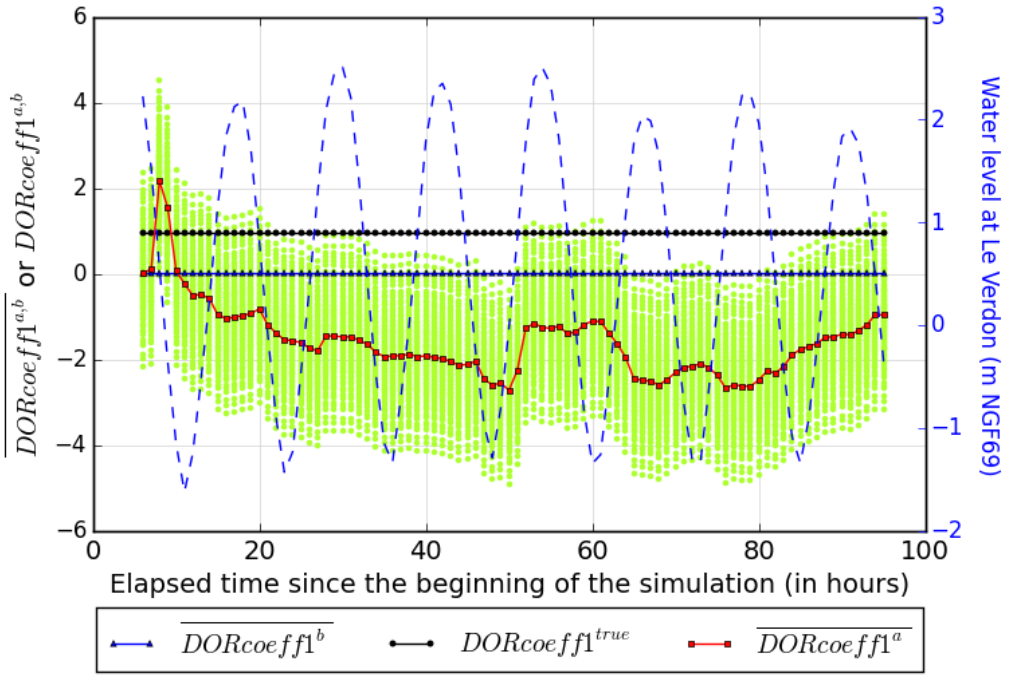



(c) QGAR - mode 3

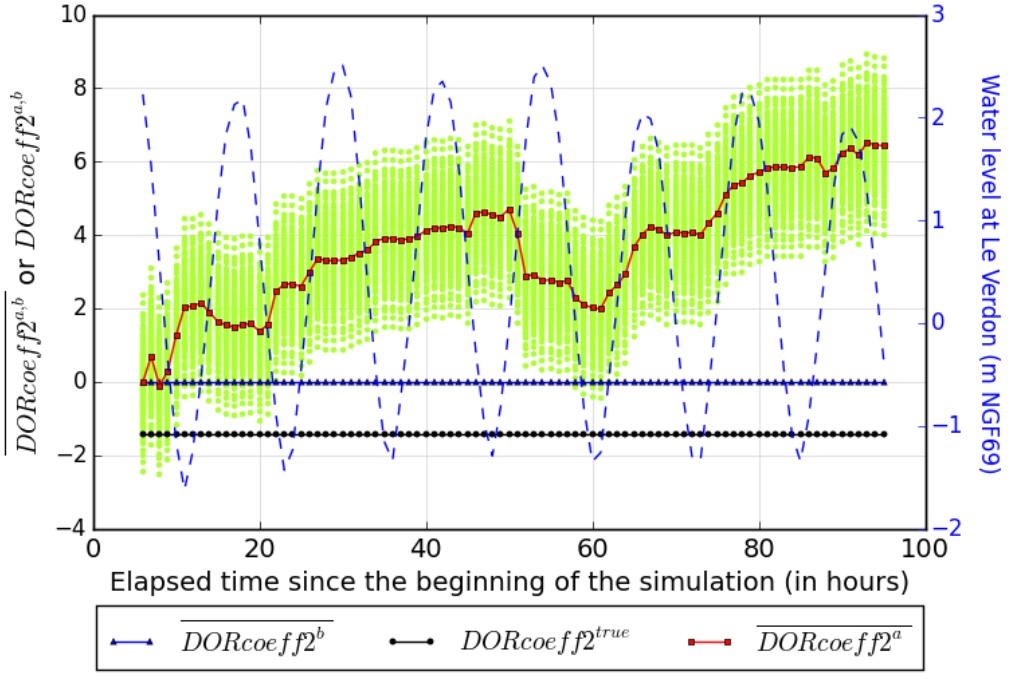



(d) QGAR - mode 4

FIGURE 5.13 : Évolution des modes de QGAR pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (fin)

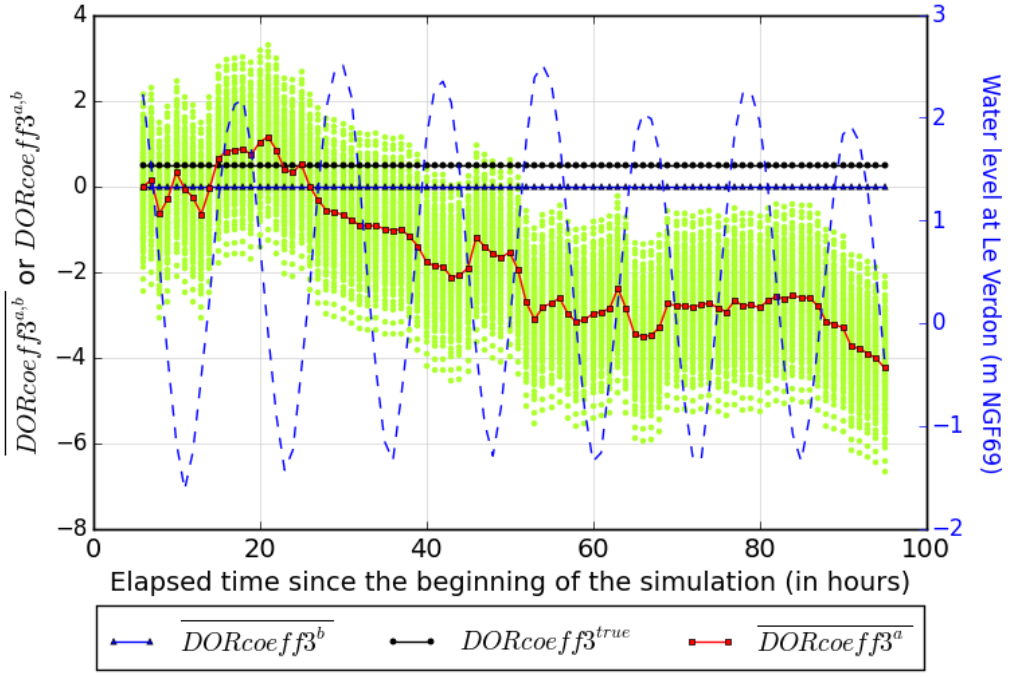



(a) QDOR - mode 1

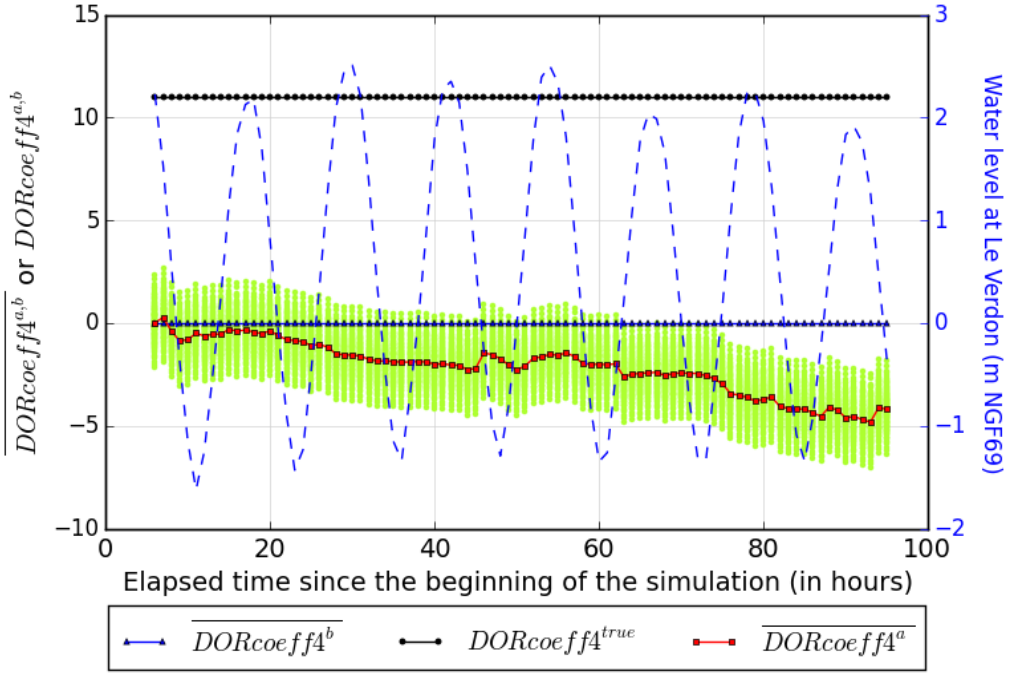



(b) QDOR - mode 2

FIGURE 5.14 : Évolution des modes de QDOR pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (début)

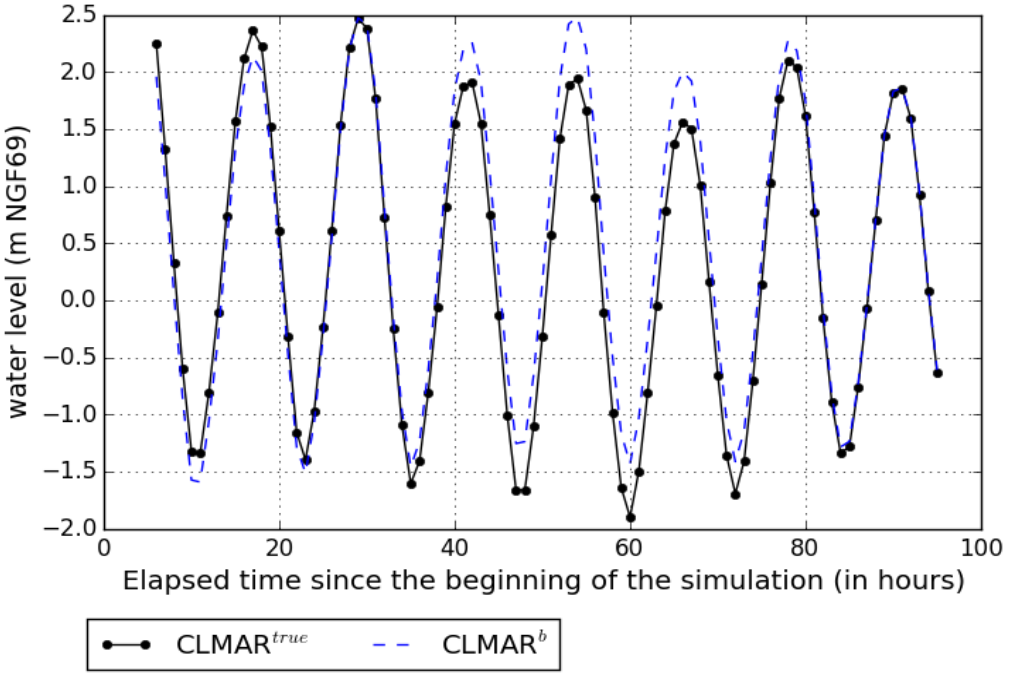



(c) QDOR - mode 3

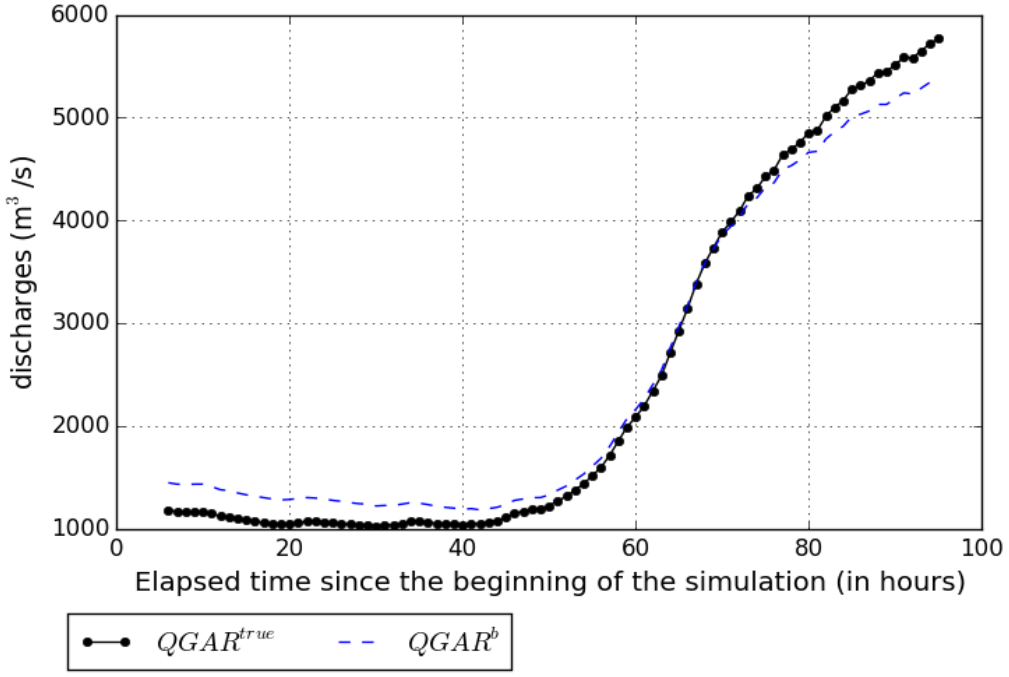



(d) QDOR - mode 4

FIGURE 5.14 : Évolution des modes de QDOR pour l'Exp. 3-b au cours des cycles d'assimilation (L=1h - S=1h). Le signal maritime au Verdon est représenté en pointillés bleus sur l'axe bleu de droite. Les points verts représentent l'ensemble des valeurs prises par la variable de contrôle à chaque pas de temps. (fin)




(a) CLMAR




(b) QGAR

FIGURE 5.15 : Évolution temporelle (depuis le début de l'événement) pour l'expérience de référence $True2$ : a) de la condition à la limite maritime $CLMAR$ ; b) de la condition aux limites fluviale sur la Garonne. $CLMAR^{true}$ et $QGAR^{true}$ sont les valeurs "vraies" dont sont issues les observations synthétiques $\mathbf{y}^0$. $CLMAR^b$ et $QGAR^b$ sont les forçages d'ébauche ("background") du run de contrôle.

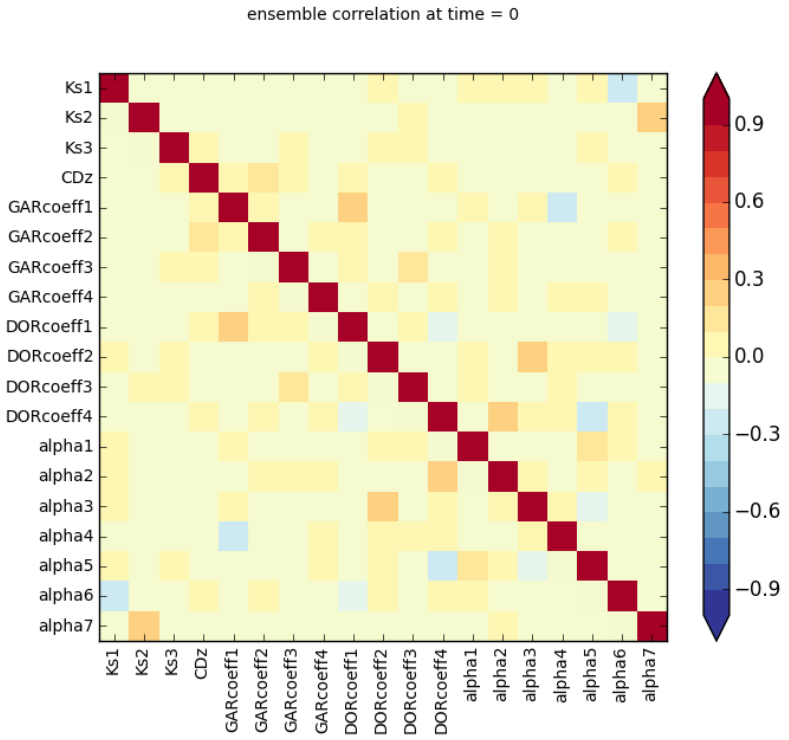
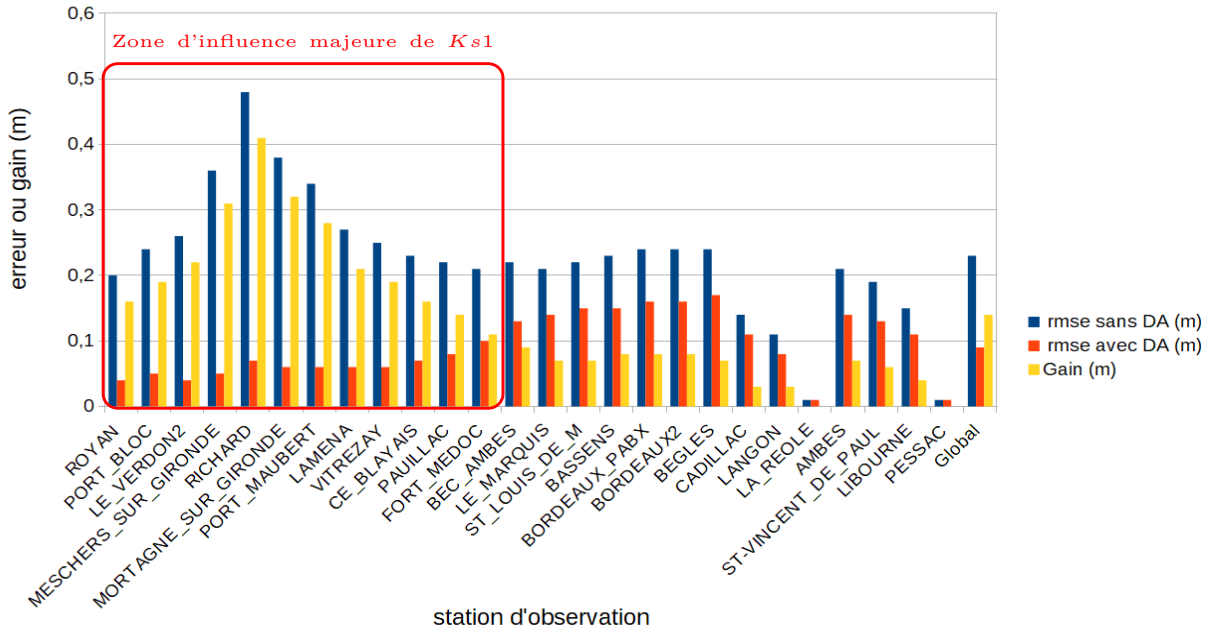
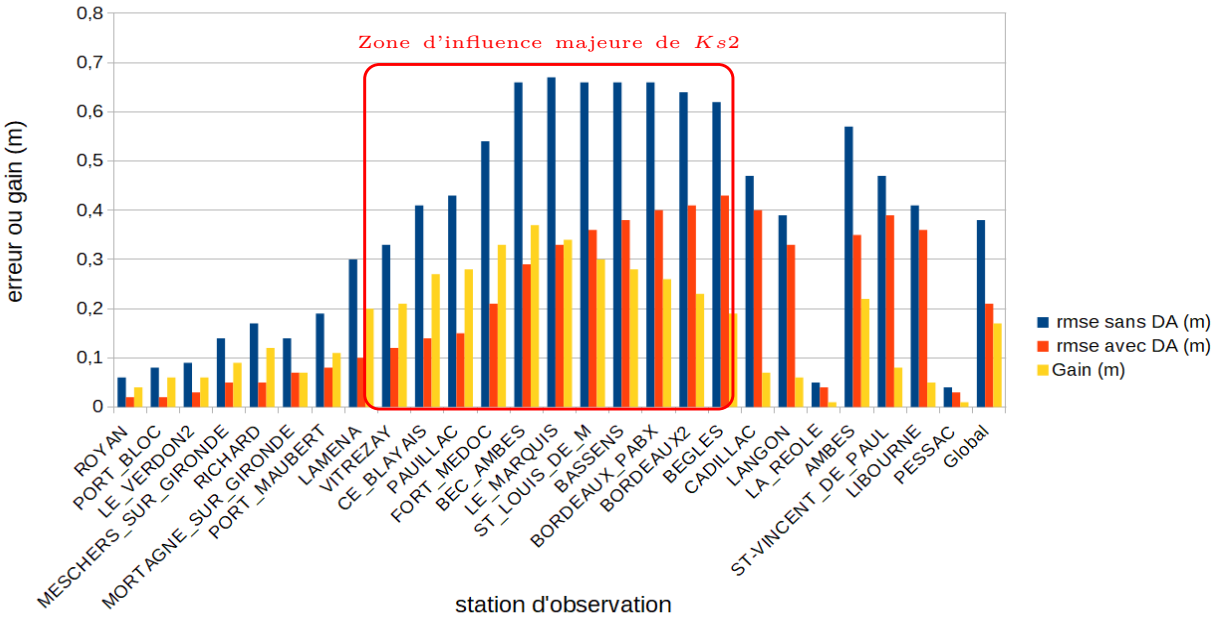



FIGURE 5.16 : Matrice de covariance d'erreur d'ébauche de l'ensemble pour l'Exp. 3

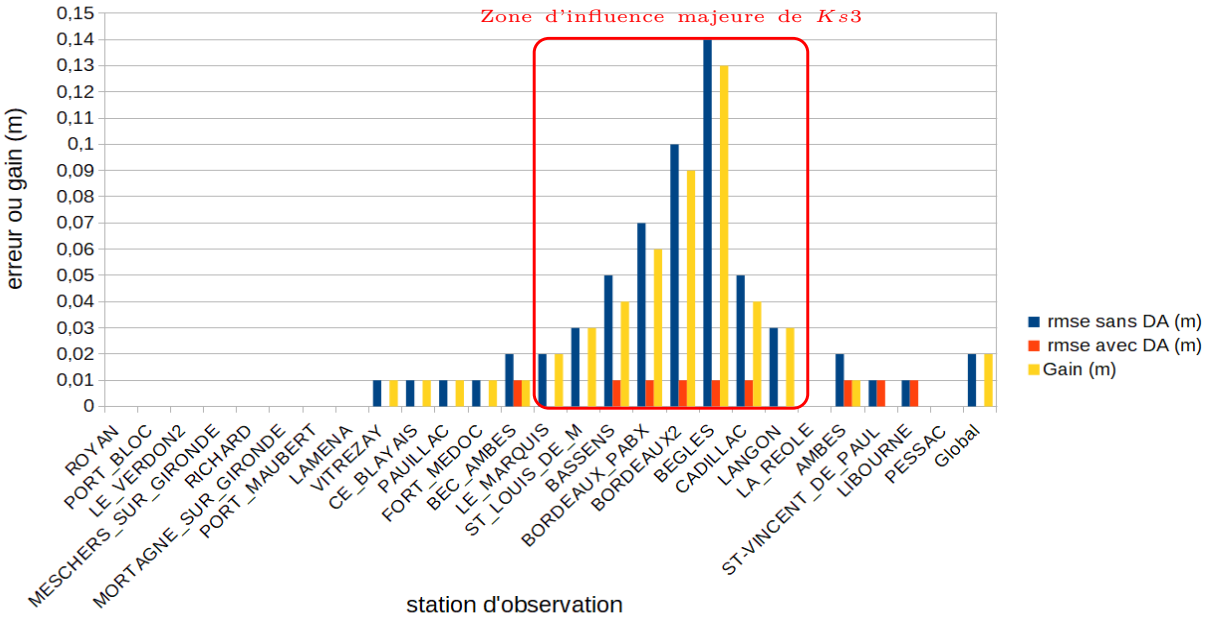



(a) 1-a

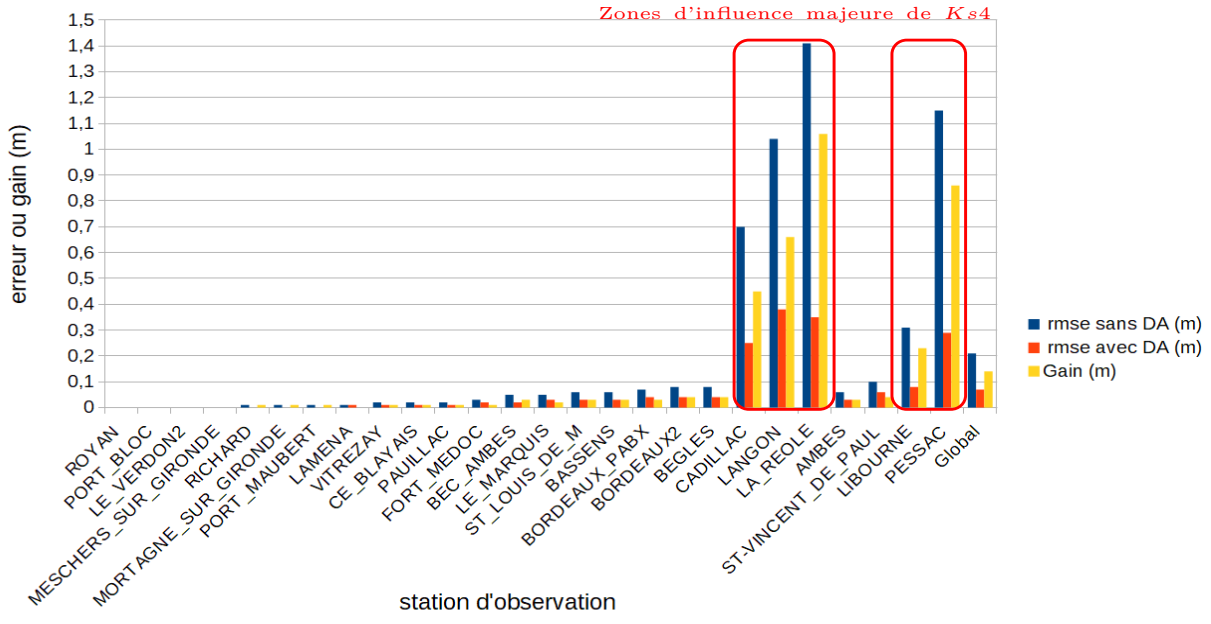



(b) 1-b

FIGURE 5.17 : Erreur quadratique moyenne (RMSE) sur la durée de l'Exp. 1 (24 h) le long de l'estuaire de la Gironde. (début)

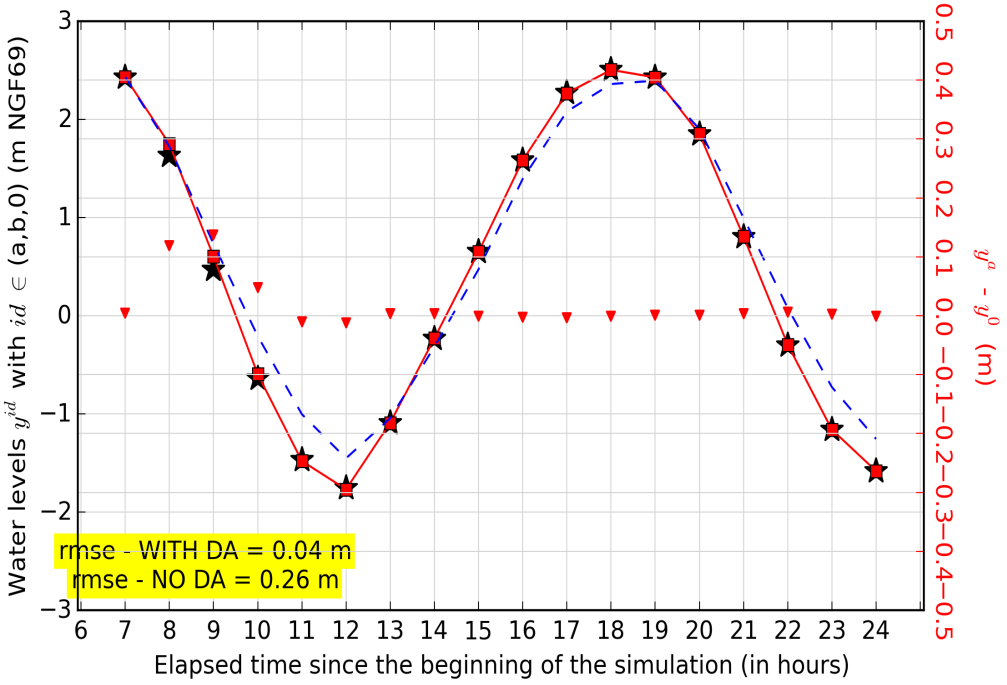



(c) 1-c

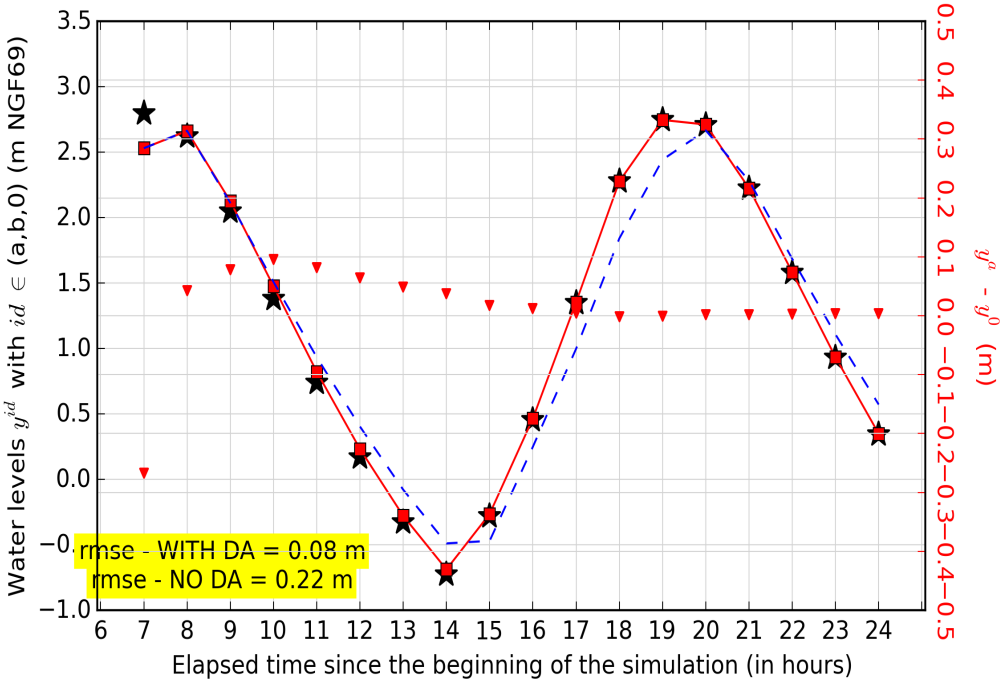



(d) 1-d

FIGURE 5.17 : Erreur quadratique moyenne (RMSE) sur la durée de l'Exp. 1 (24 h) le long de l'estuaire de la Gironde. (fin)

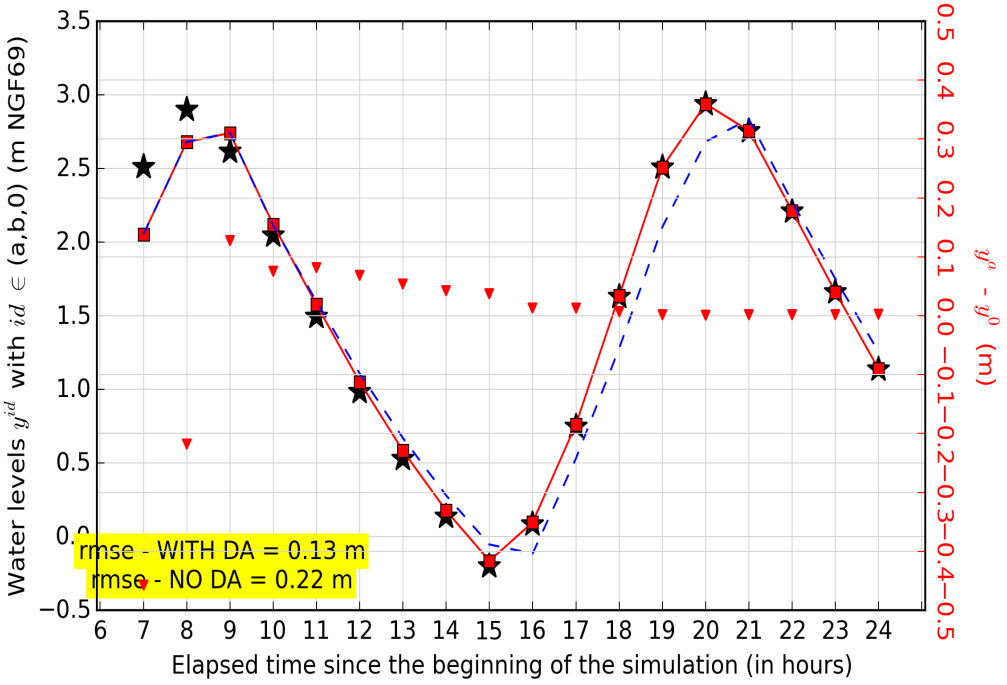



(a) Le Verdon (observations assimilées)

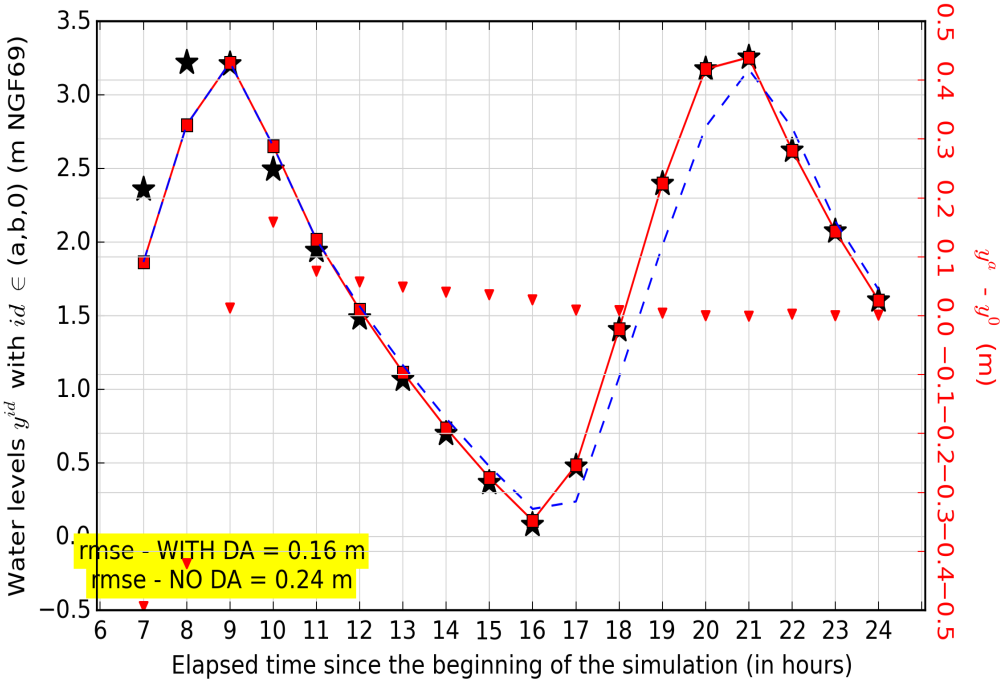



(b) Pauillac

FIGURE 5.18 : Évolution des niveaux d'eau pour l'Exp. 1-a au cours des cycles d'assimilation (L=1h - S=1h) (début)

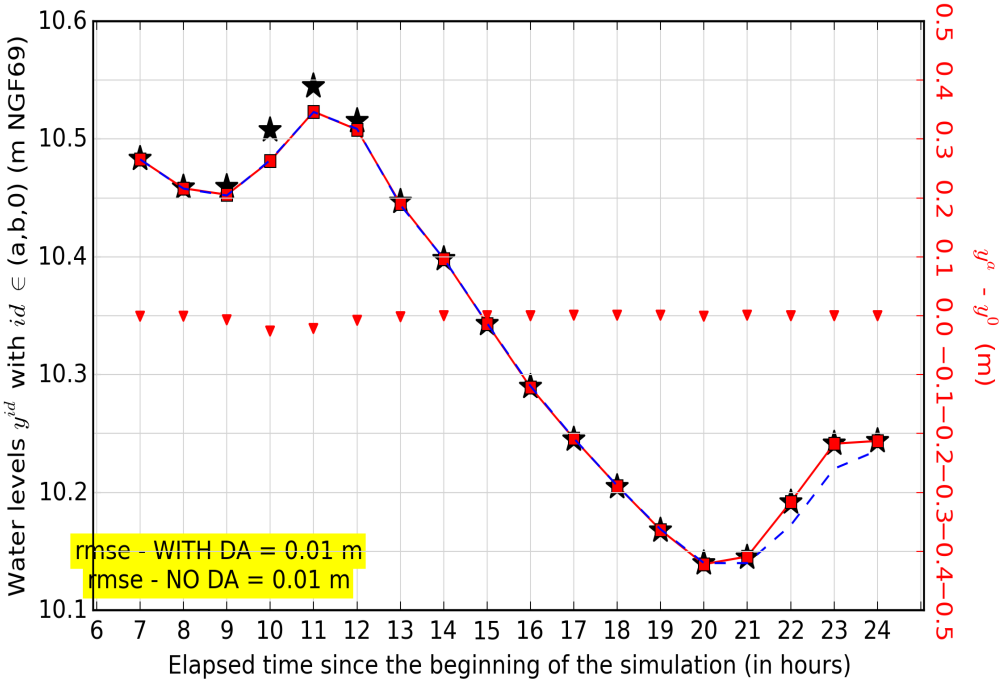



(c) Bec d'Ambès

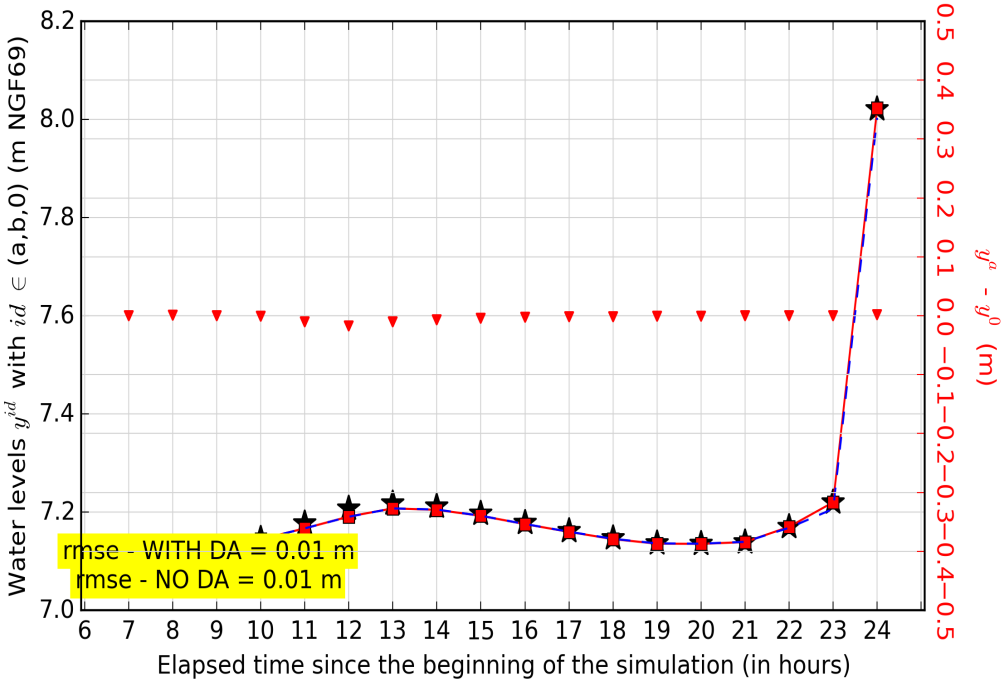



(d) Bordeaux PABX

FIGURE 5.18 : Évolution des niveaux d'eau pour l'Exp. 1-a au cours des cycles d'assimilation (L=1h - S=1h) (suite)

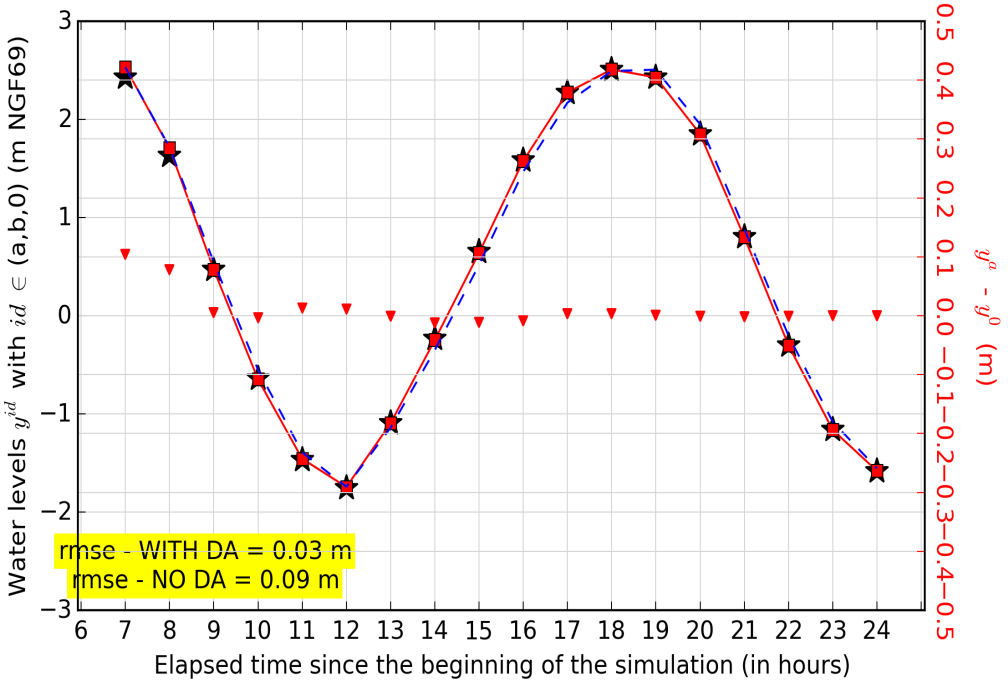



(e) La Réole

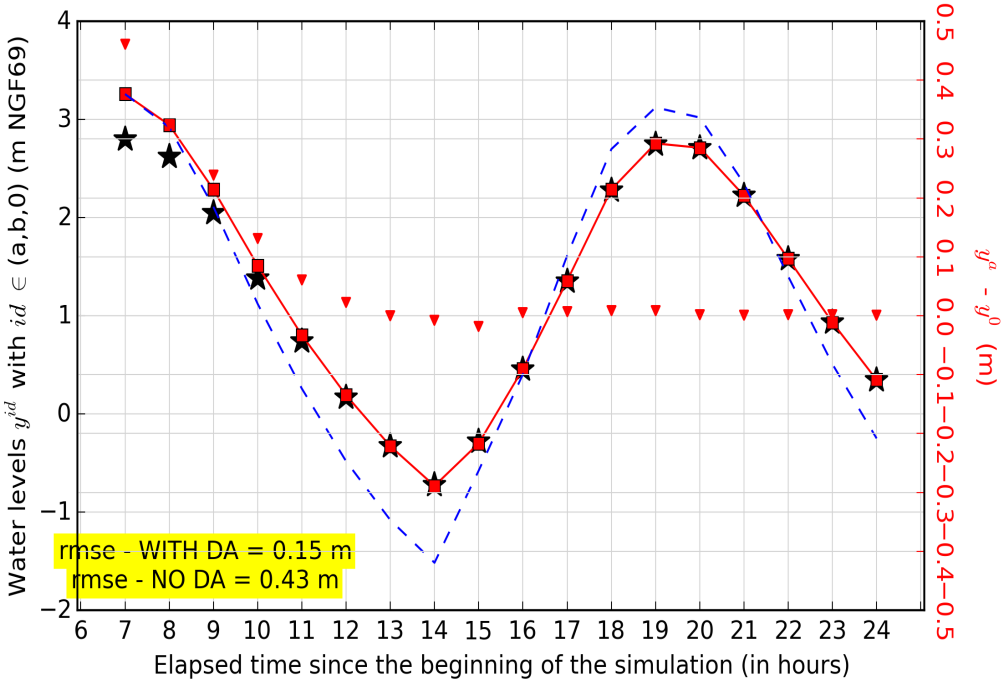



(f) Pessac

FIGURE 5.18 : Évolution des niveaux d'eau pour l'Exp. 1-a au cours des cycles d'assimilation (L=1h - S=1h) (fin)

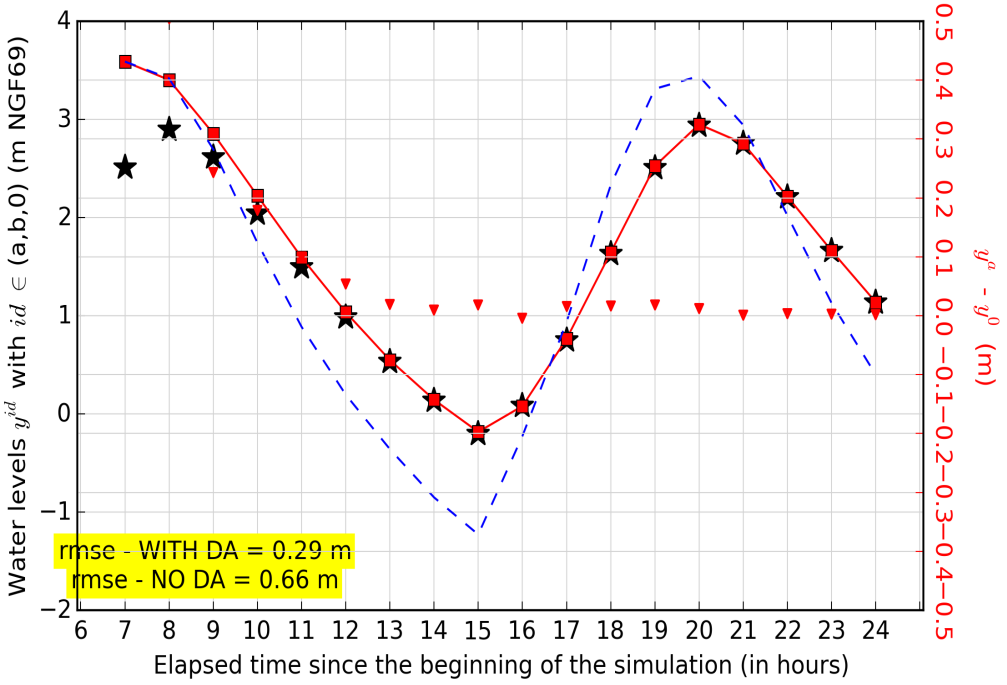



(a) Le Verdon

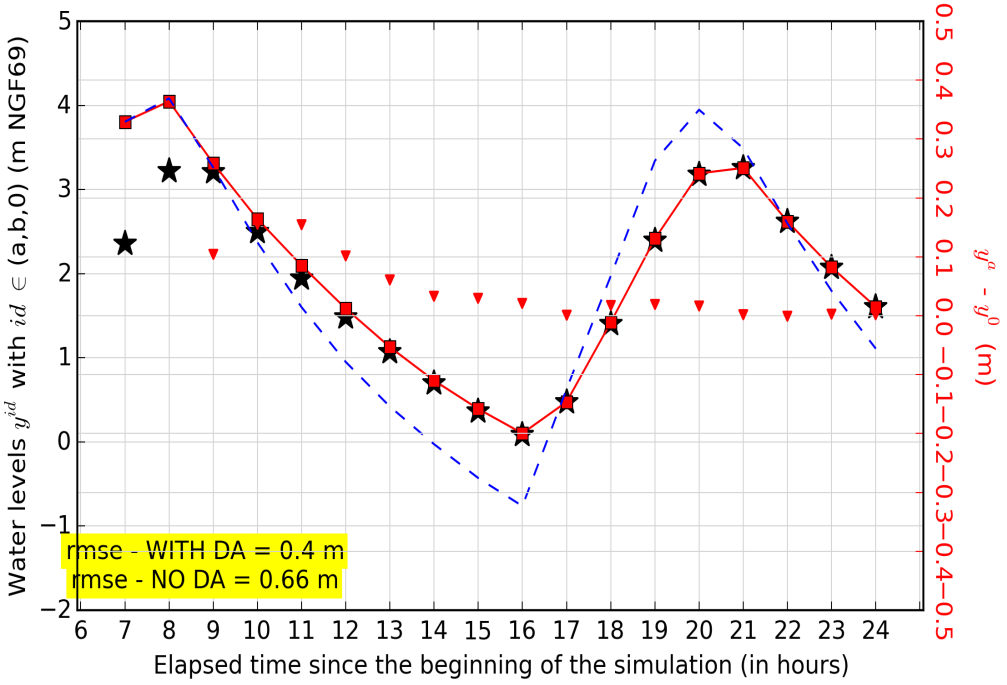



(b) Pauillac (observations assimilées)

FIGURE 5.19 : Évolution des niveaux d'eau pour l'Exp. 1-b au cours des cycles d'assimilation (L=1h - S=1h) (début)

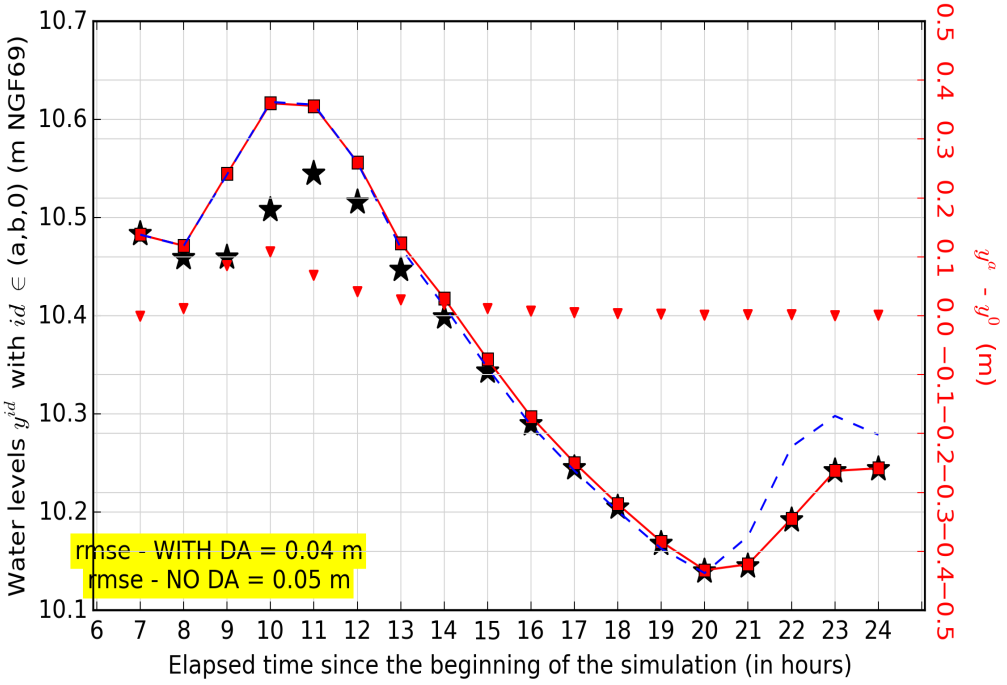



(c) Bec d'Ambès

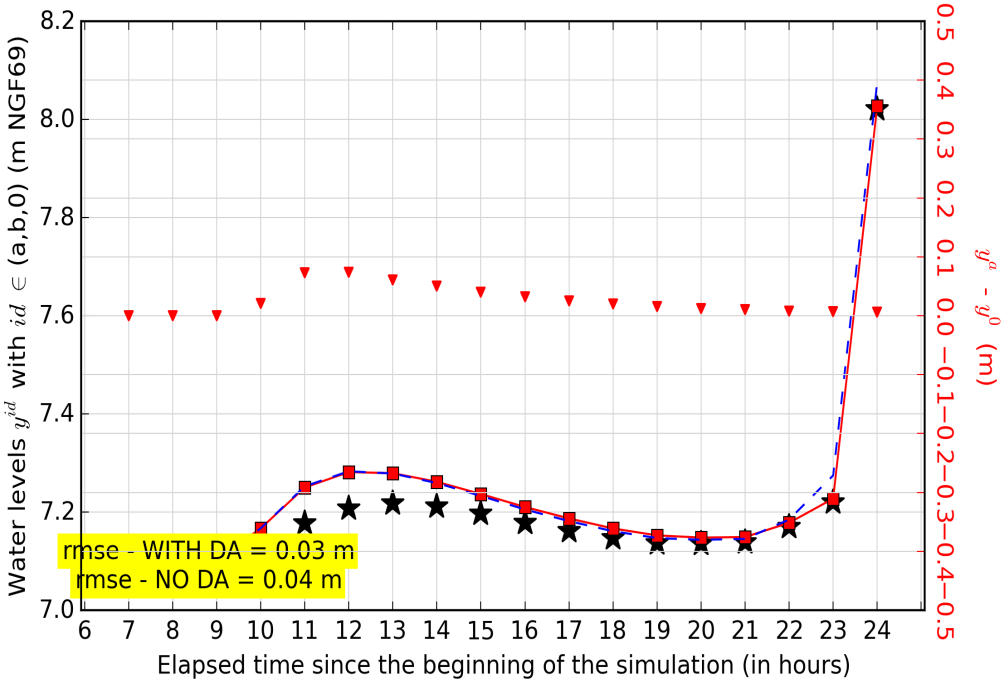



(d) Bordeaux PABX

FIGURE 5.19 : Évolution des niveaux d'eau pour l'Exp. 1-b au cours des cycles d'assimilation (L=1h - S=1h) (suite)

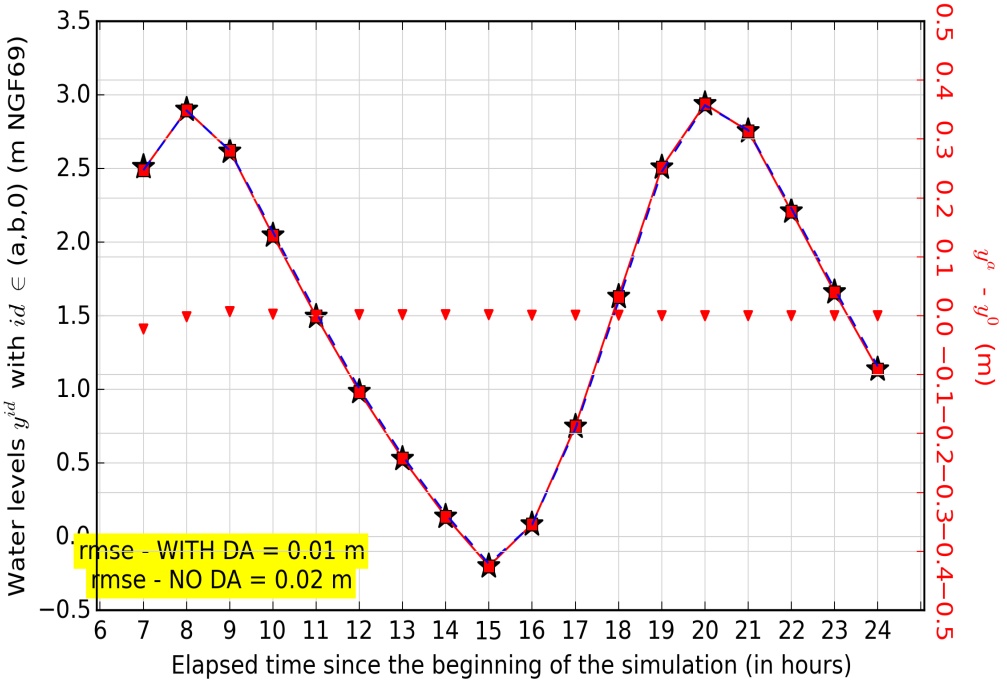



(e) La Réole

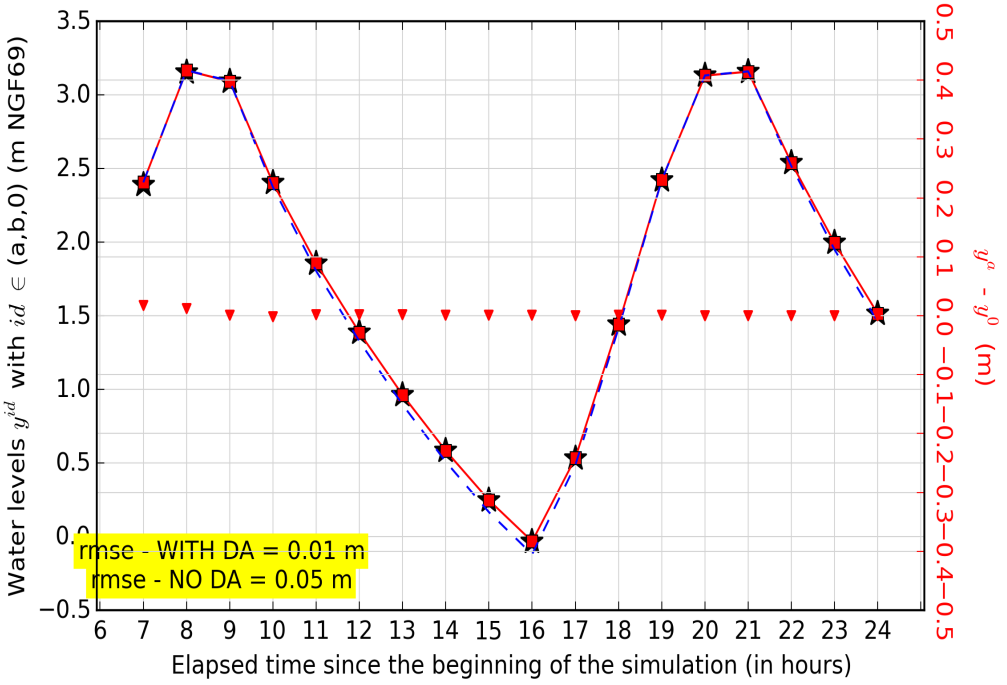



(f) Pessac

FIGURE 5.19 : Évolution des niveaux d'eau pour l'Exp. 1-b au cours des cycles d'assimilation (L=1h - S=1h) (fin)

