



Contribution de la future mission altimétrique à large fauchée SWOT pour la modélisation hydrologique à grande échelle

Charlotte Emery

► To cite this version:

Charlotte Emery. Contribution de la future mission altimétrique à large fauchée SWOT pour la modélisation hydrologique à grande échelle. Sciences de la Terre. Université Paul Sabatier - Toulouse III, 2017. Français. NNT : 2017TOU30034 . tel-01523481v2

HAL Id: tel-01523481

<https://theses.hal.science/tel-01523481v2>

Submitted on 25 May 2018

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.



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :

Charlotte Emery

le Vendredi 3 Février 2017

Titre :

Contribution de la future mission altimétrique à large fauchée SWOT pour la
modélisation hydrologique à grande échelle

École doctorale et discipline ou spécialité :

ED SDU2E : Surfaces et interfaces continentales, Hydrologie

Unité de recherche :

UMR 5566 - LEGOS

Directeur/trice(s) de Thèse :

Sylvain BIANCAMARIA (CR-CNRS)

Aaron BOONE (CR-CNRS)

Jury :

Pierre-Olivier MALATERRE (ICPEF-IRSTEA) - Rapporteur

Catherine OTTLE (DR-CNRS) - Rapporteur

Stéphane CALMANT (DR-IRD) - Président du Jury

Agnès DUCHARNE (DR-CNRS) - Examinatrice

Paul BATES (Professor-University of Bristol) - Examineur

Marielle GOSSET (CR-CNRS) - Examinatrice

Remerciements

Voilà, on y est. Il est temps d'écrire mes remerciements qui concluent plus de 3 ans de travail acharné, passionné/passionnant, parfois déprimant, éreintant mais ô combien gratifiant. Ma thèse ne fut pas un long fleuve tranquille (elle était facile celle-là). Heureusement, de nombreuses personnes m'ont accompagnées, aidées, et m'ont permis de mener ma barque à bon port. Ces quelques lignes leurs sont dédiées.

Pour commencer, je tiens à remercier chaleureusement mes directeurs de thèse, Sylvain Biancamaria et Aaron Boone. Je vous remercie pour le soutien scientifique bien sur mais aussi pour votre disponibilité sans faille dès que j'avais la moindre question, le moindre problème, et pour votre réactivité de relecture dans le rush d'avant rendu de manuscrit. Je pense qu'il est aussi important de vous remercier pour m'avoir permise d'autant bouger lors de ma thèse. J'en profite pour remercier Sophie Ricci et Mélanie Rochoux du CERFACS qui, alors que j'étais leur étudiante en stage de M2, ont appuyé ma candidature pour cette thèse. Et je vous remercie aussi pour votre soutien et suivi qui ne s'est jamais arrêté après le stage.

Je remercie aussi le CNES et la région Occitanie. Ma thèse n'aurait pas pu être menée à bien sans leur cofinancement. J'en profite pour remercier Yves Morel, puis Alexandre Ganachaud et le LEGOS pour m'avoir accueilli et hébergé professionnellement, ainsi que l'équipe informatique du LEGOS pour tout leur soutien matériel. Je n'oublie pas que le LEGOS m'a aussi permis de terminer sereinement ma rédaction en me finançant un mois de plus, merci aussi pour ça. Dans la même idée, merci au centre de calcul hautes performances de CALMIP de m'avoir permis d'utiliser leur supercalculateur, et plus particulièrement merci à Nicolas Renon, pour les interminables échanges de mails et le temps qu'il m'a accordé lorsque je galérais à installer mes programmes. Un grand merci à l'équipe support de SURFEX du CNRM, Bertrand Decharme, Stéphanie Faroux, Patrick LeMoigne et aussi Vanessa Pedinotti, pour l'aide précieuse qu'il m'ont apporté lors de ma prise en main du modèle. Je remercie finalement l'ensemble de mon jury de thèse d'avoir accepté d'évaluer mon travail.

En revenant dans l'enceinte du LEGOS, tant de rencontres et de personnes à qui je souhaite m'adresser. Pour commencer, un grand merci à Martine, Nadine, Agathe et Brigitte, pour pleins de choses. Merci pour votre disponibilité continue et votre sourire permanent. Toujours là pour nous soutenir, nous glisser un petit mot gentil dans les moments difficiles aussi bien de la thèse qu'administratifs ! J'en profite pour saluer la gentillesse de Christine, à l'accueil de l'OMP, et celle d'Audrey, qui avait toujours deux minutes entre deux coups de balai pour un petit catch-up. J'ai eu la chance d'interagir avec énormément de personnes. Je pense bien évidemment aux personnes de l'équipe ECHOS, Stéphane Calmant, Denis Blumstein, Jean-François Crétaux, Philippe Maisongrande, mais aussi Marielle Gosset, avec qui j'ai échangé autour de l'hydrologie d'une façon ou d'une autre. À toutes ces personnes, merci d'avoir fait partie de mon aventure doctorale.

Quand j'ai commencé ma thèse, j'étais dans le fameux bureau des stagiaires, en attendant de m'installer dans mon bureau définitif. J'ai pu y rencontrer des personnes formidables qui m'ont assisté dans mon début de thèse et qui m'offre encore aujourd'hui des soirées mémorables. Je pense à Elsa, Robin et ses jeux en ligne (2048 et Geoguesser), Maud et ses cocktails, Vincent et son accent, Lucien et sa folie du rap et de Jul, Sylvain et ses expressions (c'est bijoux ...), Loïc et ses histoires sur Ikea, Florian, Virgile et Baptiste. Merci à vous pour votre bonne humeur et ces bons moments et ... We are Groot !

J'ai ensuite déménagé dans le couloir du SHOM. Je n'ai jamais eu/entendu autant d'histoires dans un couloir. Un couloir tel que tout le monde nous l'envie, certain veulent même nous le prendre ! Bref, je m'égare et je dédis ce paragraphe à mes différents cobureaux. Marine, Pierre, Pierre-André et maintenant Simon et Lise, merci d'avoir fait de ce bureau ce qu'il est. Un espace d'échanges, d'observations des lapins, de potins, de rigolade, de stockage de thé, de gateaux et de posters !

Je m'adresse maintenant à tous mes fellow co-thésards, co-CDD, co-postdoc : Laurence ("koi fé ?"), Vanessa ("j'ai plus de cheveux !"), Alice ("il y aura du punch à ta thèse Charlotte ?"), Cori (les "Éclai", ses colocs et les histoires de frigo en panne), Kévin ("j'ai un tas de copies à corriger, je te dis pas !"), Lise ("le monde est petit !"), Simon ("ne me parlez pas de mariage"), Emilie ("mon ordinateur marche pas !"), Adrien ("j'ai pas dormi de la nuit !"), Guillaume C (L'Isle-Jourdain, représente !), Fifi ("quand tu enlèves tout ce qui va pas, tout va bien !"), Edward, Amanda, Aude, Marine B, Marine R, Guillaume S et les petits nouveaux ... Vous êtes une part toute entière de ma thèse. Je ne vous remercierais jamais assez de tous les bons moments que j'ai passé en votre compagnie au LEGOS, dans un bar ou en soirée. L'apothéose de ces moments est l'intoxication alimentaire que je vous ai refilé lors de la soirée de fin de thèse !!! Je retiendrais aussi les facultés de Cori/Aude/Alice à faire se vider les grandes tables au restaurant pour qu'on mange toujours tous ensemble, les sujets de conversation toujours plus improbables (pourquoi les bananes ne peuvent pas traverser l'équateur, comment faire des calins à des poules ...), les achats compulsifs à la pause café et pleins d'autres choses que je n'ai pas la place d'écrire.

Un petit paragraphe tout spécial pour ma coloc', soeur de coeur ... Laurence ! On en aura vécu, une année folle entre Frascati ("mais quel *beep* !!!"), la Réunion (haut comme la montagne...fort comme le soleil), la colocation (il y a un truc qui crame !!!), les soirées improvisées. Merci pour tout ce que tu as fait pour moi : le soutien psychologique intense, le supportage de mes sauts d'humeur et de mes caprices de fin de thèse. Je ne te remercierai jamais assez.

Je tiens à dédier un paragraphe à mes petites basketteuses. Celles que j'ai depuis mes débuts en tant que coach : mes petites Louise, Nela et Ninon et vos coéquipières Olivia, Sirine, Sara, Roro, Marie, Romane, Carla, Maelys, Inès, Maïlys et Emma. J'aurais été votre coach pendant mes trois années de thèse. Vous coacher pendant ces années m'a permis de tenir sous la pression de la thèse. Ça n'a pas toujours été facile mais vous avez été un vrai bol d'air frais. Vous avez tout naturellement votre place dans ces remerciements.

Mon dernier paragraphe s'adresse tout naturellement à ma famille : papa, maman, mamie, Alice et Pierre et ceux qui sont partis en cours de thèse. Vous avez été d'un soutien inconditionnel tout au long de ma vie et bien sur de ma thèse. Et, vous avez géré le pot de thèse d'une main de maître. Je n'ai pas toujours été la plus facile à vivre alors merci pour tout !

*À Mamie Claire et Papi Richard, je vous dédis ce manuscrit.
Et non, je ne présenterai pas la météo à la télévision.*

Résumé

L'objectif scientifique de ce travail de thèse est d'améliorer l'estimation des flux d'eau à la surface des continents, à l'échelle saisonnière et interannuelle (de quelques années à décennale). En particulier, il s'agit d'étudier l'apport de données satellites, notamment de la future mission SWOT (Surface Water and Ocean Topography, lancement prévu en 2021), pour l'étude de la partie continentale du cycle de l'eau à l'échelle globale, à l'aide du modèle global de surfaces continentales ISBA-TRIP (Interactions Sol-Biosphère-Atmosphère/Total Runoff and Integrating Pathways).

Dans ce travail de thèse, j'explore le potentiel des données d'altimétrie satellitaire, pour corriger certains paramètres du modèle de routage de rivière TRIP et aussi pour corriger ses variables d'état. Pour ce faire, une plateforme d'assimilation de données virtuelles SWOT, mais aussi de données d'altimètres nadirs actuels a été mise en place. Mais avant l'assimilation de ces données de télédétection, il a été nécessaire de faire une analyse de sensibilité du modèle TRIP à ses paramètres, pour déterminer quels paramètres ont le plus d'influence sur les observables SWOT et qui donc pourront être corrigés.

L'analyse de sensibilité (ANOVA) a alors été menée sur les principaux paramètres de TRIP. L'analyse de sensibilité a été menée sur le bassin de L'Amazonie et les résultats ont été publiés. Les résultats ont montré que les hauteurs d'eau simulées sont sensibles aux paramètres géomorphologiques locaux exclusivement tandis que les débits simulés sont sensibles à l'ensemble des paramètres amont (selon le réseau de routage TRIP) et surtout au paramètre lié au temps de résidence des eaux souterraines. Enfin, les anomalies de hauteurs présentent des sensibilités similaires aux hauteurs d'eau mais avec des variations temporelles plus marquées. Ces résultats nous ont permis de faire les choix algorithmiques dans le cadre de l'assimilation de données.

Ensuite, je me suis concentrée sur le développement de la maquette d'assimilation de données consistant en un Filtre de Kalman d'Ensemble (EnKF) et permet de faire soit de l'estimation de paramètres, soit de l'estimation d'état. La maquette en « estimation de paramètres » est testée et validée par une série d'expériences jumelles. On a assimilé des pseudo-observations de hauteurs et d'anomalies d'eau le long des traces du satellite SWOT, afin de corriger les coefficients de Manning du lit de la rivière, avec possibilité d'étendre à d'autres paramètres. Les premiers résultats montrent que la maquette est capable de retrouver la bonne distribution des coefficients de Manning en assimilant les hauteurs d'eau et les anomalies. Pour l'estimation d'état, on réalise des étapes d'assimilation journalières pour corriger le stock d'eau initial (condition initiale du modèle), en assimilant des débits estimés à partir de séries altimétriques de côtes d'eau ENVISAT. A partir de courbe de tarage hauteurs d'eau-débits calibrées sur le bassin de l'Amazonie avec le modèle hydrologique MGB-IPH, les cotes d'eau ont été transformées en « débits altimétriques » que l'on assimile alors dans la maquette. Ces expériences d'estimation d'état nous permettent de sortir du cadre idéalisé des expériences jumelles en assimilant des données réelles, mais nous permet aussi de tester l'apport d'un premier jeu de données de débits provenant de mesures satellites, qui préfigure le futur produit de débit SWOT. Les résultats montrent que les erreurs sur le débit sont globalement améliorées : le run libre donne un RMSE de $2,79 \times 10^3 \text{ m}^3/\text{s}$ (73,6 %) par rapport aux données in situ disponible sur le bassin et le run corrigé un RMSE de $1,98 \times 10^3 \text{ m}^3/\text{s}$ (53,9 %).

Abstract

Scientific objective of this PhD work is to improve water fluxes estimation on the continental surfaces, at interannual and interseasonal scale (from few years to decennial time period). More specifically, it studies contribution of remotely-sensed measurements to improve hydrology model. Notably, this work focuses on the incoming SWOT mission (Surface Water and Ocean Topography, launch scheduled for 2021) for the study of the continental water cycle at global scale, and using the land surface model ISBA-TRIP.

In this PhD work, I explore the potential of satellite data to correct both input parameters of the river routing scheme TRIP and its state variables. To do so, a data assimilation platform has been set to assimilate SWOT virtual observation as well as discharge estimated from real nadir altimetry data. Beforehand, it was necessary to do a sensibility analysis of TRIP model to its parameters. The aim of such study was to highlight what are the most impacting parameters on SWOT-observed variables and therefore select the ones to correct via data assimilation.

The sensibility analysis (ANOVA) has been led on TRIP main parameters. The study has been done over the Amazon basin. The results showed that the simulated water levels are sensitive to local geomorphological parameters exclusively. On the other hand, the simulated discharges are sensitive to upstream parameters (according to the TRIP river routing network) and more particularly to the groundwater time constant. Finally, water anomalies present sensitivities similar to those of the water levels but with more pronounced temporal variations. These results also lead me to do some choices in the implementation of the assimilation scheme and have been published.