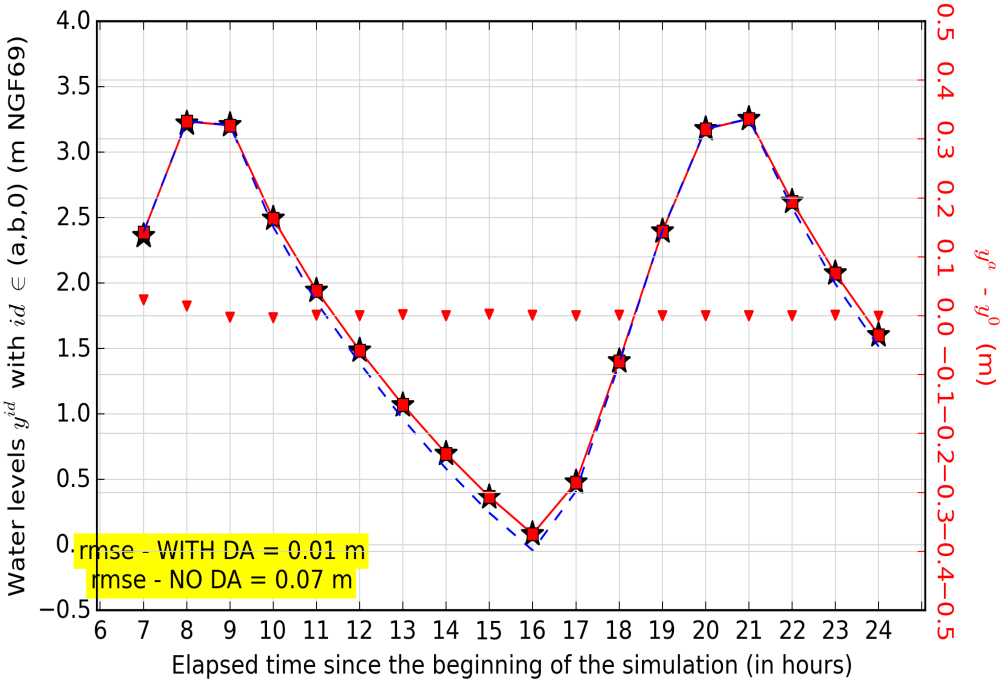



(a) Bec d'Ambès

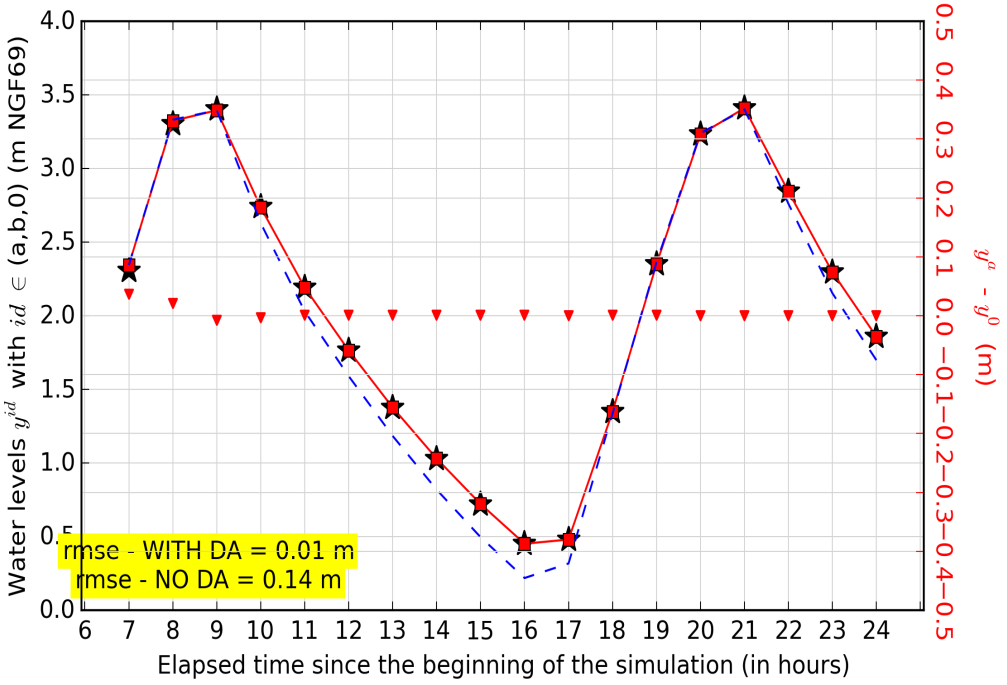



(b) Bassens

FIGURE 5.20 : Évolution des niveaux d'eau pour l'Exp. 1-c au cours des cycles d'assimilation (L=1h - S=1h) (début)

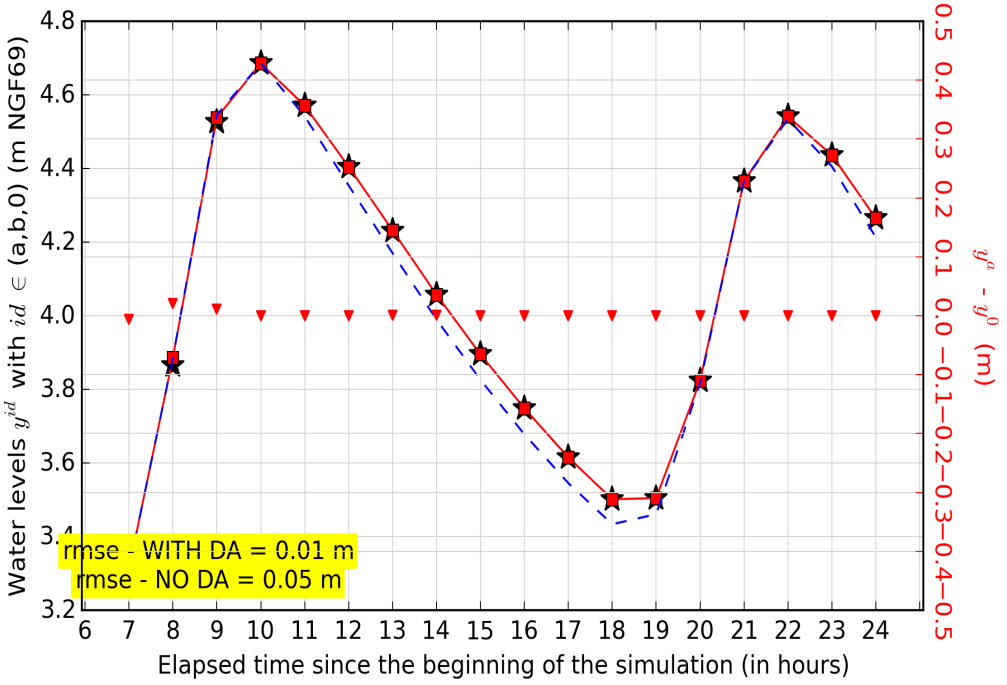



(c) Bordeaux PABX (observations assimilées)

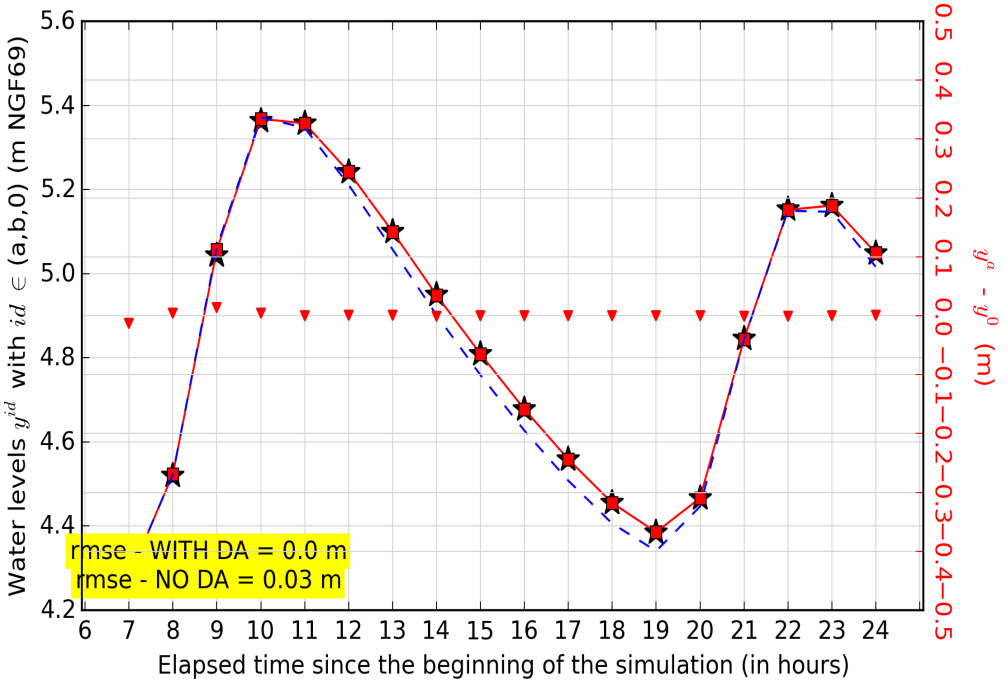



(d) Bègles

FIGURE 5.20 : Évolution des niveaux d'eau pour l'Exp. 1-c au cours des cycles d'assimilation (L=1h - S=1h) (suite)

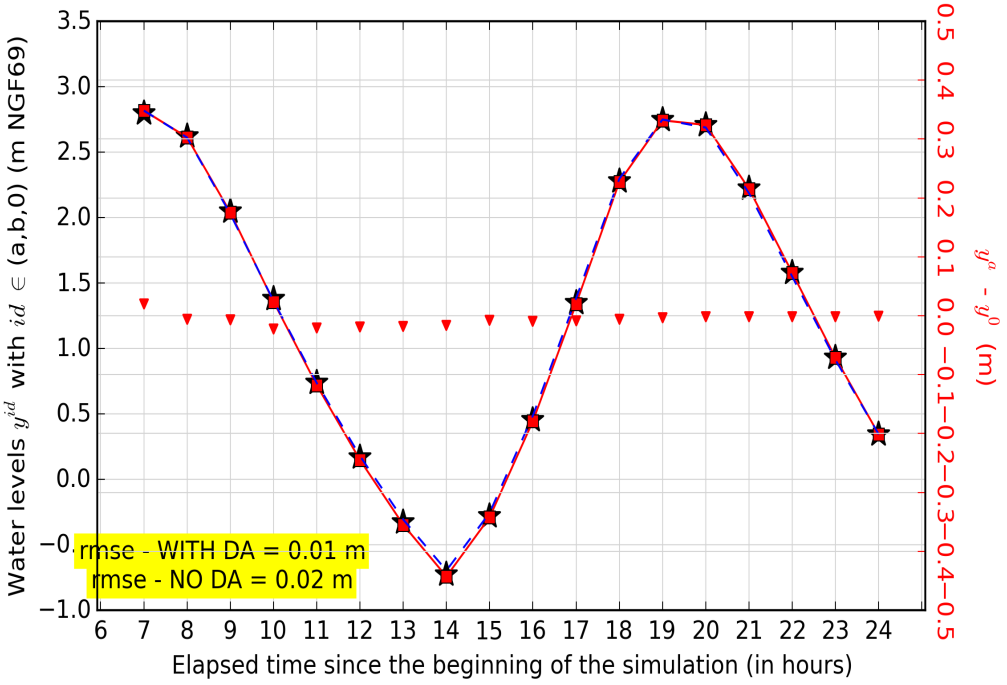



(e) Cadillac

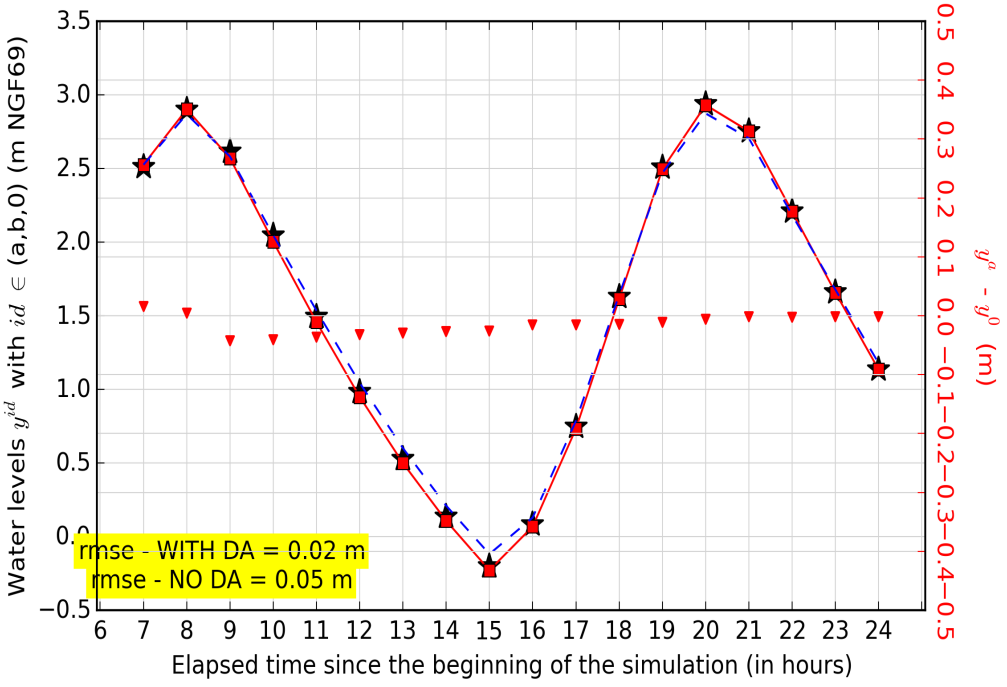



(f) Langon

FIGURE 5.20 : Évolution des niveaux d'eau pour l'Exp. 1-c au cours des cycles d'assimilation (L=1h - S=1h) (fin)

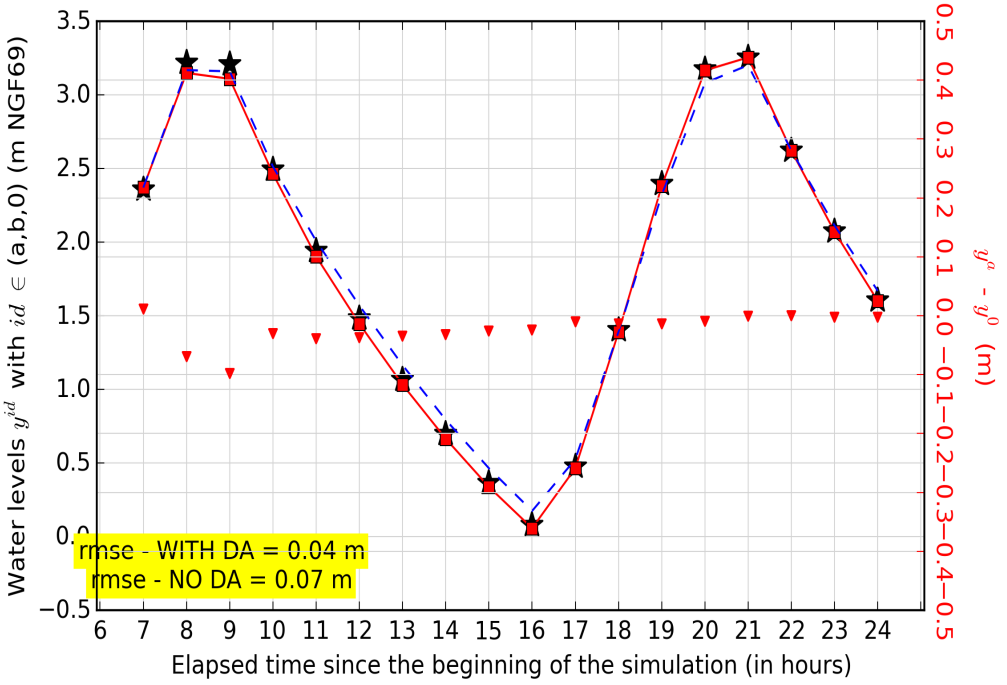



(a) Pauillac

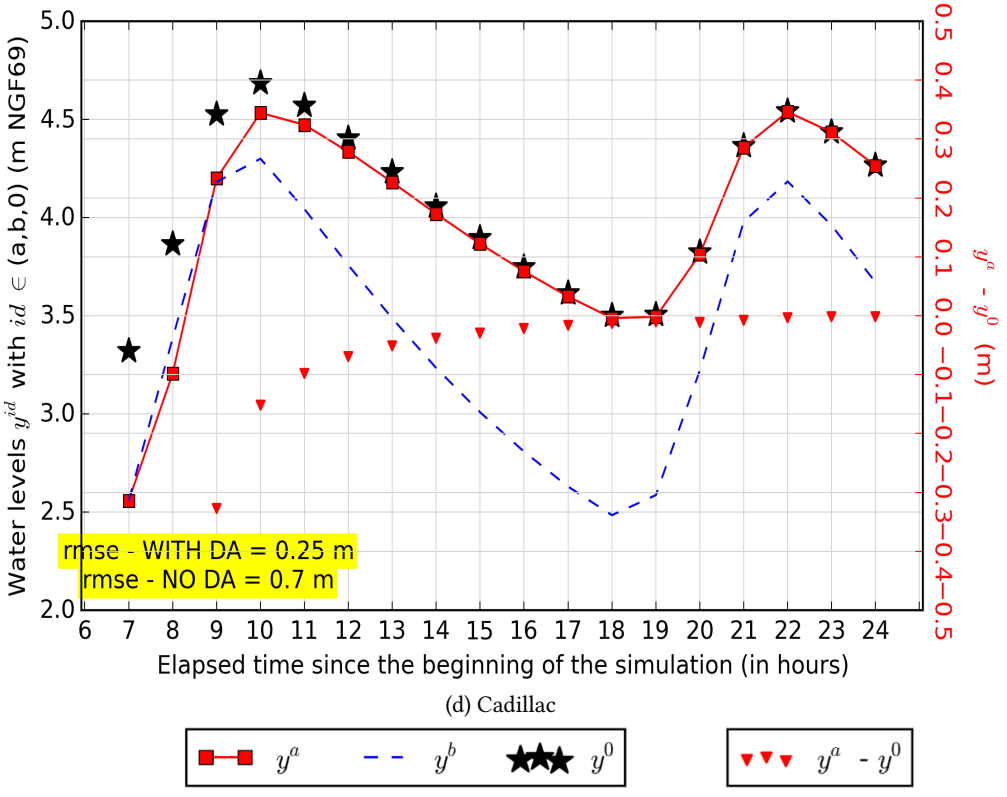



(b) Bec d'Ambès

FIGURE 5.21 : Évolution des niveaux d'eau pour l'Exp. 1-d au cours des cycles d'assimilation (L=1h - S=1h) (début)

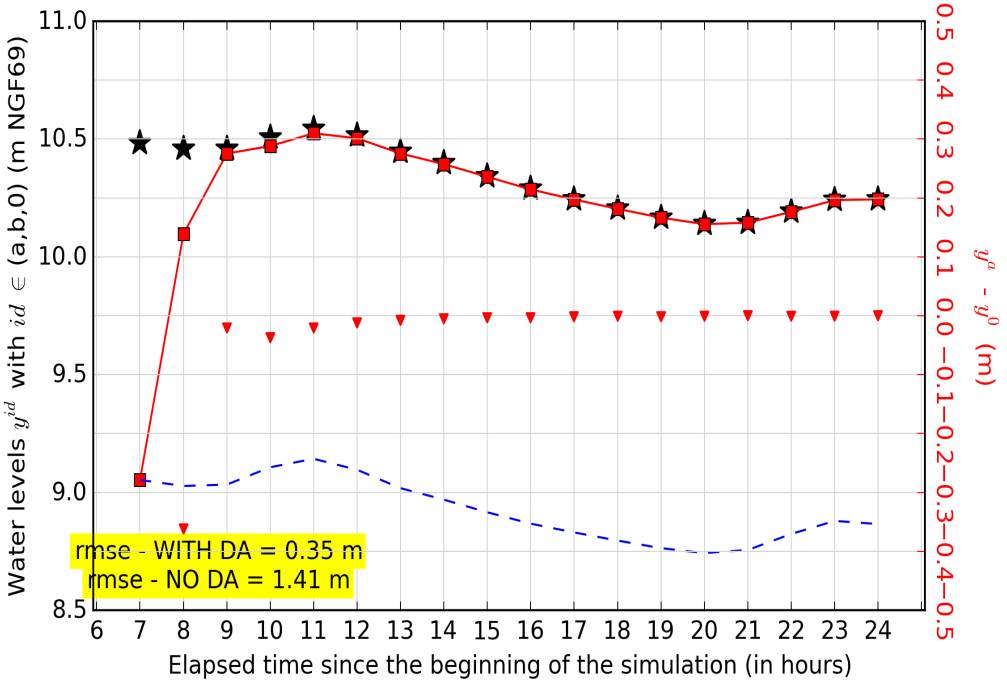



(c) Bordeaux PABX

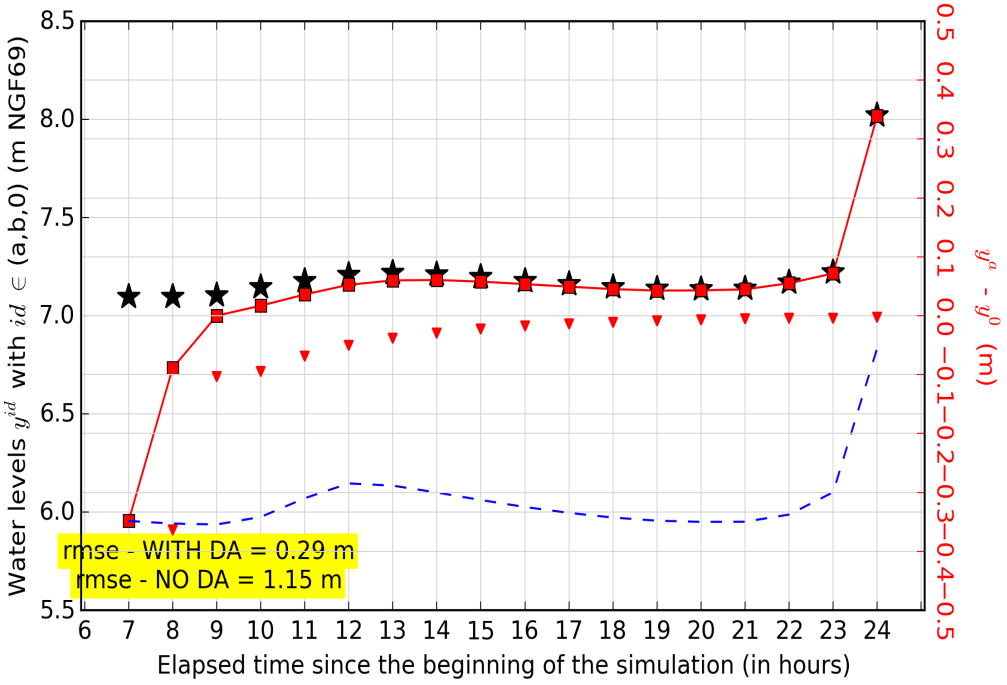



(d) Cadillac

FIGURE 5.21 : Évolution des niveaux d'eau pour l'Exp. 1-d au cours des cycles d'assimilation (L=1h - S=1h) (suite)

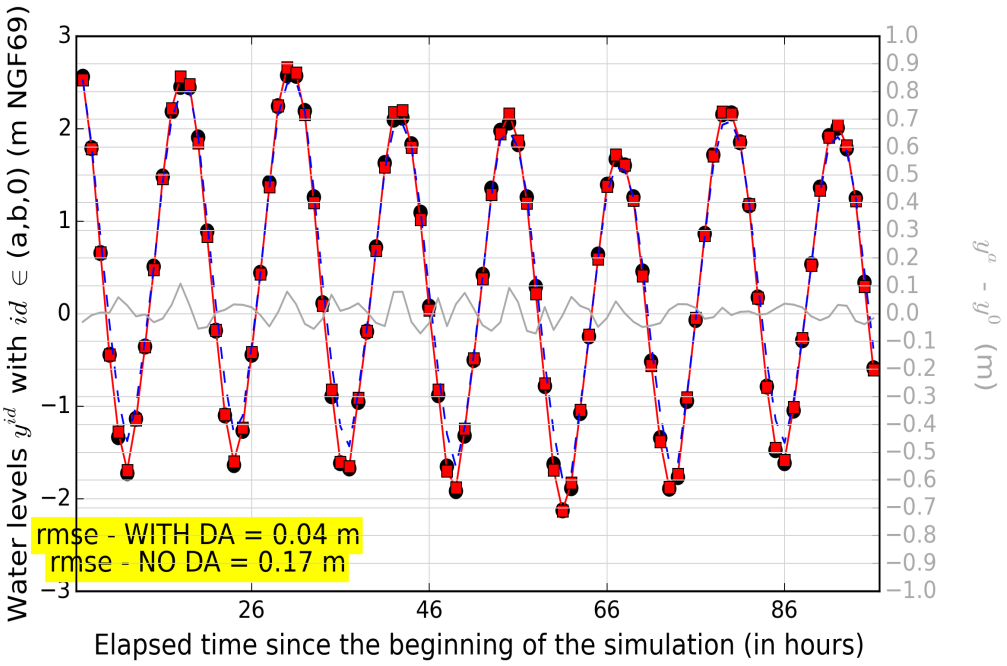



(e) La Réole (observations assimilées)

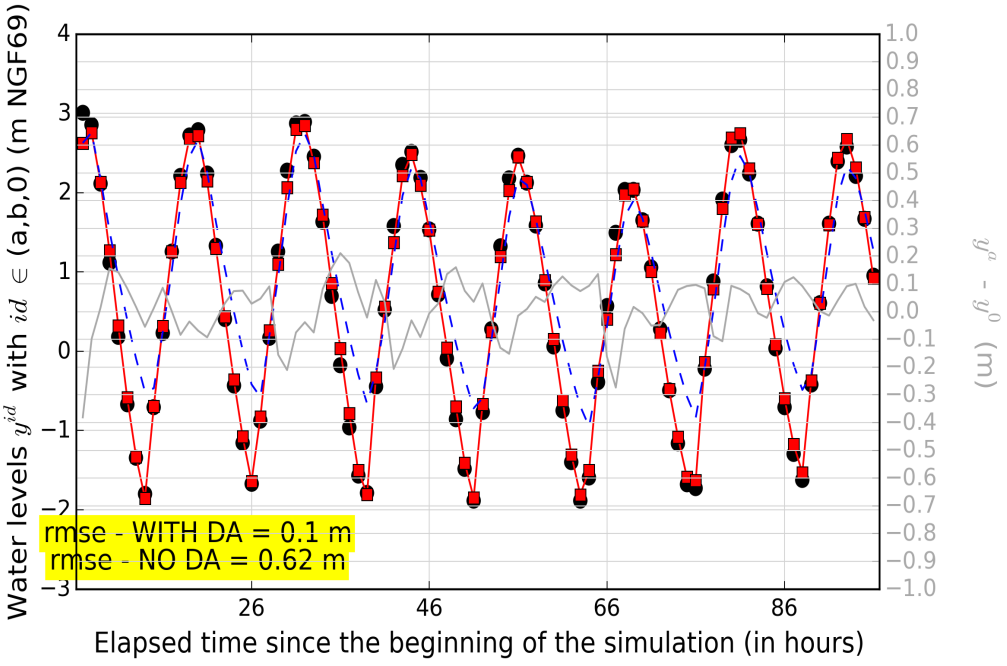



(f) Pessac

FIGURE 5.21 : Évolution des niveaux d'eau pour l'Exp. 1-d au cours des cycles d'assimilation (L=1h - S=1h) (fin)




(a) Le Verdon




(b) Pauillac

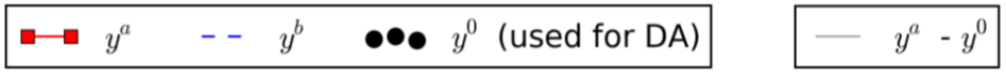



FIGURE 5.22 : Évolution des niveaux d'eau sur les stations "assimilées" pour l'Exp. 2-b (début)

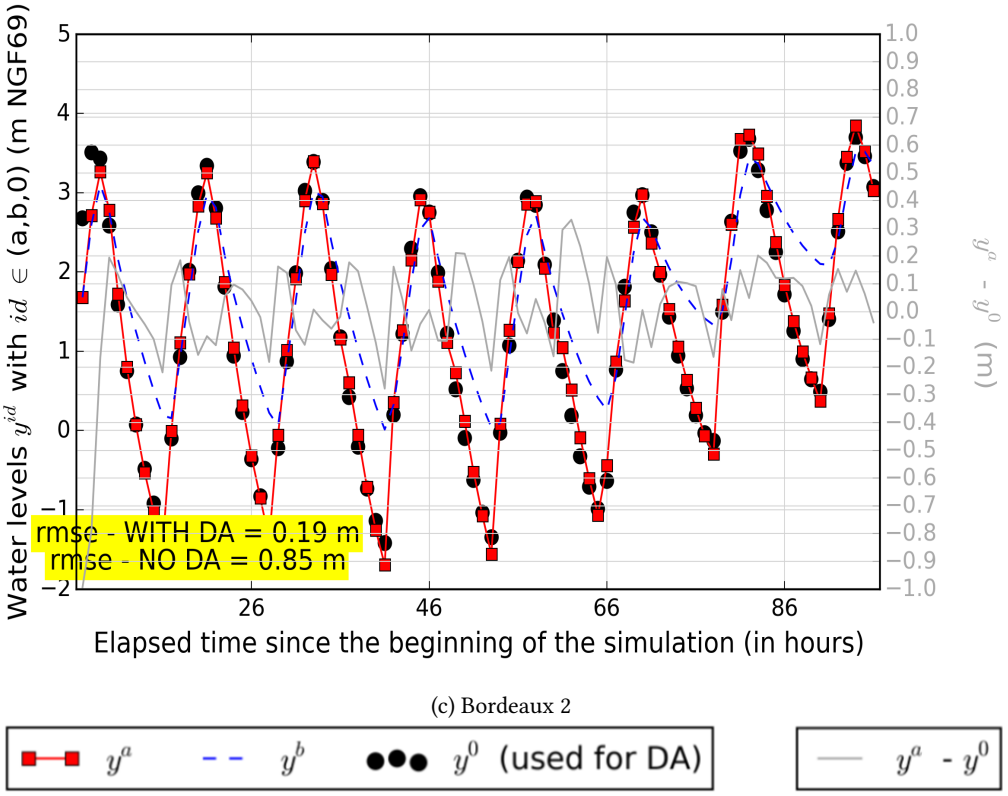
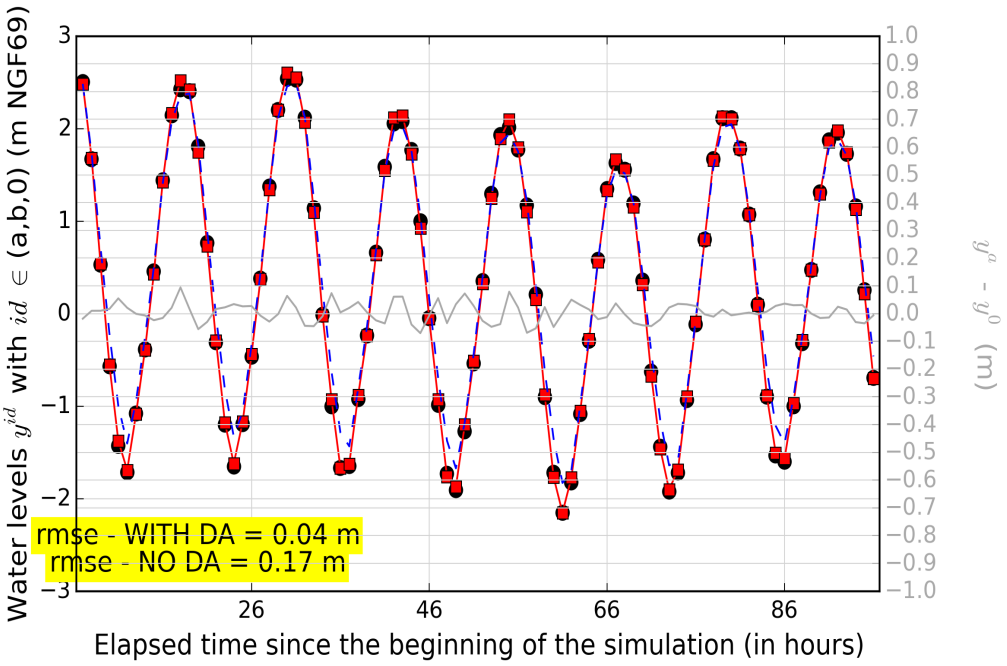
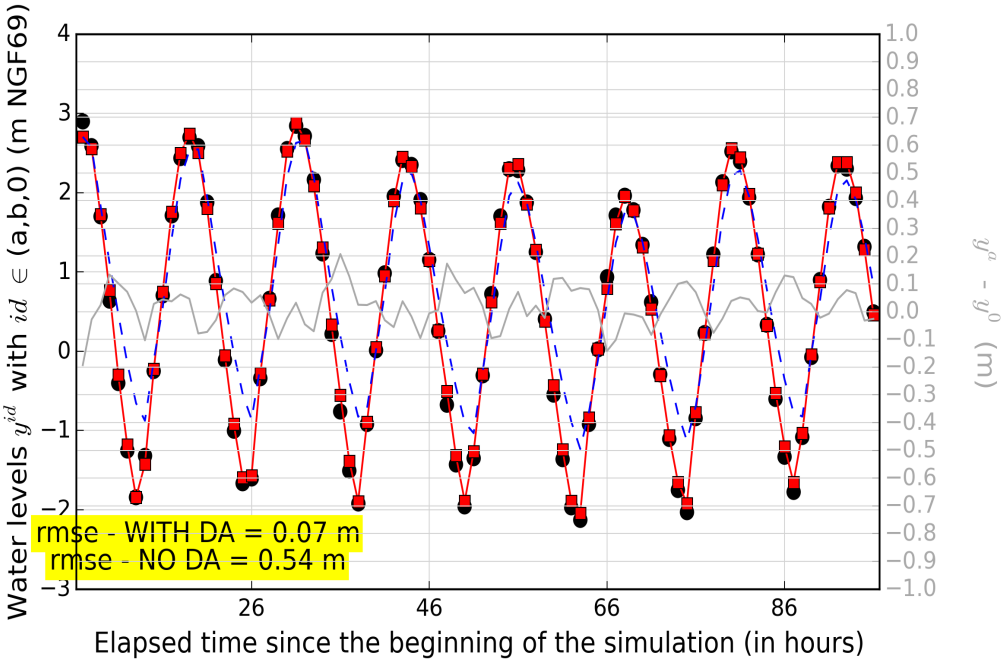



FIGURE 5.22 : Évolution des niveaux d'eau sur les stations "assimilées" pour l'Exp. 2-b (fin)




(a) Port-Bloc




(b) Laména

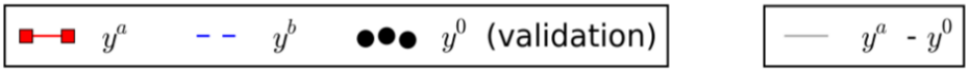
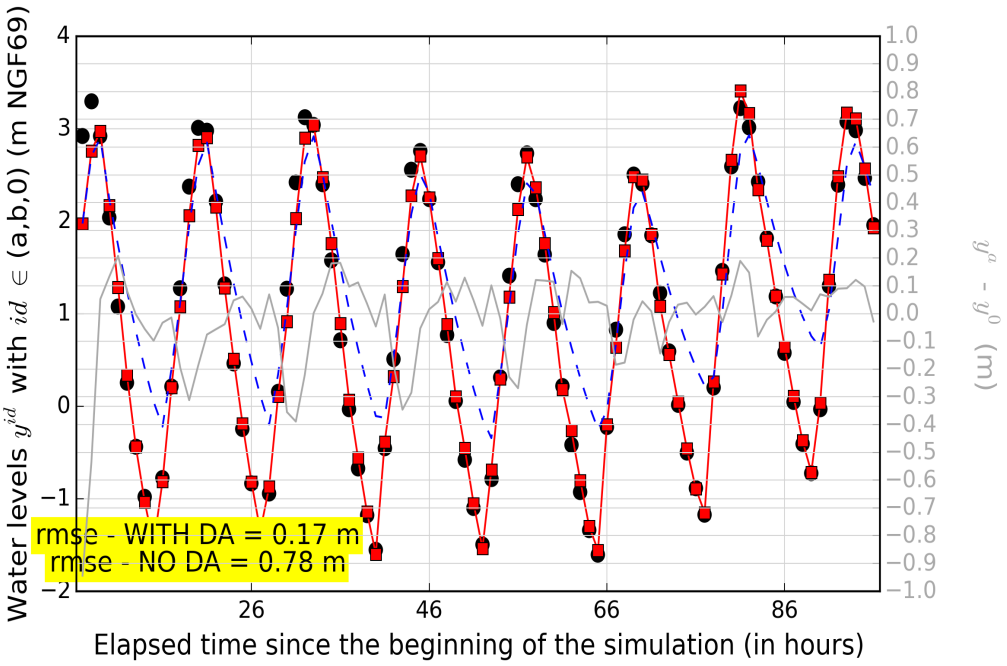
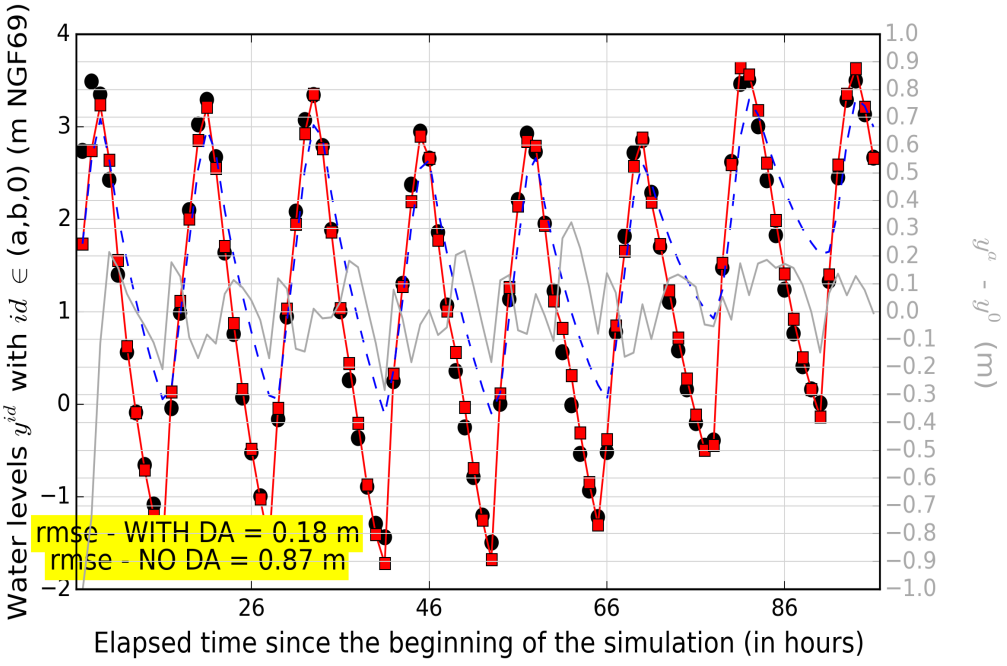



FIGURE 5.23 : Évolution des niveaux d'eau sur les stations de validation pour l'Exp. 2-b (début)




(c) Le Marquis




(d) Bordeaux PABX

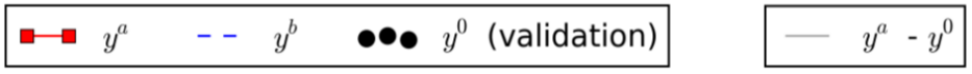



FIGURE 5.23 : Évolution des niveaux d'eau sur les stations de validation pour l'Exp. 2-b (fin)

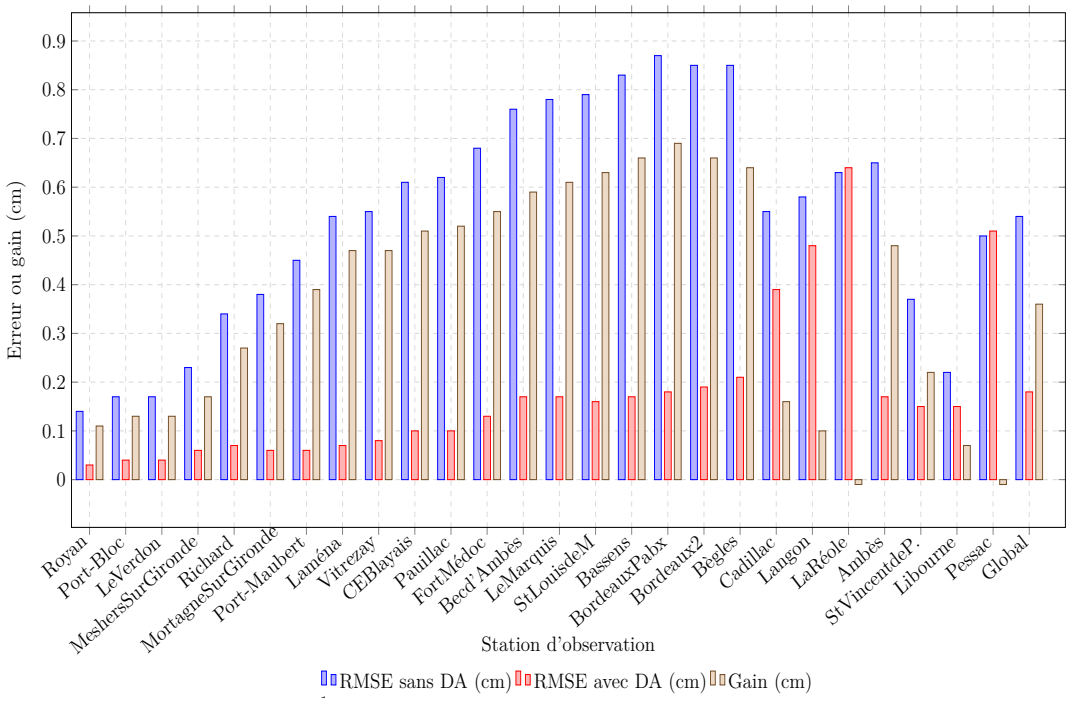



FIGURE 5.24 : Erreur quadratique moyenne le long de l'estuaire de la Gironde pour l'Exp 2-b. Les stations dont les observations synthétiques issues de $True2c$ sont assimilées sont : Le Verdon, Pauillac et Bordeaux PABx.

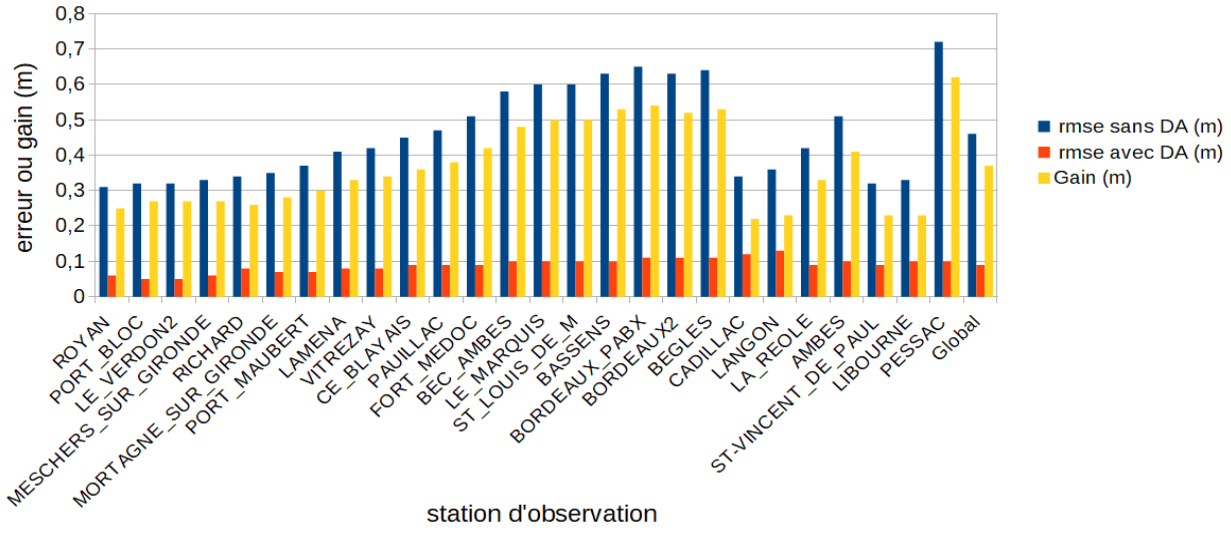
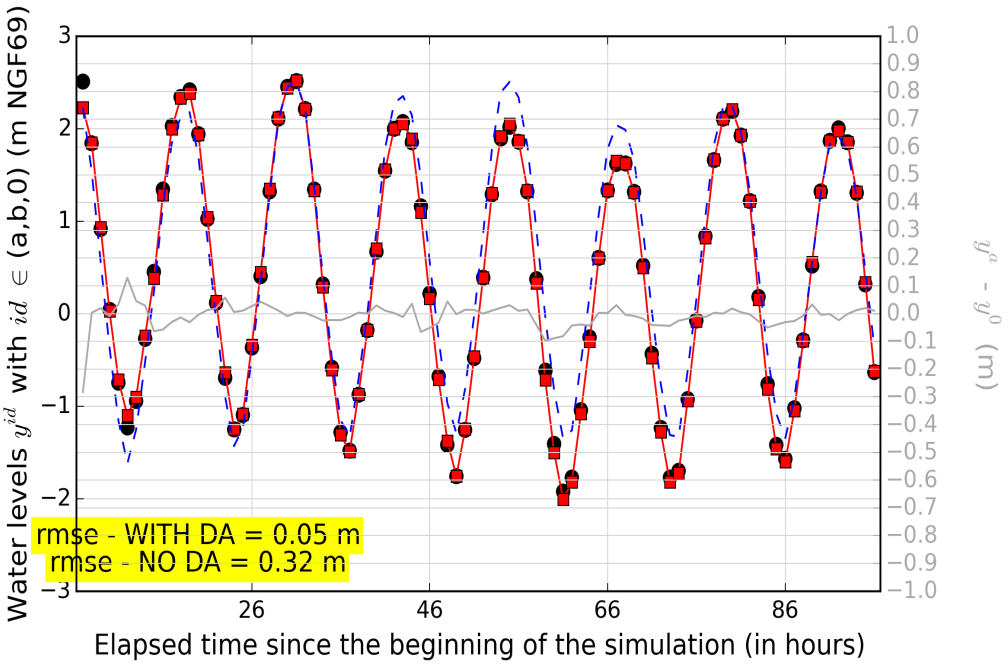
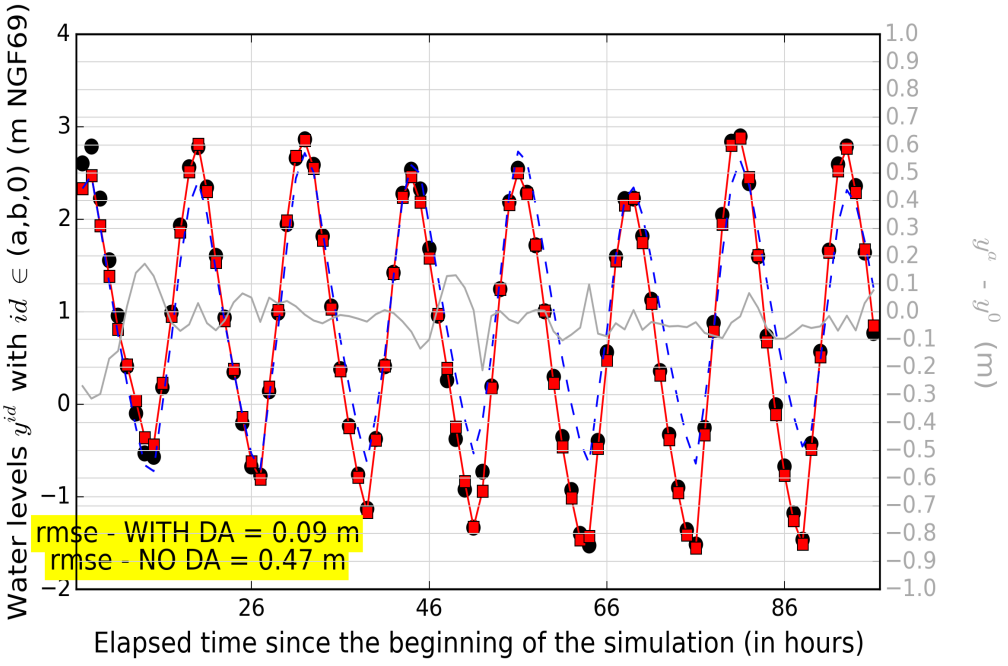



FIGURE 5.25 : Erreur quadratique moyenne le long de l'estuaire de la Gironde pour l'Exp 3-b




(a) Le Verdon




(b) Pauillac

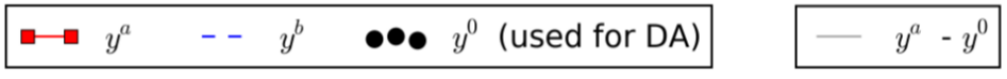
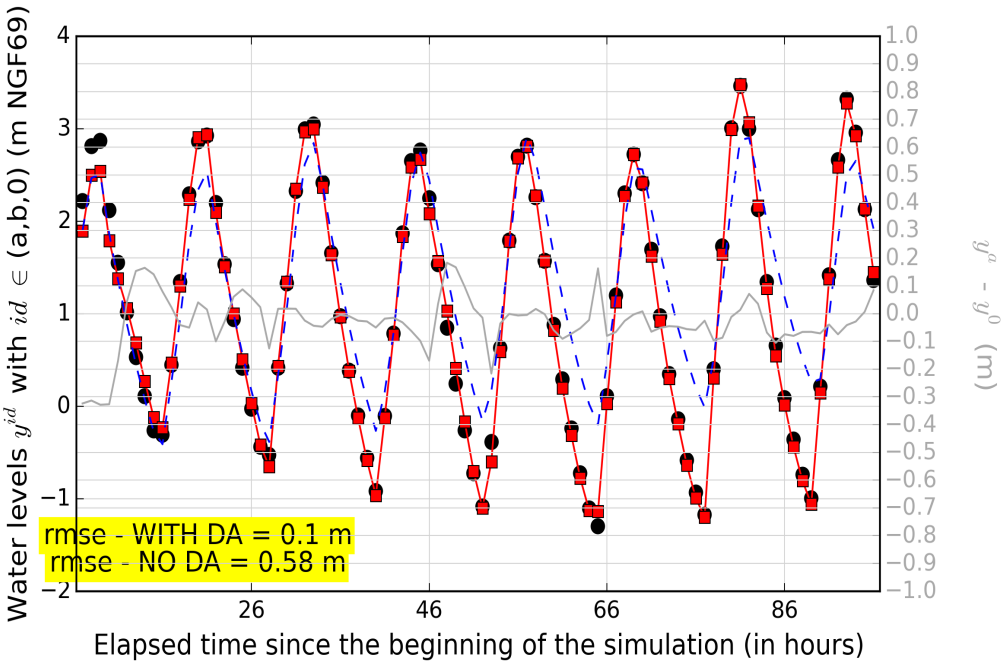
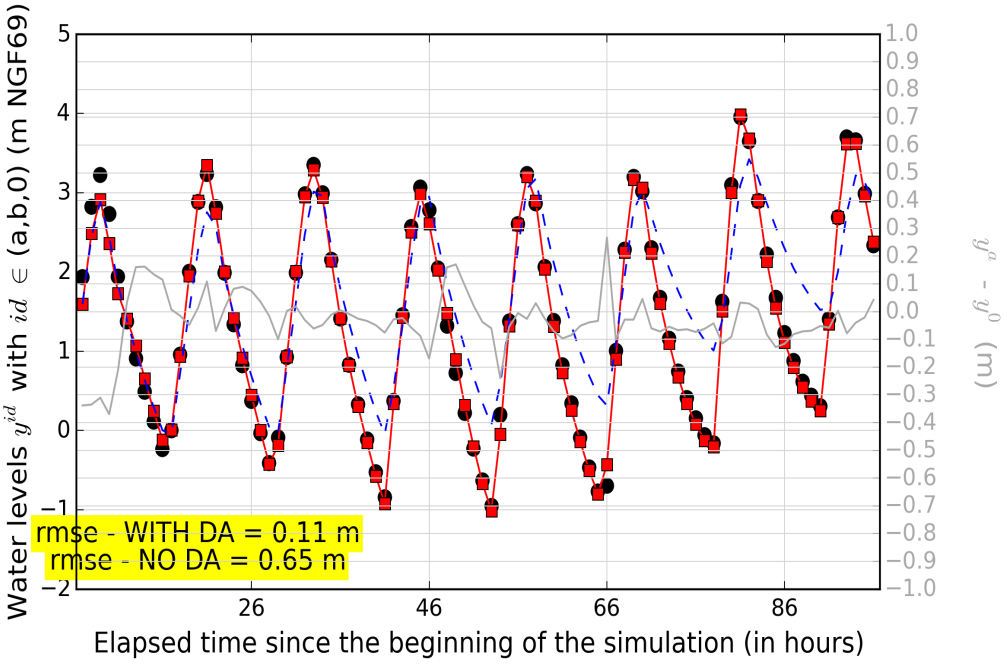



FIGURE 5.26 : Évolution des niveaux d'eau pour l'Exp 3-b (début)




(c) Bec d'Ambès




(d) Bordeaux PABX

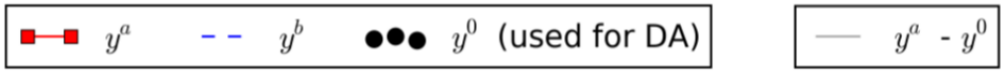
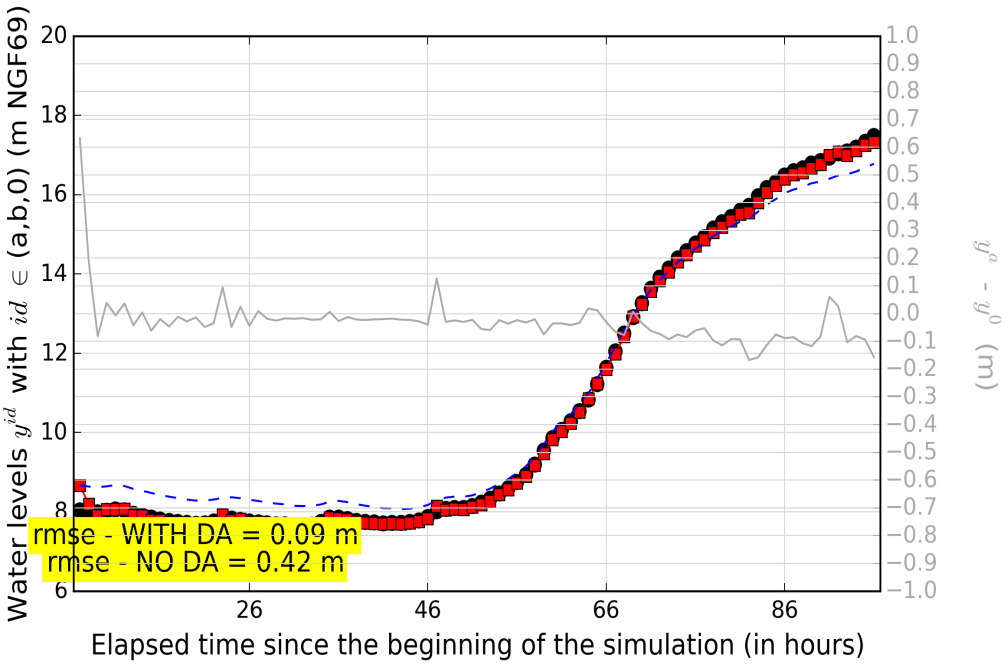
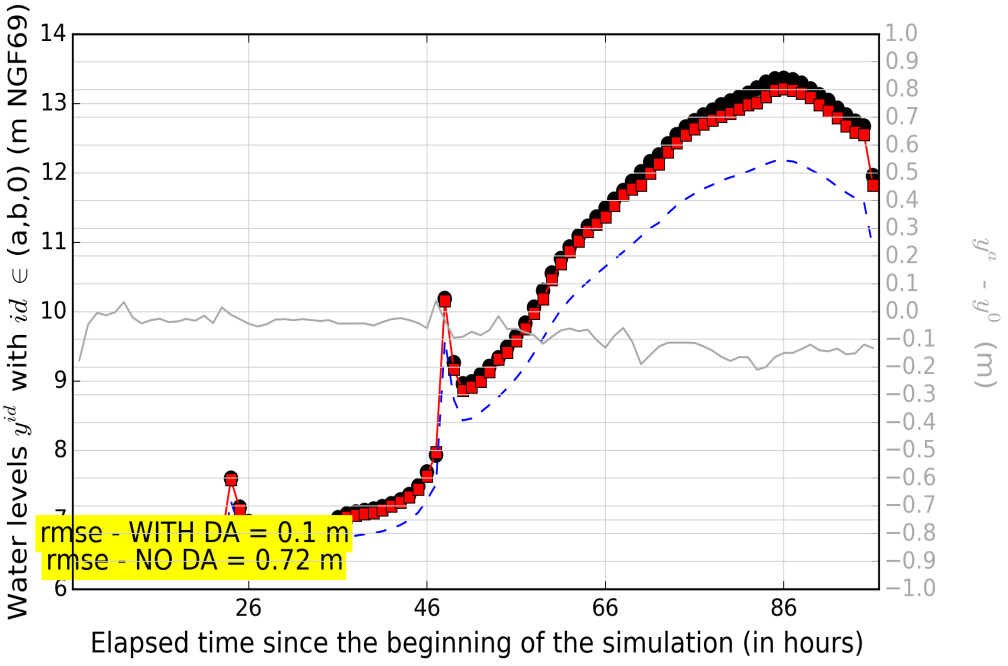



FIGURE 5.26 : Évolution des niveaux d'eau pour l'Exp 3-b (suite)