Therefore, in the second part of my PhD, I focused on developing a data assimilation platform which consists in an Ensemble Kalman Filter (EnKF). It could either correct the model input parameters or directly its state. A series of twin experiments is used to test and validate the parameter estimation module of the platform. SWOT virtual-observations of water heights and anomalies along SWOT tracks are assimilated to correct the river manning coefficient, with the possibility to easily extend to other parameters. First results show that the platform is able to recover the "true" Manning distribution assimilating SWOT-like water heights and anomalies. In the state estimation mode, daily assimilation cycles are realized to correct TRIP river water storage initial state by assimilating ENVISAT-based discharge. Those observations are derived from ENVISAT water elevation measures, using rating curves from the MGB-IPH hydrological model (calibrated over the Amazon using in situ gages discharge). Using such kind of observation allows going beyond idealized twin experiments and also to test contribution of a remotely-sensed discharge product, which could prefigure the SWOT discharge product. The results show that discharge after assimilation are globally improved : the root-mean-square error between the analysis discharge ensemble mean and in situ discharges is reduced by 28 %, compared to the root-mean-square error between the free run and in situ discharges (RMSE are respectively equal to $2.79 \times 10^3 \text{ m}^3/\text{s}$ and $1.98 \times 10^3 \text{ m}^3/\text{s}$)

Table des matières

INTRODUCTION	1
1 Modélisation de la dynamique spatio-temporelle des eaux de surface et observation du cycle de l'eau	9
Introduction	10
1.1 Le cycle de l'eau et sa composante continentale	10
1.1.1 L'évapotranspiration	11
1.1.2 Les précipitations	12
1.1.3 Le ruissellement et l'infiltration	12
1.2 Modéliser le cycle de l'eau	14
1.2.1 Développements des modèles de surface continentale	14
1.2.2 Développements des schémas de routage	18
1.2.3 Incertitudes, erreurs et limitations de la modélisation	19
1.3 Observer le cycle de l'eau	20
1.3.1 Vue d'ensemble des données in situ : atouts et limitations	20
1.3.2 Observer les réservoirs d'eau non fluviaux	22
1.3.2.1 Observer les champs de précipitations	22
1.3.2.2 Mesurer l'humidité du sol	23
1.3.2.3 Évaluer les réserves d'eau souterraine	25
1.3.2.4 Suivre les glaciers et le manteau neigeux	25
1.3.2.5 Observer la couverture végétale	27
1.3.3 Observer et suivre les fleuves et les rivières	27
1.3.3.1 Mesures in situ	27
1.3.3.2 Mesures des hauteurs par profils altimétriques	28
1.3.3.3 Mesures des hauteurs par imagerie Radar à Synthèse d'Ouverture (RSO) et Modèles Numériques de Terrain	31
1.3.3.4 Mesures par altimétrie LIDAR	34
1.3.3.5 Mesures des débits via la télédétection	35
1.3.4 Incertitudes, erreurs et limitations de l'altimétrie nadir	35
1.3.5 La mission Surface Water and Ocean Topography (SWOT)	36
1.3.5.1 Genèse du projet et questions scientifiques	36
1.3.5.2 La technologie SWOT	39
1.3.5.3 Produits et erreurs SWOT en hydrologie	42
1.3.5.4 Déroulement de la mission	43
1.3.5.5 Les attentes de SWOT en hydrologie	46
1.4 Combiner modèle et observation avec l'assimilation de données	48
1.4.1 L'assimilation de données comme un problème inverse probabiliste	49

1.4.2	Formalisme : variables d'assimilation de données	50
1.4.2.1	Modèle de propagation du système physique	51
1.4.2.2	Modèle stochastique et vecteur de contrôle	51
1.4.2.3	Les observations et l'opérateur d'observation	53
1.4.2.4	Le vecteur d'innovation	54
1.4.3	Traitement stochastique des erreurs	55
1.4.3.1	Matrices des covariance d'erreur	55
1.4.3.2	Formulation du problème d'estimation	56
1.4.4	Formule de Bayes et formulation générale d'un problème d'assimilation de données	57
1.4.4.1	Formule de Bayes	57
1.4.4.2	Formulation de l'assimilation de données comme problème d'optimisation	58
1.4.5	Le Filtre de Kalman et ses extensions	59
1.4.5.1	Le Filtre de Kalman (KF) : une solution analytique au problème d'optimisation bayésien	59
1.4.5.2	Extensions non-linéaires du filtre de Kalman	61
1.4.5.3	Le Filtre de Kalman d'Ensemble (EnKF)	63
1.4.6	Filtres non-gaussiens et non-linéaires à particules	66
1.4.7	Les méthodes variationnelles	69
1.4.7.1	Formulation du problème de minimisation	69
1.4.7.2	Méthodes séquentielles 3D-VAR et 3D-VAR incrémentale	69
1.5	Application de l'assimilation de données en hydrologie de surface	71
	Conclusions	72
2	Description et étude de sensibilité du modèle hydrologique grande échelle ISBA-TRIP	77
	Introduction	77
2.1	Le modèle de surface continentale ISBA-TRIP	79
2.1.1	Historique et différentes versions d'ISBA	79
2.1.2	Description du modèle ISBA-3L	80
2.1.2.1	Traitement de l'eau liquide	81
2.1.2.2	Traitement des flux de surface	86
2.1.2.3	Traitement de la neige	89
2.1.2.4	Traitement du sol gelé	89
2.1.3	Version originale et développements successifs du modèle de routage TRIP	90
2.1.4	Paramètres spécifiques de TRIP	93
2.2	Le domaine d'étude : le bassin de l'Amazonie	95
2.3	Analyse de Sensibilité (AS) de la composante "routage de rivière" du modèle ISBA-TRIP	96
2.3.1	Choix de la méthode d'Analyse de Sensibilité	97
2.3.2	Formulation de la méthode ANOVA	98
2.3.3	Application de l'ANOVA au modèle TRIP	101
2.3.3.1	Configuration générale	101
2.3.3.2	Choix des paramètres	102
2.3.4	Résultats	105
2.3.4.1	Résultats sur les hauteurs d'eau	106
2.3.4.2	Résultats sur les anomalies	112
2.3.4.3	Résultats sur les débits	117
	Conclusions et perspectives	123

3	Correction des paramètres de TRIP via l'assimilation de pseudo-observations SWOT	127
	Introduction	127
3.1	Description des variables observées et des variables à corriger	130
3.1.1	Description des pseudo-observations SWOT	131
3.1.2	Description des variables de contrôle	133
3.2	Choix de la méthode d'assimilation	135
3.2.1	Définition de la fenêtre d'assimilation et conséquences	135
3.2.2	Le Filtre de Kalman d'Ensemble Asynchrone en estimation de paramètres	136
3.2.3	Configuration des simulations d'estimation de paramètres	140
3.2.3.1	Description générale des expériences d'assimilation	140
3.2.3.2	Expériences avec inondations	143
3.2.3.3	Expériences assimilant les anomalies de hauteurs d'eau	143
3.3	Résultats pour l'assimilation des profondeurs d'eau	146
3.3.1	Résultats des expériences EP1F et EP1NF	146
3.3.2	Résultats de l'expérience EP2F	155
3.4	Résultats pour l'assimilation des anomalies de hauteurs d'eau	159
3.4.1	Résultats des expériences EP3-1 et EP3-2	159
3.4.2	Résultats des expériences EP4F et EP4NF	166
	Conclusions et perspectives	175
4	Préparation à l'assimilation du produit de débits SWOT : estimation d'état avec des débits altimétriques	181
	Introduction	182
4.1	Comparaison du modèle ISBA-TRIP aux débits observés	184
4.1.1	Débits in situ	184
4.1.2	Débits altimétriques	184
4.1.3	Performances d'ISBA-TRIP sur l'Amazonie	188
4.2	Méthode d'assimilation des débits ENVISAT	200
4.2.1	Définition des variables d'observations	200
4.2.2	Stratégies d'assimilation de données et choix du vecteur de contrôle	201
4.2.2.1	Option 1 : Assimilation des débits ENVISAT pour corriger... les débits ISBA-TRIP	201
4.2.2.2	Option 2 : Assimilation des débits ENVISAT pour corriger... les stocks finaux ISBA-TRIP	202
4.2.2.3	Option 3 : Assimilation des débits ENVISAT pour corriger... les stock initiaux ISBA-TRIP	203
4.2.3	Génération de l'ensemble	203
4.2.3.1	Perturbations de la condition initiale d'ISBA-TRIP	204
4.2.3.2	Perturbations des précipitations	206
4.2.3.3	Localisation des matrices de covariance d'erreurs	209
4.2.4	Présentation de la stratégie d'assimilation et des expériences d'assimilation	211
4.3	Résultats de l'expérience EE1-direct	212
4.3.1	Résultats généraux	212
4.3.2	Comportement pour certaines stations tests	215
4.3.3	Le cas de la confluence Negro-Branco	219
4.3.4	Bilan sur l'expérience EE1-direct	224
4.4	Performances de l'EnKF avec localisation	225
4.4.1	Expérience préliminaire : EE1-diag	225
4.4.1.1	Configuration de l'expérience	225

4.4.1.2	Résultats généraux	225
4.4.1.3	Comportement aux stations tests	228
4.4.2	Expérience-test du masque de localisation : EE1-local	231
4.4.2.1	Configuration de l'expérience	231
4.4.2.2	Résultats généraux	231
4.4.2.3	Comportement aux stations tests	233
4.4.3	Conclusions sur la localisation	236
4.5	Sensibilité à l'opérateur d'observations	238
4.5.1	Évaluation des incertitudes de l'opérateur Z	238
4.5.2	Comparaison des expériences EE1-local, EE2-local et EE3-local	239
	Intercomparaison des simulations, conclusion et perspectives	244
CONCLUSIONS ET PERSPECTIVES		249
A Description de la maquette d'assimilation de données réalisée avec le coupleur Open-PALM		255
A.1	Présentation du coupleur OpenPALM	255
A.2	La fonctionnalité Parasol	256
A.3	Description de la maquette d'AD	258
B Résultats et figures supplémentaires pour les expériences d'estimation d'état		260
B.1	Statistiques globales	260
B.1.1	EE2-local	260
B.1.2	EE3-local	260
B.2	Débites aux stations in situ tests	263
B.3	Erreur de formulation sur l'opérateur Z_k	266
C Article : "Temporal Variance-based Sensitivity Analysis of the River Routing Component of the large scale hydrological model ISBA-TRIP : Application on the Amazon Basin"		268
BIBLIOGRAPHIE		314

Introduction

L'eau est une ressource essentielle à la vie, qui recouvre plus de 70% de la surface de la Terre. Elle est en continuels mouvement, grâce à la circulation naturelle de l'eau sur Terre appelée "cycle de l'eau" ou "cycle hydrologique" (Oki and Kanae, 2006; Ducharne and Laval, 2015). Le cycle naturel de l'eau est piloté par le rayonnement solaire et atmosphérique incident, qui est absorbé par les surfaces continentales et les océans (Figure 1). Cette énergie chauffe les différentes surfaces et entraîne l'évaporation des océans et des sols, des lacs et des rivières. Elle permet aussi la fonte et la sublimation des neiges et des glaciers, ainsi que la transpiration des plantes. L'eau évaporée est transportée sous forme de vapeur par les courants atmosphériques, mais peut aussi condenser sous forme de nuages et retourner à la surface sous forme de précipitations liquides (pluies) ou solides (neiges). Ces précipitations vont alors alimenter les océans et les réservoirs continentaux, comme la cryosphère (calottes, glaciers et manteau neigeux), les sols en s'infiltrant, les lacs, les rivières/fleuves et le couvert végétal. La saturation de certains de ces réservoirs peut alors entraîner des phénomènes extrêmes comme des crues. L'eau infiltrée peut rejoindre les cours d'eau après un certain temps de résidence dans les couches superficielles du sol ou percoler vers des réservoirs souterrains plus profonds. Ces aquifères peuvent être soit complètement confinés, soit se déverser directement dans les océans sans remonter en surface. Similairement, l'eau s'écoulant en surface et dans les aquifères connectés alimente les lacs et les cours d'eau, ceux-ci pouvant se déverser dans les océans (bassin exoréique) ou non (bassin endoréique).

La vitesse de renouvellement des stocks d'eau est déterminée par le système climatique (Oki and Kanae, 2006) c'est-à-dire l'ensemble des interactions entre l'atmosphère et les surfaces continentales (composées des eaux de surfaces, de la cryosphère, de la lithosphère et de la biosphère). La circulation atmosphérique et le cycle de l'eau assurent alors le transfert de cette énergie calorifique des zones intertropicales vers les hautes latitudes. Ce cycle naturel est alors modulé par la variabilité naturelle du climat qui entraîne des changements constants dans l'état des ressources en eaux. Simultanément, depuis le milieu du XXème siècle, les activités humaines modifient l'environnement et puisent dans les ressources hydriques. Les activités humaines impactent ainsi directement et indirectement le cycle hydrologique.

Plus particulièrement, la partie continentale du cycle de l'eau est une interface très dynamique avec l'atmosphère d'un côté, et les océans de l'autre. L'eau douce y est continuellement et rapidement renouvelée. En effet, (Ducharne and Laval, 2015) donne un temps de résidence moyen de quelques jours pour les cours d'eau (16 jours), quelques années pour les milieux humides et dans le sol hors aquifères (1 à 5 ans) et de plus d'une décennie pour les lacs (17 ans). Les principaux flux d'eau douce vers et hors des surfaces continentales sont les précipitations (apports) les flux d'évapotranspiration (pertes) pour l'interface avec l'atmosphère et les débits fluviaux (pertes) pour l'interface avec les océans. Tous ces flux ont une très grande variabilité spatio-temporelle (Milliman and Farnsworth, 2011). Les taux de précipitations varient énormément d'une région à l'autre et d'une année sur l'autre.

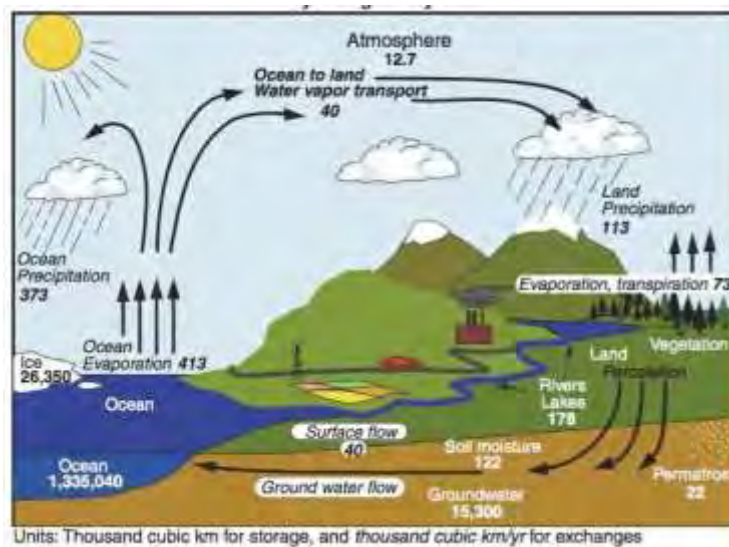


Figure 1 - Le cycle de l'eau (Trenberth and Asrar, 2014).

Les plus gros taux de précipitations ont alors lieu au-dessus des régions climatiques équatoriales et sur les zones côtières. Quant aux flux d'évapotranspiration, ils dépendent à la fois de l'humidité et de la température de l'air ainsi que la nature de la végétation. Les flux d'évapotranspiration tendent ainsi à décroître avec l'augmentation des latitudes (Milliman and Farnsworth, 2011), qui entraîne des baisses de température et des modifications de la végétation. Enfin, les débits fluviaux envoient une importante masse d'eau douce vers les océans ($\sim 36000 \text{ km}^3$). La Figure 2 détaille la répartition spatiale des pertes d'eau douce vers les océans via les débits fluviaux : les rivières drainant le nord de l'Amérique du Sud et l'Asie du Sud/Océanie compte pour la moitié de ces apports. Plus particulièrement, le débit à l'embouchure de l'Amazonie ($6300 \text{ km}^3/\text{an}$) dépasse les débits combinés des huit grands fleuves mondiaux qui le suivent (par ordre décroissant de débit : le Congo, l'Orénoque, le Chang Jiang, le Brahmapoutre, le Yenisei, le Mekong, la Lena et le Gange). (Oki and Kanae, 2006) ont estimé à 2000 km^3 la masse totale d'eau douce disponible dans toutes les rivières du monde alors que la masse d'eau douce totale puisée chaque année est de 3800 km^3 . Visiblement, la seule mesure des stocks d'eau n'est pas adaptée pour mesurer les disponibilités réelles de l'eau douce car ces stocks sont très dynamiques. Par conséquent, il apparaît plus judicieux d'évaluer les flux d'eau douce pour vraiment évaluer l'eau disponible.