(e) La Réole




(f) Pessac

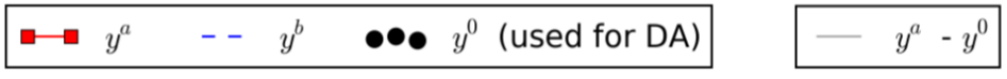
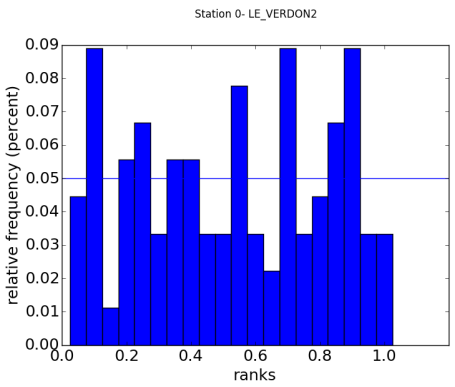
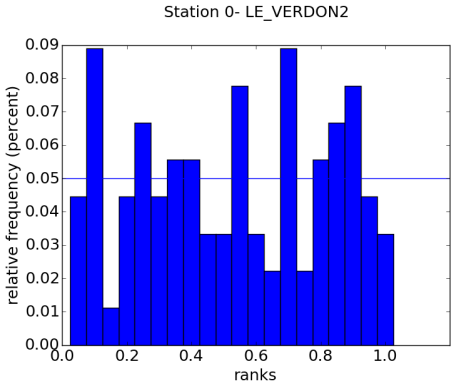


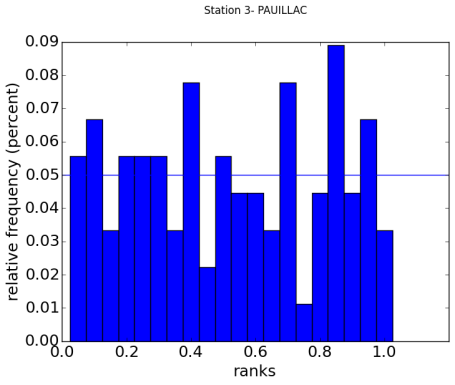

FIGURE 5.26 : Évolution des niveaux d'eau pour l'Exp 3-b (fin)



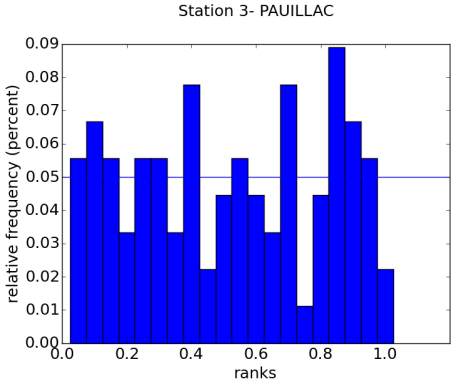

(a) Exp 2-c - Le Verdon



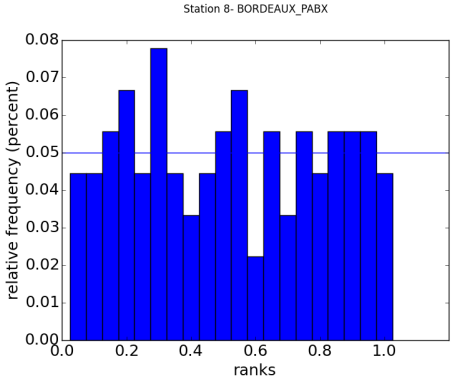

(b) Exp 2-d - Le Verdon



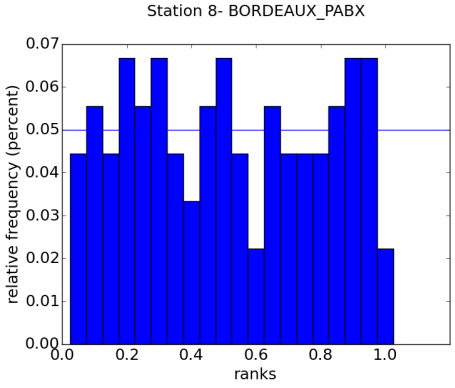

(c) Exp 2-c - Paullac




(d) Exp 2-d - Paullac

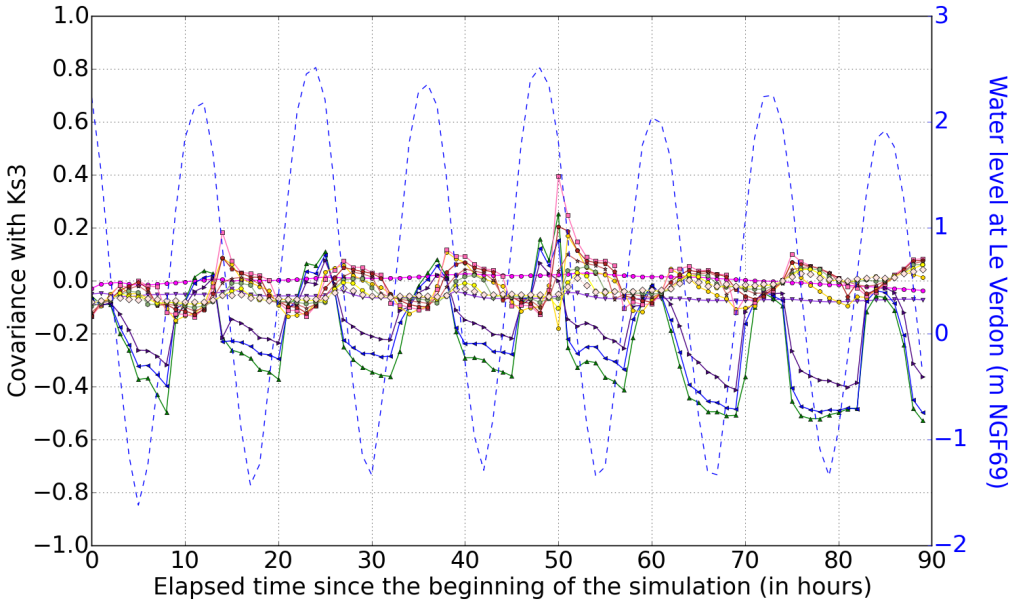



(e) Exp 2-c - Bordeaux PABx

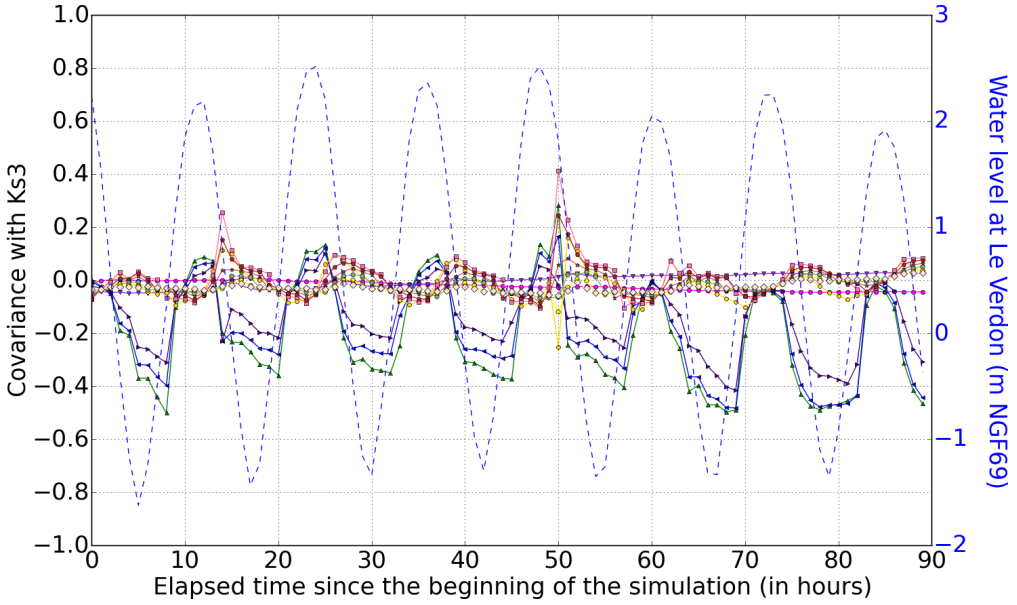



(f) Exp 2-d - Bordeaux PABx

FIGURE 5.27 : Diagramme de Talagrand après lissage en 3 stations (Le Verdon, Pauillac et Bordeaux PABX) pour les Exp. 2-c et Exp. 2-d




(a) $Ks3 - N_e = 100$




(b) $Ks3 - N_e = 200$

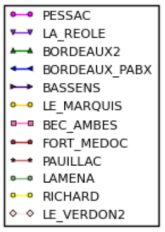
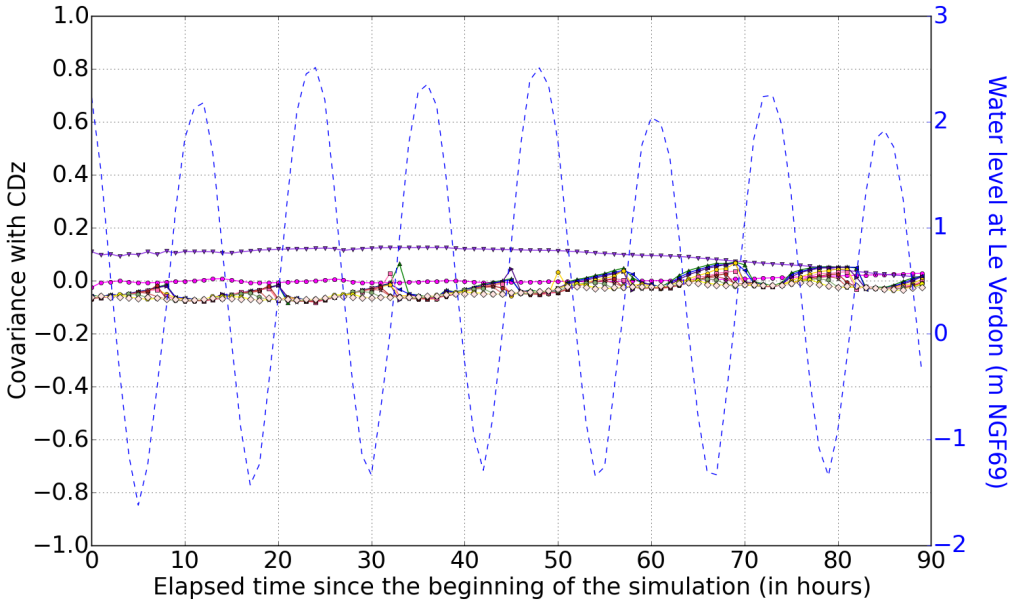
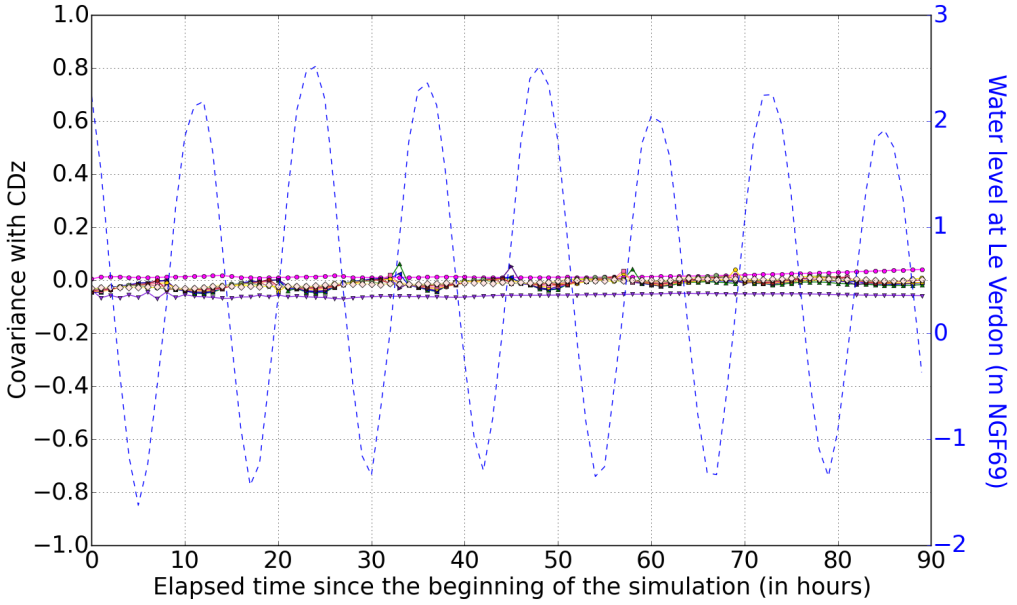



FIGURE 5.28 : Fonction de covariance de $Ks3$, $Cdz$ et $\alpha_{GAR,1}$ avec les niveaux d'eau de Telemac2D aux points de "mesures" pour l'Exp 2-c ($N_e = 100$) et l'Exp 2-d ($N_e = 200$) (début)




(c) $Cdz - N_e = 100$




(d) $Cdz - N_e = 200$

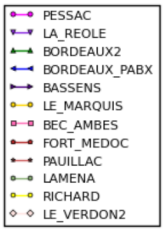
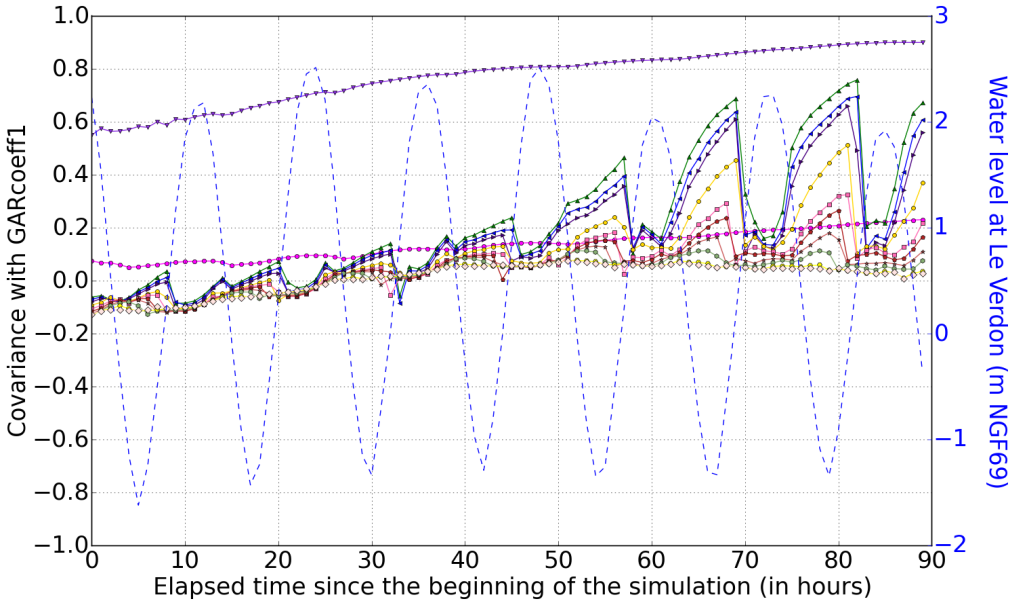
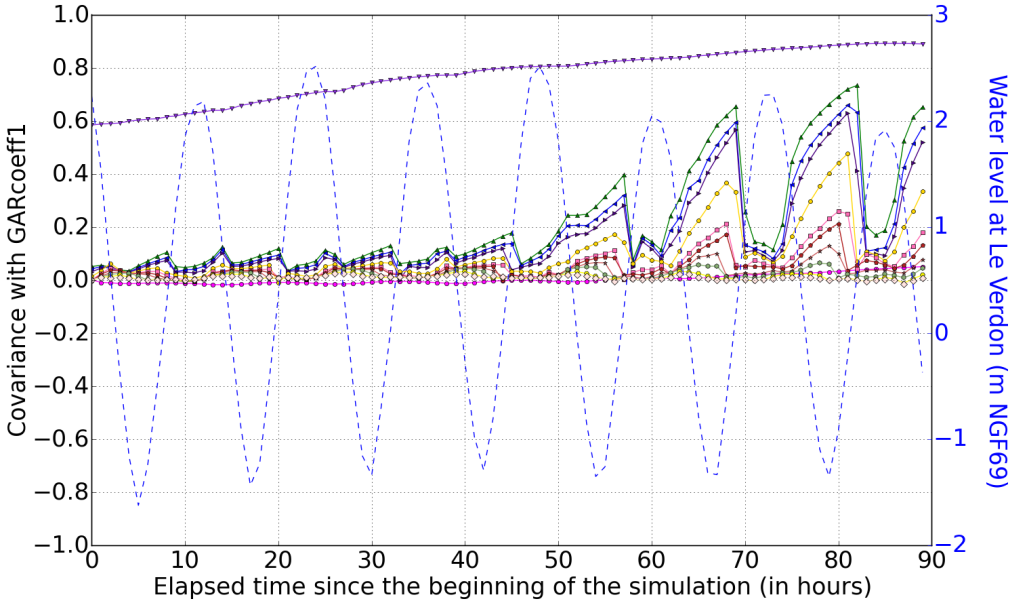



FIGURE 5.28 : Fonction de covariance de $Ks3$, $Cdz$ et $\alpha_{GAR,1}$ avec les niveaux d'eau de Telemac2D aux points de "mesures" pour l'Exp 2-c ($N_e = 100$) et l'Exp 2-d ($N_e = 200$) (suite)




(e) $\alpha_{GAR,1} - N_e = 100$




(f) $\alpha_{GAR,1} - N_e = 200$

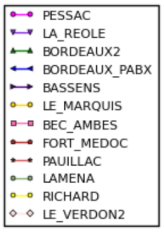
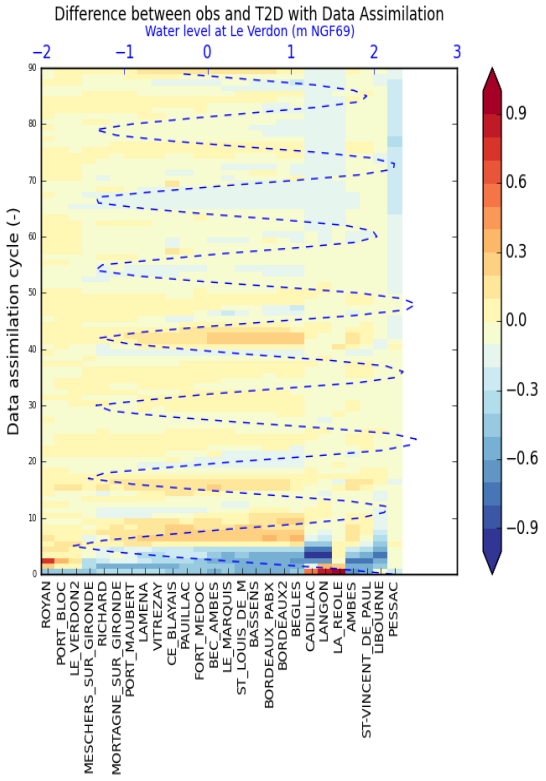
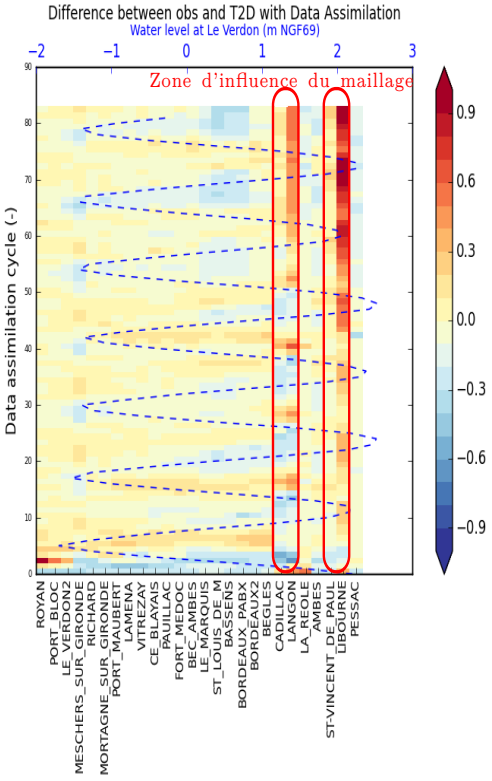


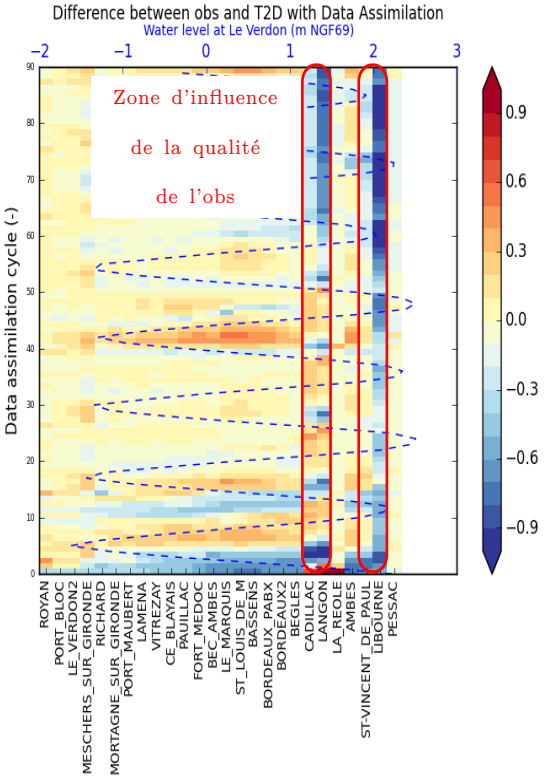

FIGURE 5.28 : Fonction de covariance de $Ks3$, $Cdz$ et $\alpha_{GAR,1}$ avec les niveaux d'eau de Telemac2D aux points de "mesures" pour l'Exp 2-c ($N_e = 100$) et l'Exp 2-d ($N_e = 200$) (fin)




(a) Exp 3-a - avec DA




(b) Exp 3-i - avec DA




(c) Exp 3-j - avec DA

FIGURE 5.29 : Différence entre niveaux d'eau calculés par Telemac2D et observations selon le raffinement du maillage utilisé pour l'assimilation ou l'observation synthétique

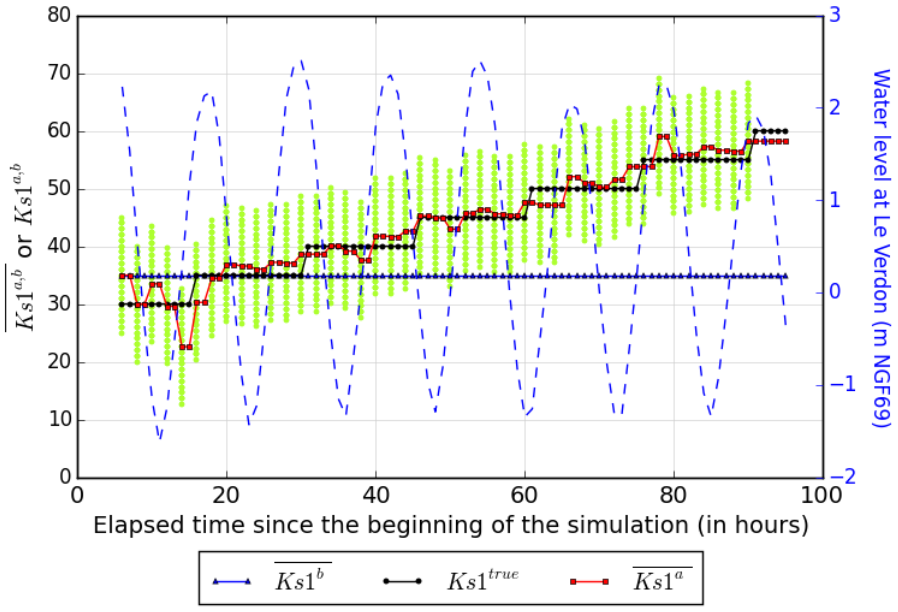
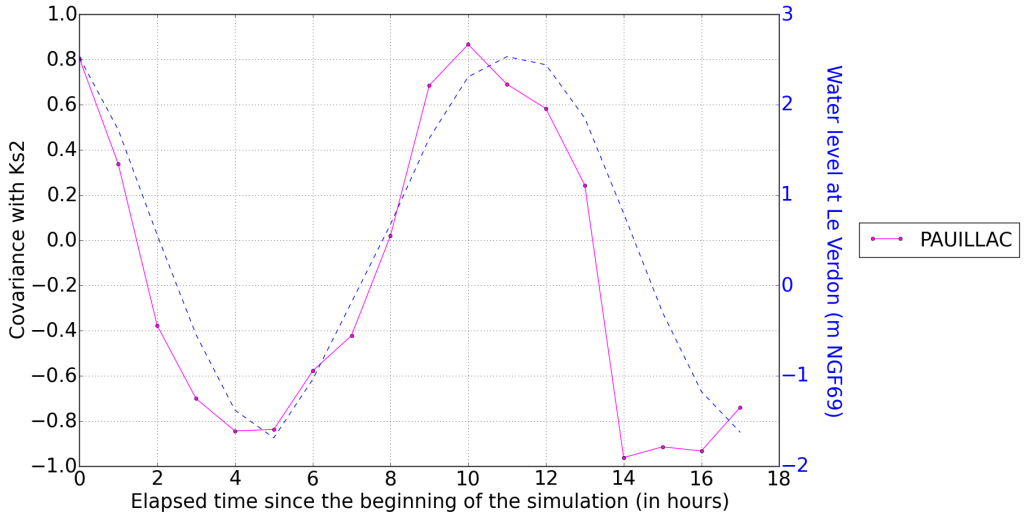
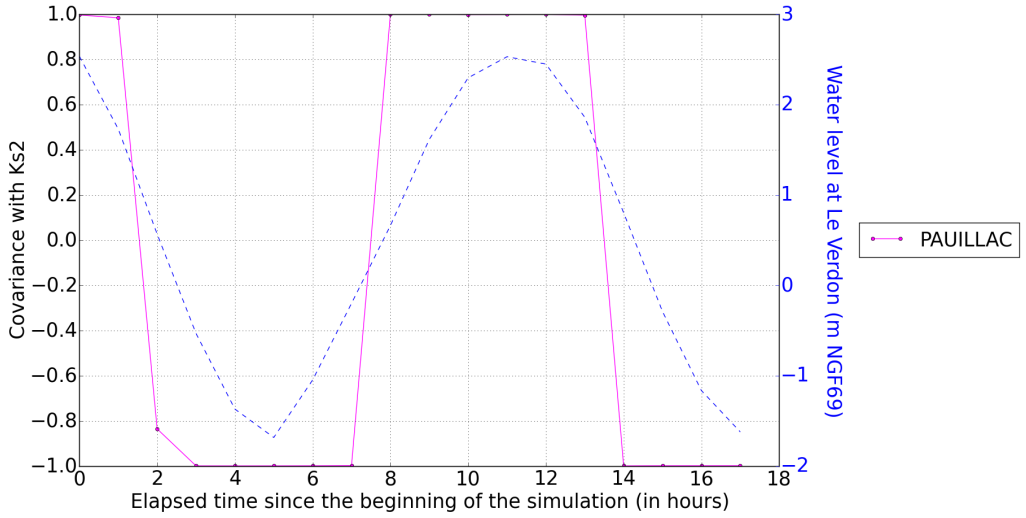



FIGURE 5.30 : Évolution de la variable de contrôle $Ks1$ pour l'Exp. 3-h ($L = 6\ h$, $S = 2\ h$)

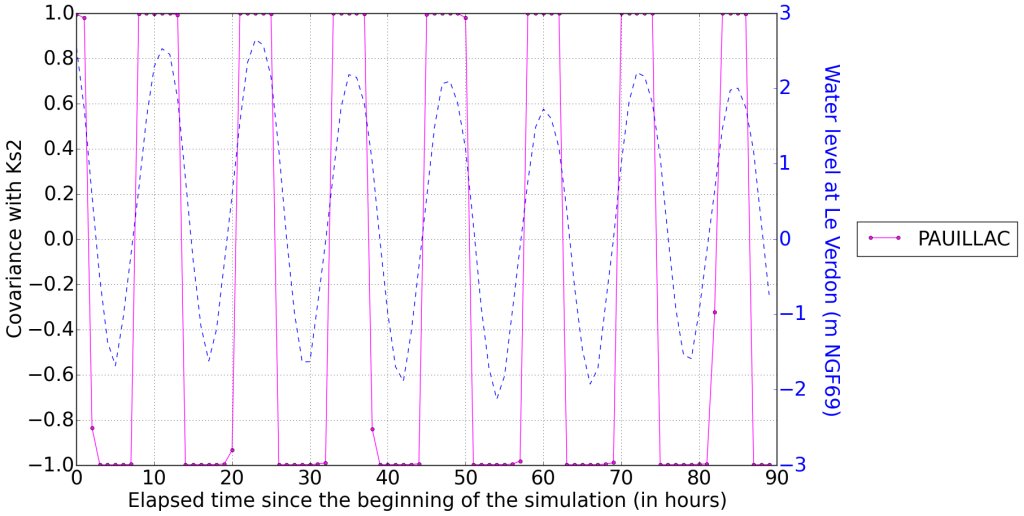



(a)

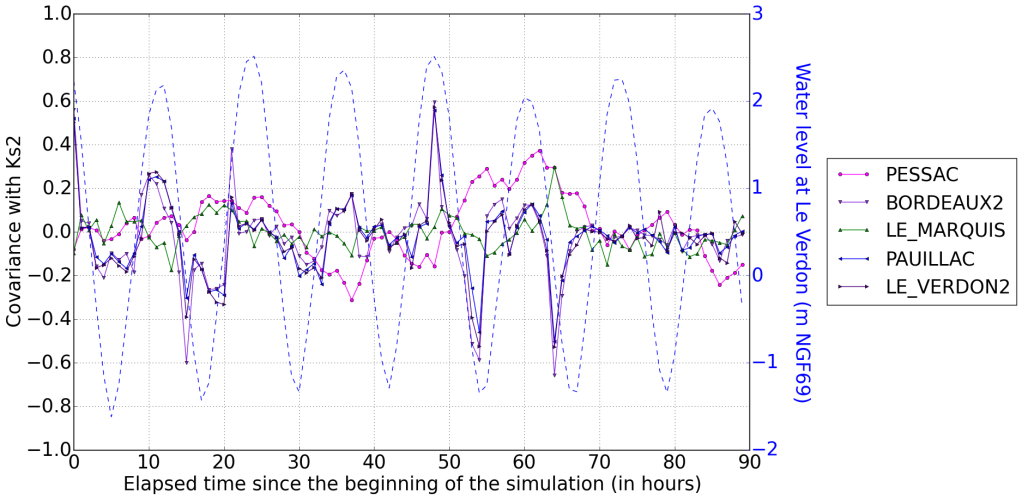



(b)

FIGURE 5.31 : Fonction temporelle de covariance entre $Ks2$ et $y$ à Pauillac pour : a) $N_e = 10$ membres sans rééchantillonnage (durée totale : 24 h) - b) $N_e = 10$ membres avec rééchantillonnage (durée totale : 24 h) - c) $N_e = 100$ membres avec rééchantillonnage (durée totale : 96 h) - d) $N_e$=100 membres sans rééchantillonnage et avec correction simultanée des autres $Ks$ et du forçage maritime dans l'espace réduit de KL (durée totale : 96 h). L'effondrement de l'ensemble a lieu autour de 18 h après le début de l'expérience. (début)

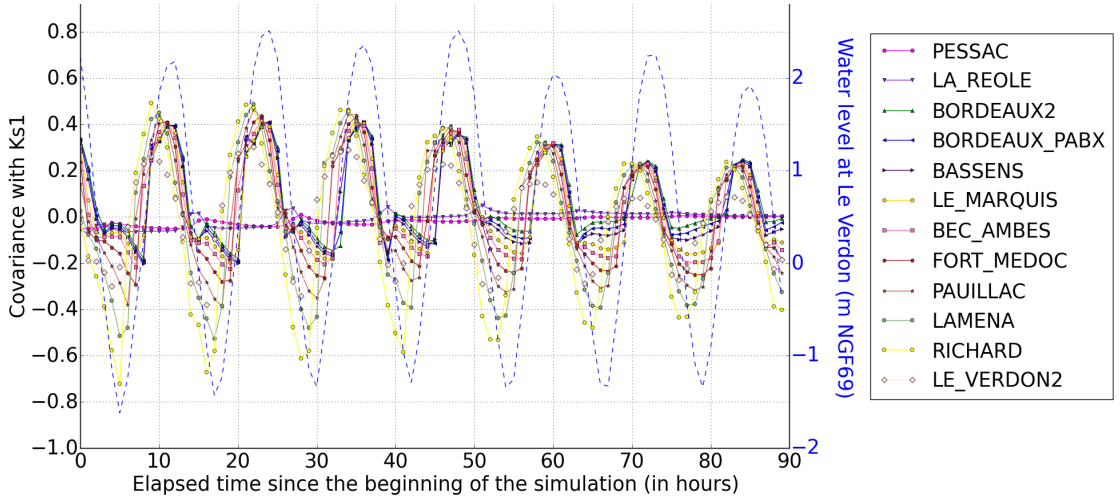



(c)

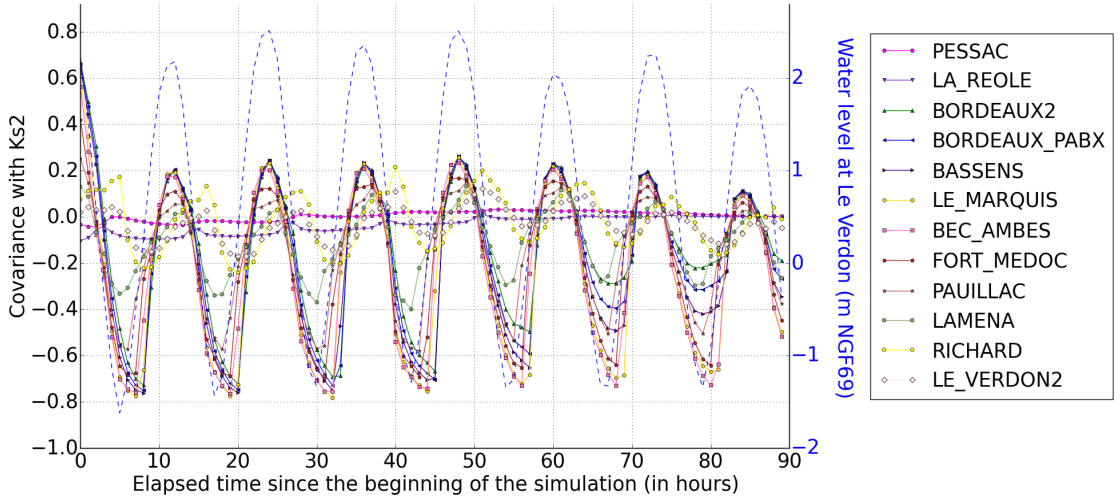



(d)

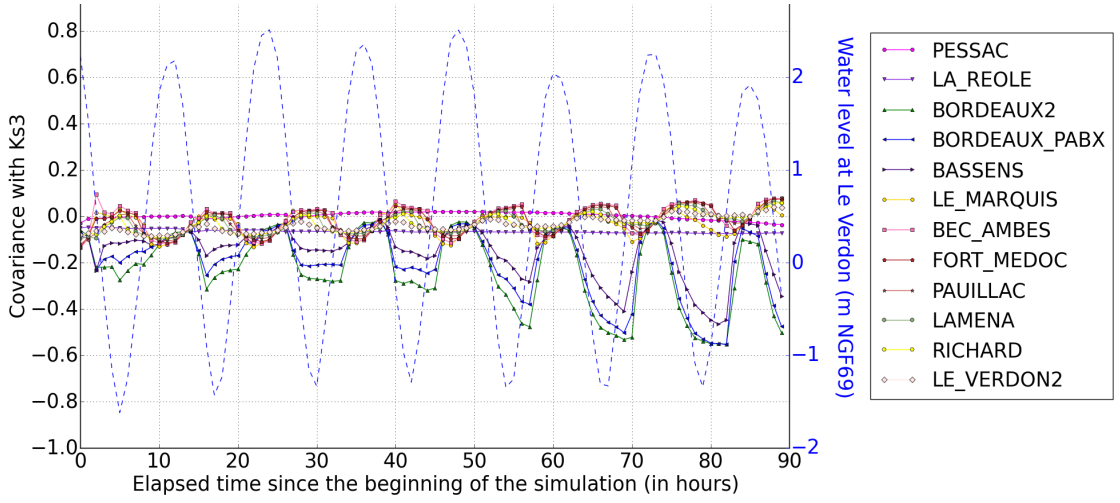
FIGURE 5.31 : Fonction temporelle de covariance entre $Ks2$ et $y$ à Pauillac pour : a) $N_e = 10$ membres sans rééchantillonnage (durée totale : 24 h) - b) $N_e = 10$ membres avec rééchantillonnage (durée totale : 24 h) - c) $N_e = 100$ membres avec rééchantillonnage (durée totale : 96 h) - d) $N_e$=100 membres sans rééchantillonnage et avec correction simultanée des autres $Ks$ et du forçage maritime dans l'espace réduit de KL (durée totale : 96 h). L'effondrement de l'ensemble a lieu autour de 18 h après le début de l'expérience. (fin)




(a) $Ks1$

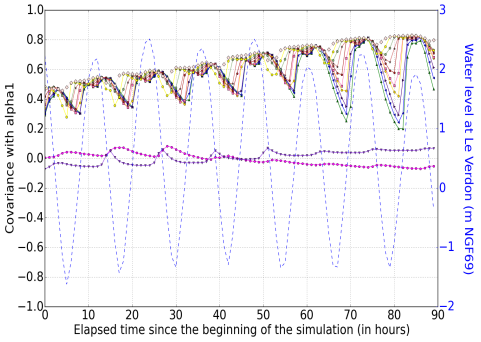



(b) $Ks2$

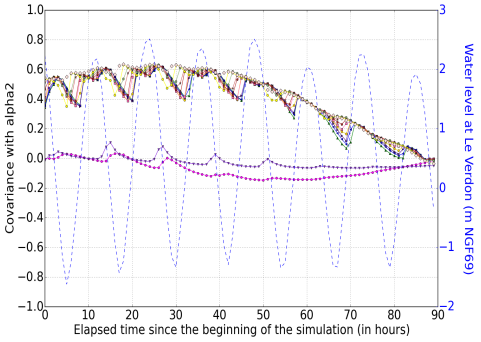



(c) $Ks3$

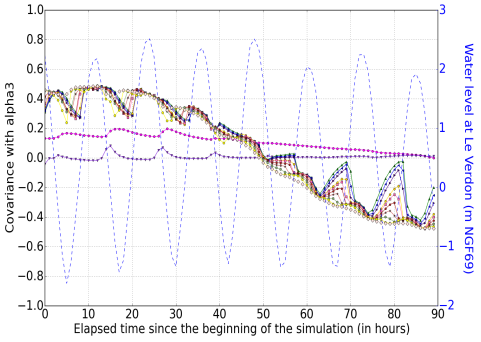
FIGURE 5.32 : Fonction de covariance de $Ks1$, $Ks2$ et $Ks3$ avec les observations pour l'Exp. 3-b



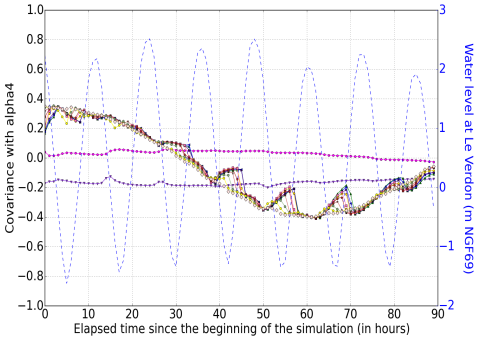

(a) CLMAR - mode 1



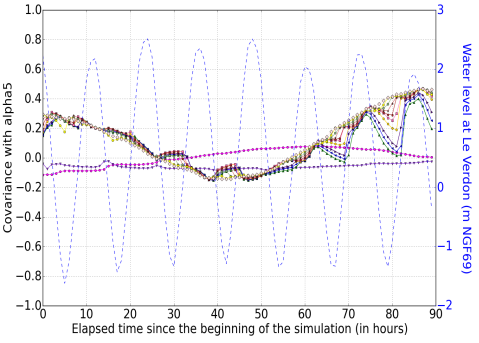

(b) CLMAR - mode 2



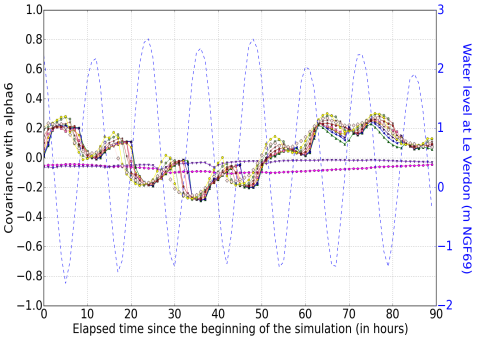

(c) CLMAR - mode 3



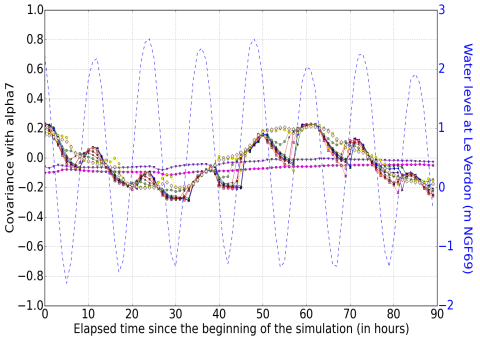

(d) CLMAR - mode 4

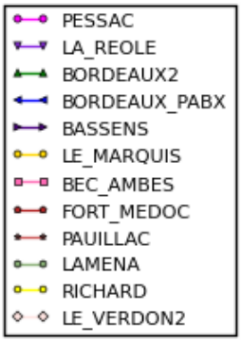



(e) CLMAR - mode 5



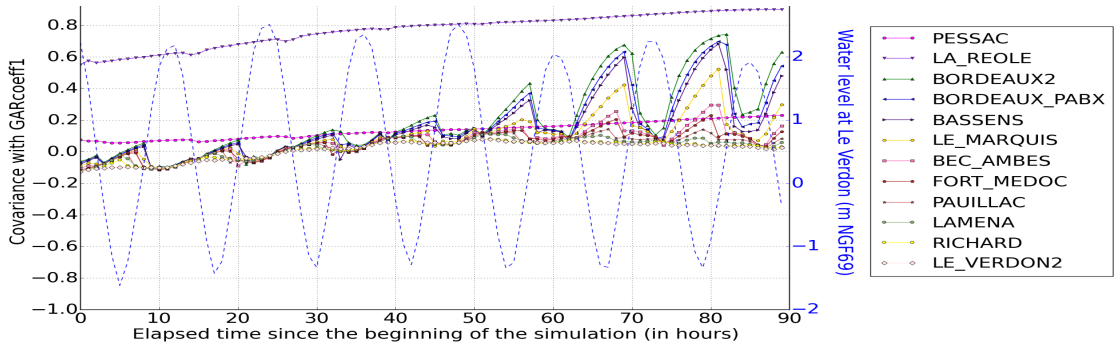

(f) CLMAR - mode 6



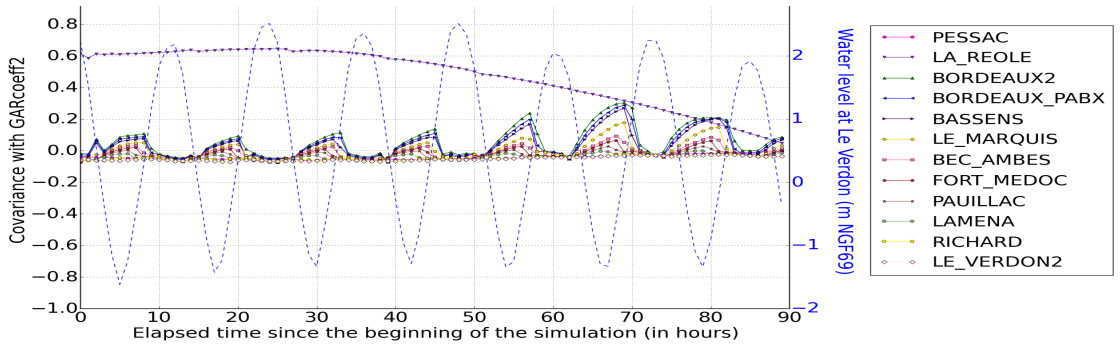




(g) CLMAR - mode 7

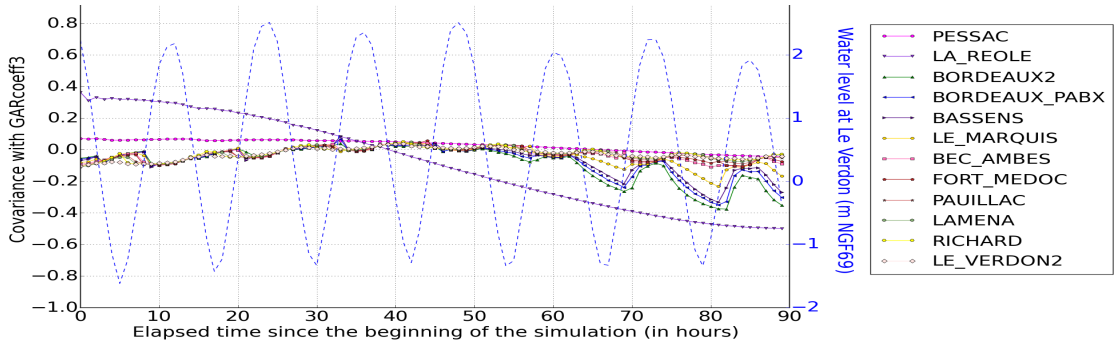
FIGURE 5.33 : Fonction de covariance de CLMAR avec les observations pour l'Exp. 3-b



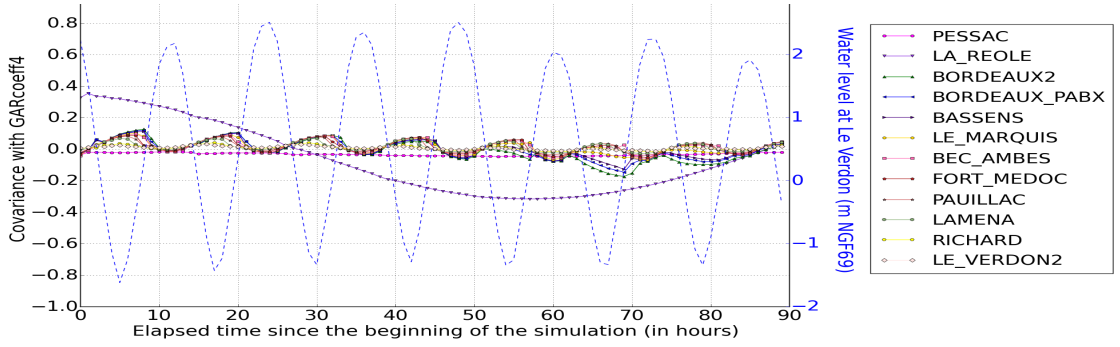

(a) QGAR - mode 1




(b) QGAR - mode 2

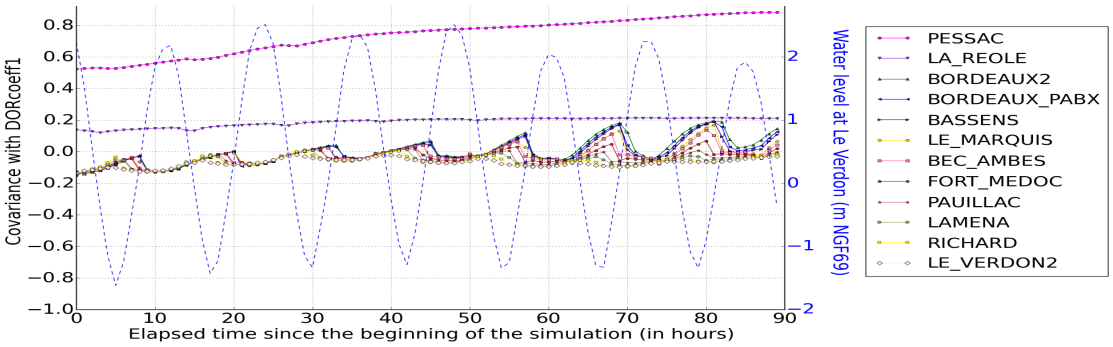



(c) QGAR - mode 3

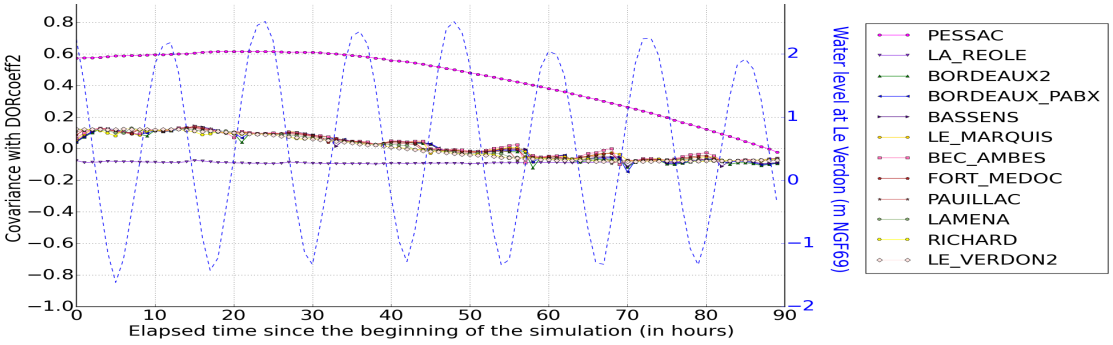



(d) QGAR - mode 4

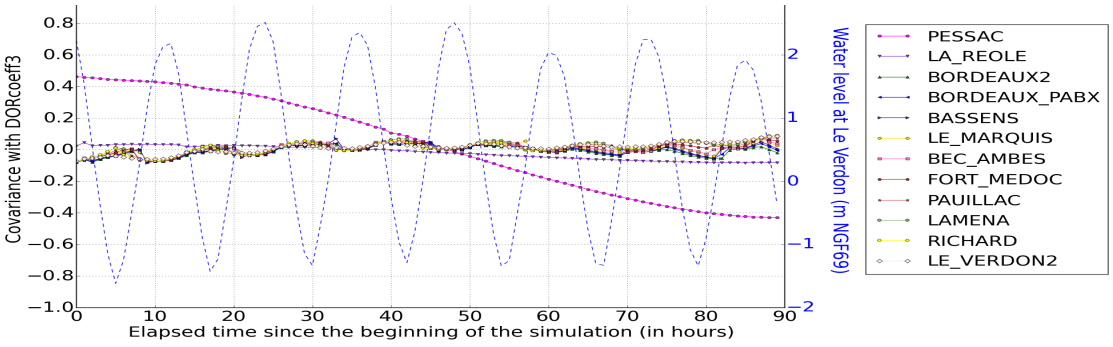
FIGURE 5.34 : Fonction de covariance de QGAR avec les observations pour l'Exp. 3-b



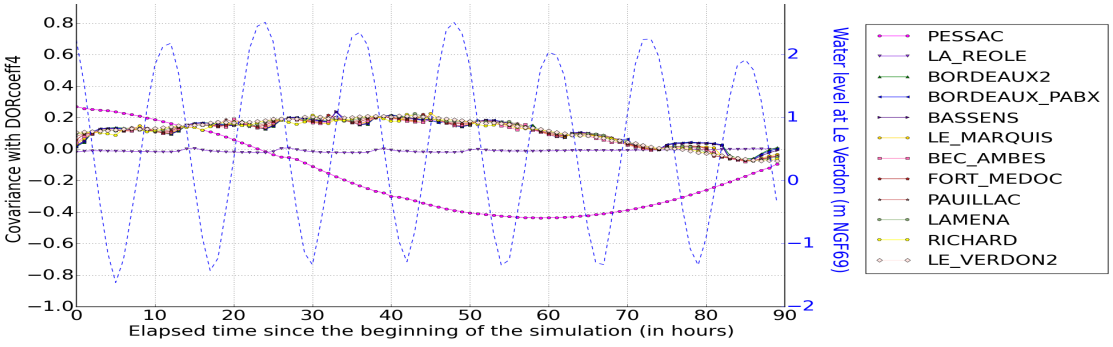

(a) QDOR - mode 1




(b) QDOR - mode 2




(c) QDOR - mode 3




(d) QDOR - mode 4

FIGURE 5.35 : Fonction de covariance de QDOR avec les observations pour l'Exp. 3-b

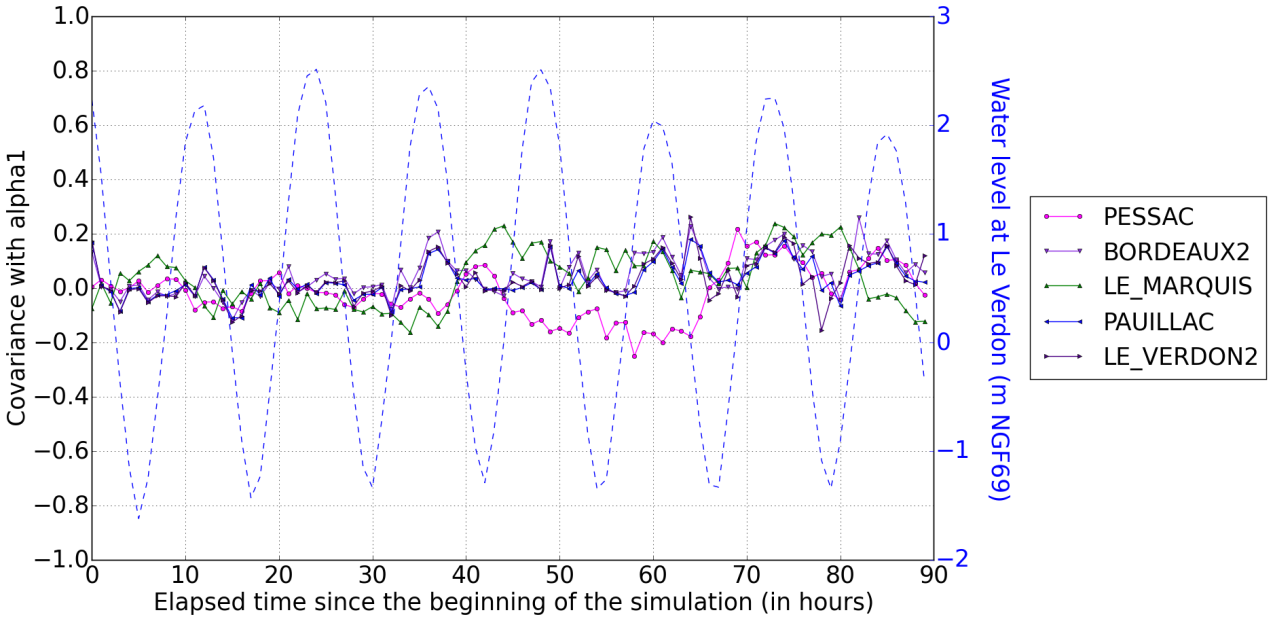
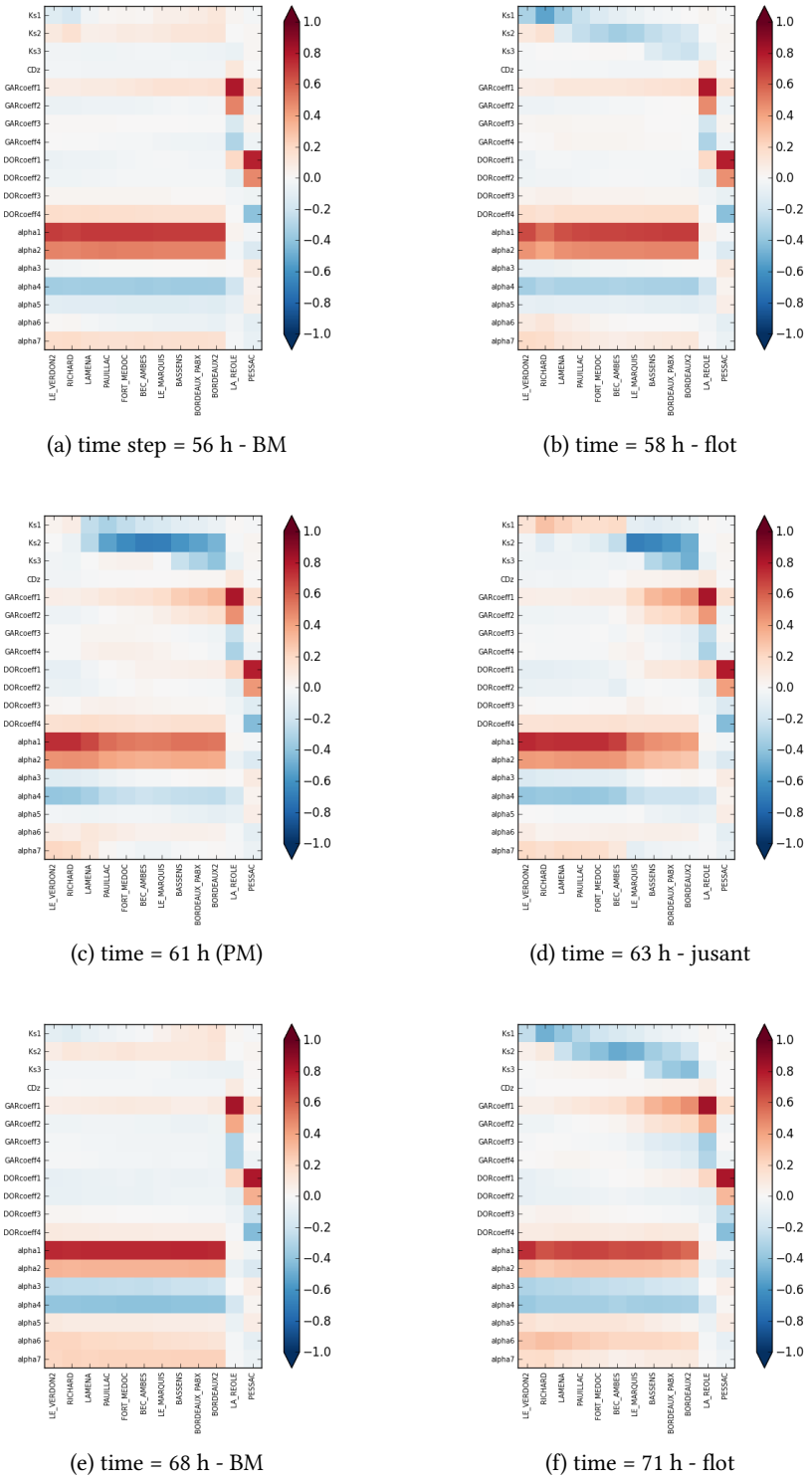



FIGURE 5.36 : Fonction de covariance croisée entre $\alpha 1$ et les stations d'observation de l'estuaire en l'absence de rééchantillonnage pour l'Exp. 3




(a) time step = 56 h - BM

(b) time = 58 h - flot

(c) time = 61 h (PM)

(d) time = 63 h - jusant

(e) time = 68 h - BM

(f) time = 71 h - flot

FIGURE 5.37 : Matrice de covariance entre vecteur de contrôle et niveaux d'eau calculés par Telemac2D dans l'espace des observations pour l'expérience 3-b au cours d'un cycle de marée

# Chapitre 6

# Évaluation de l'EnKF-$\gamma$-KLBC-MEG sur données réelles

**Sommaire**


**6.1 Caractérisation d'une expérience réelle . . . . . 223**

6.1.1 Événements réels et observations . . . . . 223

6.1.2 Vecteur de contrôle : ébauche et run de contrôle . . . . . 224

**6.2 Description du dispositif d'expériences réelles . . . . . 225**

**6.3 Recalage par assimilation de données de l'événement de Février 2003 (Exp. $R1$) . . . . . 230**

6.3.1 Estimation des coefficients de frottement par assimilation de données sur l'événement de calage de 2003 . . . . . 230

6.3.2 Niveaux d'eau prédits et scores par estimation des coefficients de frottement sur la tempête de Février 2003 . . . . . 231

**6.4 Analyse des performances de l'EnKF-$\gamma$-KLBC-MEG sur un cycle de marée (Exp. $R2$) . . . . . 233**

6.4.1 Influence de la composition du vecteur de contrôle : sensibilité à $Ks3$ et non convergence des coefficients modaux . . . . . 233

6.4.2 Influence sur les niveaux d'eau simulés par l'EnKF-$\gamma$-KLBC-MEG de l'intégration de Ks3 dans un vecteur de contrôle incluant la correction du forçage maritime dans l'espace réduit de KL . . . . . 234

**6.5 Analyse des performances de l'EnKF-$\gamma$-KLBC-MEG sur un événement réel (Exp. $R3$) . . . . . 236**

6.5.1 Correction des paramètres et des forçages temporels pour une réanalyse d'une semaine . . . . . 236

6.5.2 Prévisions de niveaux d'eau en réanalyse par estimation des coefficients de frottement et correction du forçage maritime . . . . . 237

**6.6 Matrice Gain de Kalman : réseau de mesures idéal et localisation spatio-temporelle . . . . . 239**

6.7 Conclusions sur les performances de l'EnKF-$\gamma$-KLBC-MEG en réanalyse 242

La chapitre 5 a permis de valider et d'établir des spécifications pour l'utilisation de l'EnKF-$\gamma$-KLBC-MEG dans le but d'améliorer les niveaux d'eau simulés dans l'estuaire de la Gironde dans un cadre synthétique en corrigeant les paramètres et/ou les forçages du modèle Telemac2D au cours de plusieurs cycles de marée.

L'objet de ce chapitre est d'évaluer, à l'aide de données issues d'observations mesurées pendant des événements réels, les gains apportés par l'EnKF-$\gamma$-KLBC-MEG sur la prévision des niveaux d'eau le long de l'estuaire. Plusieurs événements ont été utilisés. La tempête de Février 2003, décrite dans la section 1.2.3 et déjà utilisée pour le calage du modèle de prévision des hautes eaux de l'estuaire de la Gironde, est le premier événement. Les deux autres événements correspondent respectivement aux périodes du 01/01/2016 au 09/01/2016 et des 09-10/01/2016.

La section 6.2 décrit le dispositif d'expériences réelles qui consiste à utiliser les mesures horaires disponibles des événements cités précédemment en différentes stations de mesure pour corriger les variables de contrôle choisies.

D'une part, l'évolution temporelle des variables corrigées selon la composition du vecteur de contrôle est analysée pour chaque expérience réelle. D'autre part, les gains obtenus à l'aide de l'EnKF-$\gamma$-KLBC-MEG pour la prévision des niveaux d'eau en réanalyse sont estimés.

## 6.1 Caractérisation d'une expérience réelle

Tout comme l'expérience synthétique, l'expérience réelle se caractérise par un jeu d'observations, correspondant cette fois-ci à des mesures réellement observées ***in situ***, et par un vecteur de contrôle $\boldsymbol{\gamma}$ dont la composition détermine les paramètres et les forçages corrigés au cours de l'expérience d'assimilation.

### 6.1.1 Événements réels et observations

Trois événements réels sont utilisés pour tester les performances de l'EnKF-$\gamma$-KLBC-MEG.

Comme indiqué ci-dessus, la tempête de Février 2003, utilisée pour le calage du modèle de prévision des hautes eaux de l'estuaire de la Gironde (cf. section 1.2.3). Il est noté $Real1$. Les débits de la Garonne et de la Dordogne, représentés sur la figure 1.9 du chapitre 1, culminent respectivement à 2200 $m^3/s$ et 5900 $m^3/s$. Le pic de surcote instantanée fournie par le modèle HYCOM du SHOM est égal à 20 cm. L'expérience simulée est restreinte ici à la période du 02/02/2003 00:00 au 06/02/2003 00:00, soit une durée totale de 96 heures.

Comme le montre la figure 6.1, la période du 01/01/2016 au 10/01/2016 est caractérisée par une crue modérée de la Garonne dont le débit croît de 100 à 1046 $m^3/s$ en 8 jours et un début de crue de la Dordogne dont le débit croît de 100 à 950 $m^3/s$ sur la même période. Les périodes du 01/01/2016 00:00 au 09/01/2016 00:00 et du 09/01/2016 06:00 au 10/01/2016 06:00, de durées respectives totales 8 jours et 24 h, sont toutes deux utilisées comme événements de référence et

respectivement notées $Real3$ et $Real2$. Ces deux événements sont intéressants, car ils n'ont pas servi au calage du modèle. Cependant, les forçages météorologiques et la condition à la limite maritime associée sont uniquement disponibles sur la période du 01/01/2016 au 10/01/2016, ce qui justifie le fait d'arrêter l'expérience $Real2$ au 10/01/2016 alors que le débit de la Dordogne augmente jusqu'au 16/01/2016.

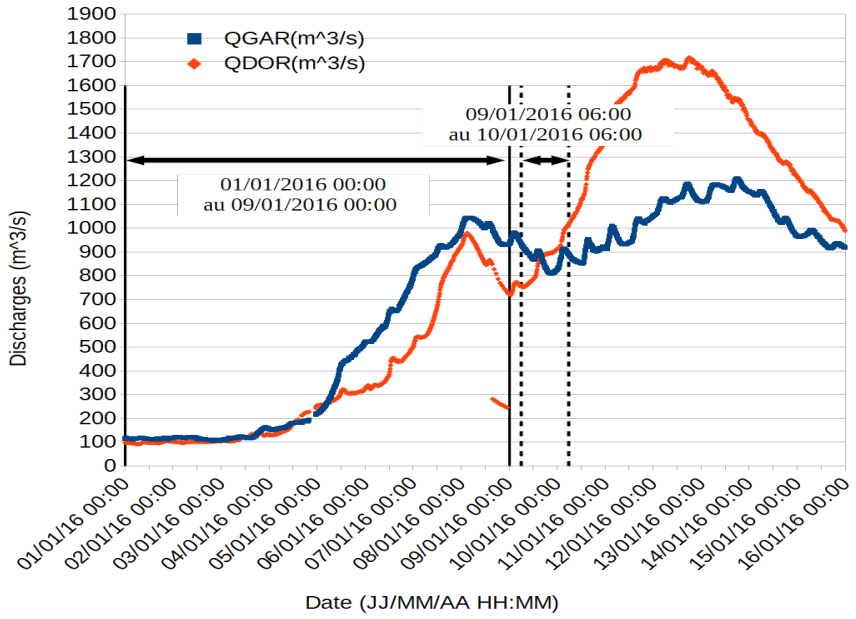



FIGURE 6.1 : Débits de la Garonne ($QGAR$) et de la Dordogne ($QDOR$) pendant les événements de Janvier 2016

Pour chacun des événements $Real1$, $Real2$ et $Real3$, les observations disponibles sur les points de mesure de l'estuaire de la Gironde sont fournies par le Service de Prévision des Crues Gironde, Adour, Dordogne (SPC-GAD) et/ou issues des bases de données du réseau REFMAR (http:\refmar.shom.fr). Il s'agit de données de niveaux d'eau disponibles à des pas de temps valant respectivement 5 mn pour Janvier 2016 ($Real2$ et $Real3$) et 10 mn pour la tempête de Février 2003 ($Real1$). Les observations peuvent être lacunaires du fait de défaillances du système d'enregistrement ou de transmission de la mesure.

### 6.1.2 Vecteur de contrôle : ébauche et run de contrôle

Compte-tenu des conclusions du chapitre 5, aucune expérience réelle n'intègre, dans son vecteur de contrôle, le coefficient d'influence du vent $Cd$ ni les débits $QGAR$ et $QDOR$ des deux rivières Garonne et Dordogne.

Les forçages maritimes fournis par le modèle HYCOM de METEO-FRANCE correspondant à

l'événement réel $Real1$ de Février 2003 sont représentés sur la figure 1.9. Ceux de Janvier 2016, notés $CLMAR^b$, sont respectivement représentés sur les figures 6.6 pour $Real2$ et 6.14 pour $Real3$.

Pour toutes les expériences réelles, les forçages fluviaux sur les limites amont du modèle, c'est-à-dire les débits au droit de La Réole sur la Garonne et de Pessac sur la Dordogne, ne sont ni perturbés ni corrigés. Ils sont issus de la réanalyse de l'événement. De fait, le maillage utilisé pour l'expérience étant celui du cadre opérationnel et le moins raffiné ($mesh0$), la courbe de tarage et les niveaux d'eau à La Réole et à Pessac, mais aussi sur les tronçons fluviaux jusqu'à Langon et Cadillac, sont mal représentés. Pour améliorer les niveaux d'eau calculés à l'amont des tronçons fluviaux, il est possible d'introduire $Ks4$ dans le vecteur de contrôle, après avoir découplé les coefficients de Strickler sur la Garonne et la Dordogne (à l'instar de la méthode utilisée pour le projet GirondeXL 3D), et/ou d'utiliser le maillage le plus raffiné. Ce n'est pas le cas ici. C'est pourquoi $Ks4$ n'est introduit dans le vecteur de contrôle que pour l'expérience de recalage par AD (Exp. $R1$ décrite en section 6.3).

Le run de contrôle est obtenu en intégrant le modèle d'évolution Telemac2D sur l'estuaire de la Gironde avec les forçages maritimes et fluviaux issus de la réanalyse pour chaque événement réel et la moyenne des paramètres d'ébauche.

## 6.2 Description du dispositif d'expériences réelles

Les tableaux 6.1, 6.2 et 6.3, qui suivent le même formalisme que les tableaux5.2 et 5.5 du chapitre 5, décrivent les observations utilisées dans le cadre réel, la composition du vecteur de contrôle $\boldsymbol{\gamma}$ et son ébauche, ainsi que les autres paramètres de l'EnKF-$\gamma$-KLBC-MEG utilisés pour les expériences réelles $R1$, $R2$ et $R3$ menées sur la base de l'assimilation de mesures réelles et non plus synthétiques.

L'expérience réelle $R1$ consiste en la réanalyse de l'événement de calage de février 2003 pour lequel tous les coefficients de Strickler sont corrigés au cours de l'événement à partir d'une ébauche dont la moyenne correspond aux valeurs de calage retenues pour chaque paramètre. Il s'agit donc d'une expérience d'optimisation du paramètre de frottement sur la base du calage du modèle sur la période de $Real1$ allant du 02/02/03 00:00 au 06/02/03 00:00, soit une durée de 4 jours. Une période d'initialisation par run libre ("spin-up") de 10 heures est mise en œuvre. Chaque cycle d'assimilation dure $L\Delta t = 6\ h$ (soit 6 observations assimilées par cycle) et la durée $S\Delta t$ entre deux cycles successifs vaut 2 h, conformément aux conclusions faites dans le chapitre 5. En revanche, toutes les stations ayant servi au calage du modèle, donc plus de stations que celles du réseau de mesures jugé "optimal" de la section 5.5.1, comme l'indique le tableau 6.1, sont utilisées ici.

L'objectif de l'expérience $R1$ est de déterminer si l'EnKF-$\gamma$-KLBC-MEG développé est capable d'améliorer les résultats en termes de RMSE par rapport au run de contrôle pour l'événement de

calage de 2003 et d'étudier l'évolution des composantes du vecteur de contrôle analysé $\overline{\gamma^a}$ par rapport à celui issu du calage du modèle ($\overline{\gamma^0}$) utilisé pour l'ébauche de $R1$. Elle montre, d'une part, une forte variabilité temporelle des coefficients de Strickler et, d'autre part, que l'EnKF-$\gamma$-KLBC-MEG tend à améliorer les BM en augmentant l'amplitude du signal. De fait, afin de ne pas mélanger les erreurs caractérisant les BM et qui diffèrent de celles des PM, elle conduit à tester un nouveau cyclage des fenêtres d'assimilation de données.

L'expérience $R2$, déclinée en 3 sous-expériences -a, -b et -c, consiste en la réanalyse de la période du 09/01/2016 06:00 au 10/01/2016 06:00 ($Real2$). La composition du vecteur de contrôle est ici testée. L'Exp. $R2a$ consiste en la correction du coefficient $Ks2$ seul, alors que $R2b$ corrige les paramètres $Ks1$, $Ks2$ et la condition aux limites maritimes $CLMAR$ dans l'espace réduit de KL. Elles utilisent respectivement les observations horaires disponibles à Pauillac pour $R2a$ et à Port-Bloc et Pauillac pour $R2b$ pendant l'événement $Real2$. L'Exp. $R2c$ permet d'évaluer les apports de la correction simultanée de $Ks1$, $Ks2$, $Ks3$ et des coefficients modaux de $CLMAR$ en assimilant les observations horaires disponibles à Port-Bloc, Pauillac et Bordeaux PABX. Ainsi, $R2b$ diffère de $R2c$ par l'introduction de $Ks3$ dans le vecteur de contrôle.
Une période d'initialisation ("spin-up") par run libre de 6 heures est mise en œuvre. Chaque cycle d'assimilation dure $L\Delta t = 4$ heures (soit 4 observations assimilées par cycle) et la durée $S\Delta t$ entre deux cycles successifs vaut 2 heures. L'exploitation des expériences d'assimilation de données $R2b$ et $R2c$ est réalisée sur 12 h. Elles permettent d'étudier finement le comportement du système (vecteur de contrôle, niveaux d'eau et correction du forçage maritime) sur un cycle de marée.

L'Exp. $R3$ présente le même vecteur de contrôle que $R2c$, soit $Ks1$, $Ks2$, $Ks3$ et les coefficients modaux de CLMAR, mais la période de l'événement sur laquelle la réanalyse est effectuée est comprise entre le 01/01/2016 18:00 et le 09/01/2016 00:00. La durée de l'expérience est donc d'environ 8 jours. Chaque cycle d'assimilation comprend $L\Delta t = 3\ h$ de réanalyse et 12 h de prédiction. La durée $S\Delta t$ entre deux cycles d'assimilation est de 2 heures. La méthodologie de construction des paramètres et des forçages de la phase de prédiction est discutée dans les perspectives du chapitre 7.

Pour chacune de ces expériences, les lois suivies par les variables perturbées sont présentées dans la partie inférieure "Ébauche" du tableau 6.2. L'option de rééchantillonnage par persistance de la matrice $\mathbf{P}_0^{\gamma}$ de covariance d'ébauche de l'ensemble (cf. section 4.1.4 du chapitre 4) est utilisée pour toutes les expériences réelles.

Par ailleurs, toutes les données d'entrée du modèle qui ne sont pas "contrôlées" et ne subissent pas de perturbation, sont fixées au cours des expériences, pour ce qui concerne les paramètres $Ks$ et $Cd$, aux valeurs issues du calage du modèle et, pour ce qui concerne les forçages maritimes et fluviaux, aux chroniques issues de la réanalyse de l'événement.

TABLE 6.1 : Description des observations réelles utilisées, du vecteur de contrôle et de l'ensemble perturbé pour les expériences réelles. Toutes les observations sont des mesures réelles horaires de niveaux d'eau.

| | **Observations** | |
|---|---|---|
| Exp. | Événement | Loc. |
| $R1$ | $Real1$<br>02/02/2003 00 :00<br>10/02/2003 00 :00 | Le Verdon, Richard, Laména, Pauillac, Fort-Médoc, Bec d'Ambès, au Le Marquis, Bassens, Bordeaux Pabx, Bordeaux2, La Réole |
| $R2a$ | $Real2$ | Pauillac. |
| $R2b$ | 09/01/2016 12 :00 | Port-Bloc, Pauillac. |
| $R2c$ | au 12/01/2016 06 :00 | Port-Bloc, Pauillac, Bordeaux PABX |
| $R3$ | $Real3$<br>01/01/2016 18 :00<br>au 10/01/2016 06 :00 | Port-Bloc, Pauillac, Bordeaux PABX |

TABLE 6.2 : Description du vecteur de contrôle et de l'ensemble perturbé pour les expériences réelles

| | $K_s$ | | | | $CLMAR$ | |
|---|---|---|---|---|---|---|
| | $Ks1$ | $Ks2$ | $Ks3$ | $Ks4$ | $dsm$ | $\alpha$ |
| **Vecteur de contrôle** | | | | | | |
| $R1$ | ✓ | ✓ | ✓ | ✓ | | |
| $R2a$ | | ✓ | | | | |
| $R2b$ | ✓ | ✓ | | | | ✓ |
| $R2c$ et $R3$ | ✓ | ✓ | ✓ | | | ✓ |
| **Ébauche** | | | | | | |
| $R1$ | $\mathcal{U}(50,70)$ | $\mathcal{U}(50,70)$ | $\mathcal{U}(40,60)$ | $\mathcal{U}(30,50)$ | | |
| $R2a$ | | $\mathcal{U}(50,70)$ | | | | |
| $R2b$ | $\mathcal{U}(50,70)$ | $\mathcal{U}(50,70)$ | | | $\pm 30\ cm$ | $\mathcal{U}^{-1}(0,1)$ |
| $R2c$ et $R3$ | $\mathcal{U}(50,70)$ | $\mathcal{U}(50,70)$ | $\mathcal{U}(40,60)$ | | $\pm 30\ cm$ | $\mathcal{U}^{-1}(0,1)$ |

TABLE 6.3 : Autres paramètres d'entrée de l'EnKF-$\gamma$-KLBC-MEG pour les expériences réelles. $N_e$=100 et le rééchantillonnage (persistance temporelle de $\mathbf{P}_0^{\gamma}$) est utilisé. Le pas de temps $\Delta t$ d'assimilation est égal à 1 h.

| Exp. | Durée Totale | cycle d'assimilation | | Durée des runs libres (h) | | |
|---|---|---|---|---|---|---|
| | | L | S | avant "spin-up" | pendant "forecast" | après "forecast" |
| $R1$ | 4 j | 6 | 2 | 10 | 0 | 0 |
| $R2a$ | 66 h | 4 | 2 | 6 | 0 | 48 h |
| $R2b$ | 18 h | 4 | 2 | 6 | 0 | 48 h |
| $R2c$ | 18 h | 4 | 2 | 6 | 0 | 48 h |
| R3 | 8.5 j | 3 | 2 | 6 | 12 h | 0 |

Les sections suivantes présentent, dans un premier temps, l'évolution temporelle des variables de contrôle pour les expériences réelles $R1$, $R2$ et $R3$, ainsi que, dans le cas d'une correction du forçage maritime dans l'espace réduit de KL, la condition aux limites analysée $CLMAR^a$.
Dans un second temps, les performances de l'EnKF-$\gamma$-KLBC-MEG sont mesurées au travers de la différence instantanée, ainsi que par les RMSE sur les durées d'expérience, entre les observations $\mathbf{y}^0$ et les niveaux d'eau $\mathbf{y}^a$ à l'issue des phases d'optimisation de l'état du système. Les figures qui représentent les niveaux d'eau du run de contrôle ($\mathbf{y}^b$), après optimisation ($\mathbf{y}^a$) et les mesures ($\mathbf{y}^0$) partagent le même formalisme que celui utilisé dans la section 5.4. La ligne d'eau correspondant au run de contrôle ($\mathbf{y}^b$) (resp. à l'état réanalysé $\mathbf{y}^a$) est représentée par une ligne pointillée bleue (resp. continue avec marqueurs rouges). Les observations sont représentées par des points noirs ($\mathbf{y}^0$) et la différence entre la ligne d'eau après assimilation et l'observation ($\mathbf{y}^a - \mathbf{y}^0$) est représentée par une ligne pointillée et des marqueurs triangulaires rouges à l'aide de l'axe des ordonnées à droite de la figure.
Les figures représentant les résultats des sections suivantes sont regroupées à la fin du chapitre.

## 6.3 Recalage par assimilation de données de l'événement de Février 2003 (Exp. $R1$)

### 6.3.1 Estimation des coefficients de frottement par assimilation de données sur l'événement de calage de 2003

L'expérience $R1$ corrige les coefficients de Strickler $Ks1$, $Ks2$, $Ks3$ et $Ks4$ à partir des valeurs de calage, utilisées comme ébauche, $\overline{Ks1_0}$=60, $\overline{Ks2_0}$=60, $\overline{Ks3_0}$=50 et $\overline{Ks4_0}$=40. L'évolution temporelle des variables de contrôle est représentée sur la figure 6.2. Rappelons qu'une fenêtre d'assimilation dure 6 h et que deux fenêtres consécutives sont décalées de 2 h. Ainsi, chaque cycle d'assimilation utilise des observations issues de différents moments de la marée.

Le coefficient de Strickler $Ks1$ atteint des valeurs supérieures à 90 en début et en fin d'expérience, mais oscille autour de 75 au cours de celle-ci. Il en est de même pour $Ks2$.
Les coefficients $Ks3$ et $Ks4$ varient peu en début d'expérience. Ils restent proches de leurs valeurs initiales respectives (50 et 40). Au bout de 48 h, en lien avec la montée de crue de la Garonne dont le débit augmente de 1200 à 5900 m$^3$/s entre le 02 Février et le 06 Février 2003 et l'assimilation des niveaux d'eau à Bordeaux et à La Réole, $Ks3$ et $Ks4$ augmentent de manière continue pour atteindre les valeurs finales de 90 pour $Ks3$ et 70 pour $Ks4$. Pendant cet événement, les crues de la Garonne et la Dordogne, dont les débits de plein bord sont respectivement évalués à 2500 m$^3$/s pour le lit mineur (et 3500 m$^3$/s entre les digues du lit majeur) à Tonneins et 1200 m$^3$/s à Pessac, conduisent théoriquement à des débordements en amont des deux rivères dans leurs lits majeurs respectifs après 60 h d'expérience. Les champs d'expansion des crues ne sont pas représentés par le modèle non-débordant qui impose une paroi verticale à la frontière solide du modèle. De fait, en situation de débordement, le modèle sur-estime les niveaux d'eau compte-tenu de la mauvaise

représentation de la section mouillée et de l'absence de débordements. C'est pourquoi $Ks4$ évolue peu pendant les premières 60 h où le débit de la Garonne (cf. figure 1.9 du chapitre 1) reste autour de 1400 $m^3/s$. Puis la Garonne déborde dans son lit majeur. L'EnKF-$\gamma$-KLBC-MEG sur-estime les hauteurs d'eau, ce qui le conduit à augmenter les valeurs de $Ks4$ pour diminuer les hauteurs d'eau du fait des mesures assimilées à la Réole.

En conclusion, il est donc possible de noter que l'estimation des coefficients de Strickler lors la réanalyse de l'événement de Février 2003 à l'aide de toutes les stations ayant servi au calage du modèle montre une variabilité temporelle autour de valeurs plus élevées que celles obtenues lors du calage pour $Ks1$ et $Ks2$. En amont de la confluence, sur la Garonne, l'assimilation des données observées au Marquis, à Bassens, à Bordeaux et à La Réole provoque une croissance continue de $Ks3$ et $Ks4$, ce qui correspond ***in fine*** à une augmentation des vitesses dans cette zone liée, pour ce qui concerne $Ks4$, à la crue de la Garonne.

### 6.3.2 Niveaux d'eau prédits et scores par estimation des coefficients de frottement sur la tempête de Février 2003

La figure 6.3 représente les niveaux d'eau sur plusieurs stations de l'estuaire (Le Verdon, Pauillac, Bec d'Ambès, Bordeaux, La Réole et Pessac) dont les mesures sont assimilées au cours de l'Exp. $R1$. Le tableau 6.4 récapitule les scores obtenus pendant la durée de l'événement de 2003, soit 8 jours.

La RMSE des stations du Verdon, de Richard, de Bassens et de Bordeaux PABX est peu modifiée par l'EnKF-$\gamma$-KLBC-MEG.
L'examen de la ligne d'eau au Verdon représentée sur la figure 6.3-a montre que les lignes d'eau avec et sans assimilation de données sont presque confondues. La solution pour diminuer la RMSE est donc ***a priori*** d'introduire les coefficients modaux de $CLMAR$ dans le vecteur de contrôle. Sur le dernier cycle de marée de l'expérience, cependant, il peut être noté que la correction de $Ks1$ tend à augmenter l'amplitude du signal en sur- (resp. sous-) estimant les pics de PM (resp. de BM) par rapport à la trajectoire fournie par le run de contrôle. Les mêmes constats peuvent être réalisés pour la station de Richard où, cependant, les pics de BM sont bien représentés par l'EnKF-$\gamma$-KLBC-MEG au détriment des pics de PM qui sont, quant à eux, sur-estimés.

L'amélioration sensible de la RMSE par l'EnKF-$\gamma$-KLBC-MEG entre Laména et Bordeaux2, y compris à Pauillac, est également liée à une meilleure représentation des BM, alors que les pics de PM sont sur-estimés par rapport à la trajectoire du run de contrôle et à l'observation du fait de l'inertie du système. Ainsi, l'amplification du signal apportée par l'EnKF-$\gamma$-KLBC-MEG permet une très bonne représentation à Bordeaux des $2^{ème}$ et $3^{ème}$ pics de PM, mais conduit à une sur-estimation du $4^{ème}$ pic de PM pendant la tempête de Février 2003.

A l'amont de la Garonne, les observations mesurées à la Réole sont assimilées et permettent de reconstituer la ligne d'eau mesurée, comme le montre la figure 6.3-e. En revanche, le fait que le coefficient de Strickler $Ks4$ soit appliqué simultanément sur les tronçons à l'amont des deux

rivières ne permet pas d'obtenir un résultat similaire à Pessac sur la figure 6.3-f. L'augmentation de $Ks4$ à partir du pas de temps 60 h correspond en réalité à la mauvaise représentation de la section mouillée à la Réole qui conduit à une surestimation du niveau d'eau lorsque le débit augmente du fait de l'utilisation du maillage le moins raffiné. Pour faire baisser le niveau d'eau et le rapprocher des observations, l'EnKF-$\gamma$-KLBC-MEG augmente alors $Ks4$. Le niveau d'eau étant également surestimé par le modèle à Pessac du fait du maillage "grossier", mais surtout de l'artéfact qui consiste à injecter les débits de la Dronne et de l'Isle à Pessac au lieu de Libourne, l'augmentation de $Ks4$ permet à partir de 60 h d'y diminuer le niveau d'eau, mais pas suffisamment pour atteindre l'observation.
Deux solutions sont à envisager si les niveaux d'eau à Pessac et La Réole doivent être corrigés. D'une part, il est possible de découpler les coefficients de Strickler des deux rivières sur leur linéaire. C'est la solution retenue dans le cadre du projet GirondeXL 3D (cf. section 1.2.1.2 du chapitre 1). D'autre part, raffiner de manière suffisante l'amont des rivières afin de représenter au mieux les sections mouillées de l'écoulement peut être une solution, de même évidemment qu'un report des débits des confluents de la Dordogne à Libourne.

TABLE 6.4 : RMSE sans et avec EnKF-$\gamma$-KLBC-MEG pour l'Exp. $R1$

| Station | RMSE sans DA (m) | RMSE avec DA (m) | Gain (m) |
|---|---|---|---|
| Le Verdon | 0.18 | 0.2 | -0.02 |
| Richard | 0.2 | 0.21 | -0.01 |
| Laména | 0.37 | 0.24 | 0.13 |
| Pauillac | 0.36 | 0.2 | 0.16 |
| Fort Médoc | 0.4 | 0.21 | 0.19 |
| Bec d'Ambès | 0.26 | 0.19 | 0.07 |
| Le Marquis | 0.32 | 0.19 | 0.13 |
| Bassens | 0.24 | 0.21 | 0.03 |
| Bordeaux Pabx | 0.21 | 0.23 | -0.02 |
| Bordeaux 2 | 0.36 | 0.18 | 0.18 |
| La Réole | 0.83 | 0.32 | 0.51 |
| Pessac | 1.73 | 1.27 | 0.46 |
| Global | 0.46 | 0.3 | 0.16 |

En conclusion sur l'Exp. $R1$, il apparaît donc que l'estimation temporelle des coefficients de Strickler permet d'améliorer la prévision des niveaux d'eau à l'aide de l'EnKF-$\gamma$-KLBC-MEG. L'amplitude du signal, ainsi que les pics de PM et de BM, sont généralement mieux représentés sur la partie centrale de l'estuaire et sur la Garonne. Il convient cependant de noter que les performances de l'EnKF-$\gamma$-KLBC-MEG sont liées au calage choisi pour le modèle numérique. Ainsi, l'EnKF-$\gamma$-KLBC-MEG appliqué sur tout le signal conduit aussi à améliorer les prévisions de niveaux d'eau sur les BM, c'est-à-dire sur les périodes de la marée pour lesquelles le comité de

pilotage du projet Gironde a volontairement choisi de dégrader les résultats du modèle en le calant sur les Nash de PM.

La durée d'un cycle d'assimilation $L\Delta t$ de 6 h incluant des observations prises aussi bien au flot qu'au jusant qu'en BM ou PM sont susceptibles de dégrader les résultats de l'EnKF-$\gamma$-KLBC-MEG. C'est pourquoi la durée de chaque cycle d'assimilation est réduite à $L\Delta = 4\ h$ (resp. 3 h) dans l'Exp. $R2$ (resp. Exp. $R3$). Par ailleurs, pour améliorer la prévision des niveaux d'eau au Verdon et à l'embouchure, l'Exp. $R1$ démontre la nécessité de corriger le forçage maritime, conformément aux résultats de l'étude de quantification d'incertitudes du chapitre 2. Les Exp. $R2$ et $R3$ incluent donc la correction du forçage maritime dans l'espace réduit de KL.

## 6.4 Analyse des performances de l'EnKF-$\gamma$-KLBC-MEG sur un cycle de marée (Exp. $R2$)

### 6.4.1 Influence de la composition du vecteur de contrôle : sensibilité à $Ks3$ et non convergence des coefficients modaux

L'objectif de l'Exp. $R2$, comprenant 6 h d'initialisation et menée sur 66 h pour $R2a$ et une courte période de 18 h (soit 5 cycles d'assimilation de données) pour $R2b$ et $R2c$, a pour but de déterminer l'influence de la composition du vecteur de contrôle lorsque les mesures horaires réelles disponibles sur la période du 09/01/2016 18:00 au 10/01/2016 06:00 sont assimilées. Les vecteurs de contrôle des expériences $R2a$, $R2b$ et $R2c$ sont indiqués dans le tableau 6.2. L'évolution sur un cycle de marée des variables de contrôle $Ks1$, $Ks2$ et $Ks3$ (pour l'Exp. $R2c$ seulement) est représentée sur les figures 6.4 et 6.5 pour les Exp. $R2b$ et $R2c$.
Au cours des deux expériences $R2b$ et $R2c$, les variables de contrôle $Ks1$ et $Ks2$ évoluent autour des valeurs d'ébauche fixées à 60 et dans des gammes dites "physiques".
En revanche, la variable de contrôle $Ks3$ prend des valeurs supérieures à 100 au cours de l'Exp $R2c$ justifiées par la nécessité d'abaisser la ligne d'eau sur l'intervalle de temps [8 h ; 14 h] à Bordeaux, à l'instar de ce qui a déjà été constaté pour l'Exp. $R1$ et l'Exp. 2 (sections 6.3 et 5.3.2).

Les conditions aux limites maritimes initiale ($CLMAR^b$) issues du modèle HYCOM de METEO-FRANCE, et corrigée ($CLMAR^a$) à l'aide de l'EnKF-$\gamma$-KLBC-MEG sont respectivement représentées en trait continu bleu et rouge sur la figure 6.6 pour chaque expérience $R2b$ et $R2c$.
Il peut être noté que l'EnKF-$\gamma$-KLBC-MEG apporte plus d'eau au système pendant la BM pour l'Exp. $R2c$, mais que, par rapport à la trajectoire initiale en bleu, le fait d'intégrer $Ks3$ dans le vecteur de contrôle et d'assimiler les mesures de Bordeaux, le conduit à retirer du volume à la fin de l'expérience, alors que la trajectoire corrigée de $R2b$ rejoint la trajectoire initiale en fin d'expérience. Par ailleurs, la correction sur le forçage maritime est principalement réalisée sur le flot (à partir du pic de BM jusqu'au pic de PM).

### 6.4.2 Influence sur les niveaux d'eau simulés par l'EnKF-$\gamma$-KLBC-MEG de l'intégration de Ks3 dans un vecteur de contrôle incluant la correction du forçage maritime dans l'espace réduit de KL

Les niveaux d'eau aux stations dont les données sont assimilées sont représentés sur la figure 6.7 (resp. 6.8) pour $R2b$ (resp. $R2c$). Pour l'Exp. $R2b$, la ligne d'eau corrigée à Port-Bloc et à Pauillac est confondue avec la mesure pendant la durée de l'expérience de réanalyse. Pour l'Exp. $R2c$, une divergence entre les mesures et la ligne d'eau corrigée apparaît à partir de t=14 h après le début de l'expérience pour les stations de Port-Bloc et de Pauillac, c'est-à-dire sur le dernier cycle d'assimilation. A Bordeaux, avant t=14 h, l'observation $\mathbf{y}^0$ que cherche à rejoindre la ligne d'eau corrigée $\mathbf{y}^a$ est en-dessous de la ligne d'eau $\mathbf{y}^b$ du run de contrôle aux stations "assimilées". $\mathbf{y}^0$ passe au-dessus de $\mathbf{y}^b$ à Bordeaux pour t=14 h. Dans une rivière, l'augmentation du coefficient de Strickler correspond à un frottement plus faible et implique une hauteur d'eau plus faible en régime stationnaire. De fait, le système cherche à baisser la ligne d'eau à Bordeaux avant t=14 h, puis à l'augmenter. L'EnKF-$\gamma$-KLBC-MEG augmente donc $Ks3$ et diminue les apports en volume à l'entrée maritime sur le dernier cycle d'assimilation. Il en résulte une dégradation des performances de l'EnKF-$\gamma$-KLBC-MEG à Port-Bloc et Pauillac, mais aussi à Bordeaux dont la ligne d'eau s'abaisse trop sous l'effet conjugué de la baisse du niveau d'eau à la frontière maritime et de l'augmentation de $Ks3$. Les gains en RMSE sur la durée des 3 sous-expériences sont récapitulés dans le tableau 6.5.

Les résultats pour l'Exp. $R2a$ montrent que la correction du coefficient de frottement $Ks2$ en assimilant les données horaires à Pauillac permet un gain maximal de 18 cm sur les 4 jours d'expérience. Par ailleurs, ce gain croît de Laména au Marquis, où il est maximal, puis décroit jusqu'à l'amont de chacune des deux rivières, de manière cohérent avec la zone d'influence de $Ks2$.

Les expériences $R2b$ et $R2c$ diffèrent par la présence au sein du vecteur de contrôle de la variable $Ks3$ et par l'assimilation des mesures horaires à la station de Bordeaux PABX .
L'analyse des résultats de $R2b$ montre que la correction simultanée des coefficients modaux de $CLMAR$ et des coefficients de Strickler $Ks1$ et $Ks2$ des zones les plus à l'aval de l'estuaire sur la durée courte de l'expérience permet d'obtenir une amélioration sensible de l'estimation des niveaux d'eau simulés ($\mathbf{y}^a$) par rapport au run de contrôle ($\mathbf{y}^b$) de l'embouchure de l'estuaire à Fort-Médoc. Le gain est maximal aux stations de Port-Bloc et Pauillac dont les données sont assimilées, mais il est aussi supérieur à 10 cm pour les stations de validation de Laména et Fort-Médoc.
Le bénéfice de l'introduction de $Ks3$ dans le vecteur de contrôle se fait sentir uniquement sur les portions fluviales des deux rivières, en particulier le gain est maximal à Bordeaux PABx et permet de réduire de 37 cm la RMSE à la station de Bordeaux2, soit 51 % d'amélioration de la RMSE sur l'expérience. En revanche, hormis aux stations de Port-Bloc et du Verdon, pour lesquelles les gains en RMSE chutent respectivement de 16 cm à 7 cm et de 4 cm à 2 cm, les résultats en RMSE sont détériorés sur les stations en aval du Marquis et, en particulier, à Pauillac, dont les mesures

TABLE 6.5 : RMSE sans et avec EnKF-$\gamma$-KLBC-MEG pour l'Exp. $R2$

| Station | ***R2a*** | | | ***R2b et -c*** | Gain | | | |
|---|---|---|---|---|---|---|---|---|
| | RMSE sans AD (m) | Gain (cm) | (%) | RMSE sans AD (cm) | ***R2b*** (cm) | (%) | ***R2c*** (cm) | (%) |
| Royan | 33 | -1 | -3 | 22 | **8** | **36** | -1 | -4 |
| Port-Bloc | 35 | -2 | -5 | 24 | **16** | **67** | 7 | 29 |
| Le Verdon | 33 | -2 | -6 | 19 | **4** | **21** | 2 | 11 |
| Laména | 34 | 0 | 22 | 0 | **12** | **55** | -8 | -36 |
| Pauillac | 39 | 6 | 15 | 21 | **12** | **57** | -10 | -48 |
| Fort-Médoc | 53 | **13** | 24 | 34 | 11 | **32** | -4 | -12 |
| Le Marquis | 56 | **18** | **32** | 46 | 9 | 20 | 4 | 9 |
| Bassens | 74 | 16 | 22 | 61 | 9 | 15 | **25** | **41** |
| Bordeaux PABX | 78 | 15 | 19 | 67 | 8 | 12 | **33** | **49** |
| Bordeaux 2 | 76 | 10 | 13 | 72 | 8 | 11 | **37** | **51** |
| Cadillac | 135 | 3 | 2 | 116 | 1 | 1 | **11** | **9** |
| Langon | 59 | 3 | 5 | 49 | 0 | 0 | **3** | **6** |
| La Réole | 48 | 0 | 0 | 29 | 0 | 0 | **1** | **3** |
| Ambès | 75 | 14 | 19 | 61 | 13 | 21 | **18** | **30** |
| Libourne | 121 | 2 | 2 | 102 | 2 | 2 | **4** | **4** |
| Global | 63 | 6 | 10 | 50 | **8** | **16** | **8** | **16** |

sont pourtant assimilées.

La comparaison simultanée des lignes d'eau avec et sans EnKF-$\gamma$-KLBC-MEG sur plusieurs stations de validation de l'estuaire pour les Exp. $R2b$ et $R2c$, représentées sur les figures 6.9 et 6.10, montre un brusque abaissement du niveau d'eau à la fin de l'Exp. $R2c$ propagé depuis la frontière maritime et accentué par la valeur trop grande de $Ks3$. Sur les portions fluviales du modèle, les valeurs supérieures à 90 de $Ks3$ , décorrélant cette variable de contrôle des niveaux d'eau simulés, conduisent à un abaissement de la ligne d'eau sur la partie fluviale.

En conclusion, l'Exp. $R2$ montre le rôle significatif joué par l'introduction dans le vecteur de contrôle de $Ks3$ qui, corrigé simultanément avec $Ks1$, $Ks2$ et le forçage maritime dans l'espace réduit de KL, reporte le bénéfice de l'EnKF-$\gamma$-KLBC-MEG en amont de la confluence entre la Garonne et la Dordogne.

## 6.5 Analyse des performances de l'EnKF-$\gamma$-KLBC-MEG sur un événement réel (Exp. $R3$)

### 6.5.1 Correction des paramètres et des forçages temporels pour une réanalyse d'une semaine

L'Exp. $R3$, similaire à $R2c$, a pour but d'étudier le comportement des variables de contrôle à l'aide de l'EnKF-$\gamma$-KLBC-MEG et de confirmer les résultats des Exp. $R1$ et Exp. $R2$ sur une période plus longue de 8 j. La valeur moyenne de l'ensemble d'ébauche pour les coefficients de Strickler $Ks1$, $Ks2$ et $Ks3$ est fixée au jeu obtenu à l'issue du calage du modèle de prévision des hautes eaux de l'estuaire de la Gironde, soit $[Ks1_0 = 60\,; Ks2_0 = 60\,; Ks3_0 = 50]$. $Ks4$ n'est ni corrigé ni perturbé et est fixé à 40. La durée $L\Delta t$ de chaque fenêtre d'assimilation de données horaires est fixée à 3 h et le décalage $S\Delta t$ entre deux cycles successifs est fixé à 2 h. L'évolution des variables de contrôle au cours de l'Exp. $R3$ est représentée sur les figures 6.11, 6.12 et 6.13.

Alors que les valeurs de $Ks1$ sont généralement comprises dans l'intervalle [20 ; 80], les coefficients de Strickler $Ks2$ et $Ks3$ atteignent, au cours de l'expérience, des valeurs supérieures aux valeurs "usuelles" généralement comprises entre 10 et 90. Une valeur trop forte du coefficient de Strickler révèle l'absence complète d'influence du frottement de la rivière sur l'écoulement et provoque une décorrélation du niveau d'eau avec ce coefficient. De telles valeurs peuvent être pertinentes dans les zones d'eau très profondes, mais le sont moins dans les rivières où le niveau d'eau résulte d'un compromis entre les apports amont, la consigne aval et les frottements. Dans ces zones, un coefficient de Strickler très grand masque un processus physique non pris en compte par ailleurs.
Les signaux temporels des coefficients de Strickler montrent, par ailleurs, des instabilités "haute fréquence" déjà mises en évidence dans le chapitre 5 et qui semblent, d'une part, être liées à l'utilisation de la technique de rééchantillonnage, mais, d'autre part, ici, à la durée $L\Delta t = 3\ h$ de la fenêtre d'assimilation seulement décalée de 2 h du cycle suivant. Les valeurs de $Ks3$ oscillent ainsi autour de 100. Ces oscillations sont également présentes sur le signal du coefficient de Strickler $Ks2$ imposé sur la partie centrale de l'estuaire. Cette variable de contrôle présente un signal avec des valeurs trop élevées pendant la première partie de l'expérience, alors que les valeurs qu'elle atteint sont comprises dans l'intervalle [60 ; 80] après 100 h d'expérience.

Comme le montrent les figures 6.12 et 6.13, les coefficients modaux du forçage maritime présentent une forte variabilité au cours de l'Exp. $R3$ et ne montrent aucun signe de convergence. Cette absence de convergence peut être liée à la méthode de rééchantillonnage et à la contrainte que celle-ci induit sur l'évolution temporelle de la covariance croisée entre la variable de contrôle et l'observation simulée ($\mathbf{P}^{\gamma,\mathbf{y}}$). $\mathbf{P}^{\gamma,\mathbf{y}}$ intervient, en effet, dans le terme d'innovation pour le calcul de l'incrément sur la variable de contrôle (cf. section 5.6) au cours des cycles d'assimilation.

L'évolution temporelle et la correction instantanée de la condition à la limite maritime correspondant aux coefficients modaux sont représentées sur la figure 6.14. La correction instantanée

apportée par l'EnKF-$\gamma$-KLBC-MEG est un signal périodique de même période que la marée qui met en évidence qu'un rephasage du signal maritime est réalisé par l'EnKF-$\gamma$-KLBC-MEG au cours de l'expérience.

Le déphasage et la correction introduits par l'EnKF-$\gamma$-KLBC-MEG par rapport au forçage maritime initial sont respectivement représentés en bleu et en rouge sur la figure 6.15 à l'aide de triangles (resp. de carrés) pour les pics de PM (resp. de BM). Le rephasage global réalisé par l'EnKF-$\gamma$-KLBC-MEG sur les pics de PM et de BM est compris entre [-90 mn ; 40 mn] avec, cependant, une moyenne de -17 mn environ et un écart-type de 32 mn, ne permettant pas de détecter une structure temporelle de l'erreur de déphasage du signal maritime.
La correction réalisée sur les pics de BM est comprise dans l'intervalle [-34 cm ; 45 cm] avec, cependant, une moyenne de 13 cm environ et un écart-type de 17 cm qui montrent que l'EnKF-$\gamma$-KLBC-MEG augmente, en moyenne, le niveau d'eau pour les pics de BM. La correction réalisée sur les pics de PM est comprise entre [-57 cm ; 29 cm] avec une moyenne de 2 cm environ et un écart-type de 20 cm qui ne permet pas de détecter une structure temporelle de l'erreur sur les pics de PM. La condition à la limite maritime n'est, en effet, pas systématiquement translatée vers le haut ou vers le bas et ne subit pas non plus de réduction systématique de son amplitude. La structuration de l'erreur n'apparaît pas non plus liée à la période de marée, ce qui conforte l'hypothèse de découplage entre les incertitudes liées à la marée et aux surcotes météorologiques.

### 6.5.2 Prévisions de niveaux d'eau en réanalyse par estimation des coefficients de frottement et correction du forçage maritime

L'Exp. $R3$ consiste à corriger, sur une durée de 8 jours, les coefficients de frottement $Ks1$, $Ks2$ et $Ks3$ ainsi que le forçage maritime dans l'espace réduit de KL à l'aide de l'EnKF-$\gamma$-KLBC-MEG. Comme pour les expériences réelles précédentes, il s'agit ici d'analyser les gains en termes d'erreur instantanée sur les niveaux d'eau simulés le long de l'estuaire, mais aussi de représentation des pics de PM et de BM. Il faut noter que les mesures aux stations de Port-Bloc, Pauillac et Bordeaux PABx, disponibles toutes les 10 mn sur la durée de l'expérience qui se déroule entre le 01/01/2016 18:00 et le 09/01/2016 06:00, sont assimilées toutes les heures. En revanche, concernant les stations de validation, les données sont lacunaires pour plusieurs d'entre elles. Cela conduit à évaluer les gains en RMSE apportés par l'EnKF-$\gamma$-KLBC-MEG pendant 3.5 jours à partir de 124 h, soit environ 5 jours après le début de l'Exp. $R3$.

Les niveaux d'eau calculés aux 3 stations de l'estuaire dont les mesures sont assimilées sont représentés sur la figure 6.16. L'EnKF-$\gamma$-KLBC-MEG, comme pour l'Exp. $R1$, corrige les erreurs réalisées sur la BM, dans la mesure où le run de contrôle, issu du calage du modèle sur les Nash de PM, représente bien les pics de PM.
Malgré cela, il peut être noté que, sur les stations de Port-Bloc et Pauillac, la différence entre le signal "corrigé" ($\mathbf{y}^a$) issu de l'EnKF-$\gamma$-KLBC-MEG et le signal observé ($\mathbf{y}^0$) est comprise entre -10 cm et +10 cm à partir de 150 h après le début de l'expérience. À Bordeaux, les pics de PM sont

mieux représentés que par le run de contrôle ($\mathbf{y}^b$) de même que les basses eaux (BE) et le pic de BM à partir de 160 h. L'amplitude du signal est également plus proche de celle du signal observé.

La figure 6.17 représente l'évolution des niveaux d'eau aux stations de validation du Verdon (-a), de Bec d'Ambès (-b), de Bordeaux 2 (-c) et de Cadillac (-d) au cours de l'Exp. $R3$.
La performance de l'EnKF-$\gamma$-KLBC-MEG au droit de la station du Verdon et de Bec d'Ambès est médiocre. La représentation des pics de PM est plutôt dégradée, sans qu'un réel bénéfice sur les BM existe sauf sur les deux derniers pics.
La station de validation de Bordeaux 2 est très proche de celle du marégraphe du Port Autonome de Bordeaux (Bordeaux PABx) dont les données sont assimilées au cours de l'Exp. $R3$. Les mêmes constats sont valables : l'amplitude du signal est mieux représentée à l'aide de l'EnKF-$\gamma$-KLBC-MEG, de même que les pics de PM et de BM.
La station de Cadillac est située entre Bordeaux et la Réole à l'amont de la Garonne. Malgré un maillage peu raffiné et une absence d'assimilation des données à la Réole, l'EnKF-$\gamma$-KLBC-MEG permet de mieux représenter les pics de PM et la cambrure du signal, ainsi que le dernier pic de BM. La RMSE en cette station est ainsi divisée par 2, passant de 68 cm à 39 cm sur la période de validation considérée (3.5 jours).

La figure 6.18 représente le déphasage entre l'observation et le signal calculé par Telemac2D sans (-a) et avec (-b) l'EnKF-$\gamma$-KLBC-MEG.
L'EnKF-$\gamma$-KLBC-MEG introduit une amplification du déphasage en retardant les pics de PM à l'amont de Bec d'Ambès sauf à Royan et à Pauillac, ainsi qu'à Bordeaux PABX. En revanche, les pics de PM sont rephasés avec l'observation à Bordeaux 2 et à Cadillac, ainsi qu'à Libourne. Cependant, le pas de temps de sortie graphique des simulations, horaire pour des raisons de stockage, conduit à utiliser une interpolation linéaire et, de fait, ne permet pas d'obtenir une précision supérieure à 30 mn sur le temps correspondant au pic observé ou simulé de PM (ou de BM) entre deux pas de temps horaires de sortie graphique.

La figure 6.19 montre les scores obtenus sans (en bleu) (resp. avec, en rouge) EnKF-$\gamma$-KLBC-MEG pour les stations de l'estuaire où des mesures sont disponibles pendant l'événement. La RMSE intègre les erreurs instantanées du signal par rapport à l'observation. De fait, elle augmente significativement en cas de déphasage du signal. L'analyse des RMSE montre ainsi des gains en RMSE médiocres (inférieurs à 5 cm) à l'aval de Bec d'Ambès. Le gain est compris entre 10 cm et 20 cm de Bordeaux à la confluence et sa valeur atteint 29 cm à Cadillac.

Afin de mieux refléter les apports de l'EnKF-$\gamma$-KLBC-MEG, la figure 6.20 représente l'erreur moyenne sur les pics de PM de l'événement, obtenus après rephasage du signal de marée simulé sur celui observé. Tout en restant inférieure à 10 cm, la RMSE sur les pics de PM est dégradée à Port-Bloc et Pauillac (dont les mesures sont pourtant assimilées), ainsi qu'à Royan. En revanche, sur la partie située en aval de la confluence, elle est améliorée de 6 cm au Verdon et à Laména et est peu influencée par l'EnKF-$\gamma$-KLBC-MEG à Fort-Médoc et à Bec d'Ambès. Sur la Garonne, l'amélioration de la représentation des pics de PM est substantielle, puisque la RMSE sur les pics

de PM décroît de 6 cm à 24 cm entre Le Marquis et Cadillac, conduisant à des RMSE sur les pics de PM inférieures à 18 cm sur l'ensemble de ces stations, alors qu'elles étaient comprises entre 30 cm et 46 cm sans EnKF-$\gamma$-KLBC-MEG.

La figure 6.21 présente le gain sur la représentation du pic de tempête le long de l'estuaire de la Gironde obtenu à l'aide de l'EnKF-$\gamma$-KLBC-MEG en réanalyse de l'événement de début janvier 2016. L'EnKF-$\gamma$-KLBC-MEG permet de calculer parfaitement le niveau d'eau au pic de tempête à Port-Bloc et à Laména. En revanche, l'erreur est augmentée de 2.5 cm à Royan, de 4 cm à Pauillac et à Bec d'Ambès, ce qui est assez faible relativement à la hauteur d'eau sur ces stations. De la station du Marquis à Cadillac, le gain est compris entre 10 cm et 20 cm, ce qui permet de représenter le pic de tempête à Bordeaux avec une erreur de 22 cm, contre 38 cm sans EnKF-$\gamma$-KLBC-MEG.

En conclusion, pour l'Exp. $R3$, l'EnKF-$\gamma$-KLBC-MEG permet de mieux estimer les pics de tempête et l'amplitude du signal sur les stations de l'estuaire. Par ailleurs, les pics de PM sont mieux évalués sur la quasi-totalité des stations en aval du Verdon.

## 6.6 Analyse de la matrice gain K du filtre de Kalman : réseau de mesures idéal et localisation spatio-temporelle de la matrice de covariance croisée

Rappelons que la technique de redispersion fondée sur la persistance de la matrice de covariance d'ébauche selon $\mathbf{P}_0^{\gamma} = \mathbf{P}_k^{\gamma}$ au début de chaque cycle d'assimilation $\mathcal{F}_k$ ($k \in \{1, \cdots, nb_{cycles}\}$) est utilisée dans le cadre de toutes les expériences réelles.

Les composantes de la matrice Gain **K** du filtre de Kalman, c'est-à-dire la contribution apportée par chaque station (en abscisse) à la correction d'une variable de contrôle donnée, sont représentées en fonction du cycle d'assimilation (en ordonnée) sur la figure 6.22 pour l'Exp. $R1$. La figure 6.23 montre, pour chaque coefficient de Strickler inclus dans le vecteur de contrôle de l'Exp. $R3$, à droite, la contribution au gain de Kalman de chaque observation assimilée (de gauche à droite à 0 h d'échéance, 1 h d'échéance et 2 h d'échéance après le début du cycle) et, à gauche, la fonction de covariance de la variable de contrôle avec l'observation en chaque station à 0 h d'échéance.

L'analyse des influences des différentes variables, réalisée dans la section 2.2.1 du chapitre 2 consacrée à l'étude de quantification d'incertitudes, montre que, pour la tempête de Février 2003, la station "idéale" pour corriger le coefficient $Ks1$ est située autour de Richard et que les niveaux d'eau des stations de Laména et du Verdon sont peu influencés par $Ks1$.
Au cours des cycles d'assimilation de l'Exp. $R1$, les contributions principales à la correction de $Ks1$ sont apportées par Le Verdon, Richard et Laména pendant les PM et par toutes les stations jusqu'à Bordeaux pendant les BM au Verdon.
Pour l'Exp. $R3$, aucune donnée n'est disponible à Richard et la station de Laména a été utilisée

comme validation. La station d'observation de Laména se situant à la fois dans la zone d'influence de $Ks1$ et de $Ks2$, il peut sembler pertinent d'assimiler les mesures relevées à Laména en l'absence de mesures à Richard. Les résultats de l'expérience synthétique Exp. 3-c indiqués dans le tableau 5.7 du chapitre 5 montrent, d'une part, que l'assimilation des données à Laména améliore la RMSE sur les stations de Port-Bloc, Royan et Le Verdon, mais de manière beaucoup moins significative que pour l'Exp. 3-a qui utilise les données de Royan et Richard. En revanche, la figure 6.24 montre une bonne reconstruction du signal de Ks1 pour l'Exp. 3-c.

Pendant l'Exp. $R1$, $Ks2$ (resp. $Ks3$) reçoit principalement les contributions des stations du Verdon, Richard et Laména (resp. Bassens, Bordeaux PABX et Bordeaux2), ainsi que de Pauillac, Fort-Médoc et Bec d'Ambès. Enfin, $Ks4$ reçoit une contribution continue de La Réole, ainsi que des stations situées sur le tronçon de la Garonne, en amont de la confluence. Ces constats sont cohérents avec les zones d'influence déterminées par l'étude de quantification d'incertitudes décrite dans le chapitre 2. Les stations citées ci-dessus contribuent sans surprise sur l'ensemble de l'expérience de manière significative à la correction des variables de contrôle. Cependant, l'analyse des matrices de gain de Kalman montre des contributions qu'il est possible de qualifier de "malvenues".

Pour l'Exp. $R2c$, la représentation pour la variable de contrôle $Ks3$ de la matrice de gain $\mathbf{K}$ du filtre de Kalman représentée sur la figure 6.25 au fil des cycles d'assimilation révèle, de la même manière que pour l'Exp. $R1$, quelques influences croisées susceptibles d'être induites par le faible nombre de membres choisi et la solution de redispersion retenue. Ces corrélations artificielles conduisent à corriger une variable de contrôle à l'aide de mesures à des stations qu'elle n'est pas censée influencer. La contribution de Pauillac à la correction de $Ks3$ semble ainsi supérieure à celle de Bordeaux au cours des cycles d'assimilation de l'Exp. $R2c$. Il convient également de noter la contribution à la correction de $Ks2$, certes faible, de l'observation à la Réole.

De la même manière, pour l'Exp. $R3$, la fonction de covariance entre $Ks1$ et les stations d'observation, représentée sur la figure 6.23-a, est un signal de même période que la marée et d'amplitude plus élevée pour les stations de Pauillac et Bordeaux que pour Port-Bloc. Le gain de Kalman correspondant, représenté sur la figure 6.23-b, montre que toutes les stations contribuent à la correction de $Ks1$. Cependant, les observations de Pauillac à toutes les échéances contribuent fortement, alors que seule l'observation de Port-Bloc à 1 h, de même que les observations à Bordeaux, contribuent sur la deuxième partie de l'Exp. $R3$. Par conséquent, il semble que l'évolution de $Ks1$ (Fig. 6.11-a) obtenue par l'assimilation des observations à Port-Bloc, Pauillac et Bordeaux PABX résulte de covariances intempestives sur de longues distances et liées au nombre de membres $N_e$. La solution radicale consisterait à supprimer $Ks1$ du vecteur de contrôle, l'effet collatéral étant d'empêcher un éventuel rephasage du signal simulé sur les observations si nécessaire. Une technique de localisation implicite faisant intervenir l'indice de Sobol' calculé sur un cycle de marée comme fonction de pondération de l'influence des stations d'observation sur la matrice de covariance $\mathbf{P}^{\boldsymbol{\gamma},\mathbf{y}}$ ou la matrice gain de Kalman $\mathbf{K}$ pourrait être envisagée.