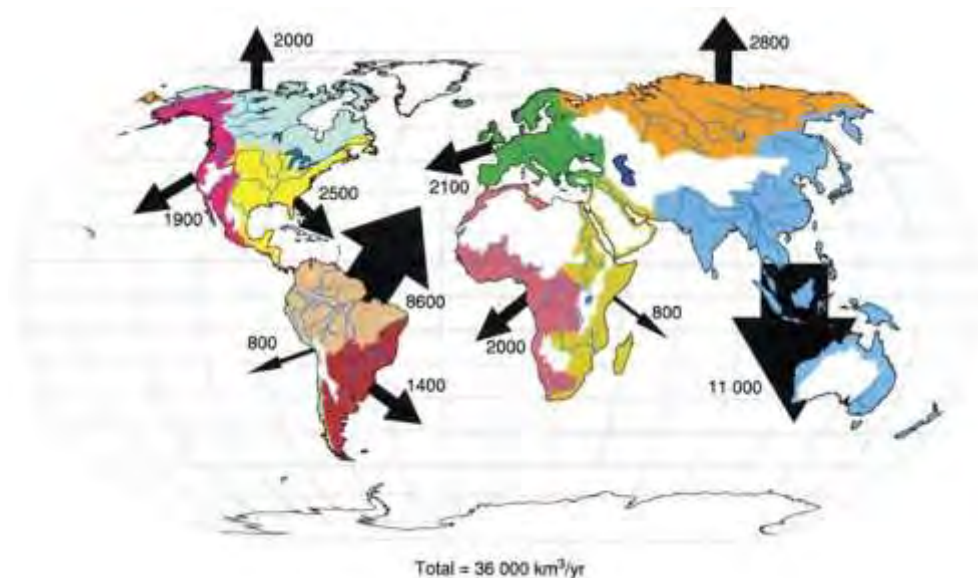


Figure 2 – Débit fluvial vers les océans à l'échelle du globe. Les nombres correspondent aux débits annuels moyens (km^3/an) et la taille des flèches est proportionnelle à ces nombres (Milliman and Farnsworth, 2011).

L'eau continentale est au centre du cycle de l'eau, à l'interface entre l'atmosphère et les océans. Parallèlement, le cycle de l'eau joue un rôle important dans la régulation thermique de la Terre. Il est donc primordial d'étudier la dynamique des eaux continentales, pour comprendre le système climatique terrestre. De plus, l'eau continentale a la particularité d'être de l'eau douce, qui est essentielle pour la plupart des formes de vies sur Terre. Toutefois, cette eau douce ne représente que 2,5% de la masse totale de l'eau présente sur Terre (Oki and Kanae, 2006) et elle est majoritairement stockée dans la cryosphère, ainsi que dans les réservoirs souterrains profonds. La Figure 3 résume la répartition de la masse totale de l'eau sur la planète. L'eau « utile », utilisée par les écosystèmes continentaux et pour les activités humaines, plus facilement accessible, n'est présente qu'en très faible quantité : seulement 0,4% de l'eau douce est présente dans l'atmosphère et les eaux de surfaces. Plus précisément sur les continents, 12,2% de cette eau douce de surface se trouve sous forme d'humidité des sols (aussi appelée eau « verte »), tandis que 67,4% sont stockées dans les lacs d'eau douce et 1,6% s'écoule dans les fleuves et les rivières (cette eau dans les lacs et rivières est appelée l'eau « bleue ») (WWDR2, 2006).

Or cette très faible quantité d'eau douce disponible est primordiale à la survie des sociétés humaines. D'une part, l'accès à l'eau potable est indispensable pour les besoins domestiques. D'autre part, l'eau est aussi primordiale pour de nombreuses activités économiques comme l'agriculture (irrigation et élevage), la production d'énergie, l'industrie ou encore les transports. Cependant, cette ressource n'est pas répartie de manière homogène d'un pays à l'autre, comme le montre la Figure 4, qui présente la quantité totale d'eau disponible par personne dans chaque pays du monde. Certains pays disposent de ressources abondantes, tandis que d'autres en manquent cruellement, notamment dans la bande équatoriale. Cet état de fait aura donc des impacts sociétaux et économiques importants et peut influencer sur les politiques intérieures et extérieures des pays. De surcroît, les besoins domestiques en eau peuvent s'opposer aux besoins industriels et mener à des problèmes de gestion

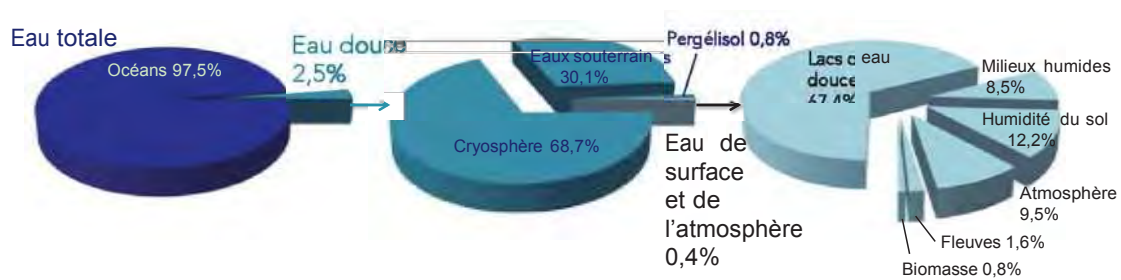


Figure 3 – Répartition de l'eau sur Terre (WWDR2, 2006; Lion, 2012)

de l'eau où l'ensemble des besoins ne peuvent pas être satisfaits. C'est le cas par exemple de l'Inde, du Nord de la Chine, de l'Ouest des États-Unis et aussi de l'Australie (Oki and Kanae, 2006). En outre, les flux d'eau ne sont pas confinés à l'intérieur des frontières politiques et la gestion de ces flux par les pays amonts peut accentuer les difficultés d'accès à l'eau et devenir un enjeu politique majeur. On estime à 148 le nombre d'États se trouvant dans des bassins hydrologiques transfrontaliers (WWDR4, 2012) et à 273 les aquifères transfrontaliers (Puri and Aureli, 2009). Selon la FAO (*Food and Agriculture Organization of the United Nations*), plus de 3600 traités existent à ce jour entre pays se trouvant dans des bassins transfrontaliers. Ces traités permettent de garantir des débits minimums ou de répartir l'utilisation de ces eaux entre les différents pays ripariens, voire d'aider à une gestion plus ou moins commune de la ressource, par la mise en place de comité de bassin interétatique. Toutefois, la demande grandissante en eau pour les populations, l'industrie, l'énergie et l'urbanisation va certainement mettre une pression supplémentaire sur ces accords transfrontaliers, ce qui pourrait conduire à leur redéfinition, ou bien même à les remettre en cause. La gestion et la connaissance des ressources en eaux sont alors nécessaires pour apporter des informations quantitatives à chaque pays et aider à définir des utilisations adaptées aux besoins et aux ressources.

Dans ce contexte, il est clair que l'observation et la compréhension de la distribution des ressources en eau et de leurs variations sont essentielles, pour répondre à des enjeux à la fois scientifiques, mais aussi sociétaux et économiques. C'est pourquoi un grand nombre de programmes internationaux ont été mis en place pour observer et étudier globalement ces ressources. Parmi ces programmes, on trouve, par exemple, l'IGOS (*Integrated Global Observing Strategy*) dont la mission est de mettre en relation les principales instances mondiales d'observations environnementales. L'IGOS a permis la mise en place du GEOSS (Global Earth Observation System of Systems) qui regroupe un ensemble de systèmes d'observations de la Terre à l'échelle globale. Ses missions sont d'améliorer la surveillance de l'état de la planète, de comprendre son fonctionnement et d'améliorer les prédictions. In fine, le GEOSS cherche à fournir aux décideurs politiques l'ensemble le plus complet possible d'information sur l'état de la planète. Le GEOSS a identifié un certain nombre de variables, appelées Variables de l'Eau Essentielles (VEE ou *Essential Water Variables*, EWV en anglais), qui sont critiques pour la gestion des ressources en eau et qui servent d'indicateurs de changement environnemental. Parmi ces variables, se trouve le débit des rivières/fleuves et les stocks d'eau de surface. Pour observer ces variables, Le GEOSS insiste sur la complémentarité des mesures in situ et spatiales et le besoin de les harmoniser au sein d'un système d'observation et d'information exhaustif (GEOSS, 2014). On peut aussi noter que le GCOS (*Global Climate Observing System*), qui est un projet commun entre l'OMM (*Organisation Météorologique Mondiale*), l'IOC (*Intergovernmental Oceanographic Commission*), l'UNESCO (*United Nations Educational Scientific and Cultural Organization*), le PNUE (Programme des Nations Unies pour l'Environnement) et l'ICSU (*International Council for Science*) et qui a pour objectif de fournir des informations exhaustives sur l'ensemble du système climatique, a défini un ensemble de

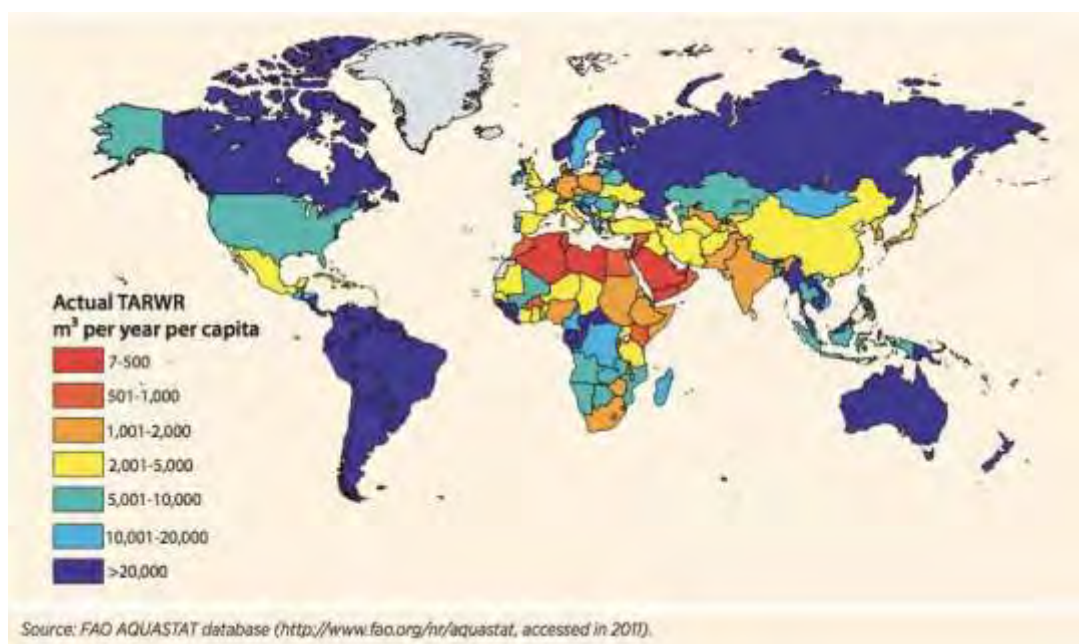


Figure 4 – Ressource totale d’eau douce renouvelable par habitant (données de 2009 ; WWDR4, 2012)

variables climatiques essentielles (*Essential Climate Variables* ou ECV) qu’il est nécessaire d’observer et mesurer pour étudier le changement climatique (Bojinski et al., 2014). Là aussi, parmi ces ECV, on trouve les flux de précipitations et d’évapotranspiration et surtout le débit fluvial et le niveau des lacs et réservoirs (Bojinski et al., 2014). Ces variables confirment la nécessité d’étudier les variations spatio-temporelles de l’eau douce à l’échelle spatiale globale et à l’échelle saisonnière et interannuelle.

Cependant, l’observation de la dynamique spatio-temporelle des flux d’eau de surface à l’échelle globale est encore sujette à d’importantes incertitudes. Relativement peu d’études (Oki and Kanae, 2006; Trenberth et al., 2007, 2011) ont essayé de produire un aperçu global et quantitatif du cycle de l’eau. Notre connaissance quantitative des différentes composantes du cycle de l’eau et surtout de leur variabilité est assez limitée à cause d’un manque de données fiables sur, par exemple, l’évapotranspiration ou le ruissellement. Il est encore difficile de fermer le bilan du cycle de l’eau à l’échelle des grands bassins hydrologiques en utilisant seulement les données disponibles (Sanoo et al., 2011; Vinukollu et al., 2011; Trenberth and Asrar, 2014). Ces études mettent en évidence les importantes limitations des méthodes actuelles d’observation et de modélisation. Les observations satellitaires fournissent des données généralement avec une bonne couverture spatiale, mais sont parfois discontinues dans le temps. Surtout, les mesures satellites brutes nécessitent un traitement préalable pour être retranscrites sous la forme d’une variable géophysique. Inévitablement, ces données présentent certaines limitations car elles doivent être validées à partir d’autres mesures indépendantes, telles que des mesures in situ. Parallèlement, de nombreux produits « hydrides », issus de la combinaison de données de plusieurs instruments et/ou avec des modèles, sont aussi développés. Une solution intéressante pour mieux étudier le cycle de l’eau continental est d’assembler les sorties d’un modèle hydrologique avec des observations (satellitaires et/ou in situ) afin de profiter des avantages de chacun. Les techniques permettant de réaliser cette combinaison incluent la calibration de modèle et

l'assimilation de données. Malheureusement, même si le nombre de produits satellites ont augmenté très rapidement ces dernières années, peu de missions sont dédiées spécifiquement à l'observation de l'eau bleue. Pour pallier ce manque et compléter les observations actuelles, la NASA (*National Aeronautics and Space Administration*), le CNES (*Centre National d'Études Spatiales*), l'ASC/CSA (*Agence Spatiale Canadienne/Canadian Space Agency*) et l'UKSA (*United Kingdom Space Agency*) ont décidé de développer conjointement la future mission altimétrique large fauchée SWOT (*Surface Water and Ocean Topography*). Cette mission devrait être lancée en 2021 et a pour objectif l'observation globale de l'eau bleue sur les continents (et les océans).

Ma thèse s'inscrit dans ce contexte et aborde la question scientifique suivante : quelle est la variabilité spatio-temporelle des flux d'eau à la surface des continents, à l'échelle saisonnière/interannuelle et à l'échelle des grands bassins versants ? En particulier, il s'agit d'avoir une meilleure connaissance de certaines VEE définies par le GEOSS (stock d'eau de surface et surtout débit fluvial) et d'estimer l'apport de la télédétection pour l'amélioration de la modélisation hydrologique à grande échelle. Pour ce faire, on utilise conjointement un modèle hydrologique grande échelle et des données spatiales d'observation du niveau d'eau dans les fleuves, dont on dérive une estimation du débit fluvial. Notamment, on se place dans le cadre de la mission SWOT qui apportera de nouvelles données sur les eaux de surfaces, en particulier les fleuves et les rivières. Pour commencer, dans le Chapitre 1 de cette thèse, je présenterai le cycle de l'eau ainsi que les différentes méthodes existantes pour l'observer et le modéliser. Je finirai ce chapitre en présentant l'ensemble des algorithmes permettant de combiner les observations avec les modèles, appelées méthodes d'assimilation de données. Ensuite, au Chapitre 2, je décrirai le modèle hydrologique grande échelle ISBA-TRIP utilisé dans la cadre de cette thèse, complété de son étude de sensibilité. Le Chapitre 3 sera consacré à l'assimilation de pseudo-observations de type SWOT en expériences jumelles pour corriger un des paramètres du modèle présenté au chapitre précédent (celui qui a le plus grand impact sur les niveaux d'eau). Enfin, au Chapitre 4, je présenterai l'assimilation de données de débits dérivées notamment de mesures satellitaires réelles (et en préfiguration du produit de débits SWOT) pour corriger l'état du modèle hydrologique.