Les valeurs trop élevées atteintes par $Ks2$ jusqu'à 100 h après le début de l'Exp. $R3$ conduisent physiquement et numériquement à une décorrélation de $Ks2$ des niveaux d'eau calculés par le modèle, ce qui peut être observé sur la figure 6.23-c jusqu'au cycle d'assimilation $\mathcal{F}_{50}$ environ, soit 100 h également après le début de l'Exp. $R3$. Sur la 2ème partie de l'Exp. $R3$, $Ks2$ est à juste titre corrélé avec les mesures à Pauillac et à Bordeaux et deux fois moins avec les observations à Port-Bloc. C'est également ce qui est observé sur la figure 6.23-d qui montre que les composantes de la matrice gain de Kalman **K** les plus élevées correspondent à Pauillac et Bordeaux.
En revanche, le fait que les mêmes conclusions puissent être faites concernant les contributions respectives des stations "assimilées" à la correction de $Ks3$ pose de nouveau la question de la pondération des contributions de chaque station d'observation par la variance censée être apportée à son niveau d'eau par $Ks3$. L'utilisation en tant que fonction de pondération spatio-temporelle de l'indice de Sobol' total pour $Ks3$ le long de l'estuaire paraît justifiée. Il peut ainsi être noté qu'alors que la limite aval de la zone d'influence de $Ks3$ est située à la confluence entre la Garonne et la Dordogne, la figure 6.23-f montre une contribution significative des mesures à Pauillac pour la correction de $Ks3$.

L'existence de ces corrélations intempestives est largement traitée dans la littérature. Elle est associée à la faible valeur de $N_e$. Evensen [2009b] confirme ainsi que l'EnKF étant une méthode de Monte-Carlo, elle converge en $\sqrt{N_e}$. De fait, l'augmentation du nombre de membres ne permet pas toujours un gain en termes de précision, comme cela a pu être constaté dans la section 5.5.2 du chapitre 5. Bien que la variabilité de l'état des systèmes et des paramètres est souvent bien représentée par un ensemble d'une centaine de membres, il reste que l'approximation de la matrice de covariance d'erreur d'ébauche par les membres de l'ensemble crée des corrélations sur de grandes distances ou bien entre des variables supposées décorrélées, comme il l'a déjà été évoqué dans la section 3.4.3.1 du chapitre 3.
Ces résultats semblent montrer la nécessité de réfléchir à l'implémentation de techniques de localisation qui sont couramment utilisées pour les variables d'état (Etala et al. [2015], Chen and Oliver [2010]), mais également pour l'estimation de paramètres dans le cadre, notamment, de la modélisation des écoulements souterrains (Nan and Wu [2011], Devegowda et al. [2010]). Deux techniques de localisation, explicite et implicite, sont décrites dans Nan and Wu [2011] qui étudient, à l'aide d'expériences jumelles d'assimilation des mesures synthétiques de charge hydraulique et de conductivité hydraulique, l'influence d'une localisation implicite. Celle-ci est gouvernée par une fonction de localisation de la covariance sur l'estimation de la conductivité hydraulique d'un système hydrogéologique à l'aide d'un EnKF déterministe (DEnKF) dissociant, dans son étape d'analyse, la moyenne et l'anomalie du vecteur de contrôle. La localisation implicite utilisée ne tronque pas les corrélations, mais applique un poids à la covariance qui est fonction d'une distance d'influence appelée longueur de filtrage. Les objectifs de la localisation sont, d'une part, de garder la structure de covariance réelle entre les variables et les stations d'observation issue de la dynamique du système, tout en supprimant les bruits induits par une taille d'ensemble trop petite. Il résulte de cette étude la difficulté de caler la longueur de filtrage pour

ne pas perdre d'information tout en supprimant la composante bruitée de la covariance.
Le modèle numérique de l'estuaire de la Gironde est construit de telle sorte que son domaine est sous influence de la condition à la limite maritime $CLMAR$, dont les perturbations sont amplifiées au fur et à mesure de la remontée de l'information dans l'estuaire du fait de son caractère hypersynchrone. Les frontières amont correspondent ainsi aux limites d'influence de celle-ci. Il apparaît ainsi délicat d'établir une méthode de localisation spatiale constante dans le temps pour filtrer l'influence de $CLMAR$ en fonction d'une longueur de filtrage. Un filtrage spatio-temporel de l'influence de $CLMAR$ paraîtrait plus indiqué. De la même manière, concernant les coefficients de frottement $Ks1$, $Ks2$, $Ks3$ et $Ks4$, les indices de Sobol' totaux, représentés sur la figure (9) du chapitre 2, pourraient permettre d'établir des poids aux stations d'observation pour pondérer leur influence dans la correction des différentes variables de contrôle. Une technique de localisation spatio-temporelle pour ces paramètres pourrait donc être testée. Il convient cependant de garder à l'esprit que l'étude de quantification d'incertitudes a été menée pour l'événement particulier de Février 2003 et qu'il est possible que les zones d'influence fluctuent pour d'autres conditions hydrométéorologiques.

## 6.7 Conclusions sur les performances de l'EnKF-$\gamma$-KLBC-MEG en réanalyse

Les performances de l'EnKF-$\gamma$-KLBC-MEG se mesurent selon différents critères. D'une part, un bon modèle de prévision des hautes eaux de l'estuaire de la Gironde doit être capable de bien représenter les niveaux d'eau de PM et les niveaux d'eau extrêmes sur les différents points d'intérêt de l'estuaire. Les résultats de l'Exp. $R3$ (cf. figures 6.20 et 6.21) montrent une amélioration du calcul des PM sur la quasi-totalité des stations en aval du Verdon, ainsi qu'une meilleure estimation des pics de tempête. Par ailleurs, l'amplitude du signal est mieux représentée. Même si le signal est principalement amélioré pendant les BM et que l'intégration de $Ks3$ conduit à privilégier une qualité des résultats sur les tronçons fluviaux et à l'embouchure au détriment de la partie centrale de l'estuaire et malgré l'équifinalité sur les variables de contrôle, l'EnKF-$\gamma$-KLBC-MEG représente mieux les niveaux d'eau que la simulation de contrôle (MEG0V7P3).

L'assimilation de données réelles au cours des Exp. $R1$, $R2$ et $R3$ met aussi en lumière plusieurs éléments concernant l'estimation des paramètres et des coefficients modaux.
D'une part, une variabilité temporelle des coefficients de Strickler, liée à la périodicité de la marée dans le cas de l'Exp. $R3$, est mise en évidence. En revanche, les expériences $R1$ et $R3$ ne permettent pas de définir de structure temporelle de correction des coefficients de frottement autour d'une valeur moyenne.
Les expériences réelles sur un cycle de marée (Exp. $R2$) et sur une semaine (Exp. $R3$) montrent la sensibilité de l'EnKF-$\gamma$-KLBC-MEG à l'intégration de la variable $Ks3$ dans le vecteur de contrôle.

La mise en place d'un espace support borné d'évolution des coefficients de Strickler semble être nécessaire pour éviter que ces derniers n'atteignent des valeurs inacceptables, les décorrélant des niveaux d'eau simulés. Il faut, cependant, noter que le fait d'atteindre des valeurs supérieures à 100 ne compromet pas la stabilité de l'EnKF-$\gamma$-KLBC-MEG qui revient vers des valeurs plus physiques sur les cycles d'assimilation suivants. Ainsi, il pourrait être envisagé dans des travaux ultérieurs de maintenir les coefficients de Strickler dans un intervalle physiquement acceptable, tel que [10 ; 90] par exemple.

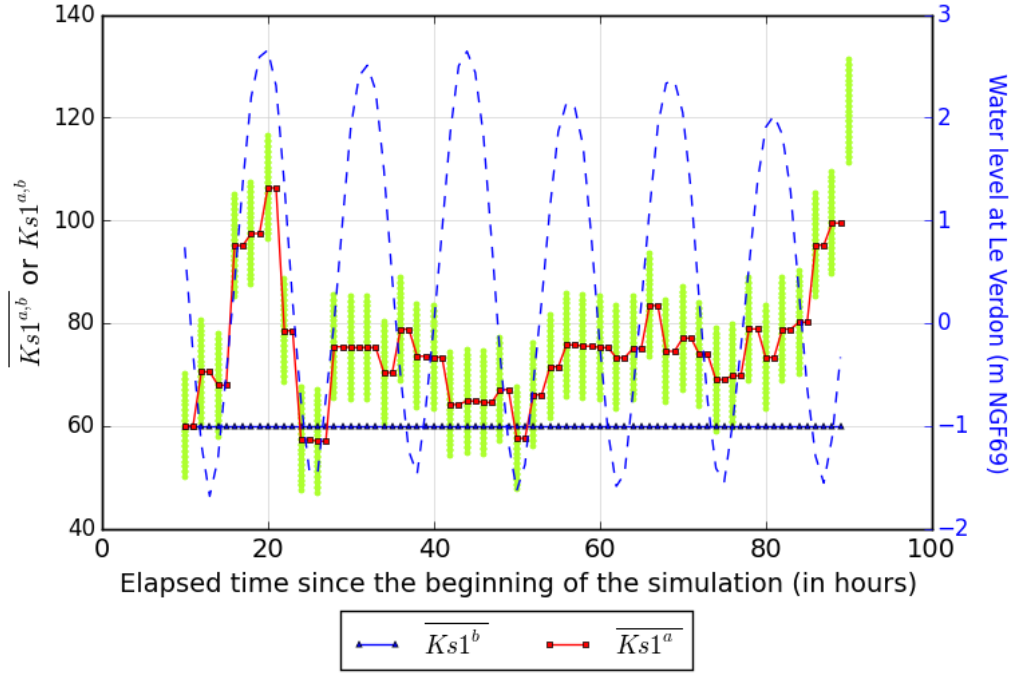
Plusieurs perspectives de travail doivent être envisagées pour améliorer ces performances.
D'une part, il semble nécessaire de considérer que le réseau de mesures jugé "idéal" à l'issue du chapitre 5 consacré aux expériences synthétiques, est susceptible de ne pas correspondre exactement à celui du cadre réel. De fait, l'Exp. R3 incite à assimiler les observations à Laména, en plus de celles à Port-Bloc, Richard, Pauillac et Bordeaux.
Une longue durée $L\Delta t$ de fenêtre d'assimilation de données contribue à lisser les oscillations haute fréquence résultant du rééchantillonnage pour les coefficients de frottement. Cependant, elle peut conduire à mélanger le nature des erreurs sur le niveau d'eau en phase de flot et de jusant. Il importe donc de trouver un compromis cohérent avec les processus physiques en jeu et permettant d'éviter la divergence de l'EnKF-$\gamma$-KLBC-MEG.

Concernant les coefficients modaux, le rééchantillonnage conduit également à imposer une fonction de covariance temporelle entre le coefficient modal et les stations d'observation proportionnelle au mode propre associé, comme cela l'a été démontré mathématiquement dans le chapitre 4. Ainsi, la structure temporelle de la corrélation des perturbations appliquées aux conditions aux limites est transportée vers la corrélation spatiale entre observations et variables de contrôle issues des forçages. Un réajustement du cyclage des fenêtres d'assimilation de données a été mis en œuvre afin de ne pas mélanger les erreurs liées au BM des erreurs liées au pleines mers (PM). Malgré cela, les Exp. $R2$ et $R3$ montrent l'absence de convergence des coefficients modaux liés au forçage maritime, ce qui ne permet pas de définir de structure temporelle d'erreur spécifique. Un rephasage du signal est réalisé par l'EnKF-$\gamma$-KLBC-MEG, mais les corrections apportées au signal maritime rephasé ne montrent pas de correction systématique sur les PM ou les BM au cours de l'Exp. $R3$.
La persistance de la corrélation temporelle entre la covariance d'erreur d'ébauche du vecteur de contrôle et les observations simulées pour les coefficients modaux peut être à l'origine d'une absence de convergence de ces derniers et de la difficulté à déterminer une structuration physique temporelle de l'erreur sur la condition à la limite maritime, telle qu'espérée en découplant le signal de marée, déterministe, et le signal de surcotes, stochastique.
Dans la mesure où, ***a priori***, la technique de rééchantillonnage est seulement nécessaire pour éviter la sous-dispersion des paramètres du modèle au fur et à mesure des cycles d'assimilation, il doit donc être envisagé de ne redisperser les membres que pour les coefficients de frottement. Des investigations supplémentaires seraient nécessaires pour déterminer si l'absence de redispersion par rééchantillonnage des coefficients modaux permettrait de favoriser leur convergence

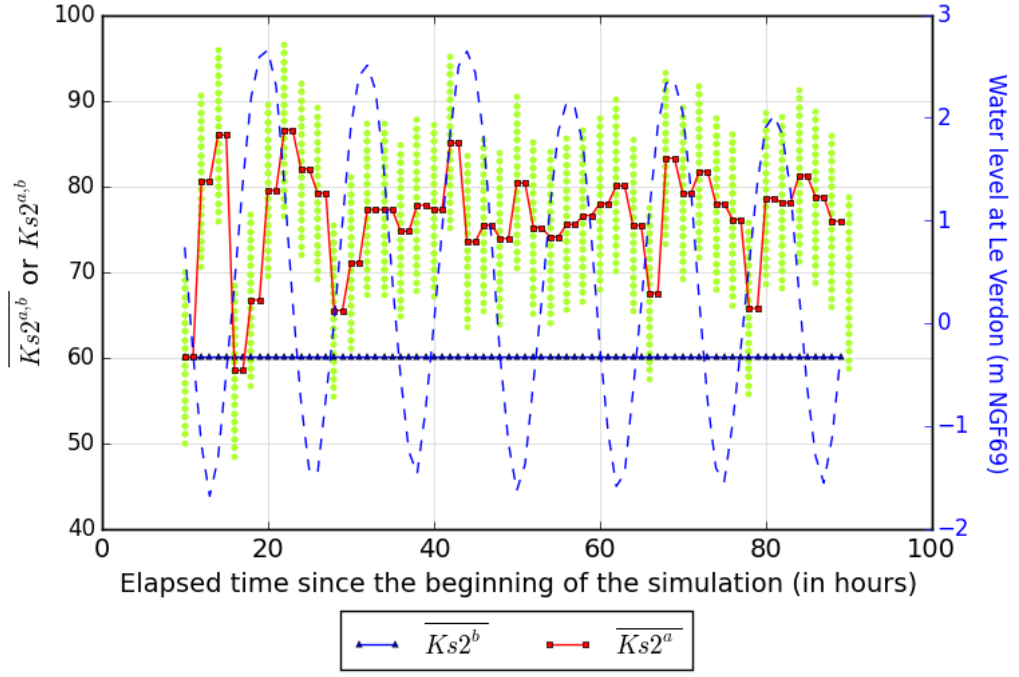
dans le temps. En effet, des expériences jumelles, non montrées ici, mettent en lumière que les coefficients modaux liés aux conditions aux limites subissent peu de réduction de leur variance au cours des cycles d'assimilation dans un cadre synthétique.

La sensibilité de l'EnKF-$\gamma$-KLBC-MEG à la présence de $Ks3$ dans le vecteur de contrôle a été mise en évidence et conforte la nécessité de considérer une méthode d'AD spécifique au traitement des processus physiques non-linéaires dans la zone de la confluence entre les rivières Garonne et Dordogne. Un compromis entre une méthode plus coûteuse en ressources informatiques et un maillage permettant de mieux représenter ces processus doit être trouvé.
Si le rééchantillonnage sur les coefficients modaux est utilisé, une solution consisterait à émuler l'EnKF-$\gamma$-KLBC-MEG par la matrice de covariance d'ébauche $\mathbf{P}^{\boldsymbol{\gamma}}_{0,lim}$ obtenue avec 10000 membres à l'issue de l'étude de quantification d'incertitudes du chapitre 2. Cette proposition peut permettre de réduire les coûts de calcul en transformant l'EnKF-$\gamma$-KLBC-MEG en filtre de Kalman déterministe dans une première étape et en supprimant ainsi les bruits introduits par l'échantillonnage de l'ensemble.
Il est alors possible d'envisager d'utiliser l'IEnKS décrit dans Bocquet and Sakov [2013], théoriquement adapté aux processus non-linéaires du système estuarien, pour descendre en profondeur sur chaque cycle d'assimilation et améliorer l'estimation des variables de contrôle à partir du vecteur de contrôle analysé issu de la première émulation.

Enfin, le faible nombre de membres de l'ensemble $N_e$ (100 ici) conduit à des covariances intempestives entre des variables de contrôle et des stations d'observation ***a priori*** non influencées par ces variables. Les performances de l'EnKF-$\gamma$-KLBC-MEG sont déterminées par la matrice gain de Kalman $\mathbf{K}$ analysée pour chaque événement réel dans la section 6.6. Cette analyse met en évidence la nécessité, particulièrement forte dans un cadre d'expérience réelle, de mettre en place une technique de localisation spatio-temporelle pour diminuer les contributions intempestives de certaines stations d'observation à la correction des variables. Il peut être intéressant par la suite de développer une méthode de localisation originale à l'aide d'une fonction de filtrage spatio-temporelle appliquée à la matrice de covariance $\mathbf{P}^{\boldsymbol{\gamma},\mathbf{y}}_{k}$ et fondée sur les indices de Sobol' totaux issus de l'étude de quantification d'incertitudes du chapitre 2.

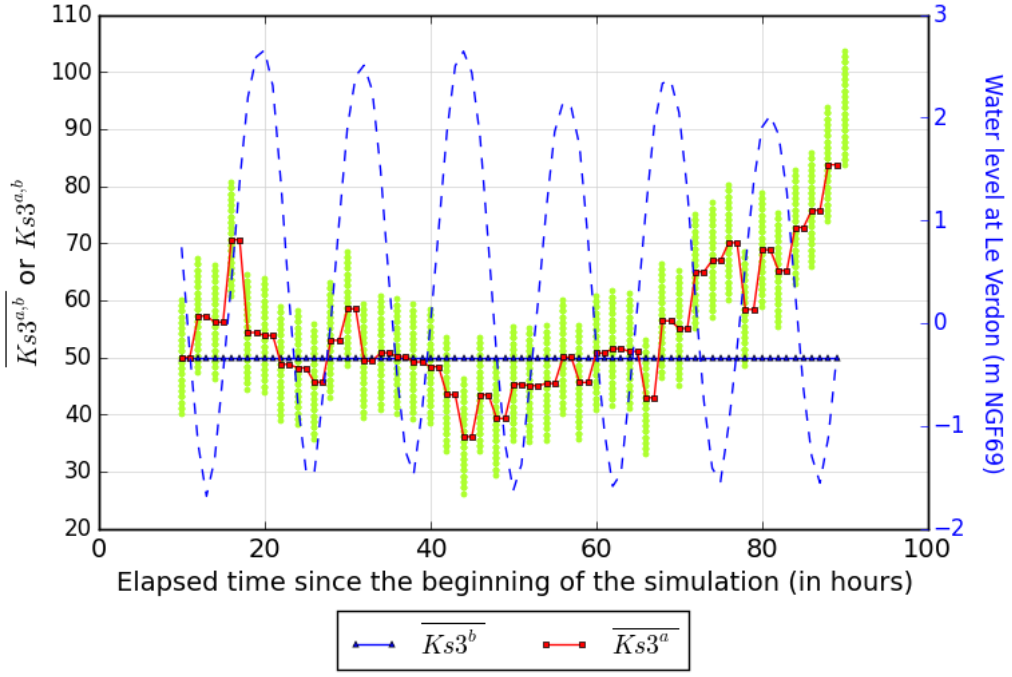



(a) $Ks1$

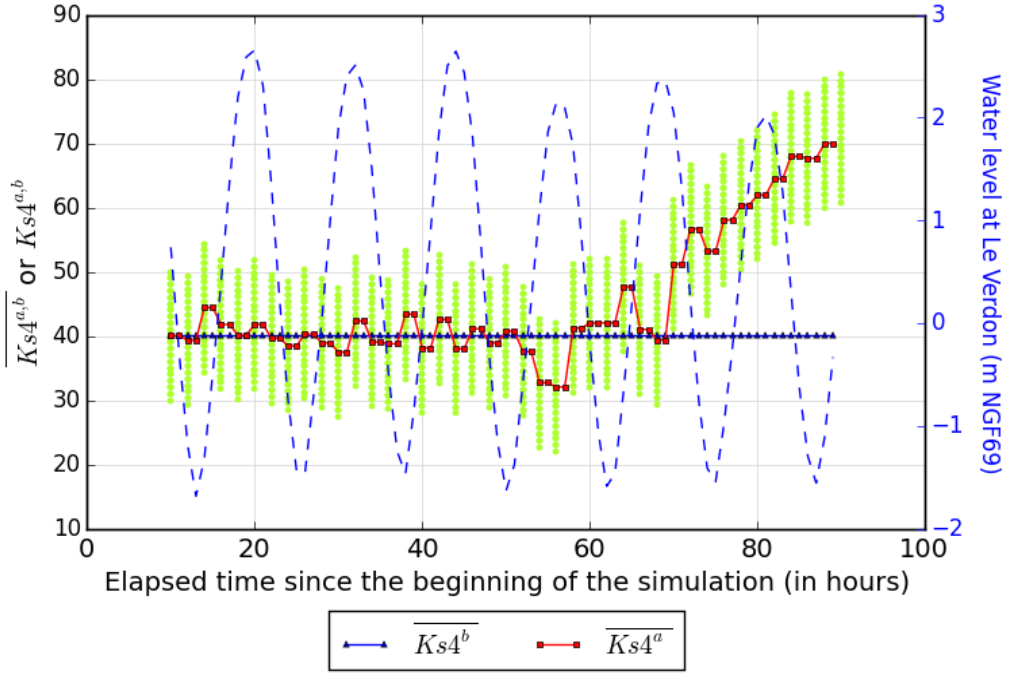



(b) $Ks2$

FIGURE 6.2 : Évolution de $Ks1$, $Ks2$, $Ks3$ et $Ks4$ pour l'Exp. $R1$ au cours des cycles d'assimilation (L=6h - S=2h). L'ensemble est représenté par des cercles verts. (début)

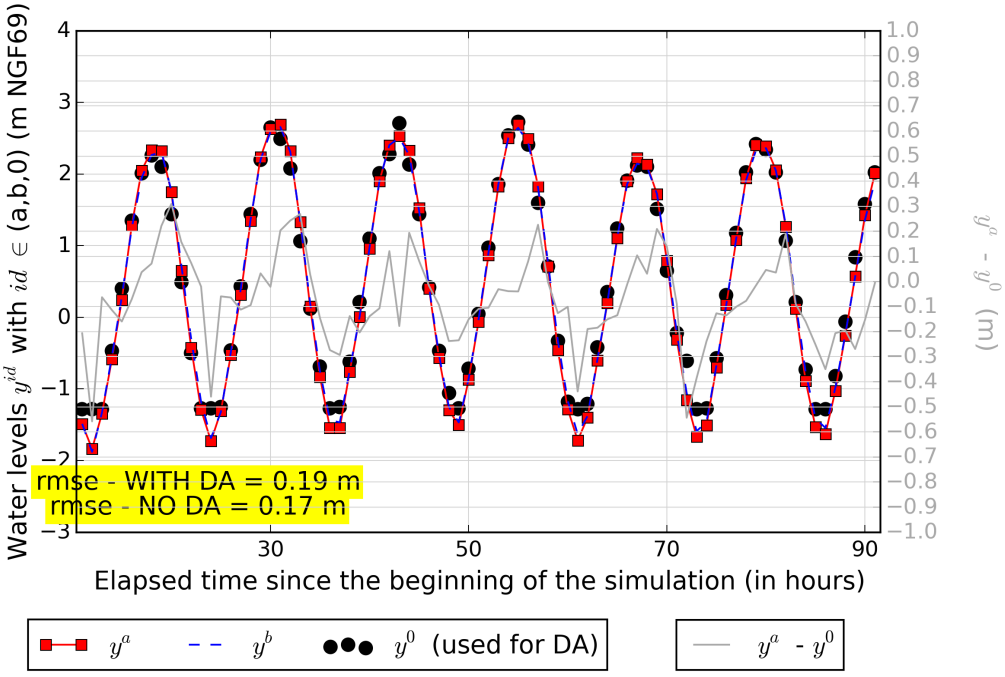



(c) $Ks3$

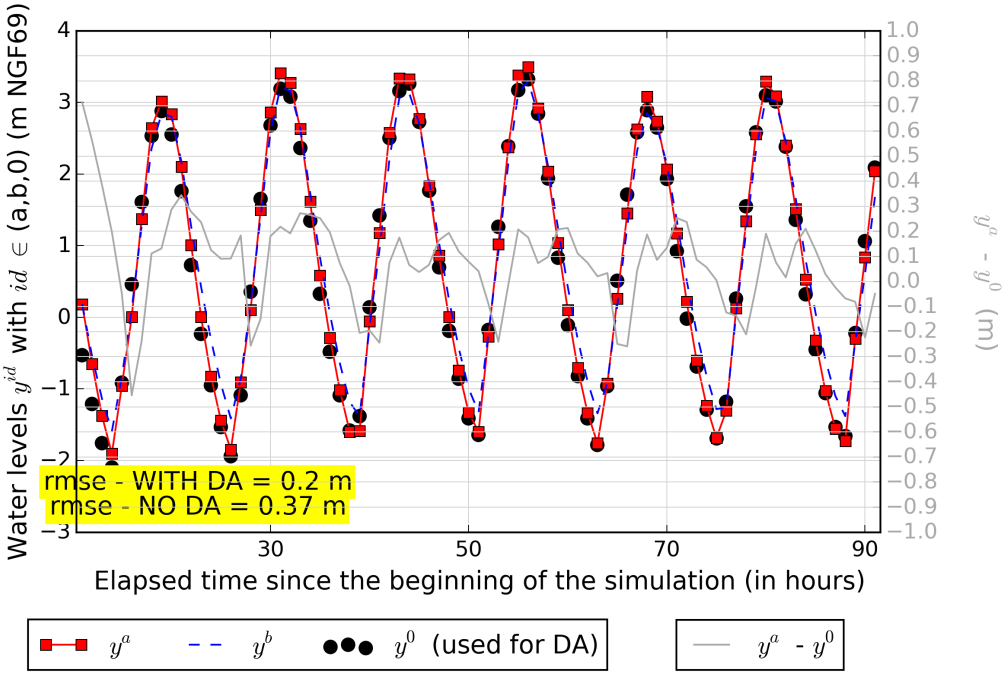



(d) $Ks4$

FIGURE 6.2 : Évolution de $Ks1$, $Ks2$, $Ks3$ et $Ks4$ pour l'Exp. $R1$ au cours des cycles d'assimilation (L=6h - S=2h). L'ensemble est représenté par des cercles verts. (fin)

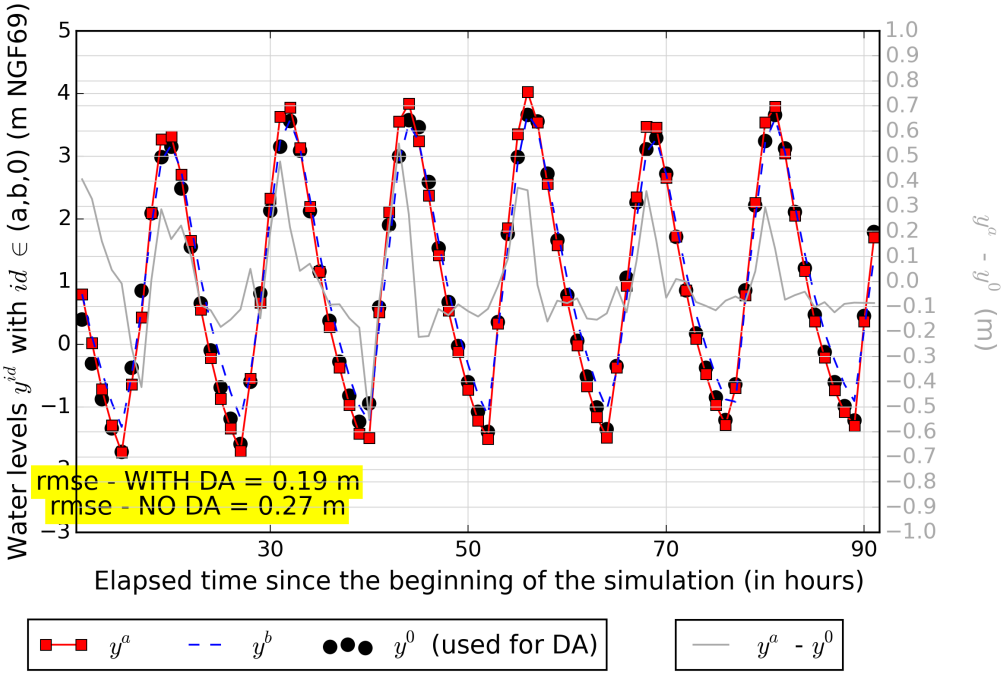



(a) Le Verdon

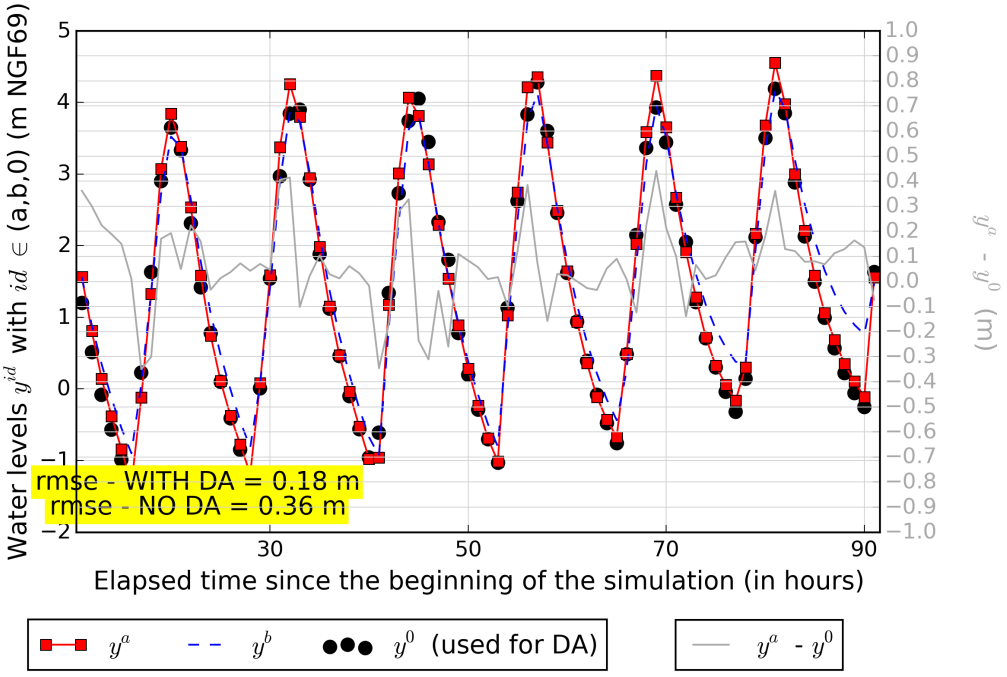



(b) Pauillac

FIGURE 6.3 : Évolution des niveaux d'eau pour l'expérience réelle Exp. $R1$ au cours des cycles d'assimilation (L=6h - S=2h) (début)

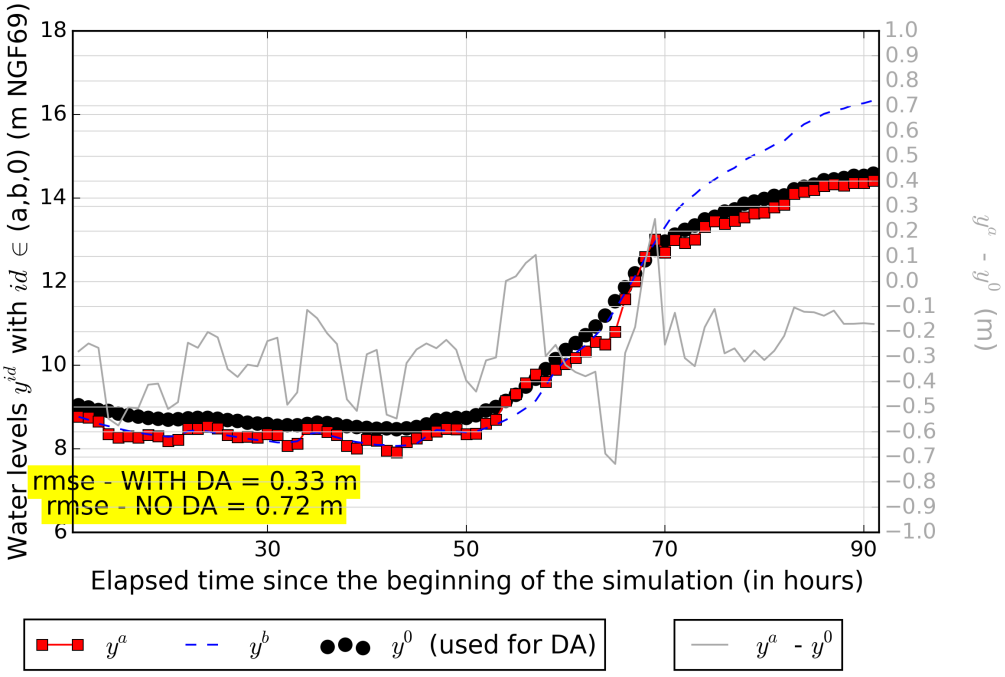



(c) Bec d'Ambès

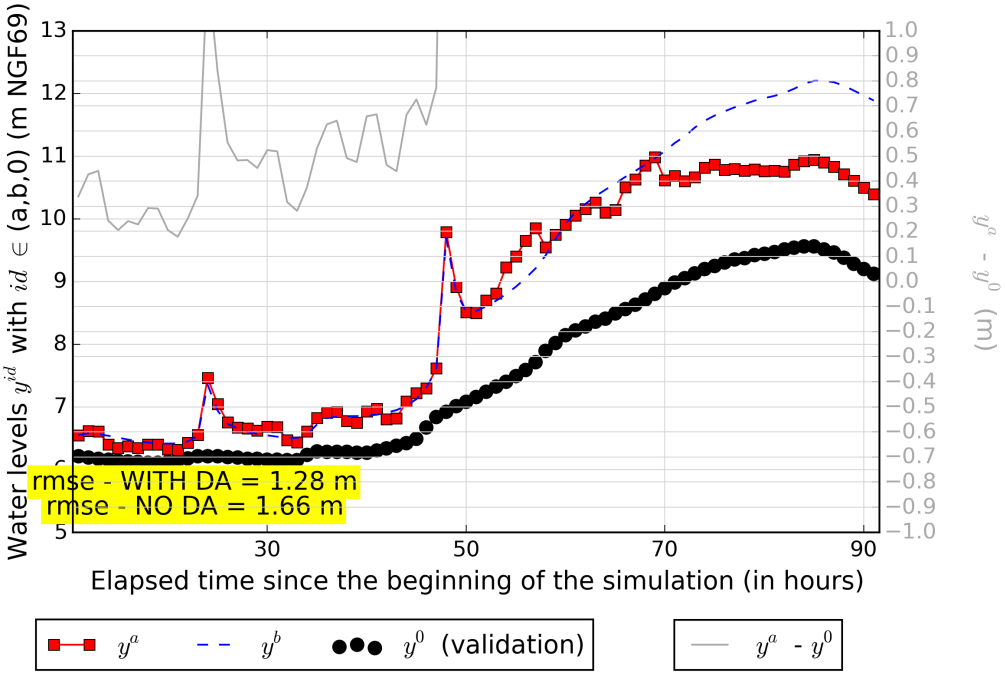



(d) Bordeaux 2

FIGURE 6.3 : Évolution des niveaux d'eau pour l'expérience réelle Exp. $R1$ au cours des cycles d'assimilation (L=6h - S=2h) (suite)

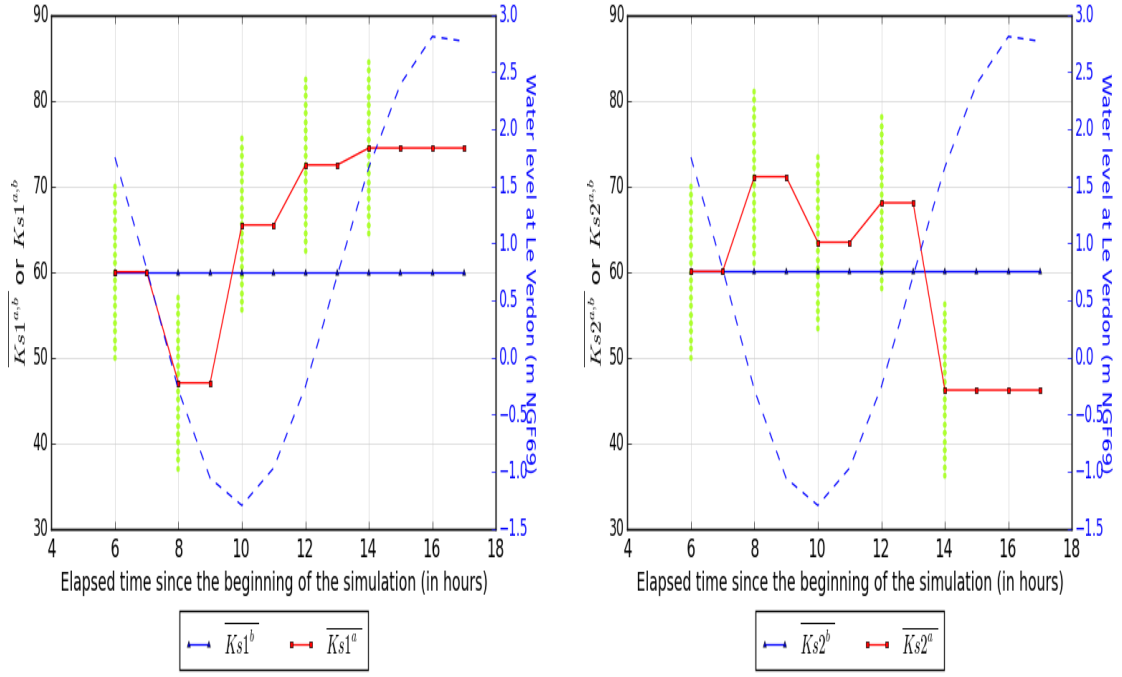



(e) La Réole



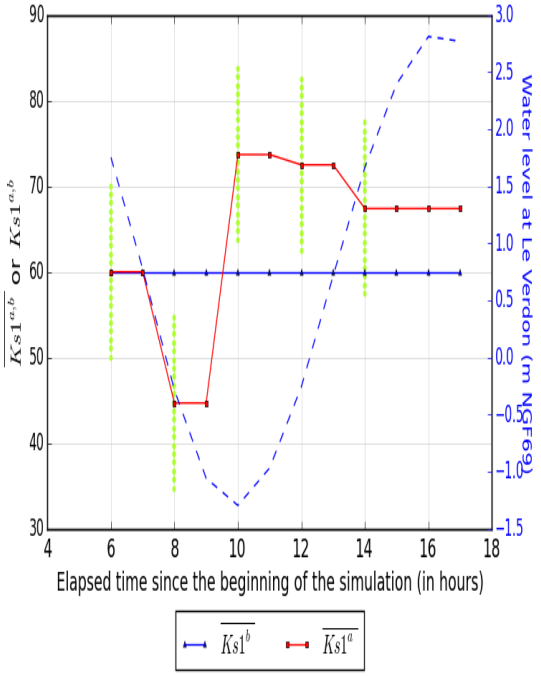

(f) Pessac

FIGURE 6.3 : Évolution des niveaux d'eau pour l'expérience réelle Exp. $R1$ au cours des cycles d'assimilation (L=6h - S=2h) (fin)

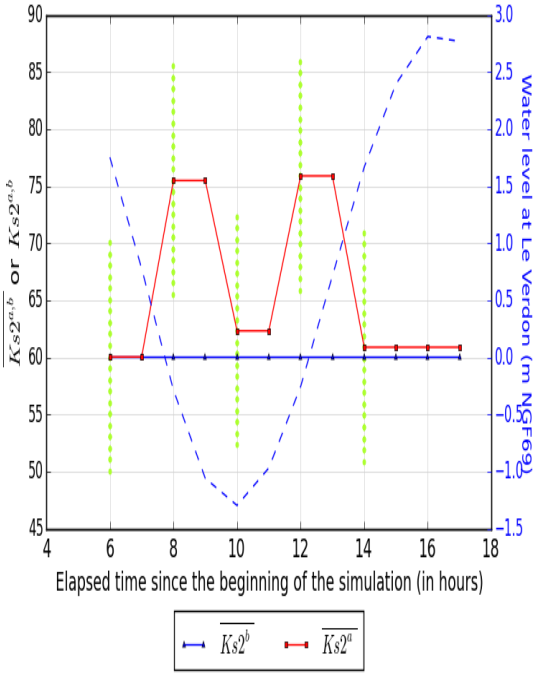



(a) $Ks1$ (b) $Ks2$

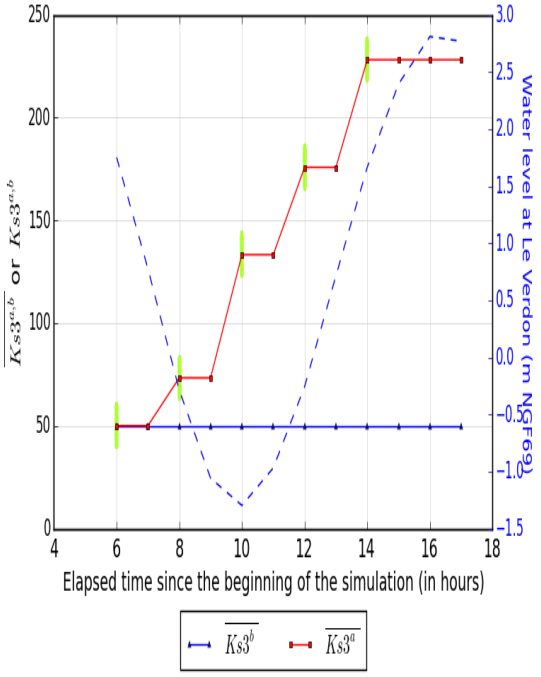
FIGURE 6.4 : Évolution de $Ks1$ et $Ks2$ pour l'Exp. $R2b$ au cours des cycles d'assimilation (L=4h - S=2h). L'ensemble est représenté par des cercles verts.




(a) $Ks1$

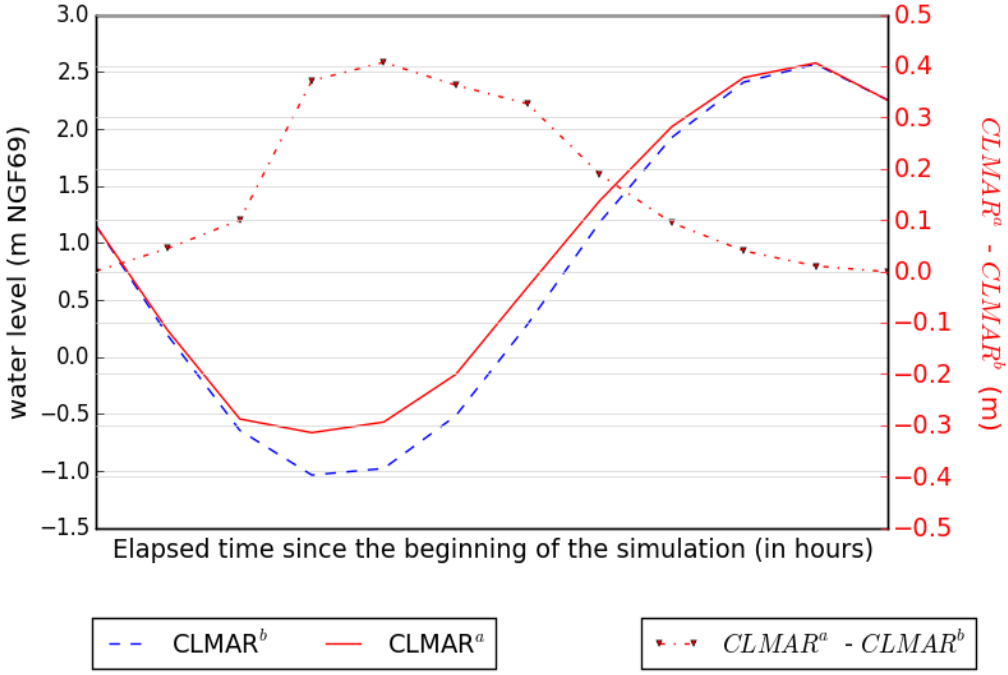



(b) $Ks2$

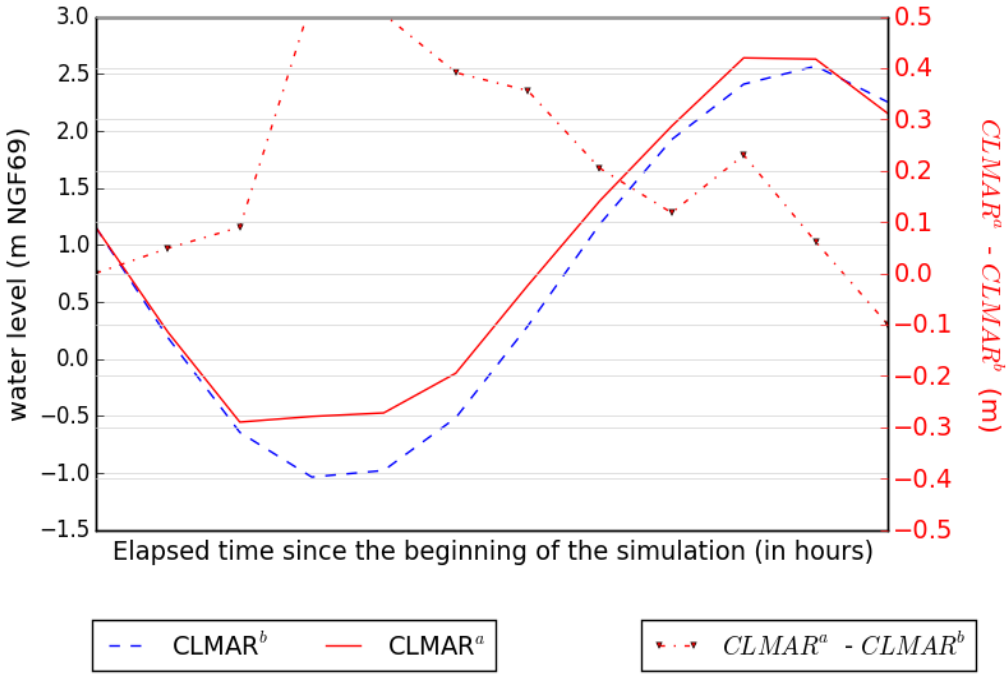



(c) $Ks3$

FIGURE 6.5 : Évolution de $Ks1$, $Ks2$ et $Ks3$ pour l'Exp. $R2c$ au cours des cycles d'assimilation (L=4h - S=2h). L'ensemble est représenté par des cercles verts.

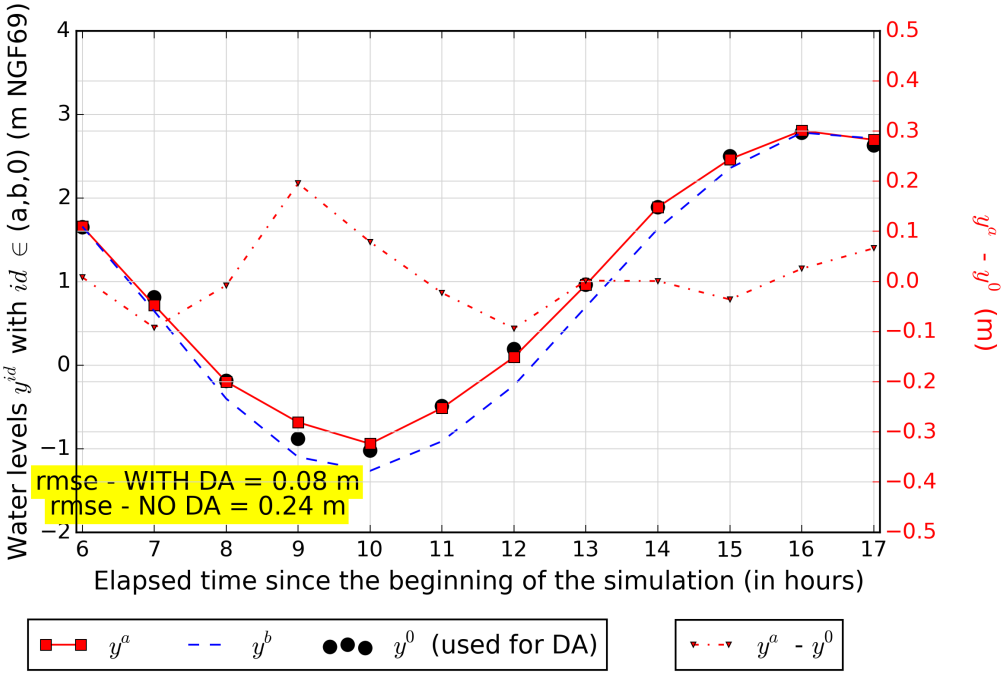



(a) $R2b$

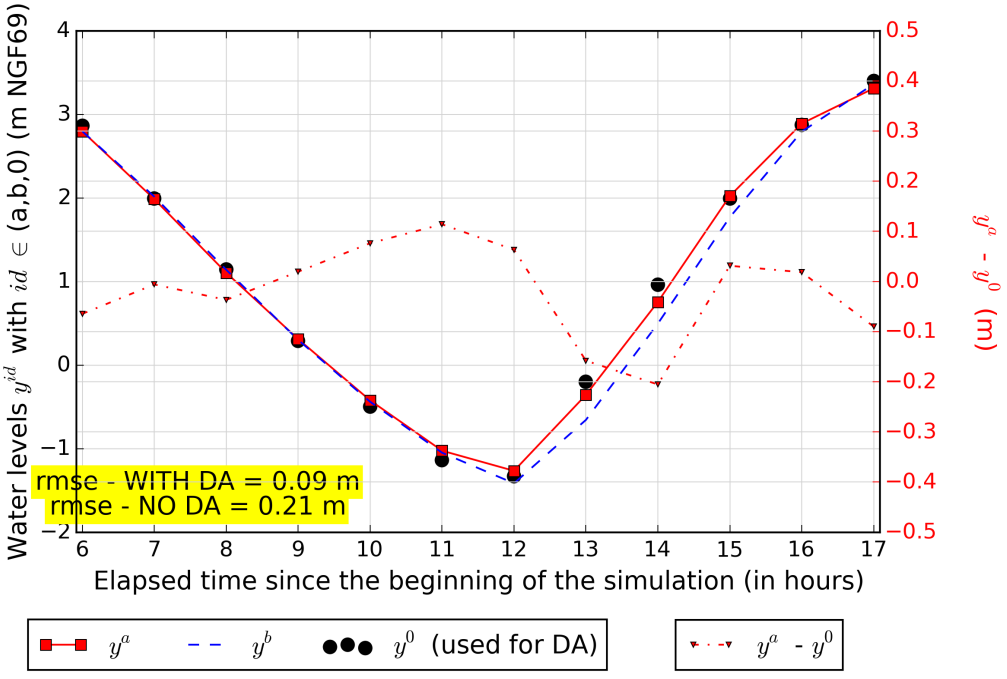



(b) $R2c$

FIGURE 6.6 : Conditions aux limites du modèle pour les expériences réelles $R2b$ et $R2c$

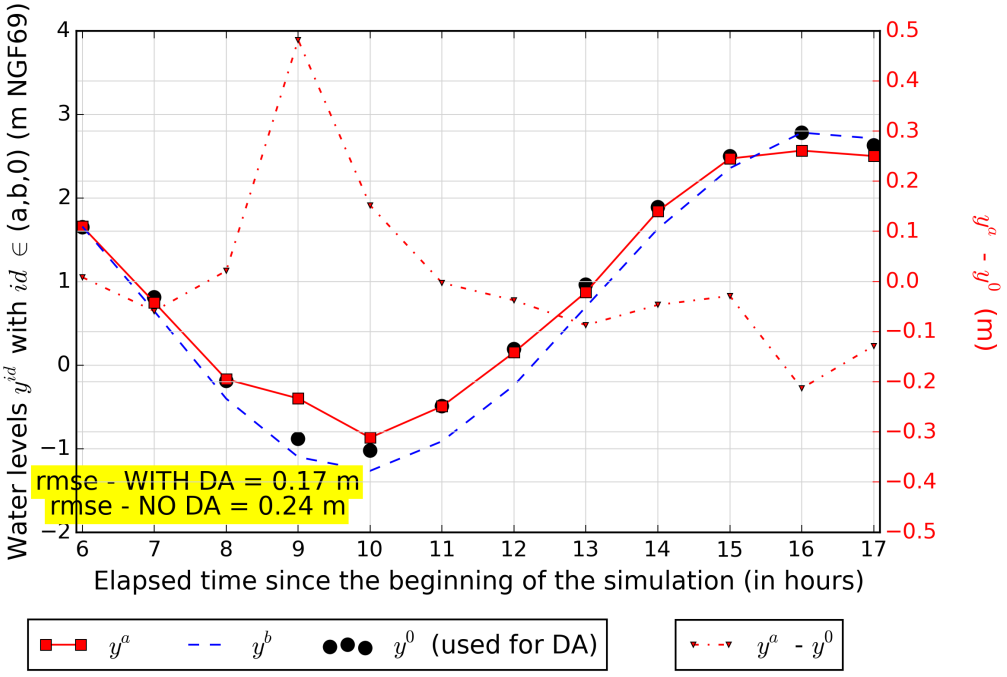



(a) Port-Bloc

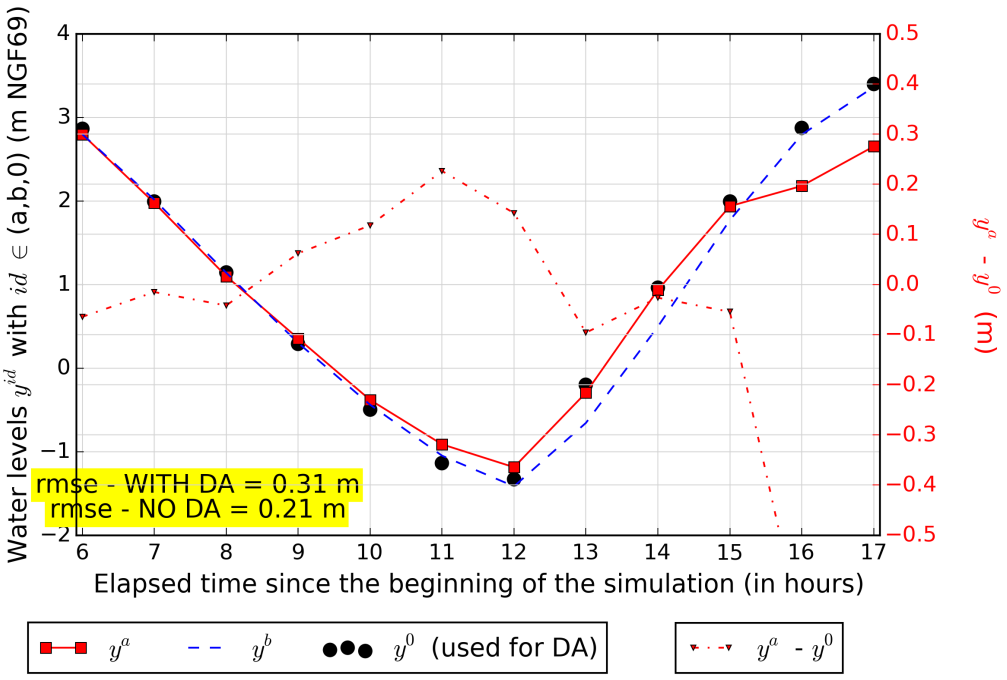



(b) Pauillac

FIGURE 6.7 : Évolution des niveaux d'eau sur les stations "assimilées" pour l'Exp. *R2b*

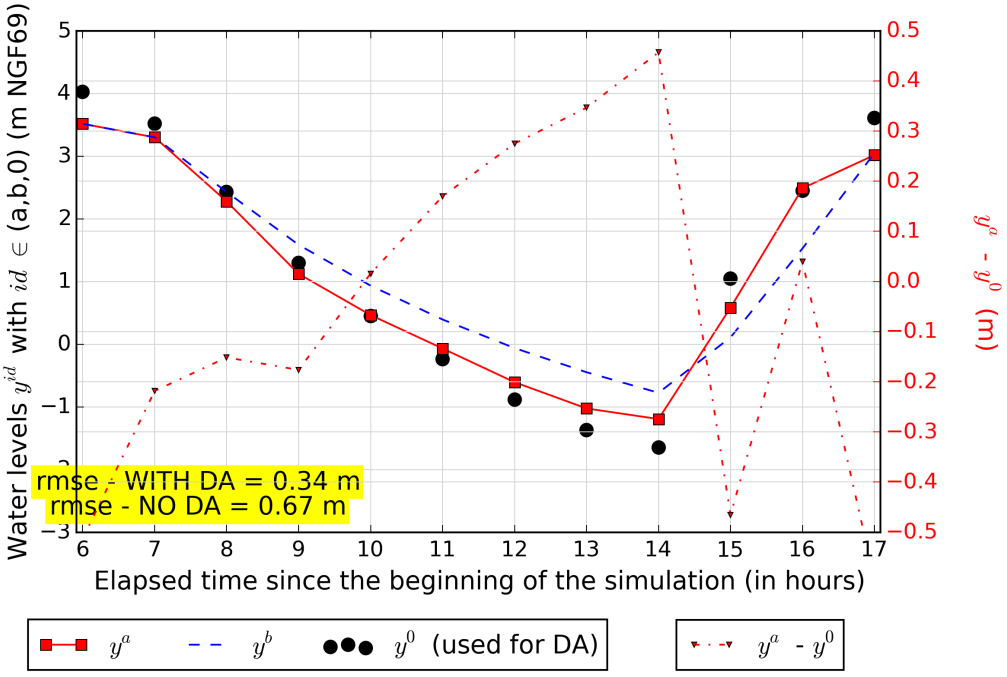



(a) Port-Bloc



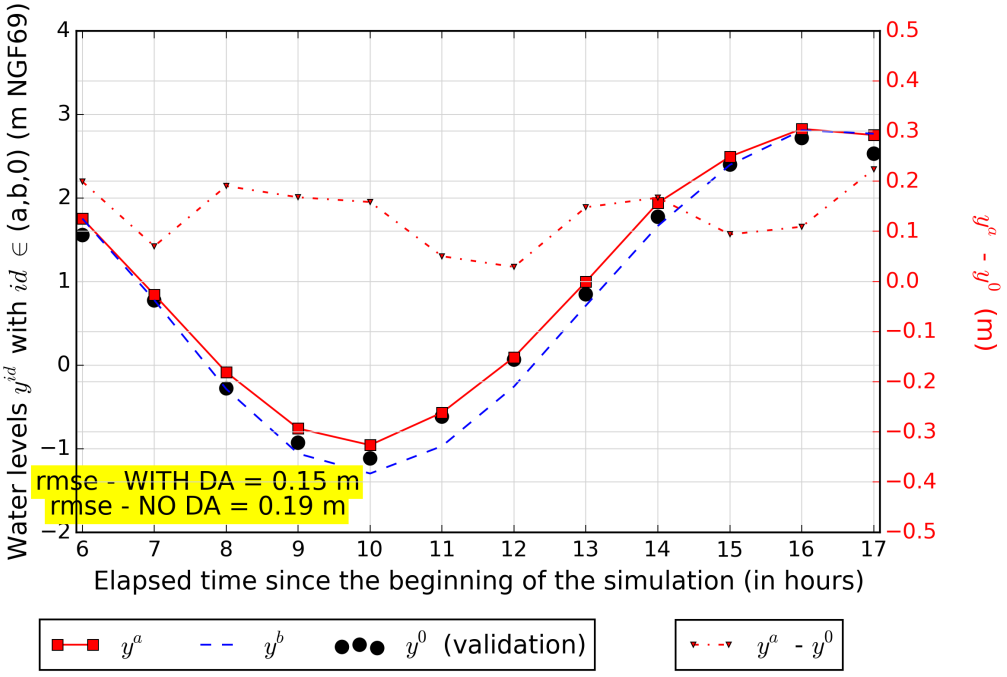

(b) Pauillac

FIGURE 6.8 : Évolution des niveaux d'eau sur les stations "assimilées" pour l'Exp. $R2c$ (début)