Chapitre 1

Modélisation de la dynamique spatio-temporelle des eaux de surface et observation du cycle de l'eau

Sommaire

Introduction	10
1.1 Le cycle de l'eau et sa composante continentale	10
1.1.1 L'évapotranspiration	11
1.1.2 Les précipitations	12
1.1.3 Le ruissellement et l'infiltration	12
1.2 Modéliser le cycle de l'eau	14
1.2.1 Développements des modèles de surface continentale	14
1.2.2 Développements des schémas de routage	18
1.2.3 Incertitudes, erreurs et limitations de la modélisation	19
1.3 Observer le cycle de l'eau	20
1.3.1 Vue d'ensemble des données in situ : atouts et limitations	20
1.3.2 Observer les réservoirs d'eau non fluviaux	22
1.3.3 Observer et suivre les fleuves et les rivières	27
1.3.4 Incertitudes, erreurs et limitations de l'altimétrie nadir	35
1.3.5 La mission Surface Water and Ocean Topography (SWOT)	36
1.4 Combiner modèle et observation avec l'assimilation de données	48
1.4.1 L'assimilation de données comme un problème inverse probabiliste	49
1.4.2 Formalisme : variables d'assimilation de données	50
1.4.3 Traitement stochastique des erreurs	55
1.4.4 Formule de Bayes et formulation générale d'un problème d'assimilation de données	57
1.4.5 Le Filtre de Kalman et ses extensions	59
1.4.6 Filtres non-gaussiens et non-linéaires à particules	66
1.4.7 Les méthodes variationnelles	69

1.5 Application de l'assimilation de données en hydrologie de surface	71
Conclusions	72

Introduction

Dans ce premier chapitre, on s'intéresse à l'étude du cycle de l'eau et, plus particulièrement, à la partie continentale de ce cycle. De manière générale dans cette thèse, on s'est exclusivement concentré sur l'étude de l'écoulement des rivières et des fleuves à grande échelle, comme indiqué en introduction. Cependant, pour bien comprendre les causes de ces écoulements et leurs interactions avec les différentes composantes du cycle de l'eau, il est nécessaire de présenter l'ensemble des réservoirs qui constituent ce cycle (du moins pour sa partie continentale) et leurs échanges. Aussi, dans les différentes Sections de ce chapitre, qui décrivent la partie continentale du cycle de l'eau, sa modélisation et son observation, on commencera par une présentation générale, puis on détaillera l'aspect fluvial.

Après une description des phénomènes physiques impliqués dans la partie continentale du cycle de l'eau en Section 1.1, on décrira les différents types de modèles globaux disponibles pour la simuler en Section 1.2 avec les différents types d'erreurs issues de la modélisation et de la résolution numérique. Ensuite, on présentera les différentes techniques de mesure et d'observation du cycle de l'eau en Section 1.3 ainsi que les incertitudes associées à ces techniques. Et enfin, on présentera en Section 1.4 les méthodes permettant de combiner ces deux sources d'informations : ce sont les algorithmes d'assimilation de données. Comme ces algorithmes se basent sur les incertitudes de modélisation et d'observation, ces dernières sont particulièrement discutées dans les sections suivantes.

1.1 Le cycle de l'eau et sa composante continentale

Commençons par décrire la partie continentale du cycle de l'eau. Le cycle hydrologique ou cycle de l'eau est l'ensemble des processus intervenant dans la circulation perpétuelle de l'eau sur et sous la surface de la Terre (Fig. 1.1). Comme l'illustre cette figure, l'eau des océans s'évapore, puis se condense dans les nuages, avant de précipiter sous forme de pluie ou de neige sur les surfaces continentales ou les océans. Par la suite l'eau provenant des pluies ou de la fonte des glaciers et du manteau neigeux peut ruisseler et rejoindre des fleuves, des lacs ou l'océan. Elle peut aussi s'évaporer ou être captée par la végétation (elle sera ensuite renvoyée dans l'atmosphère via la transpiration des plantes ou utilisée lors de la photosynthèse). Le retour de l'eau dans l'atmosphère dû à l'évaporation sur le sol ou la transpiration de la végétation constitue l'évapotranspiration. En s'insérant à travers les fissures naturelles des sols (infiltration) ou bien en migrant lentement à travers le sol (percolation), elle va rejoindre les eaux souterraines (notamment les nappes phréatiques). Cette eau souterraine n'est toutefois pas stockée indéfiniment et finit par retourner vers les fleuves, les lacs, les océans ou encore s'évaporer (ou être utilisée par les activités humaines). Ce cycle est d'une grande importance, car il permet de réguler l'énergie reçue par la Terre et influe donc sur la température de la Terre et des océans. Pendant les différentes étapes de ce cycle, considéré comme fermé, l'eau apparaît sous tous ses états - liquide, gazeux ou solide - et se déplace entre différents réservoirs constituant l'hydrosphère : les glaciers, les eaux souterraines, les océans, l'atmosphère et le sol.

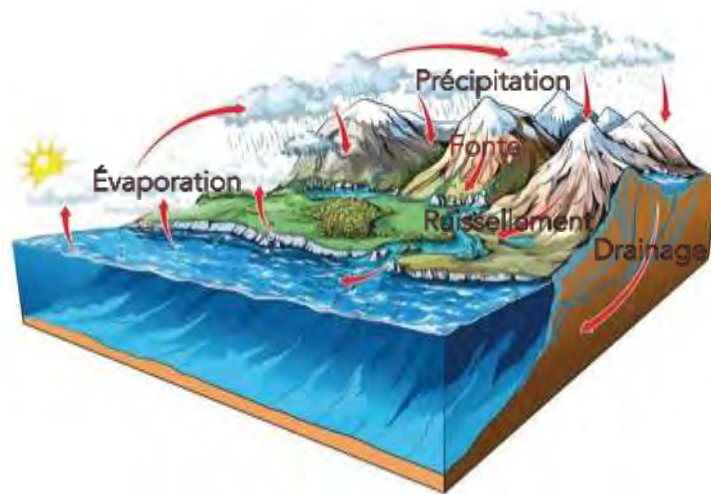


Figure 1.1 – Illustration du cycle de l'eau.

1.1.1 L'évapotranspiration

Le rayonnement électromagnétique émis par le soleil et absorbé par le système climatique sol-atmosphère-océan de la Terre est le principal moteur du cycle de l'eau. Le rayonnement solaire reçu par les couches supérieures de l'atmosphère sont d'une puissance totale de 340 W.m^{-2} . La part de cette énergie interceptée par la surface contribue à son réchauffement et provoque les différents changements d'état de l'eau, sous certaines conditions de température et de pression. Le soleil chauffe les surfaces continentales et océaniques, déclenchant de l'évaporation à la surface des océans, des rivières, des lacs, mais aussi dans la végétation et sur le sol, vers l'atmosphère. L'émission de vapeur d'eau, ou *évapotranspiration*, résulte de deux phénomènes : l'évaporation, qui est un phénomène purement physique à partir des sols humides et des différents plans d'eau, et la transpiration des plantes.

C'est par le mouvement des molécules d'eau que débute l'évaporation. A l'intérieur d'une masse d'eau liquide, les molécules vibrent et circulent de manière désordonnée. Ce mouvement est lié à la *température* : plus elle est élevée, plus le mouvement est amplifié et plus l'énergie associée est suffisante pour permettre à certaines molécules de s'échapper et d'entrer dans l'atmosphère. En théorie, dans des conditions de *pression* et de *température* données, le processus d'évaporation est possible tant que la pression de vapeur effective (pression de la vapeur d'eau dans l'air) reste inférieure à une certaine valeur. Celle-ci n'est autre que la pression de vapeur saturante, pression à laquelle la phase gazeuse est en équilibre avec sa phase liquide ou solide à la température de la surface évaporante. Le rapport entre la pression de vapeur effective et saturante, ou en d'autres termes le rapport entre le contenu en vapeur d'eau de l'air et sa capacité maximale à en contenir, se définit comme l'*humidité relative* de l'air. Ainsi, pour qu'il y ait évaporation, il faut que le gradient de pression dû à la vapeur d'eau soit positif. Ceci signifie que la pression de vapeur effective doit être inférieure à la pression de vapeur saturante (dans le cas contraire, la vapeur d'eau aura plus tendance à se liquéfier) ou que l'humidité relative de l'air soit inférieure à 1. Les conditions de température, pression et d'humidité conditionnent ainsi le phénomène d'évaporation de l'eau.

Une bonne partie de l'évaporation de l'eau s'effectue via la transpiration des plantes. Les racines des végétaux captent l'eau qui se trouve dans le sol et, une fois parvenue aux feuilles, une partie de l'eau va s'évaporer à travers les stomates des feuilles (micropores par lesquels l'eau s'évapore). L'autre partie va servir à la photosynthèse. Cette évapotranspiration dépend de deux éléments : la chaleur fournie par le rayonnement solaire et la quantité d'eau disponible dans le sol. La quantité d'énergie solaire arrivant à la surface terrestre est le facteur déterminant de la transpiration mais elle est aussi très sensible aux variations des climats et des saisons. L'évapotranspiration dépend également de l'eau disponible dans le sol. En ce qui concerne la végétation, lorsque l'eau se fait plus rare, les stomates se referment et la transpiration se ralentit. La plante évite ainsi de trop souffrir du manque d'eau. Lorsque ce manque d'eau s'aggrave et qu'il atteint le minimum vital, la plante se flétrit puis meurt. Plus généralement, l'évaporation n'est pas la même en été et en hiver. Son importance diffère également dans les pays froids ou chauds. L'évaporation est beaucoup plus importante en été lorsque le rayonnement solaire est le plus intense.

1.1.2 Les précipitations

L'eau évaporée monte dans l'atmosphère et est transportée par les flux d'air. Quand elle atteint les couches supérieures de l'atmosphère plus froides, la vapeur d'eau condense et forme les nuages. Ensuite, sous les effets de la gravité, l'eau liquide tombe à la surface de la Terre sous forme de *précipitations* liquide - pluie - ou solide - neige. Les précipitations au-dessus d'une surface continentale peuvent être interceptées par plusieurs réservoirs : la végétation ou canopée, les étendues d'eau permanentes ou temporaires que constituent les lacs, les plaines d'inondations et les rivières, ainsi que le sol lui-même sur lequel l'eau ruisselle ou s'infiltre.

Les feuilles et les branches, en agissant comme un écran, peuvent intercepter une grande partie des précipitations. Une certaine portion de cette eau est ensuite égouttée, généralement avec une énergie cinétique moindre que celle de la pluie initiale. L'autre portion est retenue assez longtemps pour être évaporée et ne jamais atteindre la surface du sol. Seule cette deuxième portion est désignée ici par le terme d'*interception*. L'interception varie énormément dans l'espace, principalement en fonction des espèces végétales et des conditions météorologiques. Les facteurs contrôlant la capacité de la canopée à intercepter et stocker de l'eau sont liés au feuillage (forme, taille, rugosité, turgidité, orientation, caducité, âge ...) et notamment sa densité, d'où l'intérêt de prendre en compte un indice foliaire (LAI, de l'anglais *Leaf Area Index*) qui correspond au rapport entre la surface de toutes les feuilles et la surface de sol située à la verticale. Si l'interception des précipitations est influencée par l'écran végétal, elle l'est aussi par de nombreux facteurs météorologiques. Ce sont surtout les caractéristiques d'intensité-durée des précipitations qui contrôlent la capacité d'interception. Ainsi, le pourcentage d'interception diminue avec l'intensité des précipitations.

1.1.3 Le ruissellement et l'infiltration

Les caractéristiques d'intensité-durée de l'épisode pluvieux impactent également la répartition ruissellement-infiltration des précipitations tombant directement sur le sol nu. *Le ruissellement* de surface constitue l'écoulement d'eau liquide sur la surface du sol. À l'opposé, *l'infiltration* désigne le processus par lequel l'eau pénètre le sol. Les premiers millimètres du sol servent d'interface entre les eaux qui s'infiltrent dans le sol et qui peuvent servir à alimenter les végétaux ou à recharger la nappe phréatique et les eaux qui ruissellent et qui sont responsables de l'érosion des sols. Cette interface est donc particulièrement importante car beaucoup de processus physiques dépendent de cette séparation des eaux de pluie entre eaux de subsurface et de surface.

Les propriétés hydriques des sols varient d'un endroit à l'autre dans le paysage. Les taux d'infiltration dépendent fortement du couvert végétal puisque les végétaux et matières organiques augmentent la porosité du sol et agissent ainsi sur sa conductivité hydraulique. En absence d'un couvert végétal, la porosité du sol près de la surface peut diminuer fortement sous l'impact des gouttes de pluie et former une croûte de battance. Dans ce cas, le taux d'infiltration est très faible. Le *ruissellement Hortonien* est provoqué lorsque l'intensité de la pluie est supérieure au taux d'infiltration du sol ; la pluie arrive à une vitesse plus importante et ne peut être absorbée par le sol. Lors de fortes pluies, il est possible que des eaux qui ont infiltré et se sont écoulées dans le sol resurgissent en surface pour provoquer un *ruissellement par retour ou exfiltration*. Il se peut également que les eaux de pluie tombent directement sur un sol complètement saturé à cause d'une remontée de la nappe phréatique en proximité d'un cours d'eau. Dans ce cas, nous parlons de ruissellement sur sol saturé ou *ruissellement de Dunne* (Dunne and Black, 1970). Le ruissellement de surface ne se constitue pas seulement des précipitations non-infiltrées. Ainsi, le phénomène de fonte des neiges et même des glaciers gonfle ce ruissellement. L'eau s'écoule sous forme de petits cours d'eau et torrents vers des canaux plus larges comme les fleuves et atteint in fine un océan ou une mer (dans ce cas on parlera de bassin exoréique). Ces cours d'eau peuvent aussi se regrouper en plans d'eau comme des mares ou des lacs dans les dépressions du sol et ne jamais atteindre un océan ou une mer (on parlera de bassin endoréique).

Parallèlement, l'eau arrivant à la surface du sol peut pénétrer dans les pores du sol grâce à l'attraction gravitaire. Concrètement, la gravité n'est qu'un des facteurs qui détermine le mouvement de l'eau dans le sol. Sur un sol sec, l'eau est fortement aspirée dans le sol par des forces de succion connues sous le nom de *succion matricielle*. Cette force d'attraction est le résultat de la forte attraction entre les molécules d'eau et les surfaces minérales d'une part, et l'attraction entre les molécules d'eau entre elles d'autre part. La première est un phénomène d'absorption, la deuxième de cohésion. La succion matricielle dépend de la teneur en eau du sol : plus le sol est sec, plus l'attraction est forte. C'est pour cela que typiquement les taux d'infiltration sont bien plus faibles sur des sols humectés ou saturés où la succion matricielle devient négligeable. Une fois infiltrées dans le sol, les eaux de *percolation (infiltration profonde)* se dirigent verticalement vers la nappe phréatique (Fox, 2008).

La *nappe phréatique* (ou nappe aquifère) est une réserve d'eau qui se trouve sous la surface de la terre au-dessus d'une poche de terre imperméable. De manière générale, toute formation géologique qui peut stocker et transmettre de l'eau est appelée un *aquifère*. Par nappe, on entend la partie saturée en eau du sol, c'est-à-dire celle où les interstices entre les grains solides sont entièrement remplis d'eau. Au-dessus, on peut trouver des terrains non saturés, dans lesquels les interstices contiennent aussi de l'air. Cette couche est appelée la zone non saturée ou encore *zone vadose*. Les nappes phréatiques sont donc alimentées par les précipitations via l'infiltration de l'eau dans le sol et s'écoulent souvent naturellement par des sources, par les berges d'un cours d'eau ou directement dans les océans. Excepté pour les canaux bétonnés, le lit d'un cours d'eau est un lieu d'échange entre l'eau s'écoulant dans le cours d'eau et les eaux de la nappe connectée. En fonction des gradients entre la surface de la nappe et le niveau du cours d'eau, la nappe peut alimenter le cours d'eau ou le cours d'eau peut alimenter la nappe.