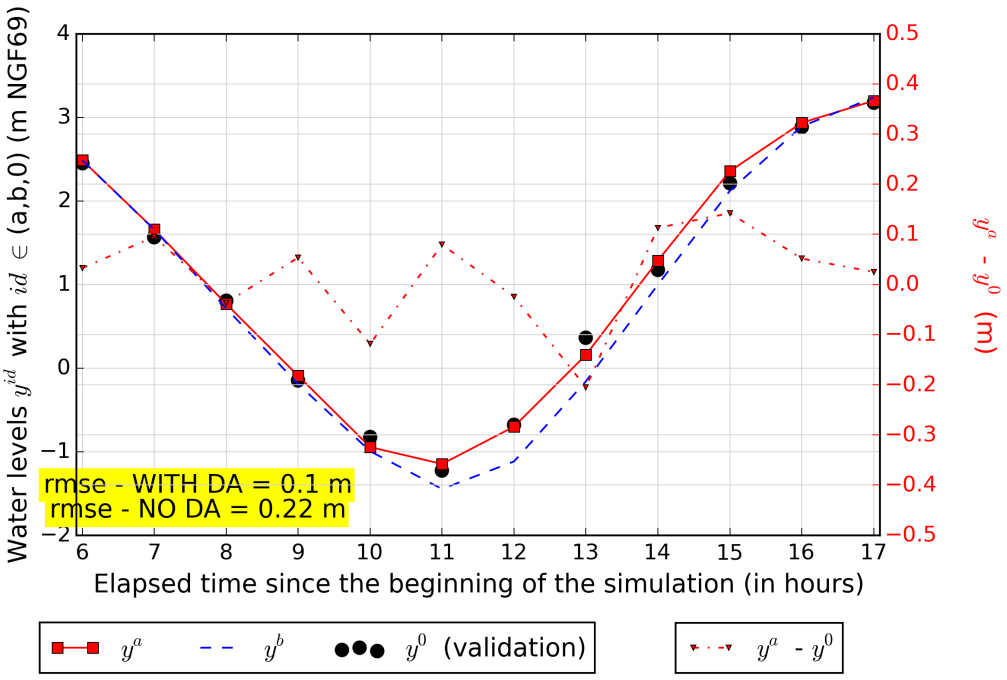



(c) Bordeaux PABX

FIGURE 6.8 : Évolution des niveaux d'eau sur les stations "assimilées" pour l'Exp. $R2c$ (fin)

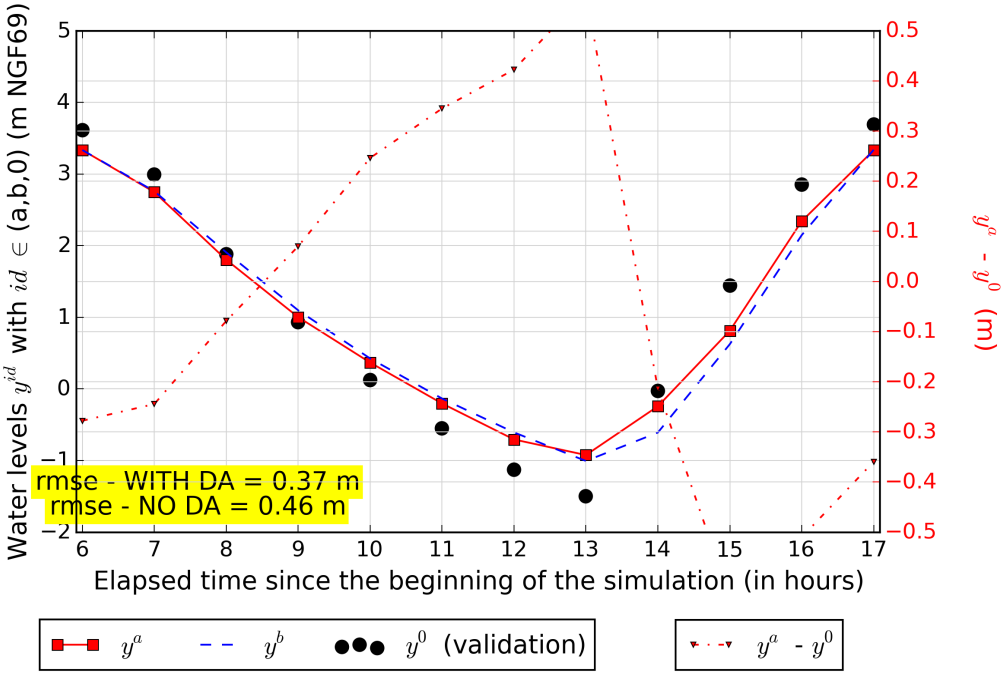



(a) Le Verdon

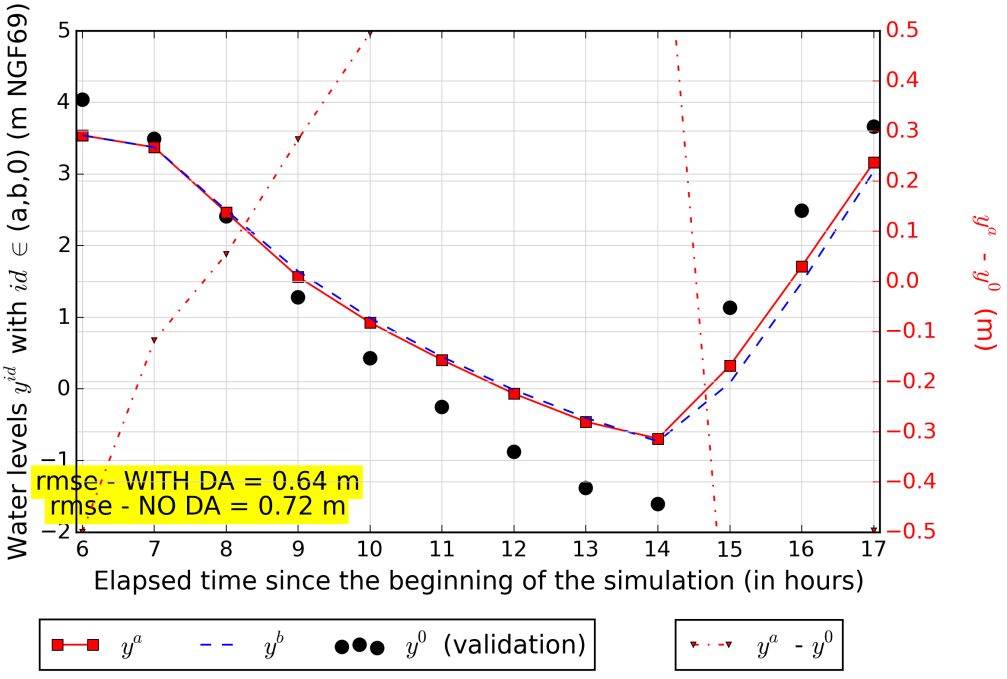



(b) Laména

FIGURE 6.9 : Évolution des niveaux d'eau sur les stations de validation pour l'Exp. $R2b$ (début)

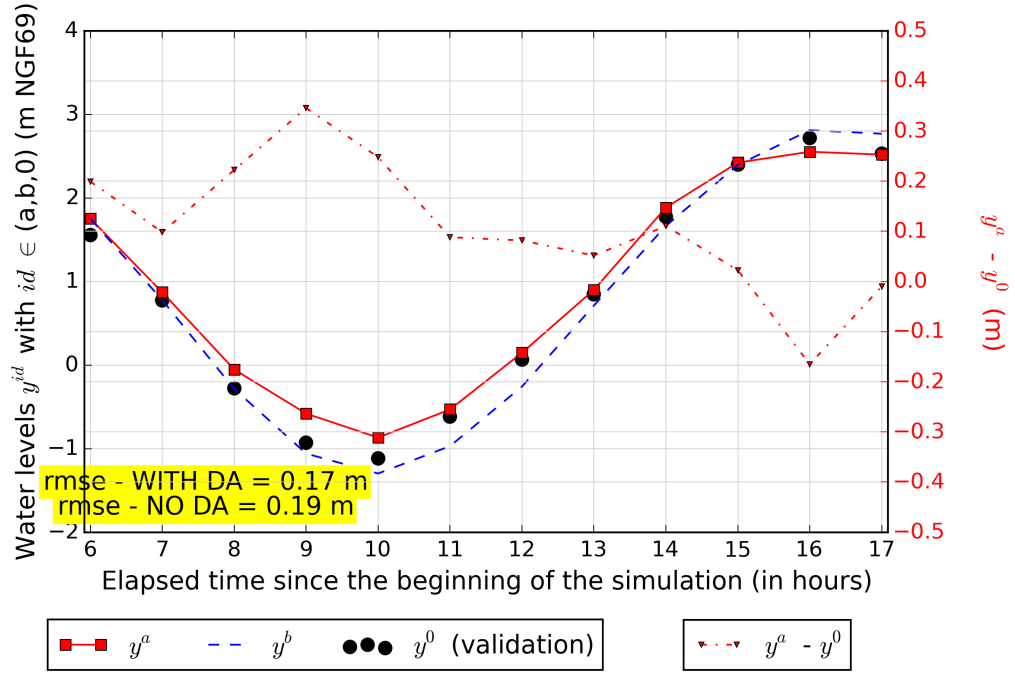



(c) Le Marquis

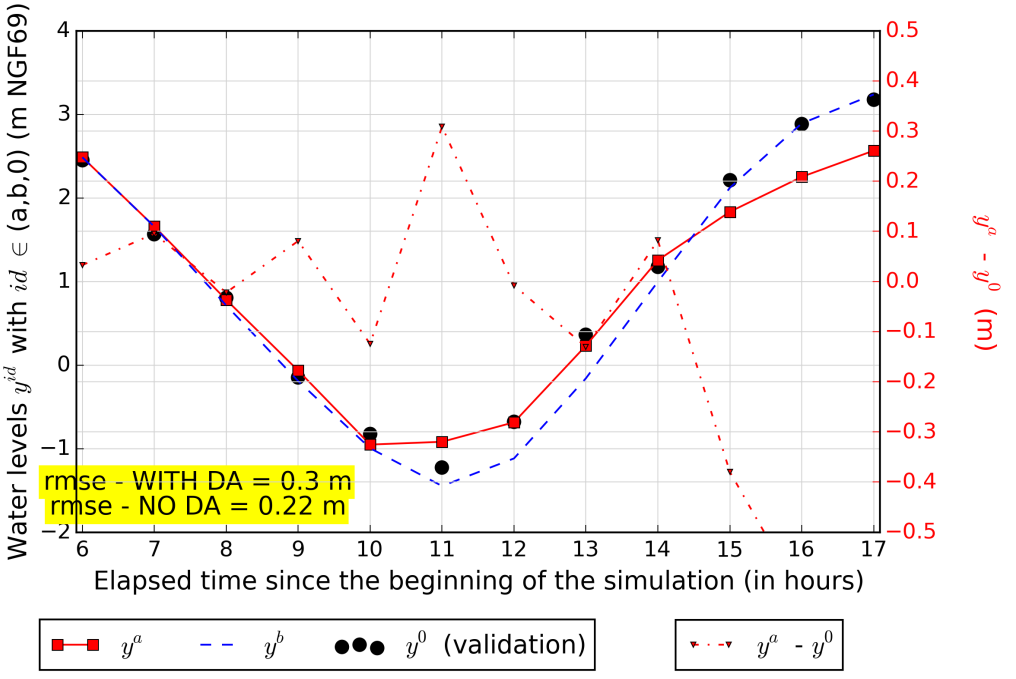



(d) Bordeaux 2

FIGURE 6.9 : Évolution des niveaux d'eau sur les stations de validation pour l'Exp. *R2b* (fin)

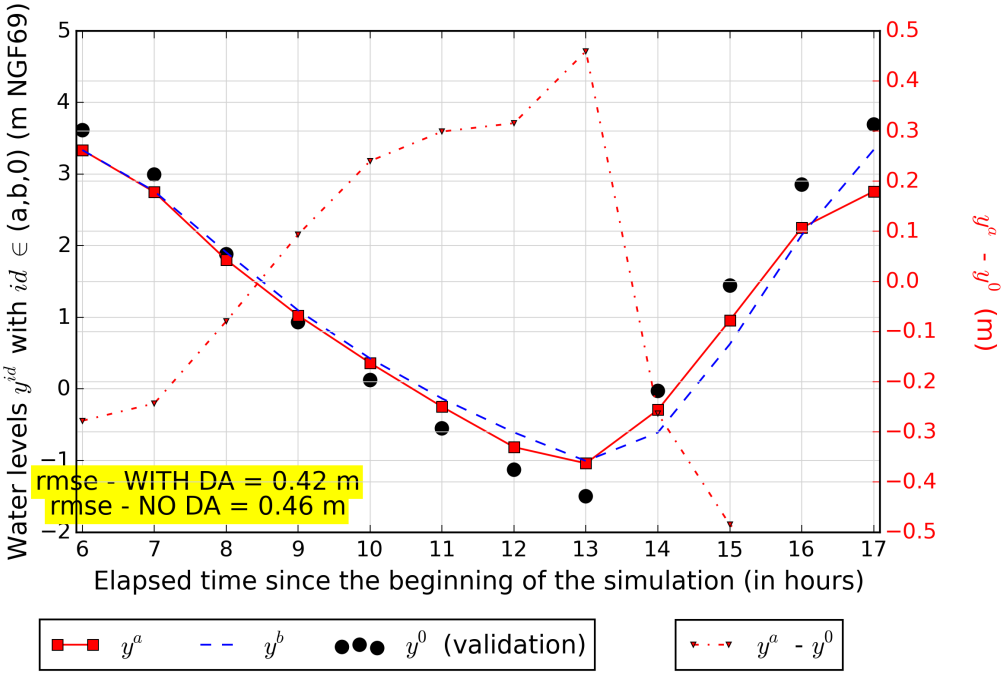



(a) Le Verdon

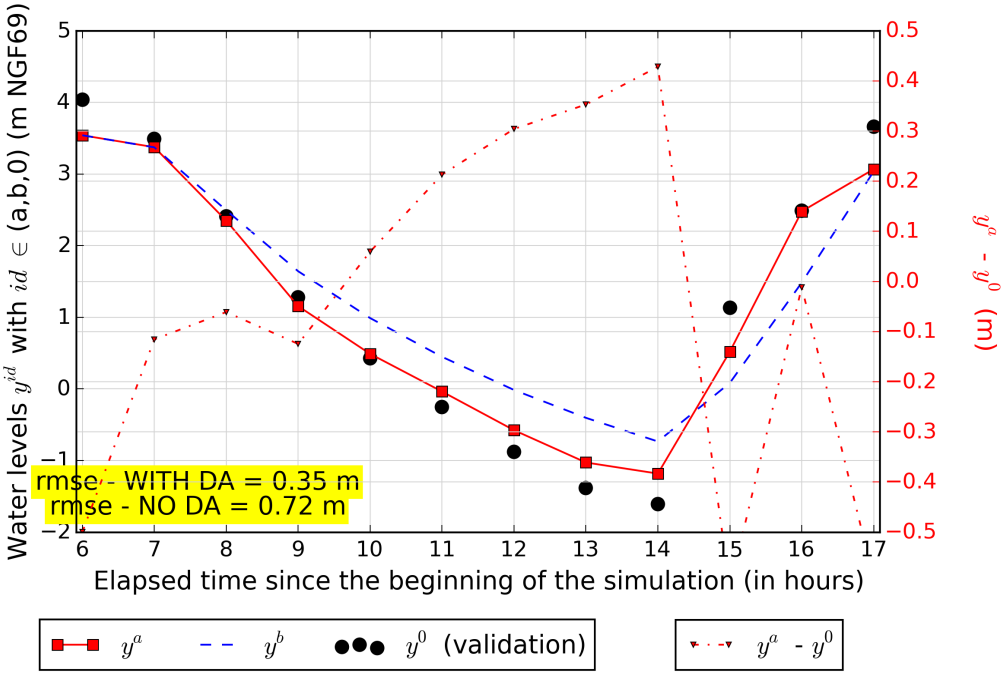



(b) Laména

FIGURE 6.10 : Évolution des niveaux d'eau sur les stations de validation pour l'Exp. $R2c$ (début)

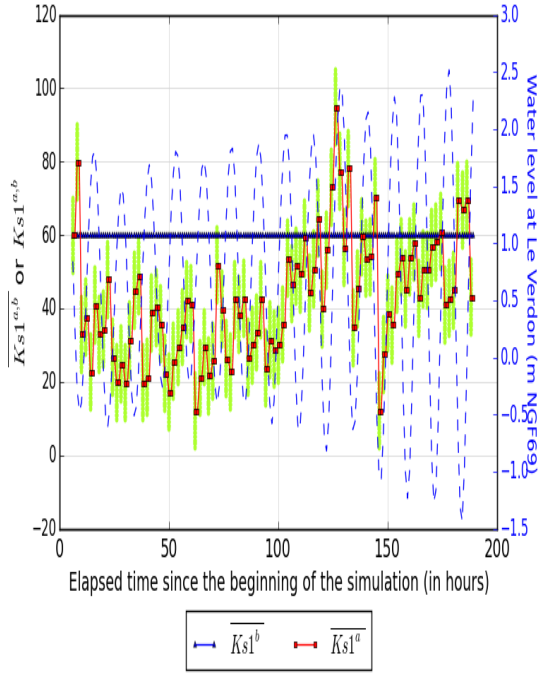



(c) Le Marquis

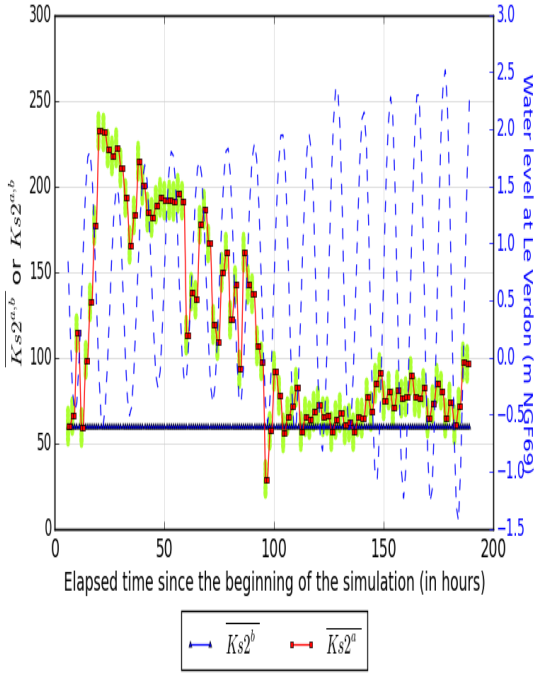



(d) Bordeaux 2

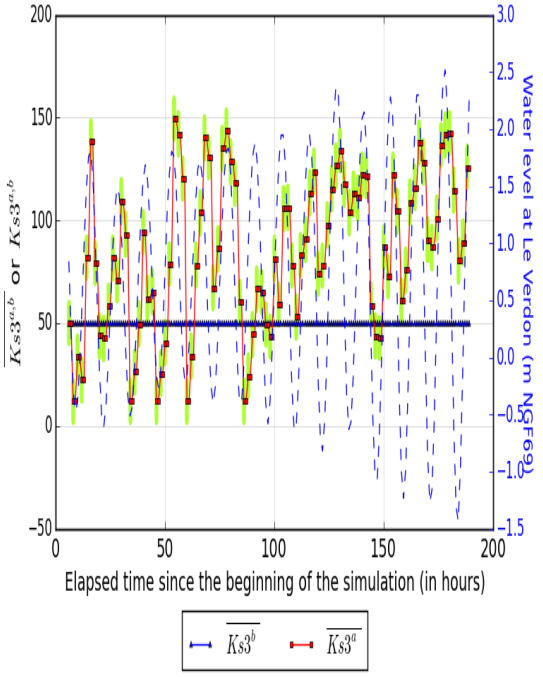
FIGURE 6.10 : Évolution des niveaux d'eau sur les stations de validation pour l'Exp. $R2c$ (fin)




(a) $Ks1$

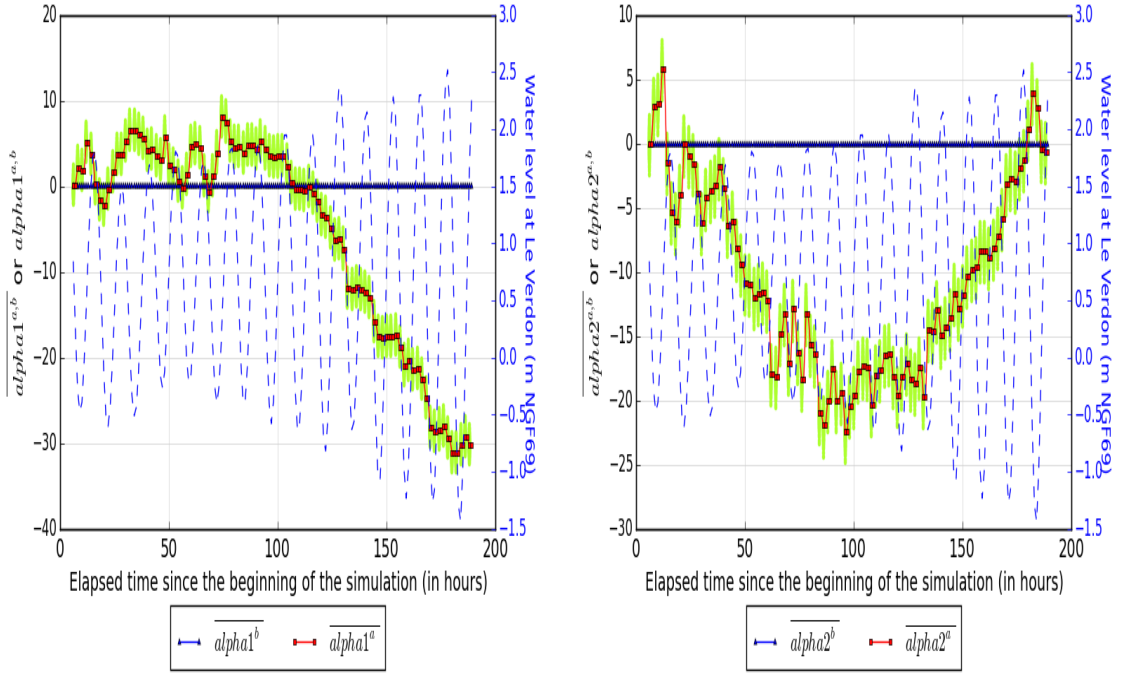



(b) $Ks2$



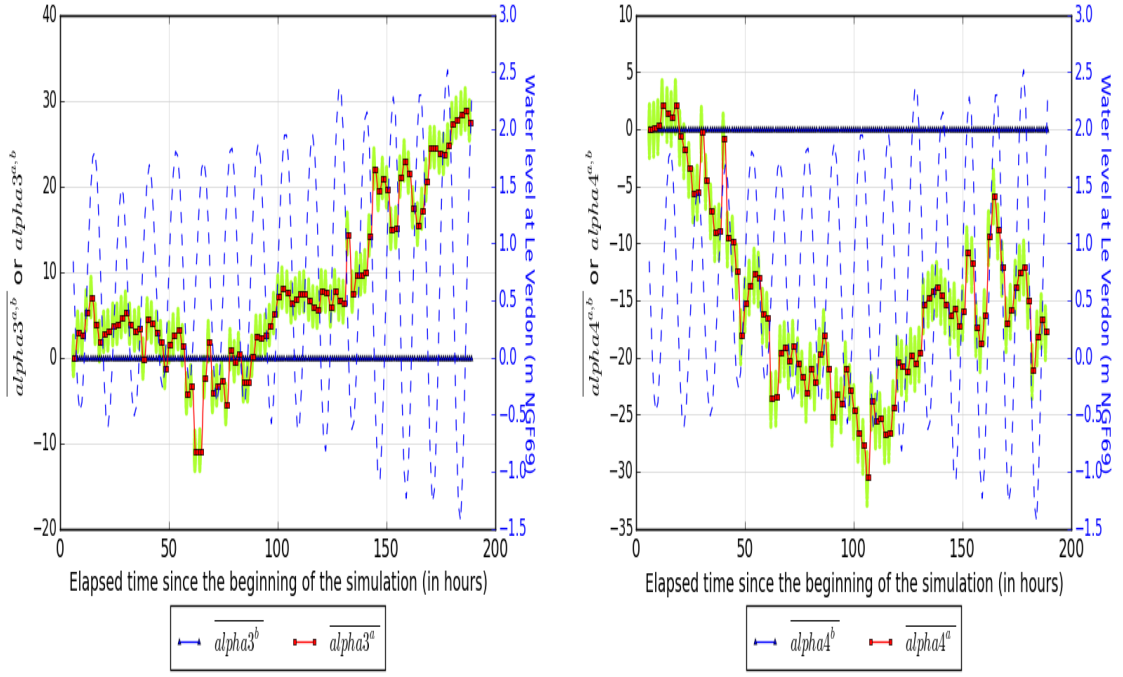

(c) $Ks3$

FIGURE 6.11 : Évolution de $Ks1$, $Ks2$ et $Ks3$ pour l'Exp. $R3$ au cours des cycles d'assimilation (L=3h - S=2h). L'ensemble est représenté par des cercles verts.




(a) CLMAR - mode 1

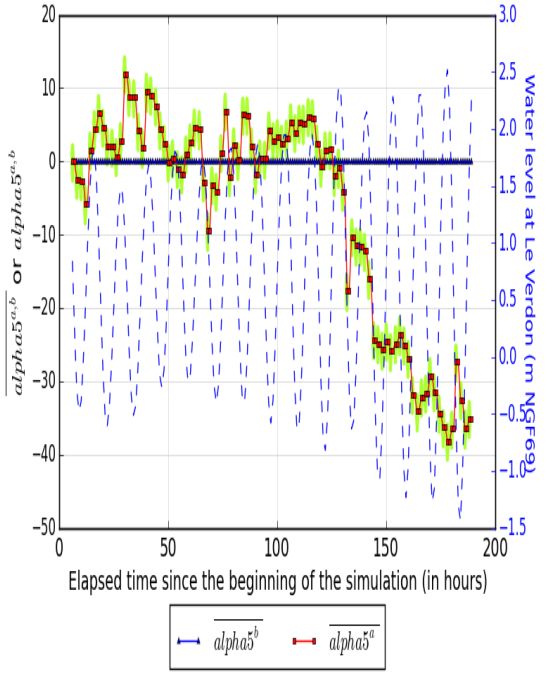
(b) CLMAR - mode 2



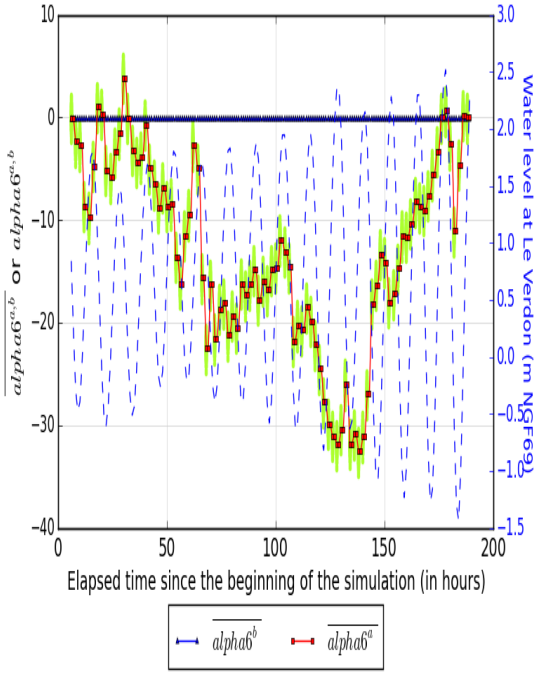

(c) CLMAR - mode 3

(d) CLMAR - mode 4

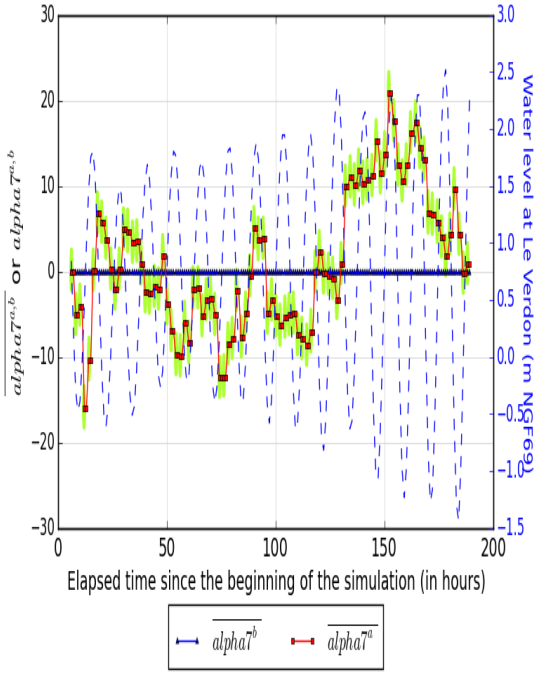
FIGURE 6.12 : Évolution des modes 1, 2, 3 et 4 de CLMAR pour l'Exp. $R3$ au cours des cycles d'assimilation (L=3h - S=2h). L'ensemble est représenté par des cercles verts.




(a) CLMAR - mode 5




(b) CLMAR - mode 6




(c) CLMAR - mode 7

FIGURE 6.13 : Évolution des modes 5, 6 et 7 de CLMAR pour l'Exp. $R3$ au cours des cycles d'assimilation (L=3h - S=2h). L'ensemble est représenté par des cercles verts.

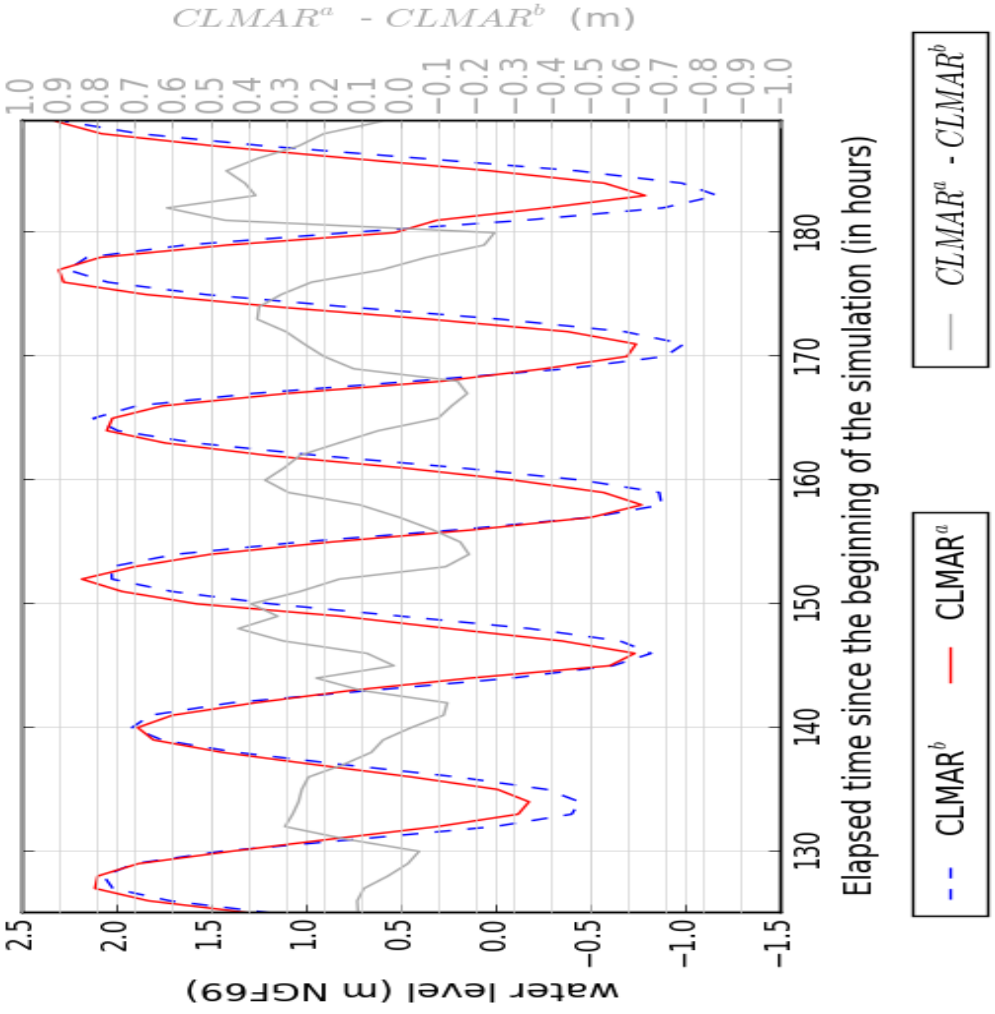
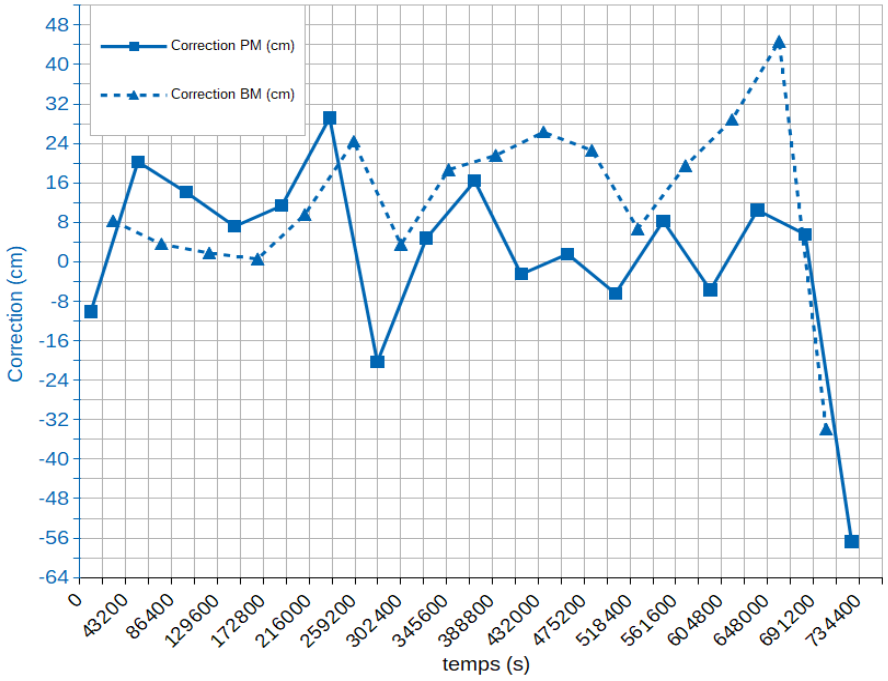
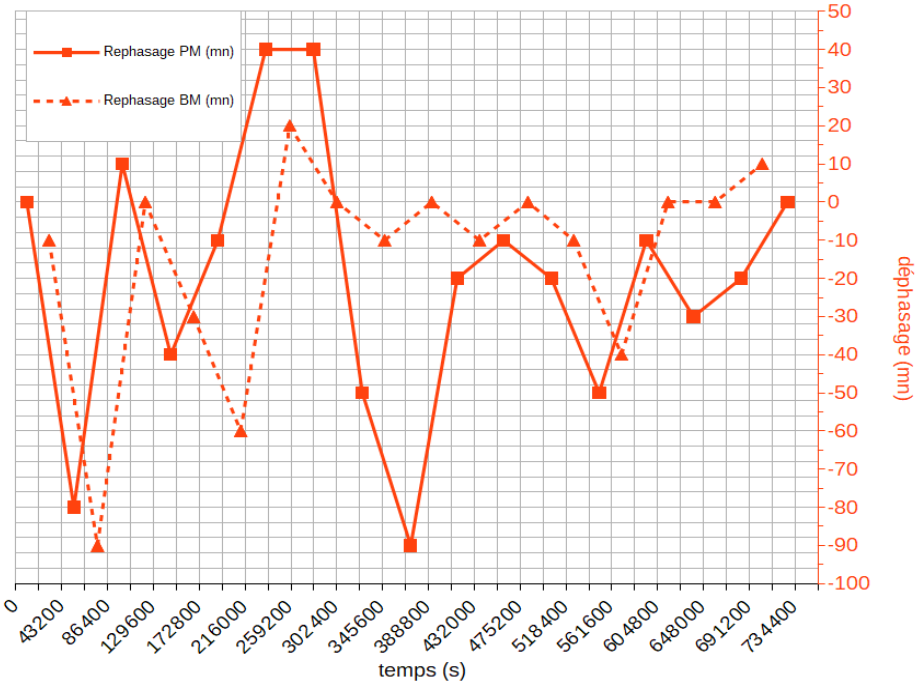



FIGURE 6.14 : Condition à la limite maritime $CLMAR$ du modèle pour l'Exp. $R3$

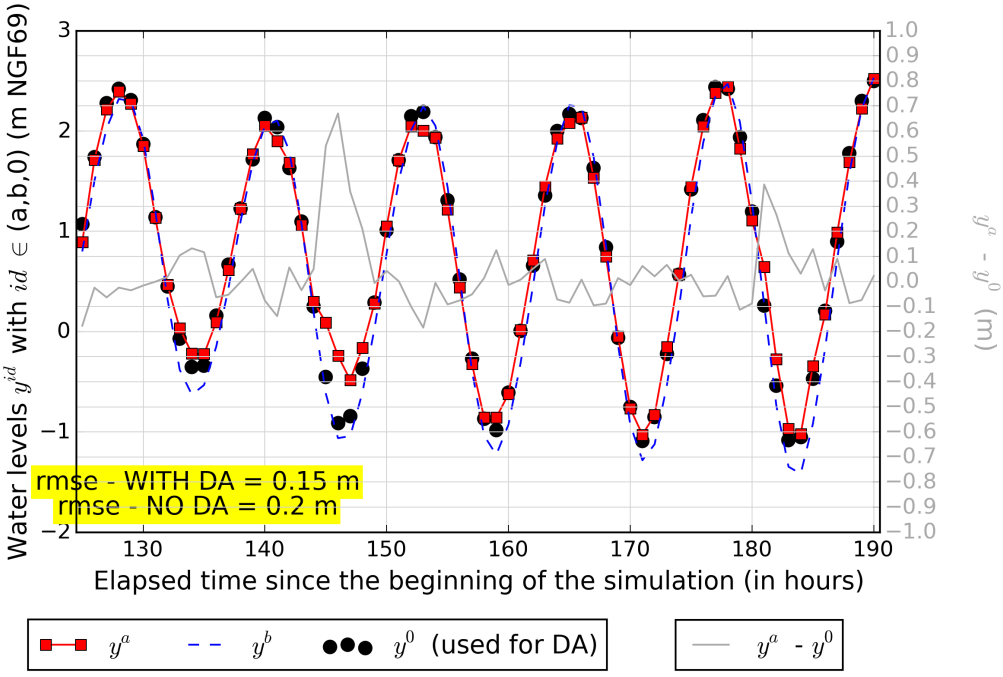



(a) Correction sur le niveau d'eau au pic de Pleine Mer

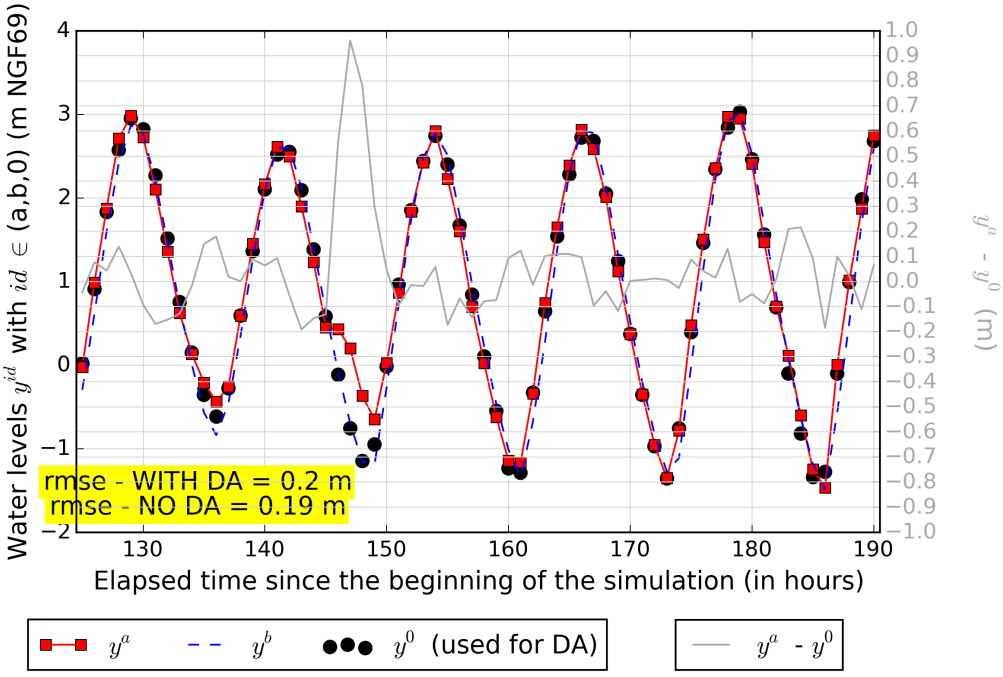



(b) Correction sur le dèphasage du pic de Pleine Mer

FIGURE 6.15 : Correction des pics de Pleine Mer et de Basse Mer de la condition aux limites maritimes $CLMAR$ pour l'Exp. $R3$

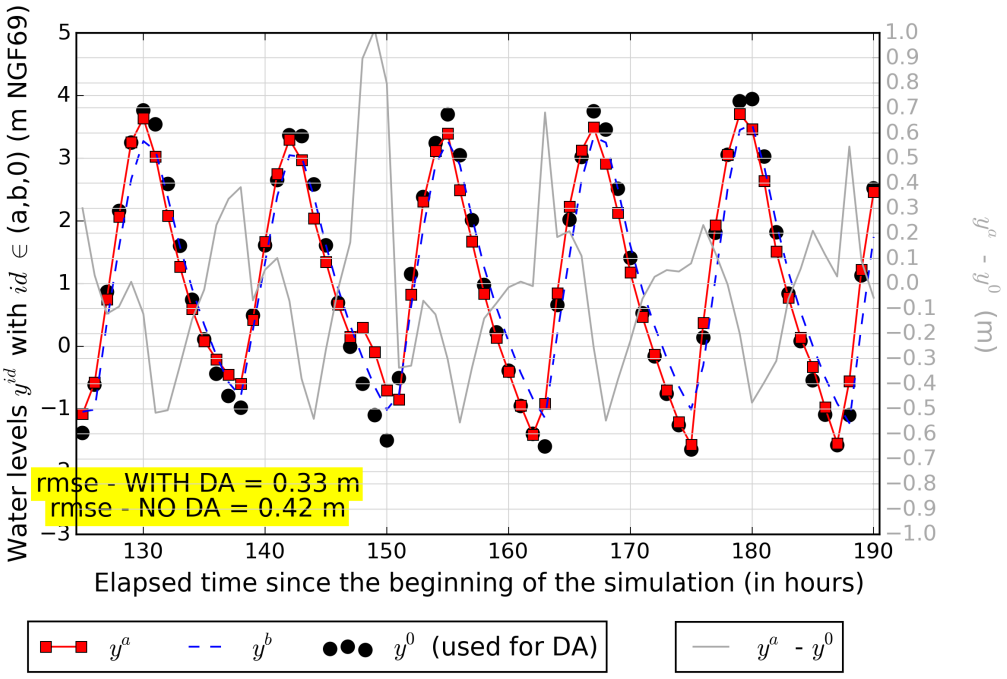



(a) Port-Bloc



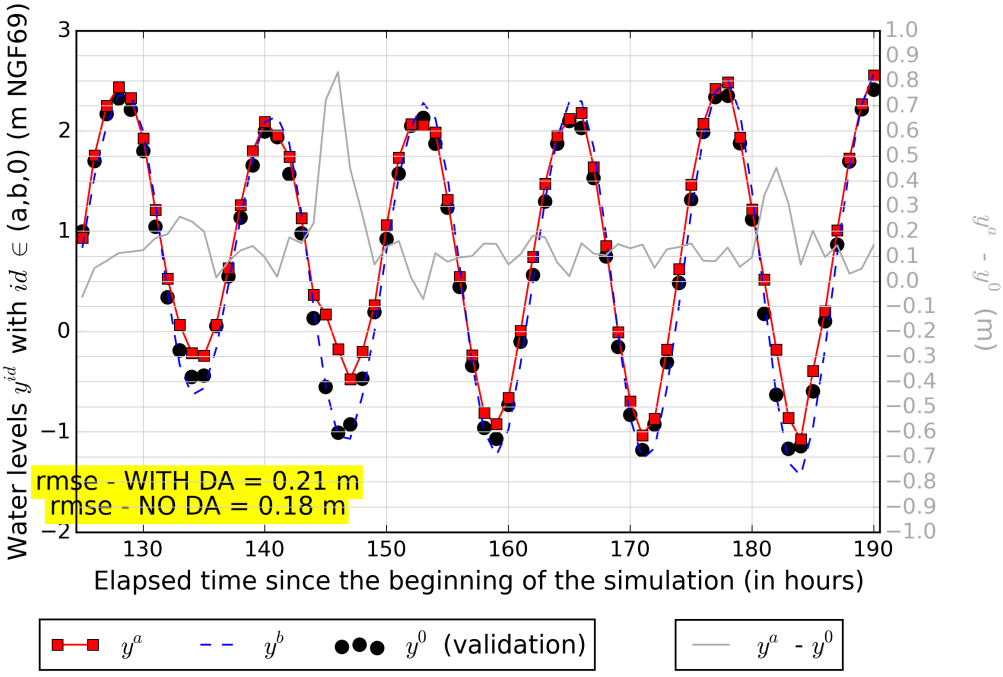

(b) Pauillac

FIGURE 6.16 : Évolution des niveaux d'eau sur les stations "assimilées" pour l'Exp. $R3$ (début)

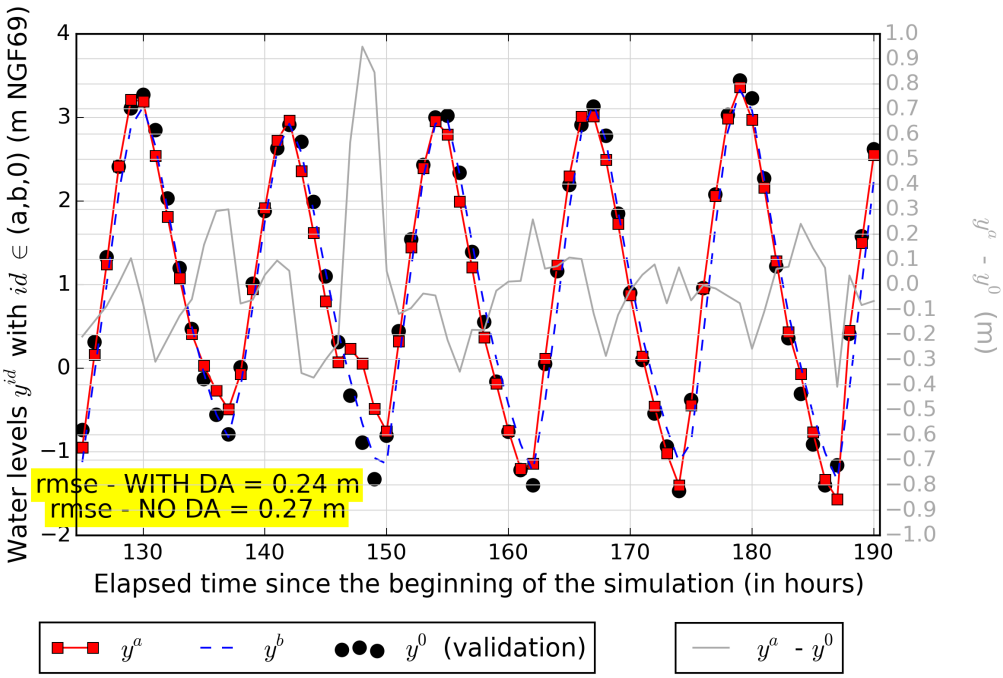



(c) Bordeaux PABX

FIGURE 6.16 : Évolution des niveaux d'eau sur les stations "assimilées" pour l'Exp. $R3$ (fin)

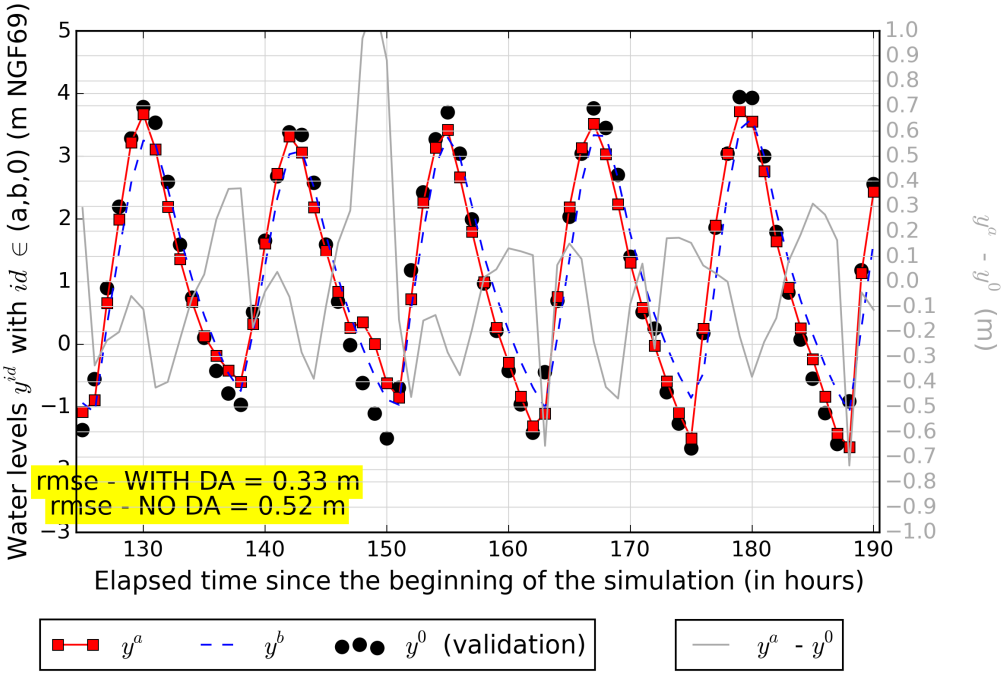



(a) Le Verdon

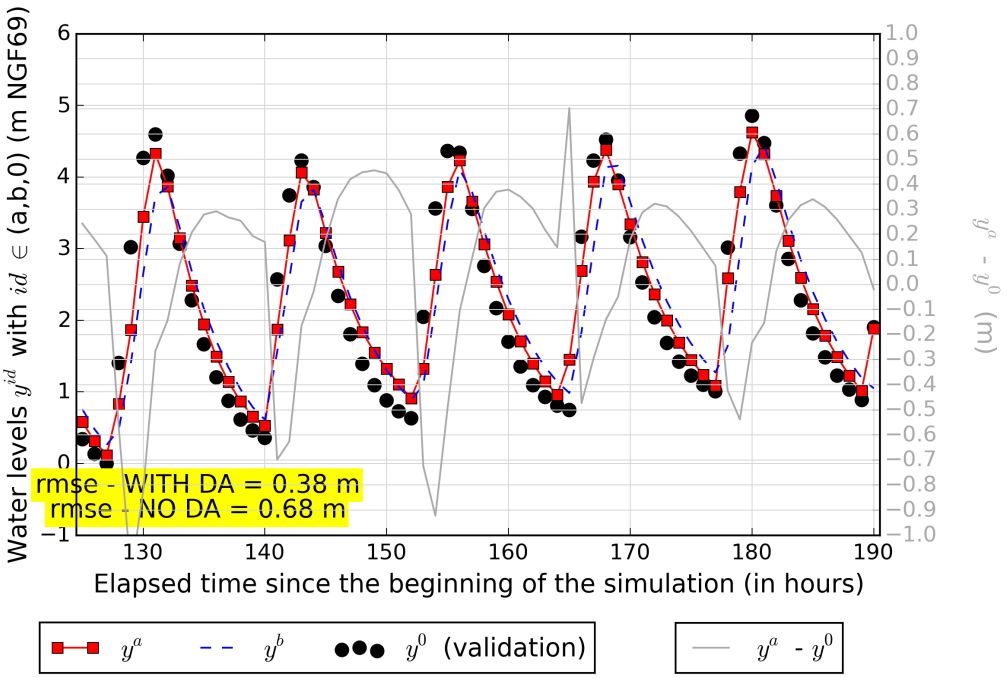



(b) Bec d'Ambès

FIGURE 6.17 : Évolution des niveaux d'eau pour les stations de validation au cours de l'Exp. $R3$ (L=3h - S=2h) (début)

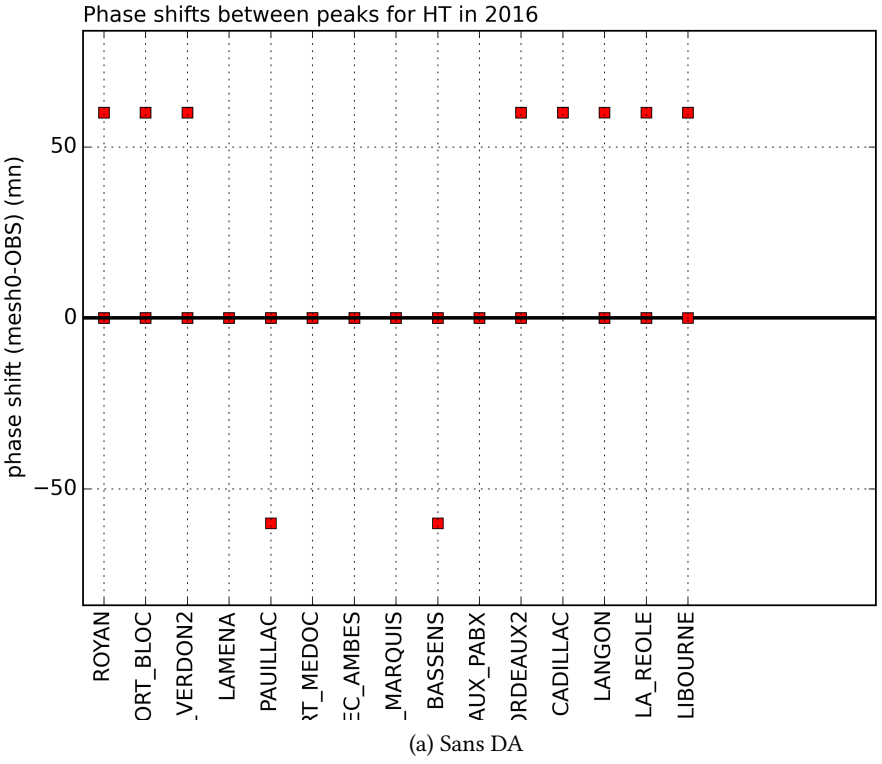



(c) Bordeaux 2

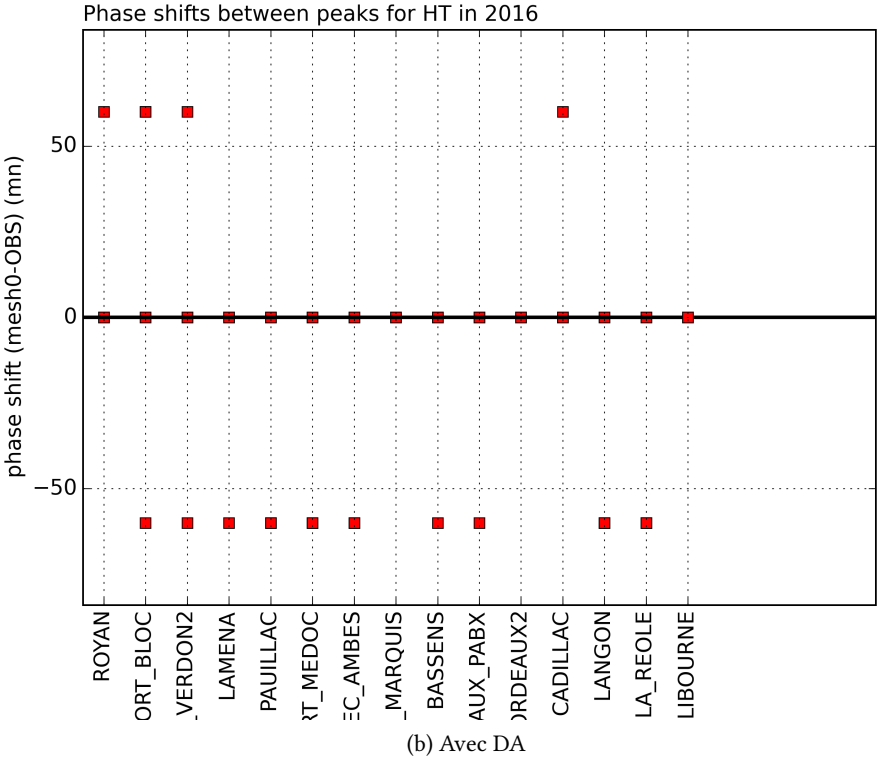



(d) Cadillac

FIGURE 6.17 : Évolution des niveaux d'eau pour les stations de validation au cours de l'Exp. $R3$ (L=3h - S=2h) (fin)