Le cycle de l'eau est ainsi un ensemble de processus permanents qui déplacent continuellement les masses d'eau entre l'océan, l'atmosphère et les surfaces continentales. Ce qui nous intéresse dans ces travaux de thèse, c'est d'être capable d'établir, de suivre et de comprendre les variations et la dynamique spatio-temporelle des eaux douces à la surface des continents. Pour cela, l'ensemble des mécanismes physiques introduits ci-dessus peuvent être théorisés sous forme de modèles hydrologiques et être simulés ou estimés numériquement.

1.2 Modéliser le cycle de l'eau

La partie continentale du cycle de l'eau peut être étudiée à grande échelle à l'aide d'un modèle de surface continentale (désigné dans la suite par l'acronyme anglais LSM, pour *Land Surface Model*) et d'un modèle de routage de rivière (désigné dans la suite par l'acronyme anglais RRM, pour *River Routing Model*). L'objectif de cette thèse étant de s'intéresser aux flux d'eau de surface à l'échelle globale et leurs interactions avec l'ensemble du cycle de l'eau, les modèles présentés dans cette section correspondent à des modèles de surfaces continentales globaux, qui peuvent être utilisés couplés avec des modèles atmosphériques et océaniques globaux (l'ensemble couplé de ces modèles constituant un *General Circulation Model* ; Döll et al., 2016). Pour ce faire, un LSM doit pouvoir simuler le bilan de masse d'eau, mais aussi d'énergie à la surface du sol. Il sert à répartir les précipitations incidentes en neige, ruissellement, eau infiltrée dans le sol, eau évaporée et à calculer la dynamique de l'humidité du sol. Les LSMs forment les conditions aux bords à l'interface surface-atmosphère, tandis que les RRM estimant le débit des fleuves, utilisé aussi comme condition aux bords aux interfaces océaniques.

Évidemment, il existe d'autres modèles hydrologiques, globaux ou non, mais qui se focalisent surtout sur l'estimation des ressources en eau et ne permettent pas le calcul du bilan d'énergie (Döll et al., 2016). Ces modèles ne peuvent donc pas être couplés avec des modèles atmosphérique et océanique et ne seront pas abordés (ou alors simplement mentionnés) dans ce document.

1.2.1 Développements des modèles de surface continentale

La Figure 1.2 présente un schéma conceptuel du fonctionnement d'un LSM. Les conditions aux bords à l'interface sol/atmosphère (c'est-à-dire les flux de masse d'eau et d'énergie et l'état de l'atmosphère au niveau du sol) sont fournis par un modèle atmosphérique ou via un produit de réanalyse de modèle atmosphérique (qui peut être corrigé en utilisant des observations in situ et/ou satellitaires). Le LSM peut être tourné soit en mode forcé, c'est-à-dire que ces conditions aux bords ont été calculées préalablement avant de lancer le LSM. Mais comme indiqué plus haut, le LSM peut aussi être couplé avec un modèle atmosphérique. Dans ce cas, les conditions aux bords du LSM sont fournis en temps réel par le modèle atmosphérique et de manière complémentaire, les conditions aux bords du modèle atmosphérique sont fournies par le LSM. Les deux modèles peuvent donc rétroagir l'un sur l'autre. Dans le cadre de cette thèse, seul le mode forcé sera utilisé.

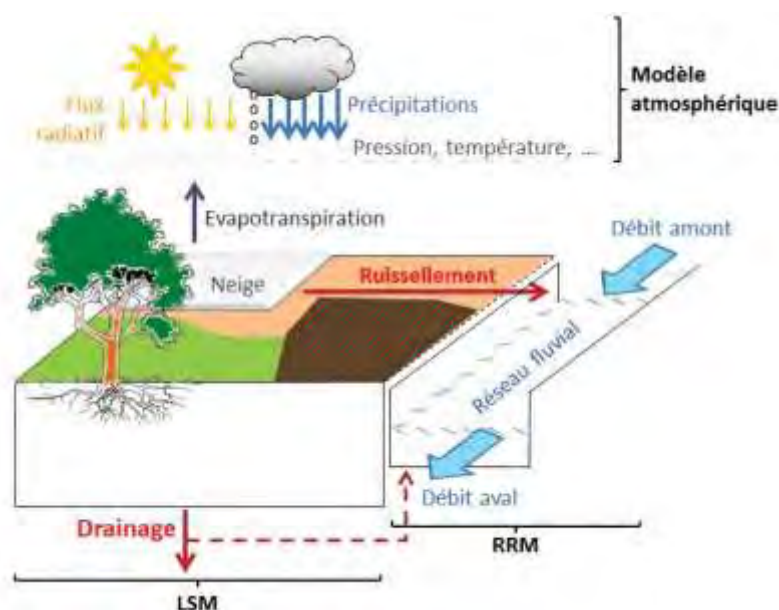


Figure 1.2 – Schéma conceptuel d'un LSM, couplé avec un RRM et de leurs conditions aux bords

Manabe (1969) propose un des premiers schémas d'hydrologie de surface basé sur le *principe du seau* (ou bucket scheme). Dans ce modèle précurseur, les phénomènes d'évaporation et de ruissellement par saturation sont estimés sur un sol ayant une épaisseur et des caractéristiques constantes. Ainsi le sol est identifié à un seau, qui peut déborder sous l'effet des précipitations lorsqu'il est plein et se vide sous l'effet de l'évapotranspiration. Depuis, les LSMs utilisés pour la prédiction météorologique et les projections climatiques se sont considérablement développés. La deuxième génération de modèles (Deardoff, 1977; Dickinson et al., 1986; Sellers et al., 1986) a inclus la représentation de la végétation, la prise en compte du cycle diurne, ainsi que l'introduction de plusieurs couches hétérogènes de sol et du phénomène de ruissellement dû à l'excès d'infiltration. Les LSMs actuels modélisent les réponses complexes de la végétation aux conditions environnementales, l'hydrologie détaillée de la surface et du sol sous-jacent, l'évolution dynamique d'un manteau neigeux éventuel et même les processus inhérents aux régions urbanisées, agricoles et lacustres.

Parmi ces LSMs de troisième génération (Bulygina, 2011), on trouve notamment le modèle CLM (Community Land Model; Lawrence et al., 2011) du National Center for Atmospheric Research (NCAR), le modèle H-TESSEL (Hydrology-Tiled ECMWF Scheme for Surface Exchange over Land; Balsamo et al., 2009) du European Center for Medium Range Weather Forecasts (ECMWF), le modèle ISBA (Interaction Sol-Biosphère-Atmosphère; Noilhan and Mahfouf, 1996) développé au Centre National de Recherches Météorologiques (CNRM)/Météo-France, le modèle ORCHIDEE (Organising Carbon and Hydrology In Dynamic Ecosystems; Krinner et al., 2005) de l'Institut Pierre Simon Laplace (IPSL) ou encore le modèle VIC (Variable Infiltration Capacity; Liang et al., 1994) de l'université de Princeton et de l'université de Washington. Les LSMs cités ci-dessus sont essentiellement utilisés pour la prévision météorologique et/ou les projections climatiques. Le Tableau 1.1, adapté de Bierkens et al. (2015), présente (de manière non exhaustive) un récapitulatif des caractéristiques de certains LSMs disponibles à l'heure actuelle et de leur RRM associé, s'il existe. Ils utilisent des mailles régulières, dont la taille peut varier de 0.25° à 1.0° (~ 25 km à 100 km à l'équateur). On peut aussi noter d'autres modèles similaires comme MARINE (Modélisation et Anticipation du Ruissellement et

des Inondations pour des évènements Extrêmes, Roux et al., 2011) ou encore MERCEDES (Maillage Élémentaire Régulier Carré pour l'Étude Des Écoulements Superficiels).

Nom LSM	Résolution spatiale ; temporelle	Sol et eaux souterraines	RRM	Référence(s)	Institut(s)
CLM	0.9°x1.25° ; 6h	Sol vertical + Rés. eaux sout.	RTM (vitesse écoulement constante)	Lawrence et al. (2011)	NCAR
H-TESSSEL	0.25° -1.0° ; 3h	Sol vertical	-	Balsamo et al. (2009)	ECMWF
ISBA	0.5° ; 3h	Sol vertical + Rés. eaux sout.	TRIP (Éq. Manning + inondations)	Noilhan and Mahfouf (1996) Decharme et al. (2008)	CNRM CEH
JULES	0.5° ; 3h	Sol vertical	Schéma d'advection (grille 1.0°)	Best et al. (2011)	Met Office Univ. Exeter
LaD	1.0° ; journalier	Sol vertical + Rés. eaux sout.	-	Milly and Shmakin (2002)	NOAA GFDL
LM3 (utilisé dans ESM2)	1.0° ; journalier (<1j pour bilan énergie)	Sol vertical + Rés. eaux sout. (dyn. basée topographie)	Rel. non-lin. stock/débit + onde diffusive 1D (faibles pentes)	Milly et al. (2014)	NOAA GFDL
MATSIRU	1.0° ; journalier (6h pour bilan énergie)	Sol vertical + Rés. eaux sout.	CaMa Flood (onde dif-fusive 1D + inondations)	Koirala et al. (2014) Yamazaki et al. (2011)	Univ. Tokyo
ORCHIDEE	0.5° ; 30min	Sol vertical	3 réservoirs lin. vit. spatialisée	Krinner et al. (2005) Polcher (2003)	IPSL LSCE
VIC	0.5° ; journalier (3h pour bilan énergie)	Sol vertical	St-Venant linéarisé	Liang et al. (1994) Lohmann et al. (1996)	Princeton Univ. Univ. Washington

Table 1.1 – Liste non-exhaustive des LSMs utilisés actuellement, ainsi que de leur RRM (quand il existe). Dans ce tableau, « sol vertical » signifie que l'écoulement dans le sol s'effectue verticalement (pas d'écoulement latéral), « Rés. eaux sout. » signifie Réservoir d'eaux souterraines et « Rel. non-lin. » signifie Relation non-linéaire. Adapté de Bierkens et al. (2015).

Comme les LSMs peuvent tourner à l'échelle globale mais aussi régionale, en mode forcé ou couplé (pour réaliser des prévisions ou pour des études climatiques), la physique simulée doit rester suffisamment simple pour ne pas nécessiter des temps de simulation trop pénalisants. C'est pourquoi certains LSMs n'ont toujours pas de RRM associés. Cependant, les évolutions techniques constantes (parallélisation des codes, calcul haute performance,...) permettent d'inclure de nouveaux processus physiques (comme la modélisation de lacs ; LeMoigne et al., 2016) et de tendre vers des résolutions de plus fines. Récemment, il a même été suggéré qu'il serait un jour possible de faire tourner des modèles hydrologiques globaux très résolus, sur des grilles de 100 m à 1 km (Wood et al., 2011; Bierkens et al., 2015). Ces modèles prendraient en compte les processus physiques, chimiques et biologiques à ces fines échelles, ainsi que l'impact des activités humaines, en utilisant toutes les observations (in situ, aéroportées et satellitaires) disponibles. Évidemment cette vision n'est pas partagée par l'ensemble de la communauté, puisque le manque de connaissance sur les hétérogénéités dans la subsurface et les processus qui y sont liés, ainsi que le manque de données d'entrée de qualité à l'échelle globale, pourraient contre-balancer l'apport d'une résolution spatiale très fine (Beven et al., 2015). Quoiqu'il en soit, les LSMs vont continuer à évoluer vers des échelles plus résolues et devront prendre en compte les processus adaptés à ces échelles.

1.2.2 Développements des schémas de routage

Comme cela a été mentionné à la section précédente, les LSMs font un bilan de masse d'eau et d'énergie à l'échelle du pixel. Cependant, chaque pixel est considéré indépendamment des autres et il n'y a pas d'échanges latéraux entre eux. Pour pouvoir modéliser la partie fluviale du cycle de l'eau, les LSMs doivent donc être combinés à un schéma de routage qui a pour rôle de diriger les flux d'eau dans le bassin, de pixel en pixel, vers l'exutoire en suivant la topographie. C'est le rôle des RRM. Ces schémas permettent alors de transformer les flux de précipitations et le ruissellement en débit fluvial. Cependant, pour s'adapter aux LSMs qui constituent leurs conditions aux bords, les RRM tournent à des résolutions assez grossières en comparaison de la largeur des rivières et des fleuves. De plus, ils ne doivent pas ralentir les temps de calculs des GCMs. C'est pourquoi, les écoulements sont en général simulés de manière assez simple.

De nombreux RRM ont été développés à l'échelle globale dans les années 90 (Vörösmarty et al., 1989; Oki and Sud, 1998; Coe, 1998; Hagemann and Dümenil, 1998; Arora et al., 1999). Ces schémas sont essentiellement basés sur la représentation d'une onde cinématique. Le réseau de routage, qui détermine le sens du flux d'un pixel à un autre, est tiré d'un Modèle Numérique de Terrain (MNT) et permet de transférer les masses d'eau d'une maille à l'autre (régulière ou non) vers l'exutoire, à l'interface continent-océan. Les RRM diffèrent les uns des autres par leur paramétrisation de la surface, la définition de la vitesse d'écoulement et la prise en compte ou non de la dynamique des eaux souterraines et des plaines d'inondations.

La description de la vitesse d'écoulement est considérée de plusieurs manières dans la littérature. Par exemple, Oki and Sud (1998) et Coe (1998) ont considéré une vitesse d'écoulement constante identique sur tout le bassin dans leur RRM respectif, TRIP (Total Runoff Integrating Pathways) et SWAM (Surface Water Area Model). Il est aussi possible d'utiliser une vitesse constante dans le temps mais distribuée spatialement selon la topographie et les caractéristiques du tronçon de la rivière (Vörösmarty et al., 1989; Hagemann and Dümenil, 1998; Ducharme et al., 2003). Cependant, la plupart des études plus récentes (par exemple Arora et al., 1999; Ngo-Duc et al., 2007; Lucas-Picher et al., 2010; Decharme et al., 2012) se basent sur une vitesse variable en temps et en espace estimée à partir de l'équation de Manning (Manning, 1891).

Les premiers RRM ne modélisaient que l'eau s'écoulant en surface, dans le cours d'eau (Vörösmarty et al., 1989; Oki and Sud, 1998; Coe, 1998). Le tronçon de rivière était (et est toujours) assimilé à un réservoir de surface qui se remplit avec le ruissellement de surface et le flux d'eau sortant des réservoirs amont et qui se vide dans le réservoir suivant selon le réseau de routage. Des études ultérieures (Decharme et al., 2003; Decharme et al., 2008; Lucas-Picher et al., 2010; Pedinotti et al., 2012) ont démontré la nécessité de simuler la contribution retardée des eaux souterraines au débit de la rivière. Par ailleurs, les plaines inondées étaient rarement prises en compte dans ces schémas. Decharme et al. (2008, 2012) ont implémenté un schéma d'inondation dans TRIP qui ne s'active que si le niveau d'eau dans le réservoir de surface dépasse une certaine hauteur seuil, tandis que Vörösmarty et al. (1989) utilise un débit seuil pour activer le schéma d'inondation. Le réservoir d'inondation peut alors échanger des masses d'eau avec la rivière adjacente.

Le Tableau 1.1 fournit une brève description des RRM utilisés avec certains LSMs. On peut constater une certaine diversité de ces schémas à l'heure actuelle : RTM (River Transport Model ; Branstetter and Famiglietti, 1999) utilisé par CLM considère toujours une vitesse constante dans le temps, la version améliorée de TRIP utilisée dans ISBA se base sur l'équation de Manning pour calculer la vitesse de l'écoulement, tandis que CaMa Flood (Catchment-Based Macro-scale Floodplain ; Yamazaki et al., 2011) utilise l'approximation diffusive de l'équation de Saint-Venant.

Les RRM continuent d'évoluer et deviennent de plus en plus précis. Notamment pour ceux utilisés dans des modèles hydrologiques régionaux, la discrétisation sur une maille régulière commence à être remplacée par des maillages irréguliers par sous-bassins hydrologiques (Paiva et al., 2013). Certains RRM comme HyMAP (Hydrological Modelling and Analysis Platform ; Getirana et al., 2012) permet aussi de prendre en compte des dynamiques d'écoulement dans les plaines d'inondation. Cependant, il n'est pas encore utilisé en mode couplé avec un LSM.

Étant donné les simplifications utilisées dans les RRM actuels, ils sont forcément moins précis que les modèles hydrodynamiques. Ces derniers résolvent, à une ou deux dimensions, les équations hydrodynamiques de Saint-Venant pour estimer le débit des rivières (Hervouet, 2007; Goutal et al., 2014). La résolution de ces équations pouvant être coûteuse en temps de calcul et la méconnaissance, à l'échelle globale, des paramètres d'entrée et de forçage de tels modèles, ils sont le plus souvent mis en place à des échelles régionales plus limitées ou sur des tronçons de fleuves (on parle de modèles débit-débit, par opposition aux modèles hydrologiques qui sont appelés modèles pluie-débit). On peut toutefois noter que Sampson et al. (2015a) ont développé une modélisation hydrodynamique globale, pour simuler des cartes mondiales de risques d'inondation.

Dans le cadre de cette thèse, le LSM et le RRM utilisés correspondent à ISBA-TRIP (voir Tableau 1.1 et Decharme et al., 2008). Ces modèles et leur fonctionnement sont décrits plus en détails dans le Chapitre 2.

Le développement constant des RRM permet aussi de réduire les incertitudes et les erreurs inhérentes. Celles-ci émanent de nombreuses sources, de la mise en équation théorique du modèle, à sa résolution numérique. Ces aspects sont discutés dans la section suivante.

1.2.3 Incertitudes, erreurs et limitations de la modélisation

Les incertitudes au sein des modèles numériques sont inévitables et émanent de plusieurs sources, de l'étape de la conception du modèle à sa résolution numérique. Tout d'abord, un modèle consiste à

exprimer, sous forme d'équations, le comportement d'un système physique donné. Cependant, dès la construction du modèle, la physique elle-même n'est pas complètement maîtrisée. Ceci impose l'utilisation d'hypothèses qui introduisent nécessairement des premières incertitudes. De plus, la physique peut être parfaitement connue, mais trop complexe pour être prise en compte dans sa totalité dans le modèle, notamment en raison des contraintes sur le temps de calcul. Les hypothèses simplificatrices qui sont prises ajoutent alors, elles aussi, leur part d'incertitudes.

Par la suite, il faut résoudre le système d'équations résultant. Il est très rare qu'une solution analytique de ces équations puisse être formulée de part la nature souvent non-linéaire du problème. Ainsi le système est résolu à l'aide de méthodes numériques. Pour cela, le problème doit être discrétisé sur un domaine de calcul défini. Plus la discrétisation est fine, plus les incertitudes engendrées sont faibles. La limite dans la finesse de la discrétisation réside dans la puissance des ordinateurs ou calculateurs utilisés pour la résolution et dans le temps de calcul qu'elle nécessitera. La résolution numérique du système engendre nécessairement des erreurs qui peuvent rester stables au fil de la résolution ou bien s'amplifier.

Par ailleurs, les modèles sont définis à l'aide de nombreux paramètres (par exemple le nombre de couches du sol ou l'humidité du sol à saturation pour un LSM). La valeur de ces paramètres n'étant pas systématiquement connue, une valeur a priori est utilisée et une incertitude liée à cette valeur a priori est introduite à son tour. Enfin, la résolution numérique du modèle requiert l'utilisation de conditions aux limites (par exemple les précipitations ou le flux radiatif incident vers le sol) et d'une condition initiale (par exemple le contenu en eau dans chaque couche du sol au début de la simulation). À l'image des paramètres d'entrée du modèle, ces conditions ne sont pas connues avec certitude et sont aussi sujets à des incertitudes. Plus particulièrement dans le cadre des modèles hydrologiques qui fonctionnent en mode forcé, les forçages atmosphériques externes constituent une des principales sources d'incertitudes.

Finalement, toutes ces incertitudes s'additionnent et se répercutent sur la sortie du modèle, ce qui donne une image approximative de la réalité qu'il tente de reproduire. La qualité des sorties des modèles ainsi que l'importance des incertitudes peuvent être quantifiées en utilisant des observations directes (ou indirectes) du phénomène physique étudié. Dans le cadre du cycle de l'eau continental, un grand nombre de processus physiques peuvent être observés, à plusieurs échelles et à partir de différentes techniques.

1.3 Observer le cycle de l'eau

1.3.1 Vue d'ensemble des données in situ : atouts et limitations

Avant l'avènement et la large utilisation des données satellites, les données in situ - du latin *sur place* - constituaient la seule source d'information fiable pour suivre les mouvements des eaux de surface (débits des rivières) et mesurer d'autres quantités d'eau comme les précipitations. Ces données sont obtenues à partir de différentes techniques de mesures : de manière continue par l'intermédiaire de stations de mesures permanentes, ou sur des périodes plus courtes lors de campagnes de mesures. Différents phénomènes du cycle de l'eau peuvent être observés via les données in situ, notamment les précipitations, l'humidité du sol, la couverture neigeuse, la hauteur d'eau dans les fleuves et les nappes souterraines (mais aussi des variables liées au bilan énergétique à la surface comme sa température ou son rayonnement).

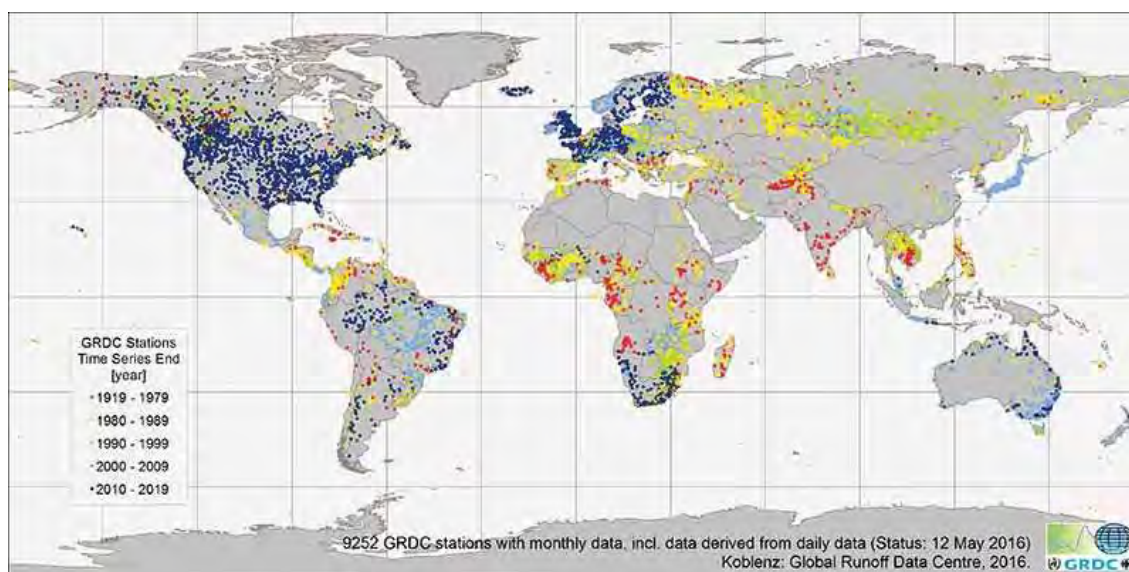


Figure 1.3 – Carte mondiale des stations de mesures in situ de débits des fleuves et rivières gérées par le GRDC (Global Runoff Data Centre). Chaque marqueur représente une station et la couleur indique la période de dernière émission de mesures : (rouge) entre 1919 et 1979, (jaune) entre 1980 et 1989, (vert) entre 1990 et 1999, (bleu clair) entre 2000 et 2009, (bleu foncé) entre 2010 et 2016.

Cependant, la première limitation des données in situ est leur faible couverture spatiale. En effet, elles ne donnent qu'un point de vue ponctuel et dans une seule dimension de phénomènes qui, bien souvent, présentent de fortes hétérogénéités dans toutes les dimensions. L'humidité du sol, par exemple, présente une forte variabilité spatiale. À l'exception de quelques zones limitées où plusieurs stations de mesure sont relativement proches, le réseau actuel de stations ne couvre qu'une très faible fraction des surfaces continentales. Ainsi, la résolution spatiale des observations est bien plus petite que l'échelle des variabilités spatiales réelles de l'humidité du sol. Par exemple, la résolution de la banque de données GSMD (Global Soil Moisture Data Bank ; Robock et al., 2000) est de l'ordre de 500 km. Similairement, le réseau de stations de mesure de débit à l'échelle globale disponible auprès du GRDC (Global Runoff Data Center, www.bafg.de/GRDC/EN/Home/homepage_node.html) a une couverture spatiale très hétérogène (Fig. 1.3). De plus, ces mesures in situ dans les fleuves sont adaptées à un écoulement confiné dans un tronçon aux rives bien définies, mais seront limitées pour des écoulements dans des plaines d'inondations.

La deuxième limitation est associée à la résolution temporelle des mesures (Fig. 1.3). Si la plupart des fleuves dans les pays développés sont continuellement surveillés et les données rendues accessibles par certains pays à la communauté internationale, ce n'est pas toujours le cas pour les grands fleuves des pays en voie de développement comme le Congo ou les fleuves en Arctique. De nombreuses stations in situ sont soit inactives soit leurs données ne sont pas divulguées : sur le Congo, mise à part la station de Kinshasa et quelques rares autres stations, aucune série temporelle, gérée par le GRDC, n'a été actualisée depuis les années 80. Autre exemple, les fleuves transfrontaliers avec de forts enjeux pour la gestion et l'accès à l'eau potable, ou se trouvant dans des zones géopolitiques instables, rendent aussi difficile la mise en place et le suivi de nouvelles stations de mesure ainsi que le partage des données recueillies.

Étant donné les limitations des réseaux de mesure in situ, depuis quelques décennies les données spatiales se sont imposées comme des moyens de mesures alternatifs et complémentaires. Elles offrent des observations répétitives, souvent à grande échelle et permettent donc d'observer les régions peu ou pas observées. Elles permettent de plus d'avoir accès à des échelles spatio-temporelles parfois différentes, mais complémentaires, des réseaux in situ. Certaines mesures (comme la variation totale de masse d'eau dans toute la colonne de sol à l'échelle d'un bassin) sont même parfois plus faciles à observer avec des satellites que sur le terrain. Cependant, il faut bien souligner que les données spatiales prennent tout leur sens et peuvent être exploitées pleinement quand elles sont combinées avec des données in situ, voir avec des modèles numériques. Malgré les limitations des mesures in situ présentées précédemment, elles sont indispensables pour la recherche hydrologique. En effet, elles demeurent des mesures directes, leurs incertitudes sont plus faibles que les estimations équivalentes faites avec des données télédétektées et sont donc cruciales pour développer et évaluer les modèles hydrologiques. Elles peuvent aussi avoir des fréquences temporelles d'observation bien plus grande que les données satellitaires ou aéroportées (par exemple, une station automatique de jaugeage peut fournir des mesures de cote d'eau à un pas de temps inférieur à l'heure, ce qui n'est pas possible par satellite). Enfin, elles servent aussi de référence pour calibrer et évaluer les actuelles missions de télédétection qui observent aussi des variables hydrologiques (Albergel et al., 2012), puisque leurs erreurs de mesure sont bien plus faibles. Cependant, l'observation de la surface de la Terre par satellite peut offrir un complément conséquent aux données in situ. L'avancée technologique en télédétection depuis ces vingt dernières années permet aujourd'hui d'observer de très nombreux processus du cycle de l'eau avec, pour la plupart, une meilleure couverture spatiale, des qualités de mesure relativement homogènes et une certaine régularité de l'échantillonnage temporel de la série de mesure, par rapport à certaines données in situ.

Un large panel des techniques d'observations des différentes composantes du cycle de l'eau va être présenté en Section 1.3.2. Dans cette section, on présentera l'observation - in situ et par satellite - des différents réservoirs de l'hydrosphère sans évoquer l'observation des fleuves, rivières et lacs. Cette dernière, étant au coeur de ces travaux de thèse, sera plus particulièrement détaillée en Section 1.3.3.

1.3.2 Observer les réservoirs d'eau non fluviaux

Pour la plupart des processus physiques que l'on cherche à observer (précipitations, humidité des sols, réserves d'eau souterraines, glaciers, manteau neigeux, couverture végétale), on présentera à la fois les techniques de mesures in situ et les techniques de télédétection couramment utilisées.

1.3.2.1 Observer les champs de précipitations

En effet, elles sont une source directe d'eau pour le bilan d'eau à la surface et influent aussi sur le bilan énergétique (Fekete et al., 2003; Gottschalck et al., 2005). Il est alors nécessaire d'en connaître l'intensité, la distribution spatiale et temporelle, pour être capable d'anticiper des événements de sécheresse ou d'inondations (Ogden and Julien, 1993, 1994; Faures et al., 1995; Nykamen et al., 2001) ou aussi pour simuler les flux passés et présents.

Il existe deux grands types d'instrument de mesure de la pluie au sol : les pluviomètres et les radars météo. Un pluviomètre collecte directement la quantité d'eau précipitée en un point donné. Sa résolution temporelle peut varier de 1 jour (typiquement pour les pluviomètres à lecture) à quelques minutes (comme par exemple pour les pluviomètres à augets basculeurs) (Cassé, 2015). Des données de précipitations, dérivées de pluviomètres, sont distribuées à l'échelle globale par des instituts comme le GPCC (Global Precipitation Climate Center; Becker et al., 2013). Le GPCC regroupe plus de 65000

stations pluviométriques actives et fournit des séries temporelles de pluies mensuelles à une résolution spatiale de $0,5^\circ \times 0,5^\circ$ sur toute la surface du globe. Les stations pluviométriques restent de facto une source conventionnelle d'informations sur l'intensité des précipitations. Les radars météo, quant à eux, ont une résolution spatiale de l'ordre de 1 km, une résolution temporelle de l'ordre de 5 minutes et une couverture spatiale de l'ordre de 100 km (Cassé, 2015). Ces instruments étant évidemment bien plus coûteux que des pluviomètres, ils y en a bien moins en service, leur distribution spatiale à l'échelle globale est très hétérogène et leurs mesures sont difficiles à obtenir.