(a) Sans DA




(b) Avec DA

FIGURE 6.18 : Déphasages des pics de pleine mer le long de l'estuaire de la Gironde pour l'Exp. $R3$ avec (en bas) et sans assimilation de données (en haut)

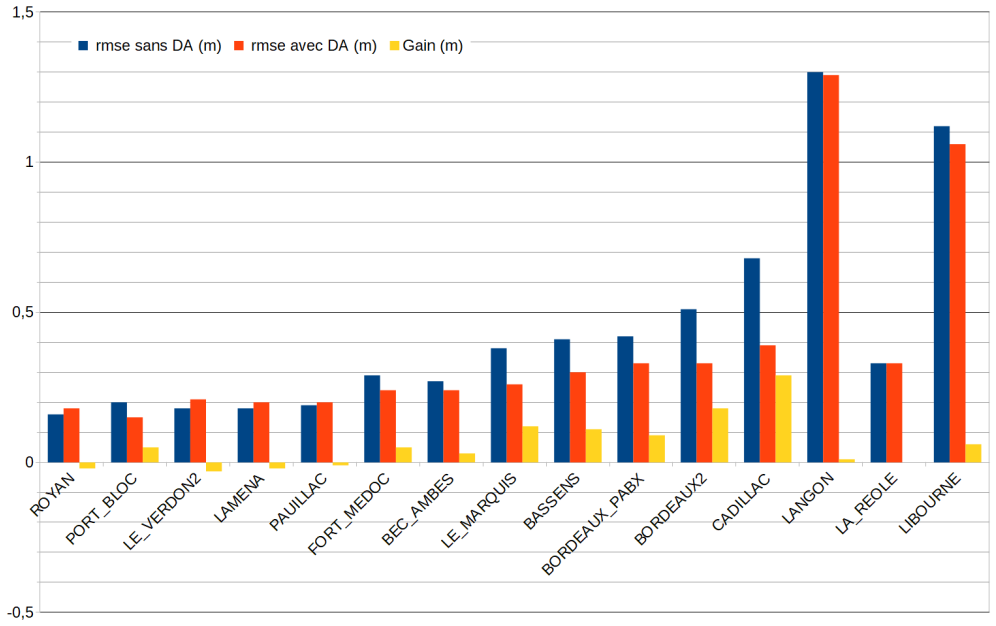



FIGURE 6.19 : Erreur quadratique moyenne le long de l'estuaire de la Gironde pour l'Exp. $R3$

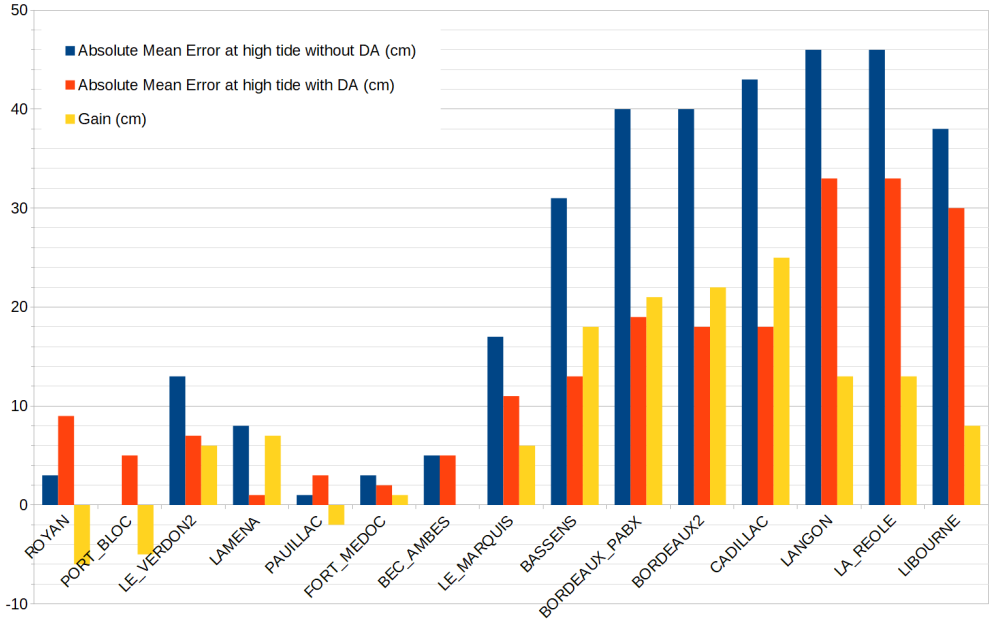



FIGURE 6.20 : Erreur quadratique moyenne sur les pics de PM le long de l'estuaire de la Gironde pour l'Exp. $R3$

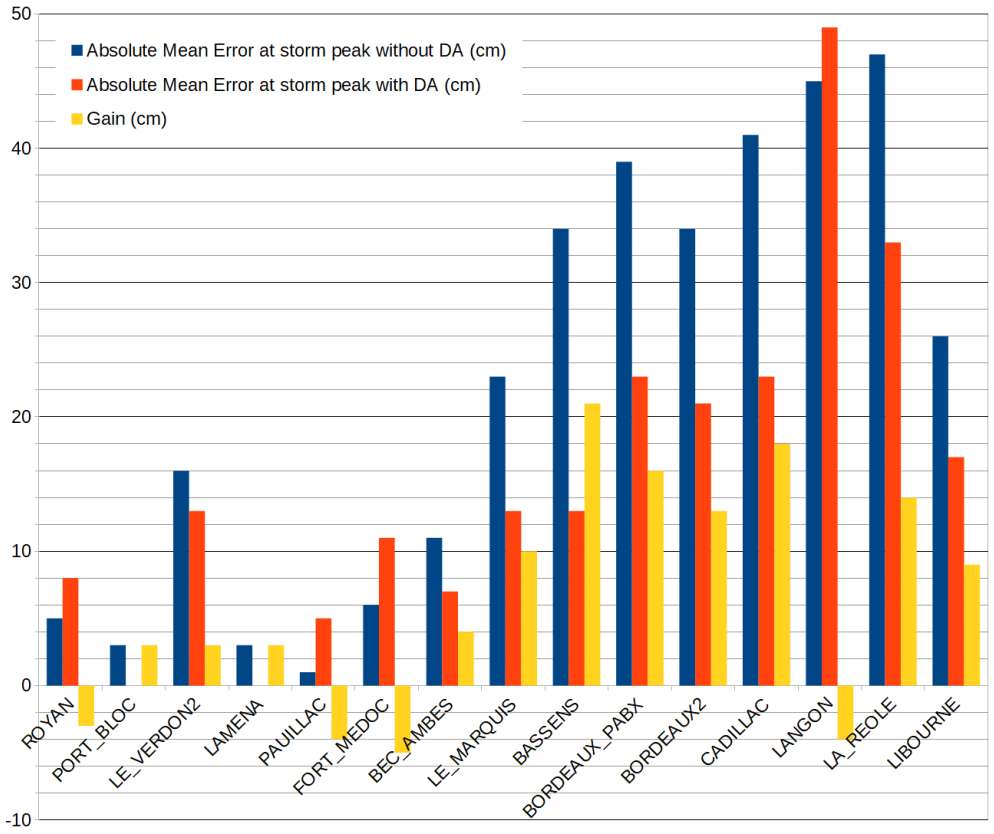
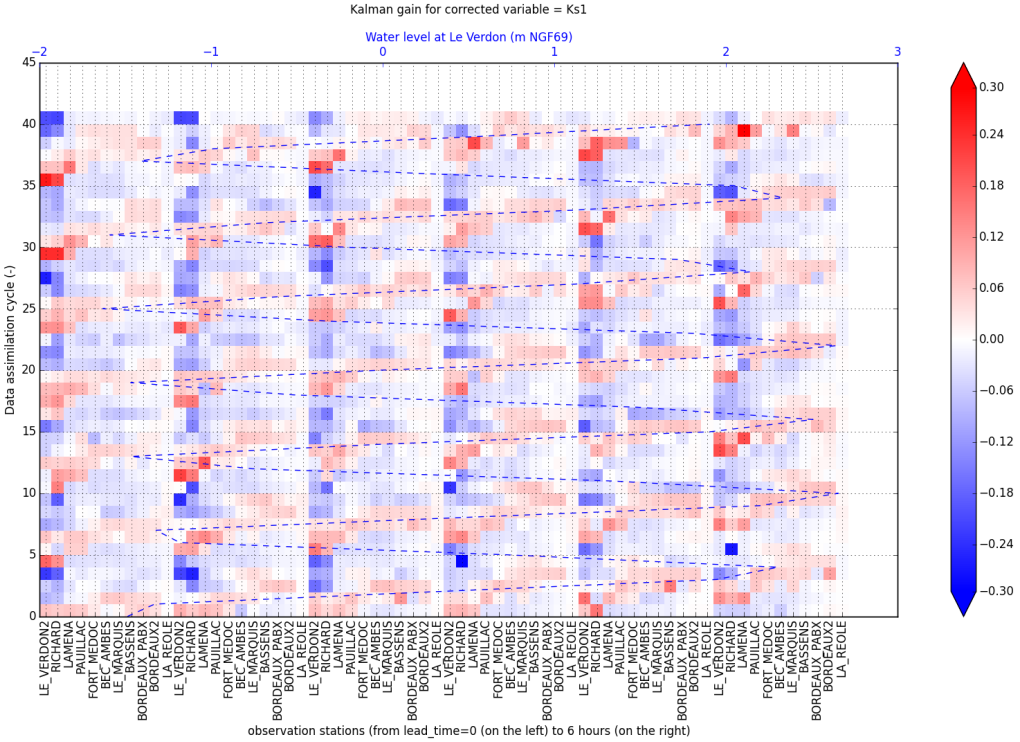
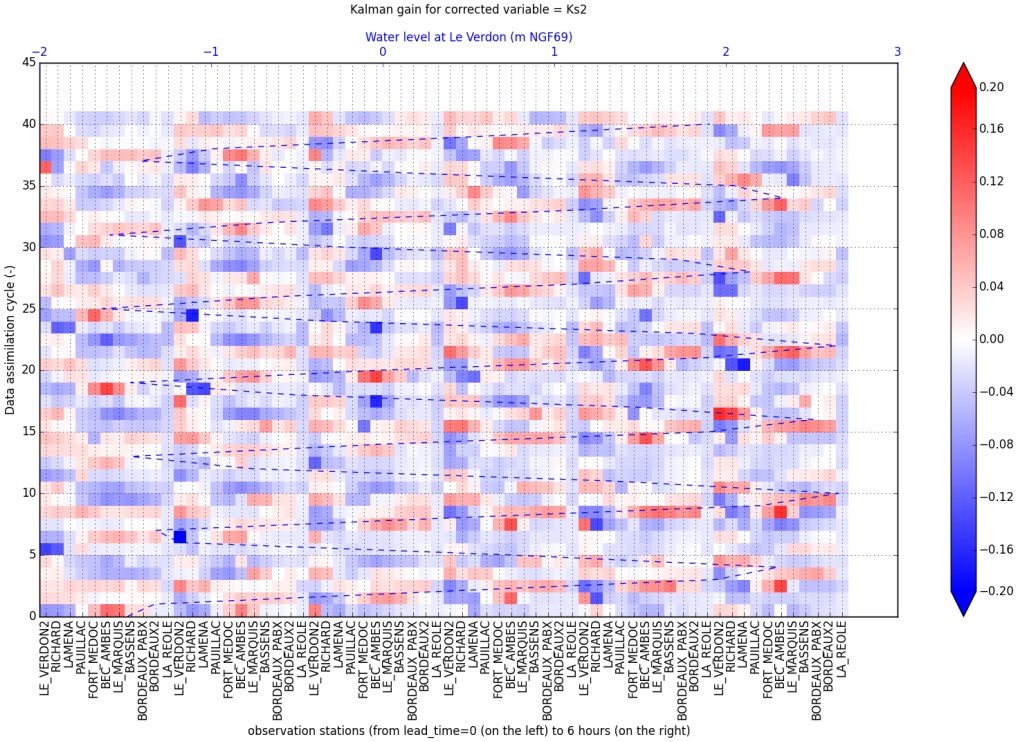



FIGURE 6.21 : Erreur sur le pic de tempête le long de l'estuaire de la Gironde pour l'Exp. $R3$

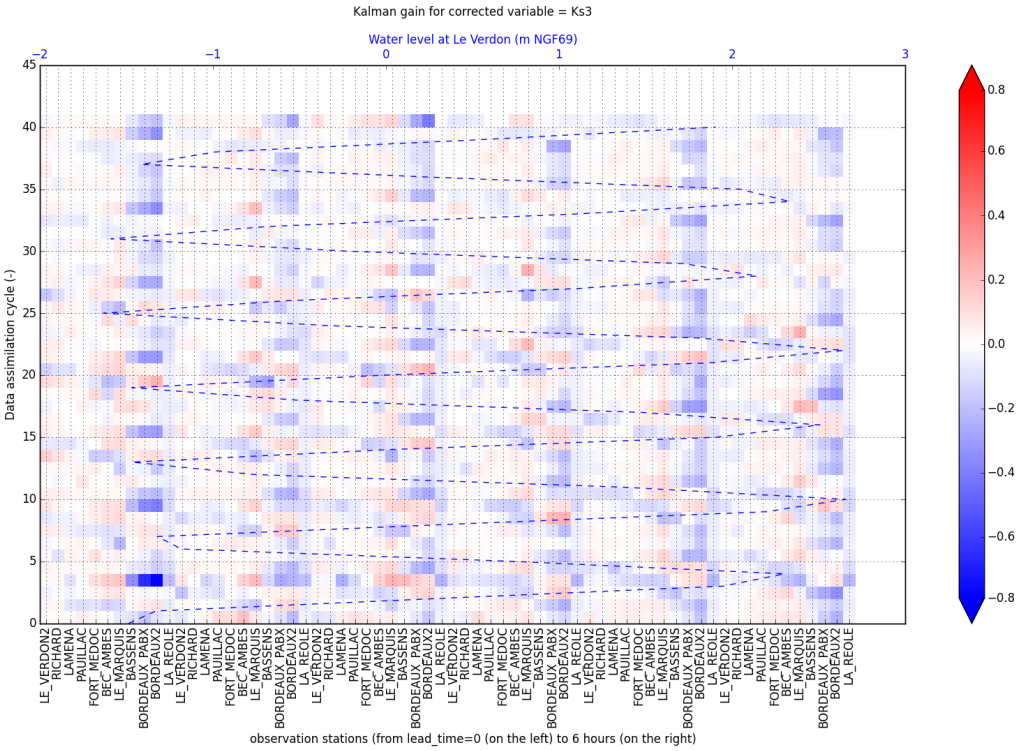



(a) $Ks1$

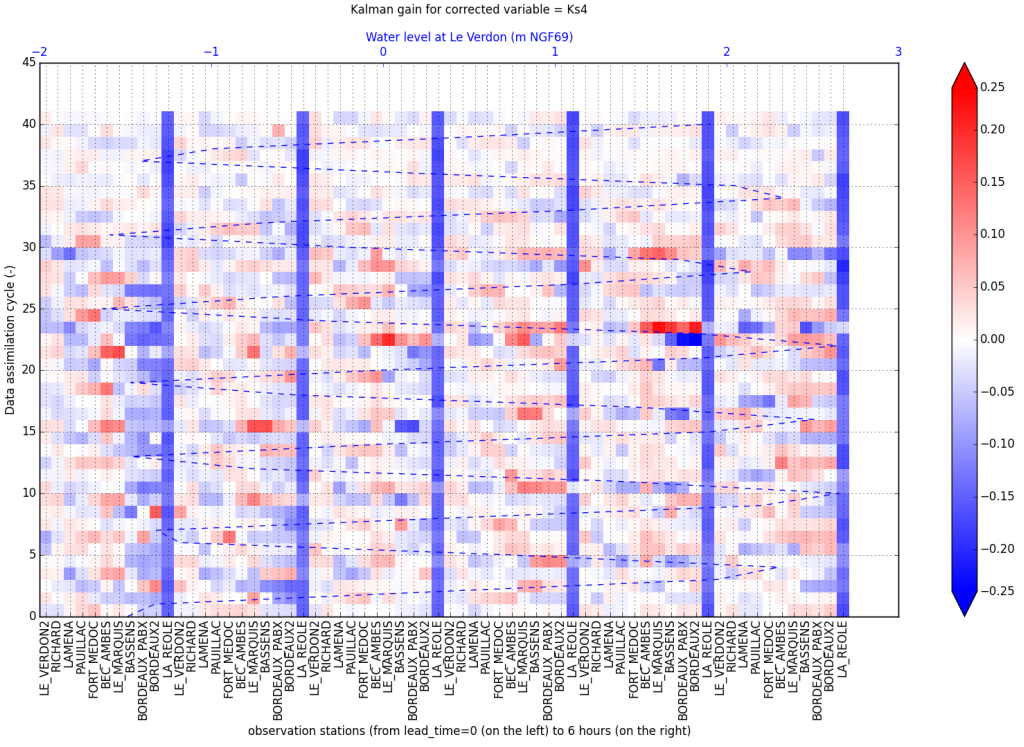



(b) $Ks2$

FIGURE 6.22 : Évolution au cours des cycles d'assimilation des composantes de la matrice gain de Kalman **K** pour chaque variable de contrôle avec les niveaux d'eau calculés par Telemac2D dans l'espace des observations (de gauche à droite de 0 h d'échéance à $(L-1)$ h d'échéance après le début du cycle) pour l'Exp. R1

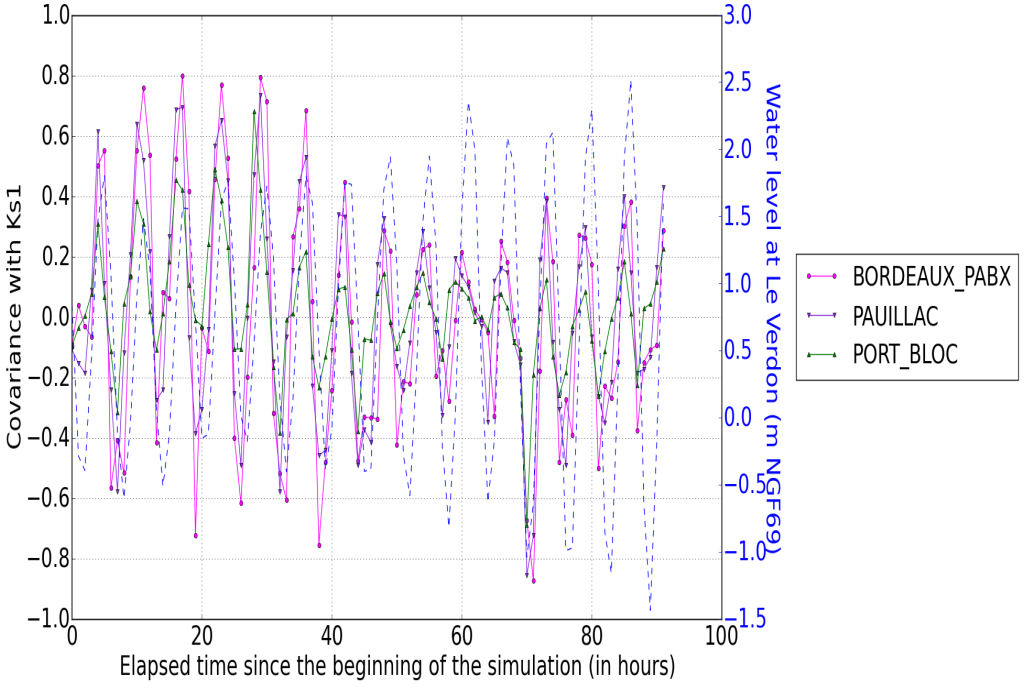



(c) $Ks3$

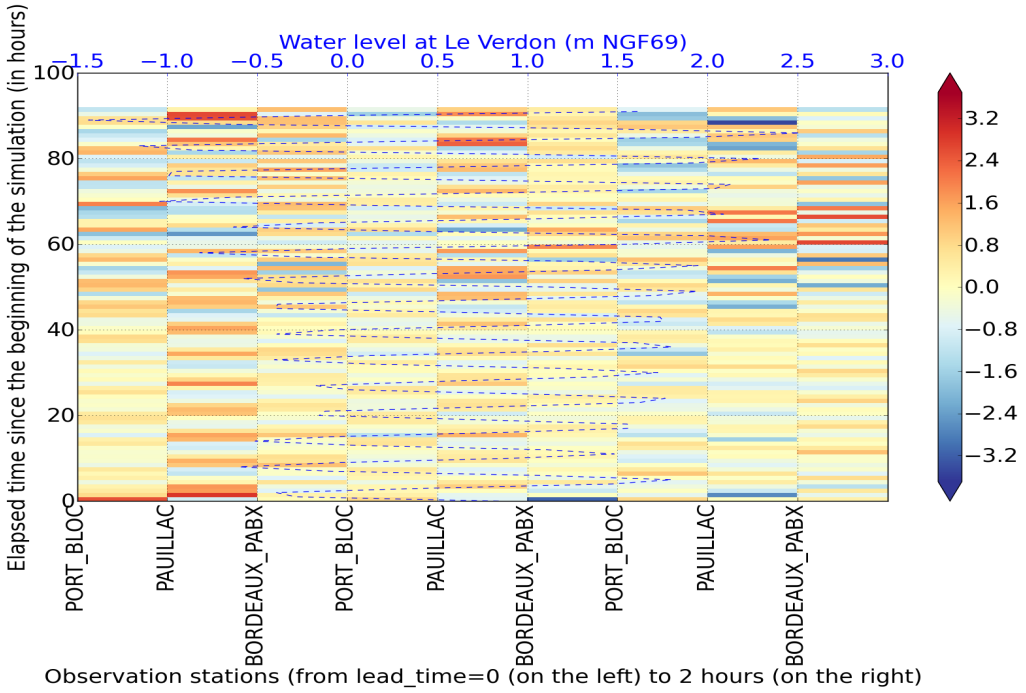



(d) $Ks4$

FIGURE 6.22 : Évolution au cours des cycles d'assimilation des composantes de la matrice gain de Kalman (**K**) pour chaque variable de contrôle avec les niveaux d'eau calculés par Telemac2D dans l'espace des observations (de gauche à droite de 0 h d'échéance à $(L-1)$ h d'échéance après le début du cycle) pour l'Exp. R1 (suite)

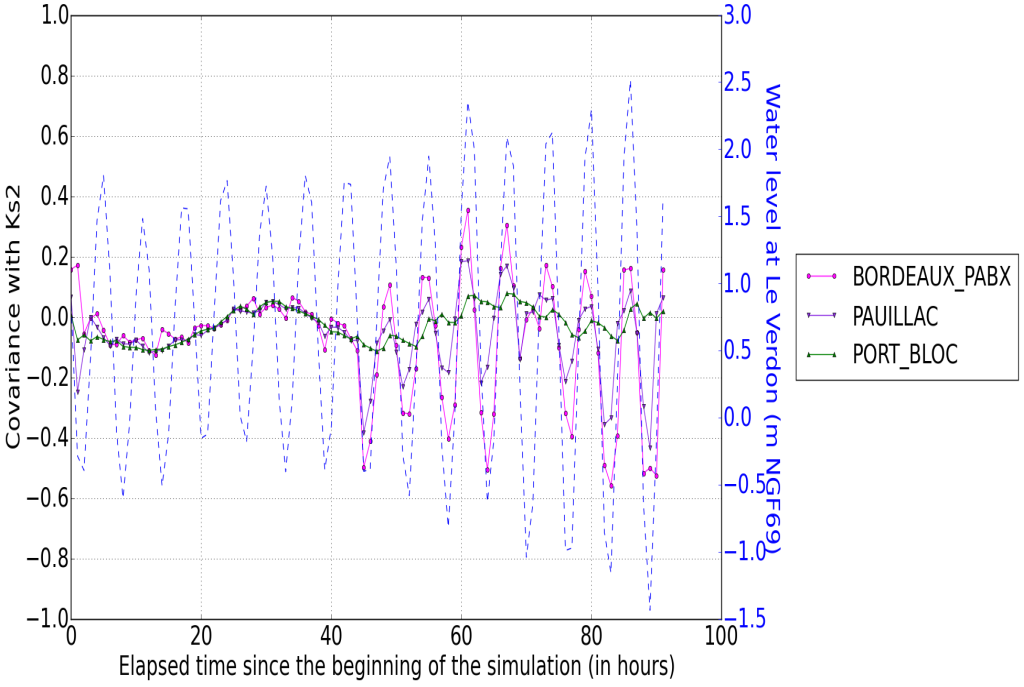



(a) $Ks1$ - $\mathbf{P}^{Ks1,y}$

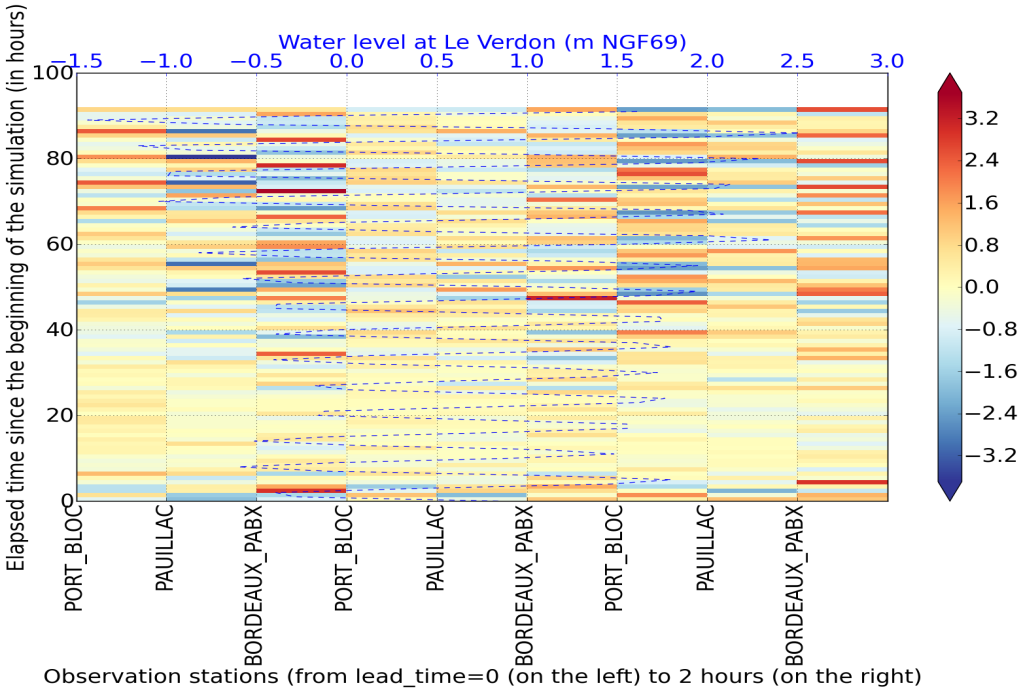



(b) $Ks1$ - gain de Kalman

FIGURE 6.23 : Évolution au cours du temps des composantes de la matrice de covariance $\mathbf{P}^{\gamma,y}$ avec chaque observation (à 0 h d'échéance) et du gain de Kalman $\mathbf{K}$ pour chaque observation (de gauche à droite de 0 h d'échéance à $(L-1)$ h d'échéance après le début du cycle) et chaque variable de contrôle pour l'Exp. $R3$ (début)

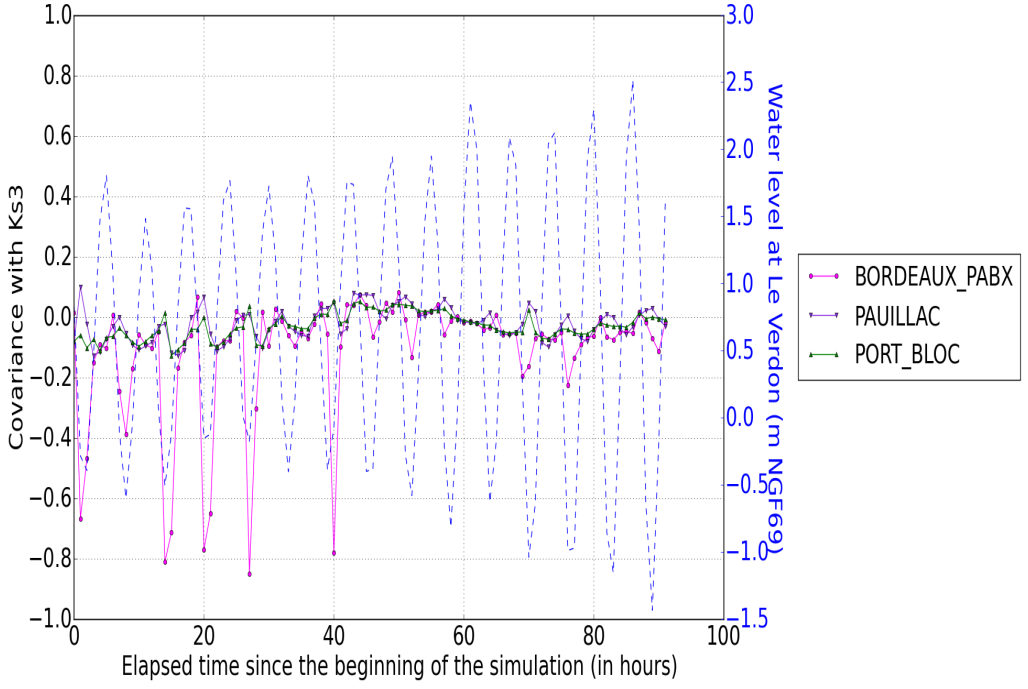



(c) $Ks2$ - $\mathbf{P}^{Ks2,y}$

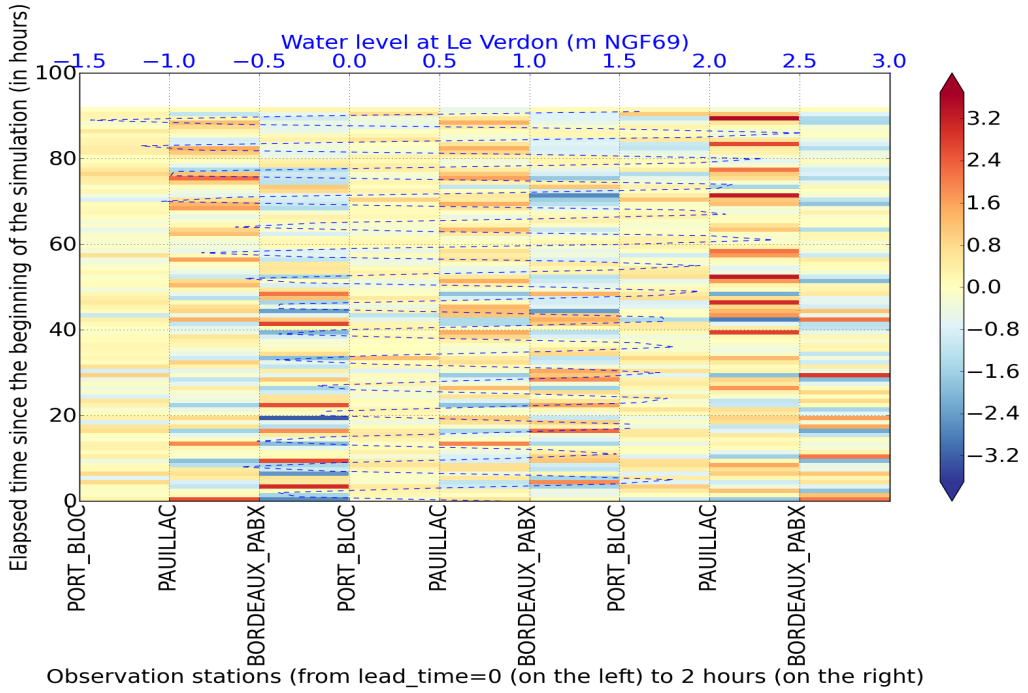



(d) $Ks2$ - gain de Kalman

FIGURE 6.23 : Évolution au cours du temps des composantes de la matrice de covariance $\mathbf{P}^{\gamma,y}$ avec chaque observation (à 0 h d'échéance) et du gain de Kalman $\mathbf{K}$ pour chaque observation (de gauche à droite de 0 h d'échéance à $(L-1)$ h d'échéance après le début du cycle) et chaque variable de contrôle pour l'Exp. $R3$ (suite)




(e) $Ks3$ - $\mathbf{P}^{Ks3,y}$




(f) $Ks3$ - gain de Kalman

FIGURE 6.23 : Évolution au cours du temps des composantes de la matrice de covariance $\mathbf{P}^{\gamma,y}$ avec chaque observation (à 0 h d'échéance) et du gain de Kalman $\mathbf{K}$ pour chaque observation (de gauche à droite de 0 h d'échéance à $(L-1)$ h d'échéance après le début du cycle) et chaque variable de contrôle pour l'Exp. $R3$ (fin)

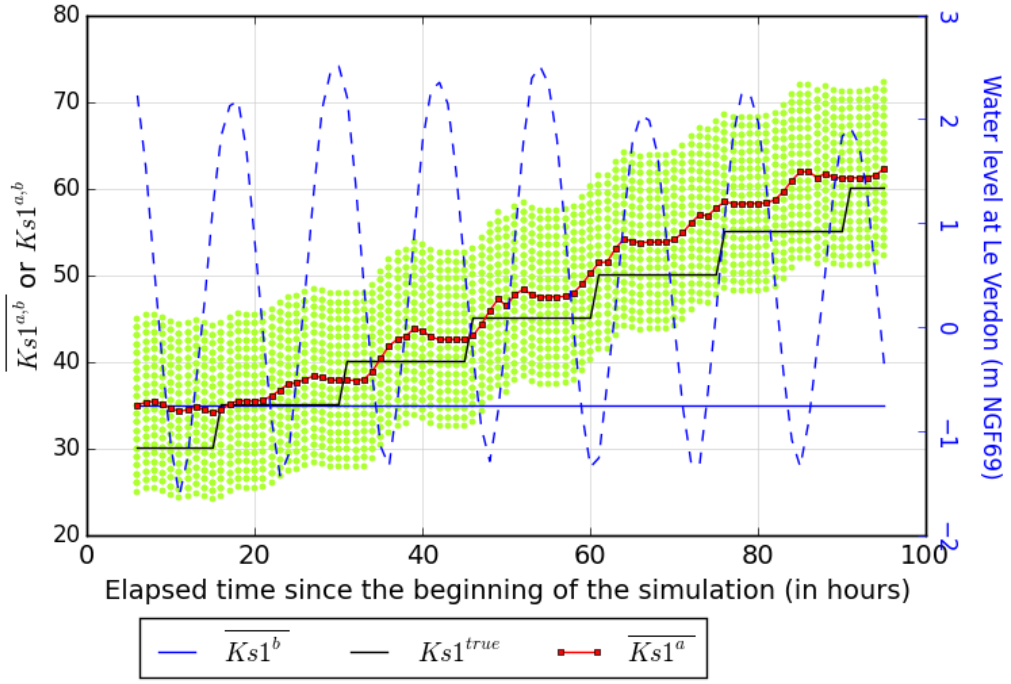



FIGURE 6.24 : Évolution temporelle de $Ks1$ pour l'Exp. 3-c synthétique au cours des cycles d'assimilation (L=1h - S=1h)

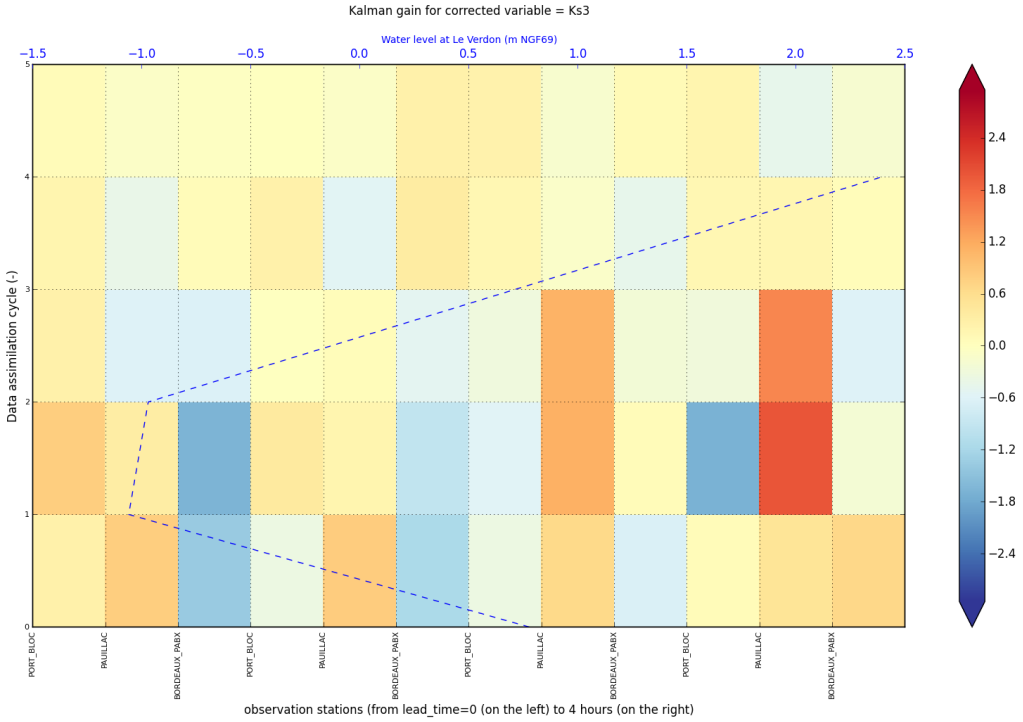



FIGURE 6.25 : Évolution au cours des cycles d'assimilation de la composante $Ks3$ de la matrice Gain de Kalman **K** avec chaque observation simulée (de gauche á droite de 0 h d'échéance á $(L-1)$ h d'échéance après le début du cycle) pour l'Exp. $R2c$

# Chapitre 7

# Conclusions et perspectives

## 7.1 Conclusions

Les travaux de thèse présentés ici ont été réalisés au sein du Laboratoire d'Hydraulique Saint-Venant, commun entre le Cerema, EDF R&D et l'ENPC, en étroite collaboration avec le CERFACS, dans le cadre du développement de techniques d'assimilation de données dans les modèles hydrodynamiques bidimensionnels avec application au modèle de prévision des hautes eaux de l'estuaire de la Gironde fondé sur le code de calcul Telemac2D (www.opentelemac.org).

Dans un premier temps, une méthodologie de traitement des incertitudes a été appliquée au modèle numérique Telemac2D utilisé pour la prévision opérationnelle des crues par le SPC-GAD dans l'estuaire de la Gironde actualisé. Une analyse de sensibilité globale (GSA) par décomposition de variance (ANOVA) a été réalisée afin d'estimer la part de variance du système liée à chaque variable pendant la tempête de 2003. La difficulté résidait dans le traitement des différentes sources d'incertitudes.
Il a été démontré que la condition à la frontière maritime détermine la dynamique de l'estuaire et que l'influence du coefficient de frottement augmente de l'embouchure de l'estuaire vers l'amont de la Garonne et de la Dordogne. Cette GSA a permis d'identifier les sources d'incertitude les plus significatives et leur évolution temporelle sur plusieurs cycles de marée.

Dans un second temps, une méthode d'assimilation de données ensembliste EnKF-$\gamma$-KLBC qui permet d'estimer de manière conjointe les paramètres d'un modèle d'évolution dynamique et de ses conditions aux limites dans un espace réduit de Karhunen-Loève a été proposée. Cette méthodologie a ensuite été appliquée au modèle de prévision des hautes eaux de l'estuaire de la Gironde fondé sur le code numérique bidimensionnel Telemac2D. Les variables de contrôle dans ce cadre précis ont été détaillées, mettant en lumière la possibilité d'utiliser l'EnKF-$\gamma$-KLBC-MEG pour la correction directe des frontières liquides via un incrément, mais aussi l'opportunité de corriger les conditions aux limites temporelles dans un espace réduit de KL afin d'estimer la structure temporelle de l'erreur aux frontières du modèle. L'estimation des coefficients modaux

et celle des paramètres du modèle peut être réalisée de manière conjointe ou indépendante.

Les performances de l'EnKF-$\gamma$-KLBC-MEG ont, tout d'abord, été validées dans un cadre synthétique et l'influence de sa configuration étudiée. Il est apparu que les paramètres $Ks1$ et $Ks2$ et les forçages "vrais" sont correctement reconstitués au cours des expériences, même dans des situations de variation temporelle "rapide" du paramètre. Une méthode de rééchantillonnage fondée sur la persistance de la variance initiale de l'ensemble a été testée pour remédier à l'effondrement de l'ensemble causé par l'EnKF-$\gamma$-KLBC-MEG pour les paramètres de frottement $Ks$ du modèle. Cependant, de possibles difficultés pour l'estimation du paramètre $Ks3$, situé dans la zone de la confluence où les écoulements sont issus de processus physiques non-linéaires, ont également été mises en évidence. Le phénomène d'"équifinalité", susceptible de poser problème lors de l'assimilation de données réelles, a été constaté lors de l'estimation de paramètres et forçages "vrais" instationnaires.
Les niveaux d'eau sont correctement estimés dans ce cadre synthétique. L'erreur est inférieure à 10 cm le long de l'estuaire pour toutes les expériences, en-dehors des tronçons amont des fleuves Garonne et Dordogne, très sensibles à la qualité de l'observation, mais également au raffinement du maillage du modèle d'évolution.
Finalement, les résultats des différentes expériences jumelles menées sont satisfaisants, mais ont mis en lumière des difficultés susceptibles d'apparaître lors de la future utilisation de mesures réelles, du fait, d'une part, du nombre de membres $N_e$ faible, responsable d'erreurs d'échantillonnage, mais aussi, d'autre part, de l'équifinalité qui peut biaiser l'estimation des variables de contrôle par l'EnKF-$\gamma$-KLBC-MEG.

Les performances de l'EnKF-$\gamma$-KLBC-MEG ont ensuite été évaluées dans un cadre d'assimilation de données réelles. Les résultats obtenus pour une réanalyse de 8 jours en janvier 2016 montrent une amélioration du calcul des PM sur la quasi-totalité des stations en aval de l'estuaire, ainsi qu'une meilleure estimation des pics de tempête. Par ailleurs, l'amplitude du signal est mieux représentée, même si le signal est principalement amélioré pendant les BM. Ainsi, l'EnKF-$\gamma$-KLBC-MEG représente mieux les niveaux d'eau que la simulation de contrôle (MEG0V7P3). Ainsi, pour les stations de Pauillac et de Port-Bloc, l'erreur instantanée entre le niveau d'eau corrigé et l'observation est inférieure à 10 cm en fin de simulation. Les RMSE sur les pics de pleine mer sont inférieures à 20 cm de l'embouchure de l'estuaire à Cadillac sur la Garonne pour des gains obtenus compris entre 6 cm et 25 cm. Les gains sur les pics de PM obtenus à l'aide de l'EnKF-$\gamma$-KLBC-MEG sont compris entre 10 cm et 21 cm sur la Garonne en amont de la confluence jusque Cadillac. L'erreur sur le pic de PM est inférieure à 13 cm en aval de la confluence.
D'une part, une variabilité temporelle des coefficients de Strickler est mise en évidence. Cependant, aucune structure temporelle de correction des coefficients de frottement autour d'une valeur moyenne n'est déterminée.
D'autre part, l'EnKF-$\gamma$-KLBC-MEG est très sensible à l'intégration de la variable $Ks3$ dans le vecteur de contrôle. La mise en place d'un espace support borné d'évolution des coefficients de Strickler semble être nécessaire pour éviter que ces derniers n'atteignent des valeurs inaccep-

tables, les décorrélant des niveaux d'eau simulés.

Compte-tenu des conclusions et recommandations apportées dans les sections précédentes, la problématique de la prévision et de la mise en œuvre d'un protocole d'expérience réelle pour déterminer l'apport de l'EnKF-$\gamma$-KLBC-MEG dans un contexte de prévision opérationnelle des niveaux d'eau est abordée dans la section 7.2.

## 7.2 Perspectives d'évaluation dans un cadre opérationnel

### 7.2.1 Contexte

En marge des simulations déterministes fondées sur le modèle numérique Telemac2D dont les résultats sont fournis par METEO-FRANCE, des méthodes "empiriques" pour la prévision des surcotes et l'estimation des niveaux d'eau dans l'estuaire de la Gironde ont été mises en place par le SPC-GAD (source : Laurent Dieval, DREAL Aquitaine, entretien du 11/06/2019).

Concernant la prévision de la surcote de PM, la méthodologie appliquée par le SPC GAD est la suivante. En temps réel, le SHOM met à disposition les mesures réelles et la marée au Verdon sur les deux dernières PM, soit 24 h. Le SPC GAD récupère de METEO-FRANCE les niveaux d'eau calculés par Telemac2D pour la simulation déterministe sur les deux dernières PM et les prévisions à 36 h. L'écart entre la surcote de PM obtenue par simulation numérique et la surcote réelle sur la PM issue des mesures du SHOM est alors calculé pour les deux dernières PM, ainsi que la moyenne de ces écarts pondérés par un poids de 75 % (resp. 25 %) accordé au coefficient de marée le plus fort (resp. faible). Cet écart est alors appliqué aux prévisions en surcotes de PM de METEO-FRANCE pour les 3 prochaines PM.
L'inconvénient majeur de la méthode réside dans le fait qu'elle fonctionne mal lorsque les fleuves sont en crue. Par ailleurs, le SPC GAD souhaiterait accéder au signal complet, comme c'est le cas sur l'Adour, et à la forme du signal pendant 4 h centrées autour de la PM.
Au droit du Verdon, 3 stations (Royan, Port-Bloc et Le Verdon) sont susceptibles d'être utilisées pour l'application de la correction. Le SPC GAD choisit la station dont le signal est le moins dégradé. Par ailleurs, le SPC a constaté une relation linéaire pour les PM entre le Verdon et Pauillac moins vraie en présence de vent dans l'estuaire. Ainsi, si une tempête est décalée de 2 h par rapport à la PM, le coup de vent est susceptible d'augmenter la PM de 15 cm et le signal de marée est alors déformé.

La prévision des débits des fleuves Garonne et Dordogne n'est pas problématique. Ainsi, l'examen de la forme du signal de la Garonne à Tonneins fourni par le SPC Garonne, Tarn, Lot fournit un niveau d'eau à La Réole sur la Garonne à 24 h entaché d'une incertitude de $\pm 10 - 15$ cm (soit 30 cm d'amplitude), c'est-à-dire une erreur sur le débit de la Garonne d'environ $\pm 20$ % par rapport au débit qui sera réellement observé. L'inertie de la Dordogne est telle que la prévision à 24 h est parfaite et qu'aucune incertitude ne doit être considérée pour l'estimation de son débit en phase

de prévision.

L'outil "idéal" pour le SPC GAD permettrait, à partir d'une réanalyse de 24 h, soit deux cycles de marée, d'améliorer la prévision des niveaux d'eau à +12 h (une marée), +24 h (2 marées) et +36 h (3 marées). Ce prototype idéal nécessite qu'une structure temporelle des variables de contrôle puisse être déterminée en utilisant les mesures des 24 h précédentes pour les réinjecter en mode prévision.

### 7.2.2 Problématique de l'estuaire de la Gironde vis-à-vis de l'état de l'art de la prévision d'ensemble dans le domaine de l'hydrodynamique côtière et estuarienne

Les techniques d'assimilation de données ne sont pas "magiques". La mise en place de prévisions d'ensemble en temps réel a déjà fait l'objet de travaux dans les domaines océaniques et côtiers, ainsi que sur les rivières.
En réalité, la littérature existante (cf. section 3.6 du chapitre 3) traite soit de la possibilité d'améliorer la prévision en montrant, comme cela a été fait dans les chapitres 5 ou 6, que l'utilisation d'un EnKF, en réanalyse sur une expérience synthétique ou réelle, permet d'améliorer les performances en RMSE d'un modèle donné (hydrologique, hydraulique, d'états de mer, etc.), soit utilise un protocole simplifié de prévision. Ainsi, la méthode usuelle consiste à utiliser l'état corrigé comme condition initiale d'une phase de prévision dont les forçages et les paramètres, corrigés ou non, ne sont plus modifiés.

Par ailleurs, les prototypes d'assimilation de données en hydraulique côtière (Zijl et al. [2015]) ou estuarienne (Barthélémy et al. [2017]) fonctionnent généralement avec des fenêtres d'assimilation de quelques heures. Zijl et al. [2015] et Barthélémy et al. [2017] utilisent des cycles d'assimilation de 6 heures. Chaque fin de cycle, c'est-à-dire de période d'analyse, correspond à un temps de base donné $T_b$ à partir duquel une période de prévision est lancée en utilisant comme condition initiale le nouvel état du système corrigé "optimisé" à l'aide de l'EnKF sur la durée du cycle. Les échéances de prévision de surcotes sont de 48 h dans le cas du modèle côtier océanique présenté dans Zijl et al. [2015], du fait de la disponibilité d'une prévision d'ensemble météorologique et de l'hypothèse que seuls les forçages météorologiques sont incertains et participent à la génération des ensembles perturbés. Barthélémy et al. [2017], sur l'estuaire de l'Adour, examine à l'aide d'expériences jumelles les performances de l'EnKF pour des échéances allant de 6 h à 10 h, ces durées étant égales aux temps de transfert de la correction vers l'aval du modèle.

Ces deux études ont ceci de commun qu'elles corrigent l'état du système (niveaux d'eau pour Zijl et al. [2015] auxquels s'ajoutent les vitesses pour Barthélémy et al. [2017]) et non les paramètres du modèle supposés calibrés. De fait, les phases de prévision correspondent à la propagation de la correction de l'état d'ébauche jusqu'à ce qu'elle soit évacuée du modèle et que l'état du système avec assimilation rejoigne l'état du système obtenu par la simulation de contrôle.

L'objectif de ce travail de thèse est ainsi de tester la possibilité d'améliorer les prévisions des niveaux d'eau sur l'estuaire de la Gironde à l'aide de techniques d'assimilation de données en corrigeant les forçages du modèle dans un espace réduit de KL, ainsi que ses paramètres les plus influents. Le nombre d'études de modélisation numérique dont l'estuaire de la Gironde a déjà fait l'objet ne permettait pas d'espérer la découverte soudaine d'un biais systématique par l'EnKF-$\gamma$-KLBC-MEG et, en effet, les premiers résultats en réanalyse d'événements réels ne montrent pas de convergence des variables de contrôle choisies permettant d'appliquer une structure temporelle d'erreur prédéfinie.

La correction de la condition aux limites fournie par METEO-FRANCE (projet HOMONIM (Jourdan et al.)) par le contrôle des coefficients modaux a été testée dans le but de corriger les erreurs issues du modèle de surcotes imposées à la frontière maritime du modèle de l'estuaire de la Gironde, afin qu'elles ne se propagent pas dans le système. La méthode de rééchantillonnage et les problèmes d'équifinalité évoqués dans la section 5.3.3.1 du chapitre 5 sont susceptibles d'empêcher la convergence des coefficients modaux et, ainsi, la détection d'un signal d'erreur déterminé temporellement par les modes de la fonction d'autocorrélation du signal. Il faut noter au passage que la durée d'auto-corrélation elle-même a été déterminée empiriquement sur la base des signaux physiques de surcotes pour plusieurs événements. Il est possible de l'intégrer elle-aussi dans le vecteur de contrôle de l'EnKF-$\gamma$-KLBC-MEG ultérieurement.
Par ailleurs, il a été montré que la correction des coefficients de Strickler est fortement corrélée au signal de marée, ce qui conduit à envisager d'imposer en phase de prévision l'évolution périodique des $Ks$ déterminée, par exemple, dans une fenêtre de 12 h ou 24 h avant le début de la prévision.

### 7.2.3 Proposition d'un protocole de création d'une prévision ensembliste sur l'estuaire de la Gironde

La nature de l'erreur entre le mode "réanalyse" et le mode "prévision" change. Le mode "réanalyse" utilise des forçages météorologiques et temporels aux frontières liquides réanalysés, c'est-à-dire optimisés par rapport aux observations passées. En mode "prévision", des hypothèses doivent être faites pour construire les forçages temporels. Le run "déterministe" de METEO-FRANCE peut fournir une prévision sur 72 h pour la surcote maritime à la frontière du modèle et les champs de vent et de pression.

Le passage d'un mode de réanalyse à un mode de prévision est ainsi caractérisé par le choix des paramètres et des forçages du modèle en phase prévisionnelle, sachant que la condition initiale correspond à l'état du système obtenu après analyse et optimisation de l'état du système $\mathbf{x}^a$ (cf. figure 4.2) au dernier pas de temps du cycle d'assimilation $t_{k,fin}$.

Une première version d'un mode "prévision" a été implémentée dans l'EnKF-$\gamma$-KLBC-MEG en appliquant une persistance du vecteur de contrôle pour la phase de prévision et ne donne aucun résultat probant. Elle permet à l'issue de chaque cycle d'assimilation de simuler une période de

prévision et, par ailleurs, à l'issue de la période complète de réanalyse de simuler également une période sans AD.

La suite de cette section propose un protocole pour permettre l'évaluation des performances de l'EnKF-$\gamma$-KLBC-MEG en mode "prévisionnel".

#### 7.2.3.1 Conditions aux limites fluviales

Compte-tenu des échanges avec le SPC GAD, le débit de la Dordogne est considéré comme parfait en phase de prévision. Il convient cependant de souligner que l'utilisation d'un maillage grossier sur les parties fluviales conduit à une erreur sur l'estimation des niveaux d'eau à Pessac faute d'un raffinement suffisant du maillage et compte-tenu de l'injection en ce point des débits de la Dronne et de l'Isle.

Les niveaux d'eau prédits à La Réole, en amont de la Garonne, sont incertains et correspondent à $\pm$ 15 cm, soit une incertitude de 17 % par rapport au débit observé $QGAR^{(obs)}$ (en bleu sur la figure 7.1). Il est alors possible de construire une chronique temporelle prédite du débit de la Garonne, $QGAR^{(pred)}$ (en rouge sur la figure 7.1). La perturbation appliquée à $QGAR^{(obs)}$ pour obtenir $QGAR^{(pred)}$ est représentée en pointillés noirs pour la période du 01/01/2016 06:00 au 10/01/2016 18:00. Au temps $t_{k,fin}$ du cycle d'assimilation $\mathcal{F}_k$ correspondant au début de la phase de prévision, le débit de la Garonne est "basculé" de la chronique temporelle réelle observée vers la chronique temporelle prédite $QGAR^{(pred)}$. Le signal en vert ($QGAR_exemple$) correspond au débit imposé pour le cycle d'assimilation dont le mode "prévision" débute à t=558000 s sur la figure 7.1. Le passage au mode prévision est donc susceptible de créer une variation brutale du débit imposé à l'amont de la Garonne.

#### 7.2.3.2 Condition aux limites maritimes

Les forçages météorologiques ne peuvent pas être ceux de la Prévision d'Ensemble ARPEGE (PEARP), qui conduirait à modifier la nature des erreurs sur la condition aux limites maritimes, mais ceux de la simulation de prévision de METEO-FRANCE associée à une condition aux limites maritimes $CLMAR^{(pred)}$. En mode "prévision", il est possible de conserver les coefficients modaux constants et d'appliquer la perturbation résultante à $CLMAR^{(pred)}$.

#### 7.2.3.3 Coefficients de Strickler

Le chapitre 6 a montré une variabilité temporelle des coefficients de Strickler dans l'estuaire de la Gironde liée à la périodicité du signal de marée pour ce qui concerne $Ks1$ et $Ks2$ et également $Ks3$ pour l'Exp. $R3$. Il paraît assez légitime, par conséquent, d'envisager une reproduction de l'évolution temporelle des coefficients de Strickler sur la dernière période de réanalyse corres-

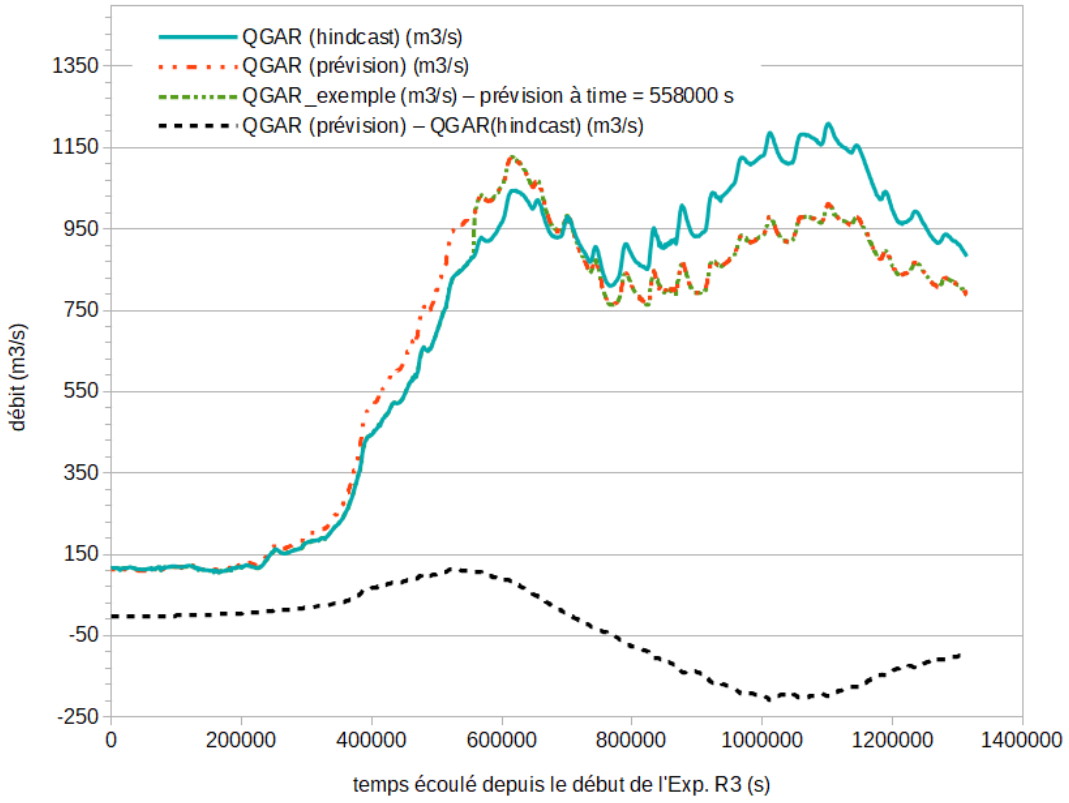



FIGURE 7.1 : Condition à la limite fluviale $QGAR$ du modèle pour le mode prévision de l'Exp. $R3$

pondant à la même portion du cycle de marée. En effet, la persistance du coefficient de Strickler conduirait, à l'issue d'un cycle d'assimilation de 3 h, à utiliser les mêmes coefficients de Strickler pendant 6 h supplémentaires, alors que sa structure d'évolution dépend du moment de la marée considéré.