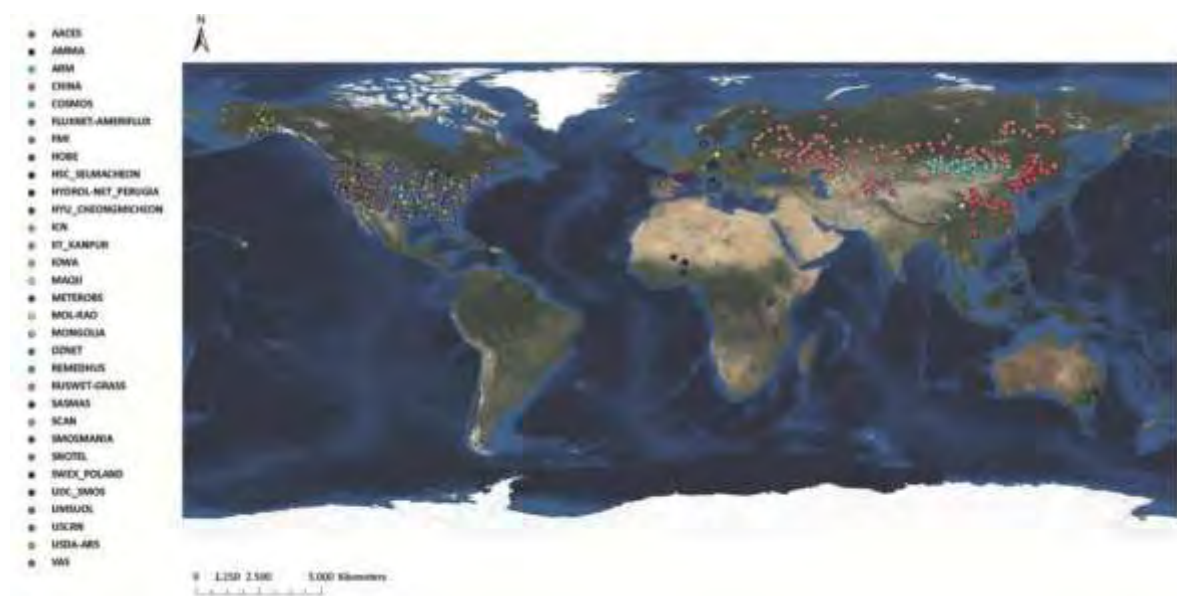
En télédétection, plusieurs techniques d'acquisition existent pour mesurer les précipitations. Elles consistent à observer les propriétés de radiance, température ou émissivité des nuages afin d'en déduire leur teneur en eau. Ce ne sont donc pas des mesures directes et il y a une grande incertitude dans l'estimation de l'intensité des précipitations. Initialement, il existe deux principaux types d'observation des précipitations : la première est basée sur l'utilisation des données infrarouge et la deuxième sur la radiométrie micro-ondes (Vicente et al., 1998; Kidd et al., 2016). Plus récemment, des techniques radar, lidar et aussi combinant infrarouges et micro-ondes sont aussi utilisées. Les techniques infrarouge thermique mesurent la température de sommet des nuages. L'intensité des précipitations correspondantes est alors déduite des mesures de température. Les techniques utilisant des données micro-ondes mesurent, elles, les propriétés de rayonnement (absorption, émission et diffusion) des nuages. Ces propriétés sont caractéristiques de la taille des gouttelettes d'eau et cristaux de glace composant les nuages et permettent d'en déduire l'intensité des précipitations.

Comme exemple de mission de télédétection pluviométrique, la mission spatiale américano-japonaise TRMM (Tropical Rainfall Measuring Mission ; Simpson et al., 1988) a permis une meilleure connaissance tridimensionnelle des systèmes pluvio-gènes mais seulement autour des zones tropicales et subtropicales (bande de latitudes entre 35° Nord et 35° Sud, de la mer Méditerranée et l'Afrique du Sud, liée à l'inclinaison du satellite) entre 1997 et 2015. Plus récemment, la constellation de satellites de la mission GPM (Global Precipitation Measurement ; Smith et al., 2007) combine plusieurs satellites polaires micro-ondes complétés par un dispositif radar comparable à TRMM, le satellite GPM Core Observatory lancé en 2014. Cette mission permet le suivi des précipitations avec une couverture totale de 65° Nord et 65° Sud (entre les deux cercles polaires) avec des données à une résolution spatiale entre 5 et 25 kilomètres toutes les 3 heures. Parmi les missions de la constellation, on trouve les missions ISRO-CNES Megha-Tropiques (<http://meghatropiques.ipsl.polytechnique.fr/>) et NASA Suomi NPP (Suomi National Polar-Orbiting Partnership <http://npp.gsfc.nasa.gov/>) lancées en 2011, ainsi que JAXA GCOM-W1 (Global Change Observation Mission - Water 1) et ESA MetOp-B (Meteorological Operational Satellite Programme - B) lancées en 2012.

1.3.2.2 Mesurer l'humidité du sol

Tout comme les précipitations, l'humidité des sols est une variable importante du système climatique. L'humidité du sol influence les flux d'évapotranspiration et la température de surface et donc les échanges d'énergie entre les surfaces continentales et l'atmosphère. Elle est aussi importante pour anticiper les risques d'inondations ou de sécheresses. Par exemple, un apport important d'eau dans un sol très saturé en eau, ou très sec, génère un fort ruissellement de surface, ce qui entraîne une augmentation du débit et peut potentiellement déclencher une crue.

L'International Soil Moisture Network (ISMN ; Dorigo et al., 2011) est consacré à la collection et à la distribution de données d'humidité du sol. Initialement appelé le Global Soil Moisture Data Bank (Robock et al., 2000), ses données fusionnées à d'autres réseaux ont été rebaptisées en ISMN. Celui-ci regroupait environ 600 stations essentiellement en Asie et aux États-Unis. La base de don-



(a)



(b)

Figure 1.4 – (a) Stations in situ d'humidité de sol, distinguées par réseaux, associées au réseau ISMN en 2012 (Dorigo et al., 2013). (b) Vue d'artiste du satellite SMOS (ESA), lancé en 2009.

nées contenait des séries temporelles mensuelles ou bi-mensuelles de mesures de gravimétrie, sur des périodes de 6 à 15 ans entre 1952 et 2000, pour une résolution spatiale d'environ 500 kilomètres. D'autres réseaux individuels ont alors rejoint le programme et donnent accès à des données provenant d'Afrique de l'Ouest, d'Australie et d'Europe. Les données sont disponibles sur les périodes de 2 à 10 ans entre 2000 et 2011, selon les réseaux. En 2013, le programme regroupait 1400 stations réparties sur 35 réseaux (Fig. 1.4a); Gruber et al., 2013; Dorigo et al., 2013).

Il a été démontré que la radiométrie en bande L et C (1-2 GHz/15-30 cm et 4-8 GHz/3,75-7,5 cm) sont des méthodes très efficaces pour surveiller l'humidité des sols à l'échelle globale (Schmugge, 1998; Wagner et al., 2007; Kerr, 2007). Ainsi, trois missions de télédétection basées sur cette technique travaillent actuellement à l'observation de l'humidité des sols : la mission ESA-CNES SMOS (Soil Moisture and Ocean Salinity, Fig. 1.4b) lancée en 2009, suivie de la mission NASA-Argentine Aquarius en 2011 et de la mission NASA SMAP (Soil Moisture Active and Passive) lancée en 2014.

1.3.2.3 Évaluer les réserves d'eau souterraine

Des études récentes ont montré l'impact important des eaux souterraines sur le bilan d'eau (par exemple Vergnes, 2012). L'amélioration des modèles hydrologiques passe alors par la prise en compte, dans les équations, de la dynamique de ces eaux souterraines et de ce fait, par l'existence et l'utilisation de mesures de la masse d'eau dans ces réservoirs.

Localement, un piézomètre va mesurer la profondeur (ou hauteur piézométrique) de la limite entre la nappe phréatique et la zone vadoze ou Zone Non Saturée c'est-à-dire zone non-saturée en eau du sol, par rapport à la surface du sol. A l'échelle globale, le réseau WHYMAP (World-wide Hydrogeological Mapping and Assessment Programme, www.whymap.org) fournit des données sur les ressources globales en eau souterraine, c'est-à-dire les quantités, la qualité et la vulnérabilité des ressources en eaux souterraines.

La mission NASA-DLR GRACE (Gravity Recovery and Climate Experiment) fournit des cartes mensuelles, à une échelle spatiale de quelques centaines de kilomètres, des variations du champ gravitationnel terrestre. Ces variations illustrent les échanges de masse entre les eaux continentales et l'atmosphère. On appelle ainsi *eau terrestre* la mesure intégrée de toutes les formes d'eau stockées sur et sous la surface de la Terre, ce qui inclut les eaux de surface, l'humidité du sol, le pergélisol (partie du sol gelé en permanence pendant au moins 2 ans), les eaux souterraines, la neige, la glace ainsi que l'eau dans la biomasse. Ainsi, GRACE s'est avéré capable de mesurer les variations d'eau terrestre ou TWS (pour l'anglais *Terrestrial Water Storage*) (Wahr et al., 1998). Depuis son lancement en 2002, GRACE a observé de manière très efficace les variations de TWS (Tapley et al., 2004; Wahr et al., 2004). Plus particulièrement, de nombreuses études se sont ensuite penchées sur l'extraction de la composante des eaux souterraines du TWS dans de nombreuses régions du globe. Rodell and Famiglietti (2002) en ont été les précurseurs en étudiant l'aquifère des Grandes Plaines au centre des États-Unis à partir de données synthétiques GRACE avant le lancement du satellite. Rodell et al. (2007); Strassberg et al. (2007) se sont concentrés sur les anomalies de variations des eaux souterraines ou GWSA (anglais pour Ground Water Storage Anomalies) sur le bassin du Mississippi ou encore Yeh et al. (2006) qui a travaillé sur l'Illinois. Pour extraire la composante GWS (Groundwater Storage) du signal GRACE, des données complémentaires de précipitations, de ruissellement et d'humidité du sol (généralement issues de modèles LSM) sont nécessaires.

1.3.2.4 Suivre les glaciers et le manteau neigeux

La couverture neigeuse est connue pour avoir d'importants effets sur les flux d'eau des surfaces continentales, essentiellement en modifiant son albedo, sa rugosité et aussi grâce aux propriétés d'isolation de la neige. Ainsi, elle agit sur les flux radiatifs, les flux de chaleur et les échanges d'humidité entre la surface et l'atmosphère. Les processus du cycle de l'eau sont aussi affectés car une surface couverte de neige retient les précipitations liquides qui atteignent alors les couches de sol bien après l'épisode pluvieux.

Deux quantités peuvent être mesurées pour caractériser la neige : l'Équivalent en Eau de la Neige (EEN ou SWE pour Snow Water Equivalent en anglais) et l'épaisseur du manteau neigeux. Une technique in situ proche du pluviomètre (Fig. 1.5a) permet d'estimer le EEN tandis que des règles graduées plantées à même la couche de neige fraîche (Fig. 1.5b) permettent de mesurer son épaisseur. La télé-détection permet elle aussi de caractériser le manteau neigeux. Des instruments opérant dans l'optique et l'infrarouge comme l'instrument MODIS (Moderate Resolution Imaging Spectroradiometers) à bord des satellites NASA *Aqua* et *Terra* fournissent des produits de masque de couverture neigeuse (par

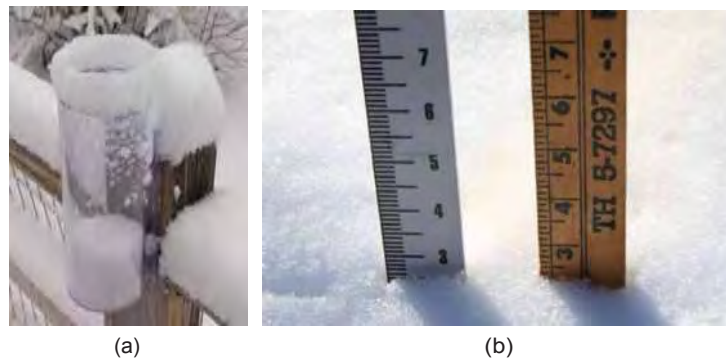


Figure 1.5 – Techniques in situ d’observation de la couverture neigeuse (source NOAA). (a) Mesure de l’Equivalent en Eau de la Neige (EEN). (b) Mesure de l’épaisseur de la couche de neige.

exemple : https://nsidc.org/data/modis/data_versions.html). Dans le cas de MODIS, les cartes sont à une résolution de 250 à 500 mètres. Parallèlement, d’autres instruments de radiométrie micro-ondes passives, comme AMSR-E, fournissent un produit de SWE. Ces instruments mesurent les températures de brillance du couvert neigeux. Ces températures de brillance sont dépendantes des caractéristiques du manteau neigeux, c’est-à-dire la nombre et la taille des grains ou encore la densité de la neige. Elles permettent alors d’en déduire la densité, l’épaisseur de la neige et une estimation du SWE. Elles ne sont en général pas utilisables dans les zones montagneuses et dans certaines zones de forêt. Il peut aussi y avoir des problèmes de localisation spatiale des zones d’accumulation et d’estimation des amplitudes des variations saisonnières (Tedesco and Narvekar, 2010).

La majorité de l’eau douce sur Terre est stockée dans les calottes polaires (Antarctique et Groenland) et les glaciers. Les glaciers résultent du tassement sous son propre poids de la neige accumulée en haute montagne. Ces glaciers peuvent fondre ou grossir de manière plus ou moins importante suivant les variations climatiques. Bien que chaque glacier soit unique, il est possible de distinguer trois zones caractéristiques sur chacun d’entre eux : la zone d’accumulation, partie du glacier où les précipitations de neige se transforment en glace ; la zone de transport, où la fonte reste limitée et où le glacier est le plus épais ; et la zone d’ablation dont la fonte importante contribue à la diminution de l’épaisseur du glacier. Ce sont ces différentes zones que l’on cherche à mesurer via les techniques d’observations afin d’établir le bilan de masse des glaciers.

On se concentre alors sur la quantification des zones d’accumulation et d’ablation. Les valeurs d’accumulation sont déterminées en forant la couche de neige et le névé (accumulation de neige permanente qui peut être à l’origine de glaciers) à l’aide d’un carottier manuel. La densité de neige est mesurée en pesant les carottes extraites et l’équivalent en eau de l’accumulation est déduite. L’ablation est mesurée à l’aide de balises étalonnées, implantées en profondeur dans la glace et dont on mesure l’émergence au-dessus de la surface, ce qui permet de déterminer la lame d’eau perdue. De ces mesures d’accumulation et d’ablation sont déduits des bilans de masse locaux qui sont ensuite extrapolés au glacier tout entier. Ces bilans de masse globaux sont affectés par le nombre limité de sites de mesure. D’autres méthodes in situ, plus indirectes, existent mais elles nécessitent énormément de données annexes pour être appliquées. Parmi les techniques de télédétection, l’imagerie optique est couramment utilisée pour tracer la ligne d’équilibre des glaciers. Pour plus de détails, se référer, par exemple, à Drolon (2016).

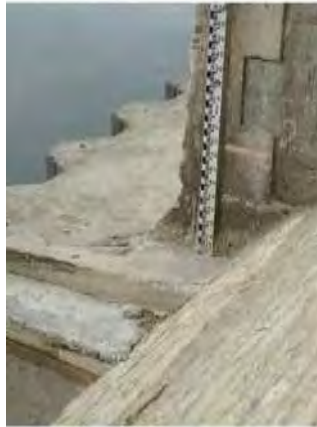


Figure 1.6 – Échelle limnimétrique au bord du Rhône à Arles.