## 7.3 Perspectives générales

Des propositions ont été formulées concernant l'évaluation des performances dans un cadre de prévision opérationnelle ensembliste pertinent compte-tenu des spécificités hydrodynamiques du modèle Telemac2D de l'estuaire de la Gironde, entièrement sous influence maritime. La simulation de prévision de METEO-FRANCE est importante dans ce cadre prévisionnel, indépendamment de la PEARP, afin d'y appliquer les corrections sur le forçage maritime issu de la phase de réanalyse de l'EnKF-$\gamma$-KLBC-MEG. Par ailleurs, dans la mesure où le dispositif d'expériences réelles a montré une évolution périodique des coefficients de frottement liée soit à la marée soit à la fenêtre d'assimilation de données, il paraît pertinent d'appliquer la structure temporelle de l'anomalie issue du dernier cycle de marée pour la prévision et de ne pas faire persister les valeurs

des paramètres de frottement à l'issue de la phase d'analyse de chaque cycle d'assimilation.

Les expériences réalisées dans un cadre d'assimilation de données réelles ont permis d'analyser l'évolution temporelle des variables de contrôle, l'influence de la composition du vecteur de contrôle obtenues avec l'EnKF-$\gamma$-KLBC-MEG. L'EnKF-$\gamma$-KLBC-MEG permet une amélioration de la simulation des niveaux d'eau dans l'estuaire en réanalyse d'événements passés.
Cependant, plusieurs perspectives d'amélioration existent concernant le cadre d'expériences réelles, à savoir une localisation de la matrice de covariance et une éventuelle émulation de celle-ci fondée sur la base de données construite dans le cadre de l'étude de quantification d'incertitudes. Par ailleurs, le coefficient de frottement $Ks3$ appliqué à la zone de la confluence entre les deux rivières Garonne et Dordogne, caractérisée par des processus hydrodynamiques non-linéaires, joue un rôle sensible, comme dans le cadre synthétique. Ce constat pourrait justifier l'utilisation d'une méthode d'assimilation de données plus adaptée à ce type d'hydrodynamique telle que l'IEnKF de Sakov et al. [2012].

Enfin, les champs météorologiques (champs de vent et de pression) sont des variables spatio-temporelles d'entrée du modèle. Elles sont bidimensionnelles et corrélées dans le temps. L'une des perspectives de ce travail est d'introduire les champs météorologiques dans l'analyse de sensibilité globale du chapitre 2 pour déterminer la part de variance des niveaux d'eau sur l'estuaire de la Gironde liée à cette variable. Une première piste, proposée par le CERFACS, est d'utiliser une analyse par composantes principales associée à une méthode par krigeage conditionnel. L'objectif est de construire des membres perturbés des forçages météorologiques associés à des chroniques de surcotes à la frontière maritime du modèle sur la base des modes propres de la PCA obtenus à l'aide des simulations de la Prévision d'Ensemble ARPEGE. La méthode de perturbation serait utilisée pour la GSA, dans un premier temps, puis intégrée dans l'EnKF-$\gamma$-KLBC. Des échanges techniques à ce sujet ont eu lieu avec METEO-FRANCE (CNRM/GMAP) et le CERFACS.

Dans un cadre plus large, compte-tenu des ressources informatiques nécessaires, une perspective à ce travail de thèse concerne la formulation d'un modèle de substitution (métamodèle) pour le modèle hydrodynamique 2D de l'estuaire de la Gironde afin de réduire le coût de l'EnKF-$\gamma$-KLBC. La combinaison d'une méta-modélisation et d'une méthode d"assimilation de données pourrait permettre de répondre à ce besoin en obtenant un algorithme d'assimilation de données ensembliste à coût de calcul plus faible pour réduire les principales sources d'incertitude et améliorer la prévision des niveaux d'eau dans l'estuaire de la Gironde.

# Bibliographie


Lorenzo Alfieri, Francesco Dottori, Richard Betts, Peter Salamon, and Luc Feyen. Multi-model projections of river flood risk in europe under global warming. ***Climate***, 6(1) :6, 2018.

S Almeida, L Rusu, C Guedes Soares, and Guedes Soares. Application of the ensemble kalman filter to a high-resolution wave forecasting model for wave height forecast in coastal areas. ***Maritime Technology and Engineering ; Guedes Soares, C., Santos, TA, Eds***, pages 1349–1354, 2015.

Mark Asch, Marc Bocquet, and Maëlle Nodet. ***Data assimilation : methods, algorithms, and applications***, volume 11. SIAM, 2016.

Sébastien Barthélémy. ***Assimilation de données ensembliste et couplage de modèles hydrauliques 1D-2D pour la prévision des crues en temps réel. Application au réseau hydraulique" Adour maritime"***. PhD thesis, 2015.

Sébastien Barthélémy, Sophie Ricci, Mélanie C Rochoux, Etienne Le Pape, and Olivier Thual. Ensemble-based data assimilation for operational flood forecasting–on the merits of state estimation for 1d hydrodynamic forecasting through the example of the "adour maritime" river. ***Journal of hydrology***, 552 :210–224, 2017.

Sébastien Barthélémy. Rapport d'avancement projet gironde xl - lot 5 : calage automatique de paramètres. Technical report, EDF R&D et Laboratoire d'Hydraulique St Venant, 2018.

Charles Troupin Beckers. Introduction to optimal interpolation and variational analysis. 2008.

Laurent Bertino, Geir Evensen, and Hans Wackernagel. Combining geostatistics and kalman filtering for data assimilation in an estuarine system. ***Inverse problems***, 18(1) :1, 2002.

Sylvain Biancamaria, Dennis P Lettenmaier, and Tamlin M Pavelsky. The swot mission and its capabilities for land hydrology. In ***Remote Sensing and Water Resources***, pages 117–147. Springer, 2016.

Marc Bocquet and Pavel Sakov. Joint state and parameter estimation with an iterative ensemble kalman smoother. ***Nonlinear Processes in Geophysics***, 20(5) :803–818, 2013.

Marc Bocquet and Pavel Sakov. An iterative ensemble kalman smoother. ***Quarterly Journal of the Royal Meteorological Society***, 140(682) :1521–1535, 2014.

Bertrand Bonan, Maëlle Nodet, Catherine Ritz, and Vincent Peyaud. An etkf approach for initial state and parameter estimation in ice sheet modelling. 2014.

Darwin Brochero, François Anctil, and Christian Gagné. Simplifying a hydrological ensemble prediction system with a backward greedy selection of members–part 1 : Optimization criteria. ***Hydrology and Earth System Sciences***, 15(11) :3307–3325, 2011.

Gerrit Burgers, Peter Jan van Leeuwen, and Geir Evensen. Analysis scheme in the ensemble kalman filter. ***Monthly weather review***, 126(6) :1719–1724, 1998.

Ângela Canas, Aires dos Santosa, and Paulo Leitãob. Implementation and validation of a sfek data assimilation application for an hydrodynamic model of the tagus estuary.

R Canižares Torres. ***On the application of data assimilation in regional coastal models***. IHE Delft Institute for Water Education, 1999.

Alberto Carrassi, Marc Bocquet, Laurent Bertino, and Geir Evensen. Data assimilation in the geosciences : An overview of methods, issues, and perspectives. ***Wiley Interdisciplinary Reviews : Climate Change***, 9(5) :535, 2018.

Patrice Castaing and George P Allen. Mechanisms controlling seaward escape of suspended sediment from the gironde : a macrotidal estuary in france. ***Marine Geology***, 40(1-2) :101–118, 1981.

Xueen Chen, Peng Zhan, Jinrui Chen, and Hongbao Qian. Numerical study of current fields near the changjiang estuary and impact of quick-enkf assimilation. ***Acta Oceanologica Sinica***, 30 (5) :33, 2011.

Yan Chen and Dean S Oliver. Cross-covariances and localization for enkf in multiphase flow data assimilation. ***Computational Geosciences***, 14(4) :579–601, 2010.

Elizabeth S Cooper, Sarah L Dance, Javier Garcia-Pintado, Nancy K Nichols, and Polly J Smith. Observation impact, domain length and parameter estimation in data assimilation for flood forecasting. ***Environmental Modelling & Software***, 104 :199–214, 2018.

Dick P Dee, SM Uppala, AJ Simmons, Paul Berrisford, P Poli, S Kobayashi, U Andrae, MA Balmaseda, G Balsamo, d P Bauer, et al. The era-interim reanalysis : Configuration and performance of the data assimilation system. ***Quarterly Journal of the royal meteorological society***, 137(656) :553–597, 2011.

Cécile Defforge. ***Assimilation de données pour des applications micro-météorologiques avec le modèle de mécanique des fluides Code_Saturne***. PhD thesis, 2019.

Cécile L Defforge, B Carissimo, M Bocquet, R Bresson, and P Armand. Improving cfd atmospheric simulations at local scale for wind resource assessment using the iterative ensemble kalman smoother. ***Journal of Wind Engineering and Industrial Aerodynamics***, 189 :243–257, 2019.

Gérald Desroziers, Loic Berre, Bernard Chapnik, and Paul Poli. Diagnosis of observation, background and analysis-error statistics in observation space. ***Quarterly Journal of the Royal Meteorological Society : A journal of the atmospheric sciences, applied meteorology and physical oceanography***, 131(613) :3385–3396, 2005.

Deepak Devegowda, Elkin Arroyo-Negrete, and Akhil Datta-Gupta. Flow relevant covariance localization during dynamic data assimilation using enkf. ***Advances in water resources***, 33 (2) :129–145, 2010.

Luc Devroye. Nonuniform random variate generation. ***Handbooks in operations research and management science***, 13 :83–121, 2006.

James Done, Christopher A Davis, and Morris Weisman. The next generation of nwp : Explicit forecasts of convection using the weather research and forecasting (wrf) model. ***Atmospheric Science Letters***, 5(6) :110–117, 2004.

Nabil El Moçayd. ***La décomposition en polynôme du chaos pour l'amélioration de l'assimilation de données ensembliste en hydraulique fluviale***. PhD thesis, 2017.

Paula Etala, Martín Saraceno, and Pablo Echevarría. An investigation of ensemble-based assimilation of satellite altimetry and tide gauge data in storm surge prediction. ***Ocean Dynamics***, 65(3) :435–447, 2015.

Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. ***Journal of Geophysical Research : Oceans***, 99(C5) :10143–10162, 1994.

Geir Evensen. The ensemble kalman filter : Theoretical formulation and practical implementation. ***Ocean dynamics***, 53(4) :343–367, 2003.

Geir Evensen. ***Data assimilation : the ensemble Kalman filter***. Springer Science & Business Media, 2009a.

Geir Evensen. The ensemble kalman filter for combined state and parameter estimation. ***IEEE Control Systems Magazine***, 29(3) :83–104, 2009b.

Roger A Flather. Results from a storm surge prediction model of the north-west european continental shelf for april, november and december, 1973. 1976.

Sergey Frolov, António M Baptista, Todd K Leen, Zhegdong Lu, and Rudolph van der Merwe. Fast data assimilation using a nonlinear kalman filter and a model surrogate : An application to the columbia river estuary. ***Dynamics of Atmospheres and Oceans***, 48(1-3) :16–45, 2009.

Guillaume Gastineau, Blandine L'hévéder, Francis Codron, and Claude Frankignoul. Mechanisms determining the winter atmospheric response to the atlantic overturning circulation. ***Journal of Climate***, 29(10) :3767–3785, 2016.

Ph Gauckler. ***Etudes Théoriques et Pratiques sur l'Ecoulement et le Mouvement des Eaux***. Gauthier-Villars, 1867.

Maryam Golnaraghi. ***Institutional partnerships in multi-hazard early warning systems : a compilation of seven national good practices and guiding principles***. Springer Science & Business Media, 2012.

Nicole Goutal, Cedric Goeury, Riadh Ata, Sophie Ricci, Nabil El Mocayd, M Rochoux, Hind Oubanas, Igor Gejadze, and Pierre-Olivier Malaterre. Uncertainty quantification for river flow simulation applied to a real test case : The garonne valley. In ***Advances in Hydroinformatics***, pages 169–187. Springer, 2018.

Bon-Ho Gu, Seung-Buhm Woo, and Sangil Kim. Improved estuaries salinity stratification at gyeonggi bay using data assimilation with finite volume coastal ocean model (fvcom). ***Journal of Coastal Research***, 91(sp1) :416–420, 2019.

Johan Habert, Sophie Ricci, E Le Pape, O Thual, Andrea Piacentini, Nicole Goutal, G Jonville, and M Rochoux. Reduction of the uncertainties in the water level-discharge relation of a 1d hydraulic model in the context of operational flood forecasting. ***Journal of Hydrology***, 532 : 52–64, 2016.

J Hartnack, HENRIK Madsen, and J Sørensen. Data assimilation in a combined 1d–2d flood model using the ensemble kalman filter. In ***4th Asia-Pacific DHI Software Conf***, 2006.

Jean-Michel Hervouet. ***Hydrodynamics of free surface flows : modelling with the finite element method***. John Wiley & Sons, 2007.

Jean-Michel Hervouet, Christophe Denis, and Eric David. Revisiting the thompson boundary conditions. ***Proceedings of the XVIIIth Telemac & Mascaret User Club 2011, 19-21 October 2011, EDF R&D, Chatou***, pages 142–147, 2011.

François Hissel, Gilles Morel, Gianluca Pescaroli, Herman Graaff, Didier Felts, and Luca Pietrantoni. Early warning and mass evacuation in coastal cities. ***Coastal Engineering***, 87 :193–204, 2014.

François Hissel. Projet gironde – calage du modèle hydraulique de l'estuaire de la gironde. Technical report, CETMEF, 2007.

François Hissel. Projet gironde – rapport final d'évaluation du modèle gironde. Technical report, CETMEF, 2010.

François Hissel and David Morellato. Projet gironde – rapport d'évaluation des scénarios débordants. Technical report, CETMEF, 2009.

François Hissel, Vanessya Laborie, and Réginald Sarralde. Projet gironde – rapport d'évaluation des performances du modèle edf/lnhe. Technical report, CETMEF, 2006.

Brian R Hunt, Eric J Kostelich, and Istvan Szunyogh. Efficient data assimilation for spatiotemporal chaos : A local ensemble transform kalman filter. ***Physica D : Nonlinear Phenomena***, 230 (1-2) :112–126, 2007.

Nicolas Huybrechts, Catherine Villaret, and Florent Lyard. Optimized predictive two-dimensional hydrodynamic model of the gironde estuary in france. ***Journal of Waterway, Port, Coastal and Ocean Engineering***, 138(4) :312–322, 2011.

Nicolas Huybrechts, Sylvain Orseau, Fatiha Chabi, Pablo Tassi, and Fabrice Klein. Rapport d'avancement sur la modélisation hydro-sédimentaire à calcul rapide (lot 3). Technical report, Cerema and EDF LNHE and GPMB, 2018a.

Nicolas Huybrechts, Hassan Smaoui, Sylvain Orseau, and Xuan Nam Do. Automatic calibration of the bed friction coefficient using optimization algorithm (lot 5 progress report december 2018 ). Technical report, Cerema and UTC Roberval, 2018b.

IPCC. Ipcc, 2019 : Summary for policymakers. In ***IPCC Special Report on the Ocean and Cryosphere in a Changing Climate***, page 1203 pages. H.–O. Pörtner, D.C. Roberts, V. Masson–Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, M. Nicolai, A. Okem, J. Petzold, B. Rama, N. Weyer (eds.), 2019. URL http://report.ipcc.ch/sr15/pdf/sr15_spm_final.pdf.

LC Jackson, R Kahana, T Graham, MA Ringer, T Woollings, JV Mecking, and RA Wood. Global and european climate impacts of a slowdown of the amoc in a high resolution gcm. ***Climate dynamics***, 45(11-12) :3299–3316, 2015.

Stat4Decision Emmanuel Jakobowicz. Le box-plot ou la fameuse boîte à moustache, 2018. URL https://www.stat4decision.com/fr/le-box-plot-ou-la-fameuse-boite-a-moustache/.

Andrew H Jazwinski. ***Stochastic processes and filtering theory***. Courier Corporation, 2007.

Nelly Jean-Baptiste, Pierre-Olivier Malaterre, Christophe Dorée, and Jacques Sau. Data assimilation for real-time estimation of hydraulic states and unmeasured perturbations in a 1d hydrodynamic model. ***Mathematics and Computers in Simulation***, 81(10) :2201–2214, 2011.

Didier Jourdan, Denis Paradis, Audrey Pasquet, Héloïse Michaud, Flavien Gouillon, Rémy Baraille, Laurie Biscara, Guillaume Voineson, and Patrick Ohl. Le projet homonim. ***COMITE SCIENTIFIQUE SCIENTIFIC COMMITTEE***, page 88.

Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. ***Journal of basic Engineering***, 82(1) :35–45, 1960.

Rama Rao Karri, Xuan Wang, and Herman Gerritsen. Ensemble based prediction of water levels and residual currents in singapore regional waters for operational forecasting. ***Environmental modelling & software***, 54 :24–38, 2014.

RR Karri and V Babovic. Enhanced predictions of tides and surges through data assimilation. ***International Journal of Engineering-Transactions A : Basics***, 30(1) :23–29, 2017.

Vanessya Laborie, François Hissel, and Philippe Sergent. Évolution de l'emprise des zones inondables de l'estuaire de la gironde sous l'influence du changement climatique. 2012.

Vanessya Laborie, François Hissel, and Philippe Sergent. Impact of climate change on gironde estuary. ***La Houille Blanche***, (6) :34–39, 2014.

Vanessya Laborie, Philippe Sergent, Florence Levy, Roberto Frau, and Jérome Weiss. The hydrodynamic, sea-state and infrastructures platform developed by saint-venant hydraulics laboratory and cerema : a special focus on the telemac2d surge levels numerical model of the atlantic ocean, the channel and the north sea. In ***Proceedings of the XXII TELEMAC-MASCARET Technical User Conference October 15-16, 2043***, pages 172–181, 2015.

Vanessya Laborie, Sophie Ricci, Matthias De Lozzo, Nicole Goutal, Yoann Audouin, and Philippe Sergent. Quantifying forcing uncertainties in the hydrodynamics of the gironde estuary. ***Computational Geosciences***, pages 1–22, 2019.

Florence Levy. Construction d'un modéle de surcotes sur la façade atlantique. Technical report, CETMEF, 2013. rapport provisoire.

Yuqiong Liu, A Weerts, M Clark, H-J Hendricks Franssen, S Kumar, H Moradkhani, D-J Seo, D Schwanenberg, P Smith, AIJM Van Dijk, et al. Advancing data assimilation in operational hydrologic forecasting : progresses, challenges, and emerging opportunities. 2012.

Henrik Madsen and Claus Skotner. Adaptive state updating in real-time river flow forecasting—a combined filtering and error forecasting procedure. ***Journal of Hydrology***, 308(1-4) :302–312, 2005.

Sandrine Marano. Rapport sur l'inondation du site du blayais survenue le 27 décembre 1999. Technical report, IRSN, 2000. URL https://www.irsn.fr/FR/expertise/rapports_expertise/Documents/surete/rapport_sur_l_inondation_du_site_du_blayais.pdf.

Amandine Marrel. ***Mise en oeuvre et utilisation du métamodèle processus gaussien pour l'analyse de sensibilité de modèles numériques : application à un code de transport hydrogéologique***. PhD thesis, Toulouse, INSA, 2008.

Peter S Maybeck. ***Stochastic models, estimation, and control***, volume 3. Academic press, 1982.

Jaak Monbaliu, Zhongyuan Chen, Didier Felts, Jianzhong Ge, Francois Hissel, Jens Kappenberg, Siddharth Narayan, Robert J Nicholls, Nino Ohle, Dagmar Schuster, et al. Risk assessment of estuaries under climate change : lessons from western europe. ***Coastal Engineering***, 87 : 32–49, 2014.

Andrew Michael Moore, Matthew James Martin, Santha Akella, Hernan Arango, Magdalena Alonso Balmaseda, Laurent Bertino, Stefano Ciavatta, Bruce Cornuelle, James Cummings, Sergey Frolov, et al. Synthesis of ocean observations using data assimilation for operational, real-time and reanalysis systems : A more complete picture of the state of the ocean. ***Frontiers in Marine Science***, 6 :90, 2019.

Hamid Moradkhani, Soroosh Sorooshian, Hoshin V Gupta, and Paul R Houser. Dual state–parameter estimation of hydrological models using ensemble kalman filter. ***Advances in water resources***, 28(2) :135–147, 2005.

Tongchao Nan and Jichun Wu. Groundwater parameter estimation using the ensemble kalman filter with localization. ***Hydrogeology journal***, 19(3) :547–561, 2011.

Hind Oubanas, Igor Gejadze, Pierre-Olivier Malaterre, and Franck Mercier. River discharge estimation from synthetic swot-type observations using variational data assimilation and the full saint-venant hydraulic model. ***Journal of Hydrology***, 559 :638–647, 2018.

Florian Pappenberger, Hannah L Cloke, Dennis J Parker, Fredrik Wetterhall, David S Richardson, and Jutta Thielen. The monetary benefit of early flood warnings in europe. ***Environmental Science & Policy***, 51 :278–291, 2015.

V Pedinotti, A Boone, S Ricci, Sylvain Biancamaria, and N Mognard. Assimilation of satellite data to optimize large-scale hydrological model parameters : a case study for the swot mission. ***Hydrology and Earth System Sciences***, 18(11) :4485–4507, 2014.

Yannick Peings, Graham Simpkins, and Gudrun Magnusdottir. Multidecadal fluctuations of the north atlantic ocean and feedback on the winter climate in cmip5 control simulations. ***Journal of Geophysical Research : Atmospheres***, 121(6) :2571–2592, 2016.

Dinh Tuan Pham. Stochastic methods for sequential data assimilation in strongly nonlinear systems. ***Monthly weather review***, 129(5) :1194–1207, 2001.

Dinh Tuan Pham, Jacques Verron, and Marie Christine Roubaud. A singular evolutive extended kalman filter for data assimilation in oceanography. ***Journal of Marine systems***, 16(3-4) : 323–340, 1998.

Serge Planton, Gonéri Le Cozannet, Anny Cazenave, Stéphane Costa, Olivier Douez, Pierre Gaufrès, François Hissel, Déborah Idier, Vanessya Laborie, Vincent Petit, et al. Le climat de la

france au xxie siècle. vol 5 : Changement climatique et niveau de la mer : de la planète aux côtes françaises. 2015.

A Poncot, C Goeury, J-P Argaud, F Zaoui, Riadh Ata, and Yoann Audouin. Optimal calibration of telemac-2d models based on a data assimilation algorithm. In ***Proceedings of the XXIVth TELEMAC-MASCARET User Conference, 17 to 20 October 2017, Graz University of Technology, Austria***, pages 73–80, 2017.

Naila F Raboudi, Boujemaa Ait-El-Fquih, Clint Dawson, and Ibrahim Hoteit. Combining hybrid and one-step-ahead smoothing for efficient short-range storm surge forecasting with an ensemble kalman filter. ***Monthly Weather Review***, (2019), 2019.

Sophie Ricci, Andrea Piacentini, Olivier Thual, E Le Pape, and Gabriel Jonville. Correction of upstream flow and hydraulic state with data assimilation in the context of flood forecasting. ***Hydrology and Earth System Sciences***, 15(11) :3555–3575, 2011.

Pamphile T Roy, Nabil El Moçayd, Sophie Ricci, Jean-Christophe Jouhaud, Nicole Goutal, Matthias De Lozzo, and Mélanie C Rochoux. Comparison of polynomial chaos and gaussian process surrogates for uncertainty quantification and correlation estimation of spatially distributed open-channel steady flows. ***Stochastic Environmental Research and Risk Assessment***, 32(6) :1723–1741, 2018.

Pavel Sakov, Dean S Oliver, and Laurent Bertino. An iterative enkf for strongly nonlinear systems. ***Monthly Weather Review***, 140(6) :1988–2004, 2012.

Pavel Sakov, Jean-Matthieu Haussaire, and Marc Bocquet. An iterative ensemble kalman filter in the presence of additive model error. ***Quarterly Journal of the Royal Meteorological Society***, 144(713) :1297–1309, 2018.

Jean-Noël Salomon. L'inondation dans la basse vallée de la garonne et l'estuaire de la gironde lors de la" tempête du siècle"(27-28 décembre 1999)/flooding in the garonne valley and the gironde estuary caused by the" storm of the century"(27-28 december 1999). ***Géomorphologie : relief, processus, environnement***, 8(2) :127–134, 2002.

Yoshikazu Sasaki. A fundamental study of the numerical prediction based on the variational principle. ***Journal of the Meteorological Society of Japan. Ser. II***, 33(6) :262–275, 1955.

Michiharu Shiiba, Xavier Laurenson, and Yasuto Tachikawa. Real-time stage and discharge estimation by a stochastic-dynamic flood routing model. ***Hydrological processes***, 14(3) :481–495, 2000.

Ehouarn Simon and Laurent Bertino. Application of the gaussian anamorphosis to assimilation in a 3-d coupled physical-ecosystem model of the north atlantic with the enkf : a twin experiment. ***Ocean Science***, 5(4) :495–510, 2009.

Ehouarn Simon and Laurent Bertino. Gaussian anamorphosis extension of the denkf for combined state parameter estimation : Application to a 1d ocean ecosystem model. ***Journal of Marine Systems***, 89(1) :1–18, 2012.

Adil Siripatana, Talea Mayo, Omar Knio, Clint Dawson, Olivier Le Maitre, and Ibrahim Hoteit. Ensemble kalman filter inference of spatially-varying manning'sn coefficients in the coastal ocean. ***Journal of hydrology***, 562 :664–684, 2018.

Jonathan R Stroud, Matthias Katzfuss, and Christopher K Wikle. A bayesian adaptive ensemble kalman filter for sequential state and parameter estimation. ***Monthly Weather Review***, 146 (1) :373–386, 2018.

Leqiang Sun, Ousmane Seidou, Ioan Nistor, and Kailei Liu. Review of the kalman-type hydrological data assimilation. ***Hydrological Sciences Journal***, 61(13) :2348–2366, 2016.

Hitoshi Tamura, Peter Bacopoulos, Dingbao Wang, Scott C Hagen, and Ethan J Kubatko. State estimation of tidal hydrodynamics using ensemble kalman filter. ***Advances in water resources***, 63 :45–56, 2014.

Pierre Tandeo, Pierre Ailliot, Marc Bocquet, Alberto Carrassi, Takemasa Miyoshi, Manuel Pulido, and Yicun Zhen. Joint estimation of model and observation error covariance matrices in data assimilation : a review. ***arXiv preprint arXiv :1807.11221***, 2018.

M Verlaan and AW Heemink. Nonlinearity in data assimilation applications : A practical method for analysis. ***Monthly Weather Review***, 129(6) :1578–1589, 2001.

Jasper A Vrugt, Hoshin V Gupta, BreanndánÓ Nualláin, and Willem Bouten. Real-time data assimilation for operational ensemble streamflow forecasting. ***Journal of Hydrometeorology***, 7(3) :548–565, 2006.

AH Weerts, HC Winsemius, and JS Verkade. Estimation of predictive hydrological uncertainty using quantile regression : examples from the national flood forecasting system (england and wales). ***Hydrology and Earth System Sciences***, 15(1) :255–265, 2011.

Tim Woollings, Jonathan M Gregory, Joaquim G Pinto, Mark Reyers, and David J Brayshaw. Response of the north atlantic storm track to climate change shaped by ocean–atmosphere coupling. ***Nature Geoscience***, 5(5) :313, 2012.

Jin Wu. Wind-stress coefficients over sea surface near neutral conditions—a revisit. ***Journal of Physical Oceanography***, 10(5) :727–740, 1980.

Margaret Yelland and Peter K Taylor. Wind stress measurements from the open ocean. ***Journal of Physical Oceanography***, 26(4) :541–558, 1996.

Firmijn Zijl, Julius Sumihar, and Martin Verlaan. Application of data assimilation for improved operational water level forecasting on the northwest european shelf and north sea. ***Ocean Dynamics***, 65(12) :1699–1716, 2015.

Matteo G Ziliani, Rabih Ghostine, Boujemaa Ait-El-Fquih, Matthew F McCabe, and Ibrahim Hoteit. Enhanced flood forecasting through ensemble data assimilation and joint state-parameter estimation. ***Journal of Hydrology***, 577 :123924, 2019.

# Annexe A

# Conséquences d'un rééchantillonnage sur la matrice de covariance croisée d'erreurs prédites entre le vecteur de contrôle et les observations simulées

La technique de rééchantillonnage proposée dans la section 4.1.4 correspond à une redispersion de l'ensemble de contrôle à l'issue de la phase d'analyse du cycle d'assimilation $\mathcal{F}_k$ selon : $P_k^{\gamma} = P_0^{\gamma}$.

Cette section montre que le choix de cette méthode implique que l'évolution temporelle de la covariance $\mathbf{P}_k^{\boldsymbol{\gamma},\mathbf{y}} = (p_{i_1,i_{obs}})_{1\leq i_1\leq nvar;1\leq i_{obs}\leq n_{obs}}$ entre les niveaux d'eau calculés par Telemac2D et un coefficient modal choisi comme variable de contrôle est alors proportionnel au mode propre correspondant. Dans le cas d'une valeur de $N_e$ faible, comme c'est le cas ici, l'erreur sur la covariance entre un coefficient modal et une station d'observation peut alors être calculée à chaque pas de temps.

La démonstration suppose l'existence d'une fonction de transfert $\beta_{i_{obs}}$ de toute perturbation appliquée à la condition aux limites à la frontière liquide vers chaque station $i_{obs}$ de l'estuaire de la Gironde sous l'influence de cette condition aux limites. Il a ainsi été vérifié que, pour la condition aux limites maritimes, $\beta_{i_{obs}} = 1 + O(1)$ pour les stations les plus proches de l'embouchure de l'estuaire de la Gironde, comme le montre la figure A.1. Sur les frontières fluviales, pour les stations de Pessac et de La Réole, il vient immédiatement $\beta_{i_{obs}} = 1$. $\beta_{i_{obs}} \neq 1$ correspond physiquement aux non-linéarités du système lors de la propagation de la perturbation vers l'intérieur de l'estuaire.

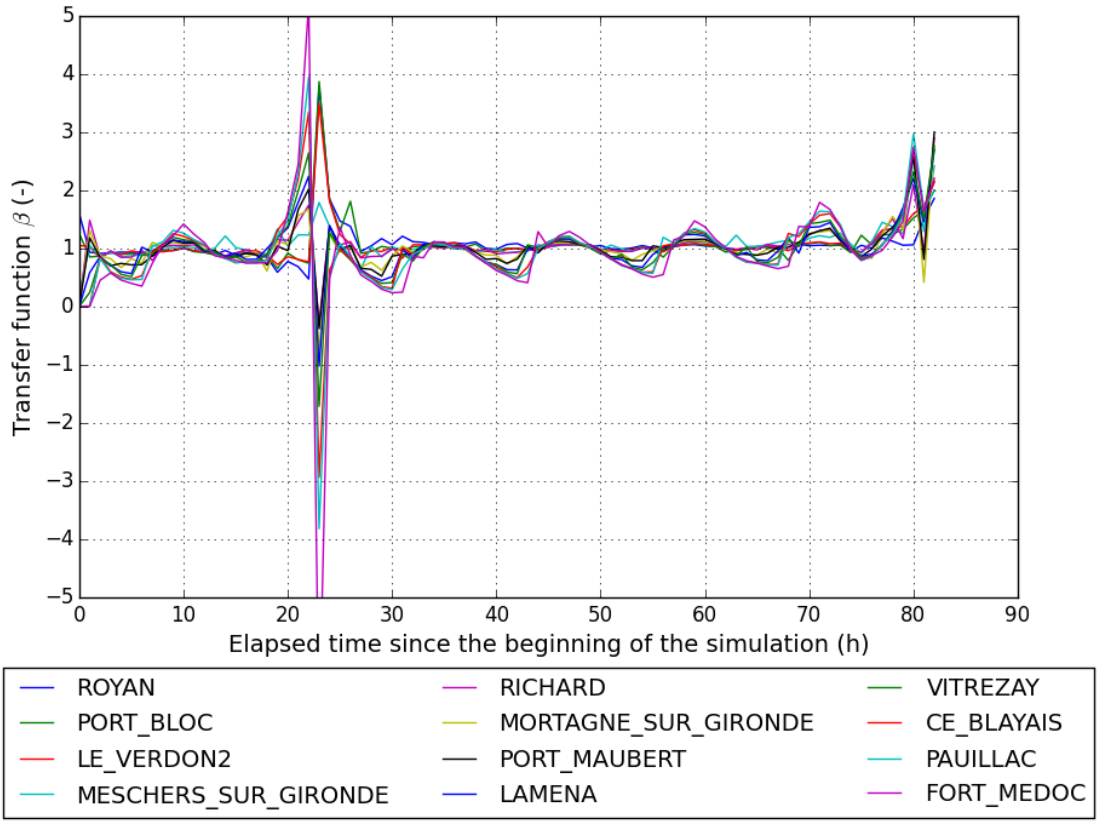



FIGURE A.1 : Fonction de transfert $\beta$ pour les stations de l'estuaire de la Gironde.

En décomposant en composantes principales la matrice d'autocorrélation temporelle $\mathbf{R}_{dsm}^{(0)}$ du vecteur de perturbation **dsm** appliqué à la frontière liquide et discrétisé sur le vecteur temporel $[t_1\,;\,t_{nt}]$ au début de l'expérience selon la méthodologie décrite dans le chapitre 2 et en notant $n_{modes} \in \{p_{CLMAR}, p_{DOR}, p_{GAR}\}$ le nombre de modes retenu pour la frontière liquide considérée, $\mathbf{R}_{dsm}^{(0)}$ s'écrit :

$$\mathcal{R}^{(0)}(t_i,t_j) = \sum_{np=1}^{n_{modes}} \lambda_{np}\,\phi_{np}(t_i)\,\phi_{np}(t_j) \tag{A.1}$$

avec :

$$\mathbf{R}_{dsm}^{(0)}\Phi_{np} = \lambda_{np}\Phi_{np},$$
$$\Phi_{np} = (\phi_{np}(t_1)...\phi_{np}(t_{nt})),$$
$$\mathbf{R}_{dsm}^{(0)} = (\mathcal{R}^{(0)}(t_i,t_j)_{1\leq i,j\leq n_p})$$

$(\lambda_{np},\,\phi_{np})$ sont les $np^e$ valeur propre et vecteur propre de $\mathcal{R}^{(0)}$, solution de l'équation de Fredholm, comme indiqué dans le chapitre 2.

Par ailleurs, le membre $l$ de la perturbation $\mathbf{dsm}^{(k)}$ appliquée à la condition aux limites pendant la fenêtre temporelle $\mathcal{F}_k$ s'écrit :

$$\mathbf{dsm}_l^{(k)}(t) = \sum_{np=1}^{n_{modes}} \sqrt{\lambda_{np}}\, \boldsymbol{\alpha}_k^{np,l}\, \phi_{np}(t) \tag{A.2}$$

où $\left\{\boldsymbol{\alpha}_k^{np,l}\right\}_{l\,\in\,\{1,\cdots,N_e\}}$ désigne l'ensemble des coefficients modaux liés à la $np^e$ composante dans l'espace réduit de KL sur la fenêtre d'assimilation $\mathcal{F}_k$. La technique de rééchantillonnage utilisée implique la réinjection de la matrice de covariance initiale $\mathbf{P}_0^{\boldsymbol{\gamma}}$ pour chaque variable de contrôle et, en particulier, pour chaque coefficient modal, ce qui revient à écrire pour la fenêtre d'assimilation $\mathcal{F}_k$ :

$$\boldsymbol{\alpha}_k^{np,l} = \overline{\boldsymbol{\alpha}_k^{np}} + \boldsymbol{\alpha}_0^{np,l} - \overline{\boldsymbol{\alpha}_0^{np}} \tag{A.3}$$

où

$$\overline{\boldsymbol{\alpha}_k^{np}} = \frac{1}{N_e}\sum_{l=1}^{N_e}\boldsymbol{\alpha}_k^{np,l} \tag{A.4}$$

Il est alors possible d'écrire la perturbation du membre $l$ :

$$\mathbf{dsm}_l^{(k)} = \sum_{np=1}^{n_{modes}} \left(\sqrt{\lambda_{np}}\, \boldsymbol{\alpha}_0^{np,l}\phi_{np}\right) + \sum_{np=1}^{n_{modes}} \left(\sqrt{\lambda_{np}}(\overline{\boldsymbol{\alpha}_k^{np}} - \overline{\boldsymbol{\alpha}_0^{np}})\phi_{np}\right) \tag{A.5}$$

L'équation (A.5) s'écrit alors :

$$\mathbf{dsm}_l^{(k)} = \sum_{np=1}^{n_{modes}} \sqrt{\lambda_{np}}\, \boldsymbol{\alpha}_0^{np,l}\, \phi_{np} + \mathbf{C}^{(k)} \tag{A.6}$$

avec :

$$\mathbf{C}^{(k)} = \sum_{np=1}^{n_{modes}} \sqrt{\lambda_{np}}\, \left(\overline{\boldsymbol{\alpha}_k^{np}} - \overline{\boldsymbol{\alpha}_0^{np}}\right) \phi_{np} \tag{A.7}$$

$\mathbf{C}^{(k)}$ dépend du cycle $\mathcal{F}_k$, de $n_{modes}$ et non du membre $l$.

Ainsi :

$$\mathbf{dsm}_l^{(k)} = \mathbf{dsm}_l^{(0)} + \mathbf{C}^{(k)} \tag{A.8}$$

Si $inp$ est la ligne de $\mathbf{P}_k^{\boldsymbol{\gamma},\mathbf{y}}(t_i)$ correspondant au coefficient modal $\boldsymbol{\alpha}^{np}$ et $iobs$ la colonne correspondant à la station d'observation de la matrice $\mathbf{P}_k^{\boldsymbol{\gamma},\mathbf{y}}(t_i)$, alors la composante de $p_{inp,iobs}^{(k)}(t_i)$

s'écrit :

$$p^{(k)}_{inp,iobs}(t_i) = \frac{1}{N_e - 1}\sum_{l=1}^{N_e}(\boldsymbol{\alpha}^{np,l}_k - \overline{\boldsymbol{\alpha}^{np}_k})(y^{(k)}_{iobs,l}(t_i) - \overline{y^{(k)}_{iobs}}(t_i)) \tag{A.9}$$

En supposant l'existence d'une fonction de transfert $\beta_{iobs}$ entre la frontière liquide et la station d'observation qui dépend du modèle et pas de la perturbation elle-même, alors :

$$p^{(k)}_{inp,iobs}(t_i) = \frac{1}{N_e - 1}\sum_{l=1}^{N_e}(\boldsymbol{\alpha}^{np,l}_0 - \overline{\boldsymbol{\alpha}^{np}_0})\beta_{iobs}(t_i)\mathbf{dsm}^{(k)}_l(t_i) \tag{A.10}$$

En utilisant l'équation (A.8) :

$$p^{(k)}_{inp,iobs}(t_i) = \frac{1}{N_e - 1}\sum_{l=1}^{N_e}\beta_{iobs}(t_i)(\boldsymbol{\alpha}^{np,l}_0 - \overline{\boldsymbol{\alpha}^{np}_0})\left(\mathbf{dsm}^{(0)}_l(t_i) + \mathbf{C}^{(k)}(t_i)\right) \tag{A.11}$$

et :

$$p^{(k)}_{inp,iobs}(t_i) = \frac{1}{N_e - 1}\sum_{l=1}^{N_e}\left(\beta_{iobs}(t_i)(\boldsymbol{\alpha}^{np,l}_0 - \overline{\boldsymbol{\alpha}^{np}_0})\mathbf{dsm}^{(0)}_l(t_i)\right) + \beta_{iobs}(t_i)\frac{\mathbf{C}^{(k)}(t_i)}{N_e - 1}\sum_{l=1}^{N_e}(\boldsymbol{\alpha}^{np,l}_0 - \overline{\boldsymbol{\alpha}^{np}_0}) \tag{A.12}$$

Le dernier terme étant nul et sachant que : $\beta_{iobs}(t_i)\ \mathbf{dsm}^{(0)}_l(t_i) = (y^{(0)}_l - \overline{y^{(0)}})$,
il vient :

$$p^{(k)}_{inp,iobs}(t_i) = p^{(0)}_{inp,iobs}(t_i) \tag{A.13}$$

Il est alors possible de démontrer que $p^{(0)}_{inp,iobs}(t_i)$ est proportionnel au mode propre $\phi_{np}$ du coefficient modal $\boldsymbol{\alpha}^{np}$ correspondant à la ligne $inp$, ce qui implique que l'enveloppe globale de la fonction de covariance croisée pour les coefficients modaux des forçages temporels repose sur les modes propres correspondants de la fonction d'autocorrélation temporelle du signal.
$p^{(0)}_{inp,iobs}$ s'écrit pour chaque coefficient modal et avec $\beta = \beta_{iobs}$ correspondant à une station donnée $iobs$ :

$$p^{(0)}_{inp,iobs}(t_i) = \frac{1}{N_e - 1}\ \beta(t_i)\sum_{l=1}^{N_e}\left(\boldsymbol{\alpha}^{np,l}_0 - \overline{\boldsymbol{\alpha}^{np}_0}\right)\mathbf{dsm}^{(0)}_l(t_i) \tag{A.14}$$

Alors :

$$p^{(0)}_{inp,iobs}(t_i) = \frac{1}{N_e - 1}\ \beta(t_i)\sum_{l=1}^{N_e}\left(\boldsymbol{\alpha}^{np,l}_0\right) - \frac{1}{N_e - 1}\ \beta(t_i)\sum_{l=1}^{N_e}\left(\overline{\boldsymbol{\alpha}^{np}_0}\right)\mathbf{dsm}^{(0)}_l(t_i) = Q_1 + Q_2 \tag{A.15}$$

où :

$$Q_1 = \frac{1}{N_e - 1}\,\beta(t_i) \sum_{l=1}^{N_e} \left(\boldsymbol{\alpha}_0^{np,l}\right) \tag{A.16}$$

et

$$Q_2 = -\frac{1}{N_e - 1}\,\beta(t_i) \sum_{l=1}^{N_e} \left(\overline{\boldsymbol{\alpha}_0^{np}}\right) \mathbf{dsm}_l^{(0)}(t_i) \tag{A.17}$$

Il vient directement de l'équation (A.17) :

$$Q_2 = -\frac{1}{N_e - 1}\,\beta(t_i)\,\left(\overline{\boldsymbol{\alpha}_0^{np}}\right) \sum_{l=1}^{N_e} \mathbf{dsm}_l^{(0)}(t_i) \tag{A.18}$$

et donc :

$$Q_2 = -\frac{N_e}{N_e - 1}\,\beta(t_i)\,\overline{\boldsymbol{\alpha}_0^{np}}\;\overline{\mathbf{dsm}^{(0)}(t_i)} \tag{A.19}$$

En introduisant l'équation (A.2) dans l'équation (A.16) pour $t_k = t_0$, il vient :

$$Q_1 = \frac{1}{N_e - 1}\,\beta(t_i) \sum_{l=1}^{N_e} \left(\boldsymbol{\alpha}_0^{np,l}\right) \left( \sum_{np'=1}^{n_{modes}} \sqrt{\lambda_{np'}}\;\boldsymbol{\alpha}_0^{np',l}\;\phi_{np'}(t_i) \right) \tag{A.20}$$

$$Q_1 = \frac{1}{N_e - 1}\,\beta(t_i) \sum_{np'=1}^{n_{modes}} \sqrt{\lambda_{np'}}\;\phi_{np'}(t_i)\;\left( \sum_{l=1}^{N_e} \left(\boldsymbol{\alpha}_0^{np,l}\boldsymbol{\alpha}_0^{np',l}\right) \right) \tag{A.21}$$

En notant :

$$D_{np,np'} = \frac{1}{N_e} \sum_{l=1}^{N_e} \left(\boldsymbol{\alpha}_0^{np,l}\boldsymbol{\alpha}_0^{np',l}\right) \tag{A.22}$$

L'équation (A.21) s'écrit alors :

$$Q_1 = \frac{N_e}{N_e - 1}\,\beta(t_i) \left( \sum_{np'=1}^{n_{modes}} \sqrt{\lambda_{np'}}\;D_{np,np'}\;\phi_{np'}(t_i) \right) \tag{A.23}$$

Par contruction, la distribution $\left(\boldsymbol{\alpha}_0^{np,l}\right)_{\{l\in(1,..,N_e)\}}$ est gaussienne pour $Ne$ "grand". Dans le cadre de l'utilisation de l'EnKF-$\theta$-KLBC appliqué au modèle Telemac2D de l'estuaire de la Gironde, $N_e$ est inférieur à 200. Le tableau A.1 ci-dessous indique pour chaque $\boldsymbol{\alpha}^{np}$ et chaque $\boldsymbol{\alpha}^{np'}$ le rapport $\frac{D_{np,np'}}{D_{np,np}}$ pour $N_e = 100$.

TABLE A.1 : Rapport $\frac{D_{np,np'}}{D_{np,np}}$ pour chaque coefficient modal $\alpha_{np}$ avec (np , np') $\in \{1, .., n_{modes}\}^2$

| $np$ | $\alpha_{np}$ | $D_{np,np}$ | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ | $\alpha_5$ | $\alpha_6$ | $\alpha_7$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | $\alpha_1$ | 0.934 | 100 % | -2.4 % | -0.9 % | -1.7 % | 19.4 % | 1.5 % | -1.4 % |
| 2 | $\alpha_2$ | 0.906 | -2.4 % | 100 % | 5.5 % | -1.3 % | 1.1 % | -1.1 % | 0.7 % |
| 3 | $\alpha_3$ | 0.910 | -0.9 % | 5.4 % | 100 % | 6.7 % | -10.6 % | -1.9 % | -3.2 % |
| 4 | $\alpha_4$ | 0.912 | -1.7 % | -1.3 % | 6.6 % | 100 % | -3.7 % | -0.9 % | -3.1 % |
| 5 | $\alpha_5$ | 0.914 | 19.9 % | 1.1 % | -10.5 % | -3.7 % | 100 % | -4.5 % | -2.3 % |
| 6 | $\alpha_6$ | 0.906 | 1.6 % | -1.1 % | -1.9 % | -0.9 % | -4.5 % | 100 % | -1.6 % |
| 7 | $\alpha_7$ | 0.903 | -1.5 % | 0.7 % | -3.2 % | -3.1 % | -2.3 % | -1.6 % | 100 % |

$\forall np' \neq np$, les termes croisés $D_{np,np'}$ qui pondèrent la contribution des modes propres $\{\phi_{np'}\}$ sont inférieurs à $(0.2\ \% \times D_{np,np})$ quand $N_e = 5000$ et tendent vers 0 quand $N_e$ tend vers l'infini. Ainsi, sachant que, par construction, $\overline{\boldsymbol{\alpha}_0^{np}} = 0$, il vient :

$$p^{(0)}_{inp,iobs}(t_i) \underset{N_e \to +\infty}{\longrightarrow} \beta_{iobs}(t_i)\sqrt{\lambda_{np}}\ \phi_{np}(t_i) \tag{A.24}$$

L'erreur introduite par l'échantillonnage est alors :

$$e^{(0)}_{inp,iobs}(t_i) = \frac{N_e}{N_e - 1}\ \beta_{iobs}(t_i) \left( \left[ \sum_{\substack{np'=1 \\ np' \neq np}}^{n_{modes}} \sqrt{\lambda_{np'}}\ D_{np,np'}\ \phi_{np'}(t_i) \right] - \overline{\boldsymbol{\alpha}_0^{np}}\ \overline{\mathbf{dsm}^{(0)}(t_i)} \right) \tag{A.25}$$

Pour $N_e = 100$, $\forall np$, $\overline{\boldsymbol{\alpha}_0^{np}} < 3\ \%$ et, $\forall t_i$, $\overline{\mathbf{dsm}^{(0)}(t_i)} < 0.01$, le second terme peut donc être négligé. Il reste alors :

$$e^{(0)}_{inp,iobs}(t_i) = \frac{N_e}{N_e - 1}\ \beta_{iobs}(t_i) \left( \sum_{\substack{np'=1 \\ np' \neq np}}^{n_{modes}} \sqrt{\lambda_{np'}}\ D_{np,np'}\ \phi_{np'}(t_i) \right) \tag{A.26}$$

# Annexe B

# Développement informatique de l'EnKF-$\gamma$-KLBC

## B.1 Exemple de fichier de configuration

```
{
    "architecture":{
        "archi": 1,
        "mesh" : "mesh0"
    },
    "experiment":{
        "repsauv" : "
           fobs3600_mainstat_real20160109_Ne_5_DAW12h_L1hS1h"
    },
    "no_forcings":{
        "fichierREFOBSnometeo": "/T2RES/
           SF_MF_Q2003_trunk_mesh0_190h.res",
        "fichierREFOBS": "/T2DRES_REF/
           MEG0_PERT_993_CLMAR_t_0_68400_Kspallier.res"
    },
    "observations":{
        "locnatobs":[["SURFACE LIBRE   ","ROYAN"],
                ["SURFACE LIBRE   ","PAUILLAC"],
                ["SURFACE LIBRE   ","BEC_AMBES"],
                ["SURFACE LIBRE   ","BORDEAUX2"],
                ["SURFACE LIBRE   ","LA_REOLE"],["SURFACE LIBRE
                     ","LIBOURNE"]],
```

```
19        "dic_optionobs":{"insitu":true,"resfile":false,"
             satellite":false},
20        "dt":300.0,
21        "fobs":12,
22        "EVENT":15,"EVENT_DATE":[20160109]
23    },
24    "ensemble":{
25        "Ne": 5,
26        "dic_lxcontr": {"Ks1":true,"Ks2":true,"Ks3":true,"Ks4":
             false,"CDz":true,
27            "QGAR":true,"QDOR":true,"CLMAR":true,"METEO":false
                 },
28        "xpert_full":{"Ks1":true,"Ks2":true,"Ks3":true,"Ks4":
             false,"CDz":true,
29            "CLFLUV":true,"CLMAR":true,"METEO":false},
30        "xcontr_opt_clmar":{"CLMAR":"alpha" ,"QGAR":"alpha", "
             QDOR":"alpha"},
31        "opt_dsm":"gp1d_sampler","opt_qgar":"gp1d_sampler","
             opt_qdor":"gp1d_sampler",
32        "n_var_csttes" : 5, "n_modes_CLFLUVDOR":4, "
             n_modes_CLFLUVGAR": 4, "n_modes_CLMAR" : 7,
33        "std_Ks": 10.0,
34        "KSinit":[60.0,60.0,50.0,40.0],
35        "KStarget":[60.0,60.0,50.0,40.0],
36        "CDzinit":[2.57e-06],
37        "CDztarget":[2.57e-06]
38    },
39    "T2Dsettings":{
40        "studyFiles":{"t2d.f"    :"user_fortran_mesh0",
41                "t2d.cas" :"
                     cas_AF_trunk_mesh0_newbathy_chemabs_6heures.
                     txt",
42                "t2d.geo" :"./geometrie/
                     estuaireGir8_t2d_newbathy.geo"},
43        "dt_t2d"     : 10.0,
44        "fsave_t2d" : 360
45    },
46    "Assim":{
47        "OptInflaGlob": true,
```

```
        "test_R_desroziers": false,
        "OptInflaGlob_desroziers" : false,
        "OptInflaCLMAR": false,
        "OptInflaQGAR": false,
        "OptInflaQDOR": false,
        "OptInflaKs": false,
        "OptInflaCDz": false,
        "sigma_obs": 0.01,
        "sigma_t2d_Ks": 0.01,
        "sigma_t2d_CDz": 5e-7,
        "sigma_t2d_QFLUV": 0.20,
        "sigma_t2d_sm" : 0.01,
        "sigma_t2d_QFLUV_dQ": 1.0,
        "sigma_t2d_QFLUV_alpha":  0.001,
        "lbda": 1.0,
        "dtassim" : 3600.0,
        "duree_DAW" : 237600.0,
        "F" : 6,
        "L" : 1,
        "FrDAW" : 0,
        "FrAfter" : 48,
        "S" : 1
    }
}
```

## B.2 Description du paramétrage

Le fichier de configuration s'appuie sur le format de données JSON (Javascript Object Notation) et contient plusieurs blocs décrits ci-dessous

#### B.2.0.1 "architecture"

Le bloc "architecture" permet de choisir la machine sur laquelle l'expérience va tourner (option "archi"). Dans le cas présent, il s'agit de choisir soit le cluster du Cerema, soit le cluster EOLE d'EDF R&D.

#### B.2.0.2 "experiment"

Le bloc "experiment" (option "sauv") permet de définir le nom générique de l'expérience pour la sauvegarde des résultats.

#### B.2.0.3 "no_forcings"

Le bloc "no_forcings" (option "fichierREFOBSnometeo") désigne le fichier de résultats de la simulation correspondante sans forçage météorologique. L'option "fichierREFOBS" désigne le fichier de référence en cas d'expérience jumelle.

#### B.2.0.4 "observations"

Le bloc "observations" (option "locnatobs") désigne, en cas d'expérience jumelle, les points d'observation et la variable hydrodynamique à extraire du fichier de référence.

L'option "dic_optionobs" permet de choisir la nature de l'expérience, jumelle ou sur données réelles ou encore utilisant des observations satellitaires.

L'option "dt" désigne la période de disponibilité des observations (en secondes) et "fobs" la fréquence d'utilisation des données d'observation pour l'expérience compte-tenu de leur disponibilité.

Les options "EVENT" et "EVENT_DATE" caractérisent l'événement simulé et permettent, dans le cas d'une réanalyse, de pointer sur la durée de corrélation du signal correspondant à chacune des chroniques temporelles pour l'événement considéré.

#### B.2.0.5 "ensemble"

L'option "Ne" du bloc "ensemble" désigne le nombre de membres de l'ensemble, tandis que l'option "dic_lxcontr" (resp. "xpert_full") désigne l'ensemble des variables contenues dans le vecteur de contrôle (resp. perturbées).

Pour les forçages temporels, les variables contrôlées (incrément ou coefficients modaux) sont choisis en renseignant les options "xcontr_opt_clmar" et "opt_dsm" pour chacune des frontières liquides, ainsi que le nombre de modes pour chacune d'entre elles dans l'option "n_var_csttes".

Les valeurs de la moyenne initiale et de l'écart-type de la distribution uniforme des coefficients de frottement sont renseignées dans les options "Ks_init" et "std_Ks".

L'ébauche du coefficient d'influence du vent est renseignée dans l'option "CDz_init".

En cas d'expérience jumelle et pour l'exploitation, les champs "Ks_target" et "CDz_target" renseignent les valeurs "vraies" respectivement pour les coefficients de frottement et le coefficient d'influence du vent vers lesquelles les variables de contrôle correspondantes sont censées converger.

#### B.2.0.6 "T2Dsettings"

Le bloc "T2Dsettings" permet de préciser les fichiers d'entrée de référence du modèle numérique dans les champs "studyFiles" (fichiers fortran "t2d.f", des paramètres "t2d.cas" et de géométrie "t2d.geo"), ainsi que le pas de temps du modèle "dt_t2d" et la fréquence de sauvegarde des résultats hydrodynamiques "fsave_t2d".

#### B.2.0.7 "Assim"

Le bloc "Assim" permet d'activer (true) ou non (false) l'option de redispersion de l'ensemble selon la séquence de Sobol' initiale ("optInflaGlob"), d'utiliser les formules d'inflation de l'ensemble et de calcul de la matrice de covariance d'erreur des mesures d'observation de Desroziers ("test_R_desroziers" et "optInflaGlob_desroziers").

Les options "optInflaCLMAR", "optInflaQGAR", "optInflaQDOR", "optInflaKs" et "optInflaCDz" permettent de procéder à la redispersion des variables de contrôle relatives à $CLMAR$, $QGAR$, $QDOR$, $Ks$ et $CDz$ de manière uniforme autour de $\overline{\boldsymbol{\gamma}^a}$ en utilisant les écart-types définis dans les champs "sigma_t2d_sm", "sigma_t2d_QFLUV_dQ" ou "sigma_t2d_QFLUV_alpha", "sigma_t2d_Ks" et "sigma_t2d_CDz" en utilisant un coefficient de relaxation égal à "lbda".

La durée totale de l'expérience, comprenant une fenêtre de free run d'initialisation, dont l'objectif est la dispersion des projections de l'ensemble dans l'espace des observations, $n_cycles$ cycles d'assimilation (comprenant chacun une phase d'analyse et une phase de run libre) et de run libre après l'expérience, ainsi que la durée entre le début de deux cycles successifs sont indiqués dans les champs "duree_DAW", "F", "L", "FrDAW", "FrAfter" et"S". La fréquence d'assimilation des données vaut "dtassim" (en secondes).

## B.3 Outils d'exploitation des résultats

Un programme d'exploitation des résultats a été développé dans le but de pouvoir visualiser et évaluer les apports de l'EnKF-$\gamma$-KLBC par rapport au run déterministe. Ce dernier consiste à calculer l'état du système à l'aide du modèle numérique en utilisant le membre moyen initial de l'ébauche sur toute la durée de l'expérience.

Ce programme d'exploitation, développé en langage python, permet de visualiser :

- l'évolution de la moyenne de l'ensemble pour les variables de contrôle au fur et à mesure des cycles d'assimilation ;
- les forçages temporels corrigés sur les frontières maritimes et fluviales du modèle ;
- les matrices de corrélation croisées et des matrices de covariance dans l'espace de contrôle et dans l'espace d'observation, ainsi que la matrice "Gain" sous forme de graphes temporels pour chaque variable de contrôle ou graphique de type colorbar à chaque pas de temps ou chaque variable de contrôle ;

- les lignes d'eau calculées en chaque station avec et sans assimilation de données, ainsi que les erreurs par rapport aux observations ;
- les diagrammes de rang des observations vis-à-vis de l'ensemble projeté dans l'espace des observations.
- le vecteur innovation

Il comprend également des outils de calcul des scores avec et sans assimilation de données en termes de RMSE et de Nash de pleine mer, ainsi que la moyenne des erreurs sur les pleines mers. Ces deux derniers scores nécessitent, en particulier, de rephaser les signaux, ce qui permet de dégager les déphasages introduits par l'EnKF-$\gamma$-KLBC en comparaison avec le run de contrôle.