1.3.2.5 Observer la couverture végétale

La couverture végétale et surtout les propriétés de la canopée jouent un rôle important dans la répartition des eaux de pluies à la surface et dans l'évaporation totale. Les propriétés du couvert végétal sont estimées à partir d'imagerie optique et proche-infrarouge comme celles issues des instruments Vegetation (du Satellite Pour l'Observation de la Terre ou SPOT) ou encore MODIS. Ces images servent alors à estimer les indices adimensionnels combinant des réflectances multispectrales tels que le NDVI (de l'anglais *Normalized Difference Vegetation Index*) ou l'EVI (*Enhanced Vegetation Index*) qui sont corrélés aux propriétés des plantes.

1.3.3 Observer et suivre les fleuves et les rivières

On passe maintenant à la présentation des techniques d'observation des fleuves et rivières. Pour l'observation de leur dynamique, on se concentre sur la mesure des hauteurs d'eau et des débits. Dans la Section 1.3.3.1 on présente les méthodes in situ de mesure de ces deux variables. Les sections suivantes abordent la mesure de ces variables par télédétection : la mesure de la cote d'eau par altimétrie nadir satellitaire (Section 1.3.3.2), en combinant imageur radar et MNT (Section 1.3.3.3) et par lidar satellitaire (Section 1.3.3.4), ainsi que l'estimation du débit par télédétection (Section 1.3.3.5).

1.3.3.1 Mesures in situ

Une station de mesure dite *limnimétrique* consiste en une échelle fixée en bord de rivière, voir Figure 1.6, et dont on connaît la position exacte, c'est-à-dire sa longitude et sa latitude. Les mesures limnimétriques donnent la hauteur d'eau par rapport au 0 de l'échelle. On construit alors des courbes de tarage hauteur-débit pour transformer les hauteurs d'eau en débits équivalents. Pour certaines stations de jaugeage, on connaît le 0 de l'échelle, c'est-à-dire la distance entre le 0 de l'échelle et une surface de référence comme un géoïde ou un ellipsoïde que l'on appelle aussi cote. La connaissance du 0 de l'échelle est utile pour définir le profil de la rivière et interpréter les mesures entre les stations. On peut alors convertir la hauteur d'eau en cote d'eau, c'est-à-dire la distance entre la surface de l'eau et le géoïde/ellipsoïde de référence.

On ne peut pas mesurer directement le débit. Cependant, partant de l'hypothèse que le débit s'obtient en multipliant l'aire d'une section transversale par une vitesse d'écoulement intégrée le long de profils verticaux, le débit peut être indirectement obtenu à partir de la mesure simultanée de la bathymétrie de la section (largeur et profondeur) et de la vitesse d'écoulement. Parmi les techniques existantes, les profileurs acoustiques de vitesse par effet Doppler (ADCP) déterminent "presque directement" le débit en mesurant le champ de vitesse dans toute la section (la vitesse est mesurée dans les trois dimensions et la bathymétrie est actualisée). L'ADCP émet des ondes sonores dans plusieurs directions qui sont réfléchies par les particules en suspension dans la rivière et se déplaçant à la même vitesse que l'eau. La mesure de l'écart de fréquence entre l'onde émise et l'onde réfléchie permet d'estimer la vitesse, alors que la mesure du temps entre l'émission et la réception permet d'estimer la distance. D'autres techniques de mesure de la vitesse d'écoulement existent comme l'utilisation d'un moulinet basé sur la relation entre vitesse de rotation de l'hélice du moulinet et vitesse de l'eau, les courantomètres électromagnétiques, l'utilisation de flotteurs se déplaçant à la surface de l'eau, ou encore la dilution de traceurs pour des torrents de montagne rapides, au lit sinueux avec une forte rugosité et une pente soutenue.

Les institutions distribuant ces mesures partagent principalement les données de débits plutôt que les données de cotes d'eau. Les débits, mesurant le flux d'eau, ont un sens physique et global alors que les hauteurs d'eau, locales, n'ont de sens qu'avec la référence du 0 de la règle utilisée pour les mesurer et la bathymétrie du fleuve. Parmi elles, le GRDC, créé en 1988, comprend les séries temporelles de débits in situ journalières ou mensuelles de plus de 9200 stations réparties sur tous les continents (Santini and di Paola, 2015), voir Fig. 1.3. Aussi, à l'échelle d'un bassin hydrologique ou d'un pays, les agences de l'eau peuvent elles aussi distribuer des séries temporelles de débits, de cotes d'eau et autres données relatives à la charge sédimentaire de la rivière par exemple. Parmi elles, l'ANA (Agencia Nacional de Agua) au Brésil distribue en libre accès (via sa plateforme hidroweb : www.hidroweb.ana.gov.br) les données de débits et de cotes d'eau sur l'ensemble du bassin de l'Amazone et ses affluents, soit plus de 9000 stations. Sur le bassin du Niger, l'ABN (Autorité du Bassin du Niger) gère les données hydrologiques provenant des pays traversés par le fleuve et rassemble une dizaine de stations le long du cours principal du fleuve. L'USGS (U.S. Geological Survey) gère le Mississippi et tous les fleuves des États-Unis en général. En France, le SCHAPI (Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations) est chargé de collecter et d'organiser l'ensemble des données d'hydrométrie (hauteurs et débits) à l'échelle nationale. Ces données sont alors regroupées dans la base de données Banque-Hydro (www.hydro.eaufrance.fr).

Les techniques de mesures in situ communément utilisées pour le suivi des fleuves et rivières consistent ainsi à mesurer directement le niveau des eaux par rapport à une référence connue et indirectement les débits à partir de mesures de la vitesse d'écoulement et de la connaissance de la bathymétrie du lit du cours d'eau. Parallèlement, la télédétection permet, elle aussi, l'obtention de certaines de ces mesures via l'altimétrie et l'imagerie radar.

1.3.3.2 Mesures des hauteurs par profils altimétriques

La mesure de profils d'eau par altimétrie radar nadir a initialement été développée pour l'océanographie avec les missions GEOS-3, d'avril 1975 à décembre 1978, et SEASAT (MacArthur, 1980), lancée en 1978 par la NASA. Le satellite SEASAT se composait, entre autres, d'un altimètre nadir qui mesure la distance altimétrique entre le satellite et la surface d'eau libre à la verticale sous le satellite (nadir du satellite), et d'un système de localisation haute précision pour connaître la position du satellite. La différence entre la distance altimétrique et la position du satellite donne alors la cote d'eau. Les cotes d'eau sont référencées par rapport à un ellipsoïde de référence et les mesures sont

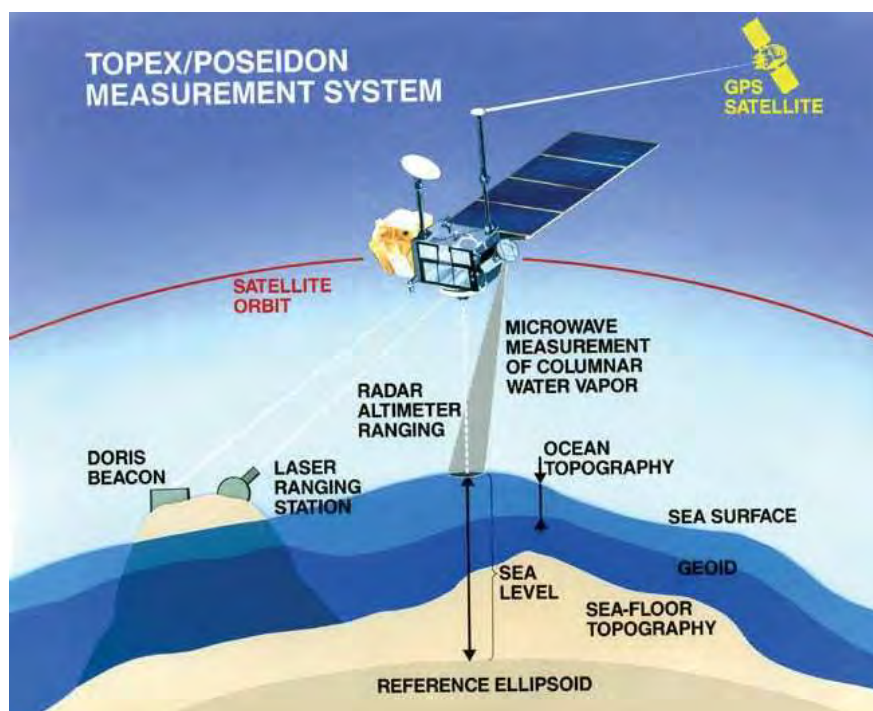


Figure 1.7 - Principe de l'altimétrie nadir (crédits : NASA-JPL : <http://topex-www.jpl.nasa.gov/technology/technology.html>).

faites en 1D le long de la trace du satellite (Fig. 1.7).

Ces missions furent les premières dédiées à l'observation et à la mesure du niveau de la mer. Leur succès a permis la mise en orbite d'un grand nombre de missions d'altimétrie nadir, telles que Geosat en 1985 puis surtout la mission CNES/NASA TOPEX/Poseidon en 1992 et ses successeurs Jason-1 en 2001, Jason-2 en 2008 et Jason-3 en 2016 ; ainsi qu'ENVISAT en 2002 et la mission Saral/AltiKa en 2013 qui permet les premières mesures en bande Ka. Le Tableau 1.2 donne un descriptif de ces différentes missions.

L'altimètre nadir utilise des ondes radar dans la gamme des micro-ondes et plus précisément dans les bandes S (bande du spectre électromagnétique couvrant les fréquences allant de 2 à 4 GHz), C (bande allant de 4 à 8 GHz), Ku (12 à 18 GHz) et plus récemment Ka (26.5 à 40 GHz), en fonction de l'instrument considéré (Tableau 1.2). Ces différentes bandes de fréquence n'ont pas les mêmes propriétés de pénétration et de rétrodiffusion selon les milieux. Certaines missions combinent ainsi les propriétés de plusieurs bandes, comme TOPEX/Poseidon et les missions Jason qui utilisent simultanément la bande C (5,3 MHz) et la bande Ku (13,575 MHz). Ces deux bandes étant affectées différemment par la traversée de l'atmosphère, leur utilisation permet d'en déduire les corrections à apporter à la mesure de l'instrument pour corriger les effets atmosphériques.

Mission	Durée de vie	Fréquence	Inclinaison	Distance traces à l'équateur	Répétitivité
Seasat	06/1978-10/1978 (105 jours)	Ku (13, 5 GHz)	108°	-	17 jours
ERS-1	1991-2000	Ku (13, 8 GHz)	98, 5°	80 km	35 jours
TOPEX/ Poseidon	1992-2005	Ku (13, 5 GHz) C (5, 3 GHz)	66°	315 km	10 jours
ERS-2	1995-2003 (2011)	Ku (13, 8 GHz)	98, 5°	80 km	35 jours
Jason-1	2001-2013	Ku (13, 5 GHz) C (5, 3 GHz)	66°	315 km	10 jours
ENVISAT	2002-2012	Ku (13, 5 GHz) S (3, 2 GHz)	98, 55°	80 km	35 jours
Jason-2	20/06/2008 -	Ku (13, 5 GHz) C (5, 3 GHz)	66°	315 km	10 jours
Saral AltiKa	25/02/2013 -	Ka (35, 75 GHz ± 250) MHz	98, 55°	80 km	35 jours
Jason-3	17/01/2016 -	Ku (13, 5 GHz) C (5, 3 GHz)	66°	315 km	10 jours
Sentinel-3	16/02/2016 -	Ku (13, 5 GHz ±175 MHz) C (5, 3 GHz ±160 MHz)	98, 65°	104 km	27 jours

Table 1.2 – Principales missions d'altimétrie nadir et leur caractéristiques. La barre horizontale sépare les missions terminées (haut) des missions en cours (bas).

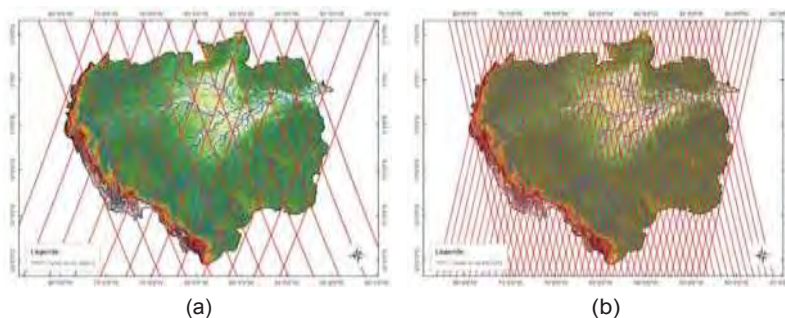


Figure 1.8 – Traces au sol au-dessus du bassin de l'Amazone de (a) Jason-2 et (b) ENVISAT.

L'idée d'utiliser les mesures altimétriques nadir pour l'observation des eaux continentales est apparue dès les lancements de Seasat et Geosat. Rapley et al. (1987) ont montré que l'altimètre de Seasat pouvait observer 32% des surfaces continentales et ont mené une première étude des échos radar sur les zones humides et les rivières. L'utilisation des données Seasat a été poursuivie par Cudlip et al. (1990) pour tracer des profils altimétriques. Une première série temporelle d'évolution des niveaux d'eau dans le bassin de l'Amazone avait été construite. Cette idée fut poursuivie par Koblinsky et al. (1993) en utilisant les données Geosat toujours sur le bassin de l'Amazone. Cependant, la faible quantité de données disponibles (Seasat a opéré seulement pendant 3 mois), les importantes erreurs sur l'orbite satellitaire et la faible précision des observations ont beaucoup dégradé ces résultats. C'est avec le lancement de TOPEX/Poseidon que l'utilisation de l'altimétrie nadir va fortement se développer pour la surveillance des lacs (Birkett, 1995; Hwang et al., 2005), des plaines inondées (Birkett et al., 2002) et des rivières (Birkett, 1998; Kouraev et al., 2004). Dans cette même optique, les missions lancées après TOPEX/Poseidon ont pu être exploitées pour le suivi des lacs et des rivières, comme les missions ERS-2, JASON-2, ENVISAT et SARAL.

La principale limitation de l'altimétrie nadir réside dans sa faible couverture spatiale et temporelle. En effet, les altimètres nadir ne fournissent que des mesures ponctuelles lorsque la trace du satellite croise le cours d'une rivière (sauf rares cas de traces parallèles au cours d'une rivière). On parle alors de *stations virtuelles* en écho aux stations in situ qui mesurent les hauteurs en un seul point de la rivière. Aussi, deux traces adjacentes peuvent être éloignées de plusieurs centaines de kilomètres. De plus, la fréquence temporelle de ces mesures dépend entièrement de la répétitivité de l'orbite du satellite qui n'est pas journalière (e.g 10 jours pour JASON-2 ou encore 35 jours pour ENVISAT), voir Figure (1.8).

1.3.3.3 Mesures des hauteurs par imagerie Radar à Synthèse d'Ouverture (RSO) et Modèles Numériques de Terrain

Contrairement aux altimètres nadir qui observent une tâche ponctuelle directement sous le satellite, les satellites radar imageurs utilisent une visée oblique dans la direction perpendiculaire, appelé axe distance (ou portée), à la direction d'avancée du satellite appelée direction azimuth (Fig. 1.9). La visée oblique est alors caractérisée par un angle d'incidence qui va définir la taille de l'image en deux dimensions vue par le satellite et aussi sa résolution. Les premiers instruments d'imagerie radar patissaient d'une mauvaise résolution azimutale et l'amélioration de celle-ci nécessitait le développement d'antennes dont la taille n'était pas physiquement réalisable. Pour limiter la taille des antennes tout en améliorant la résolution des images, la technique du Radar à Synthèse d'Ouverture (RSO) fut

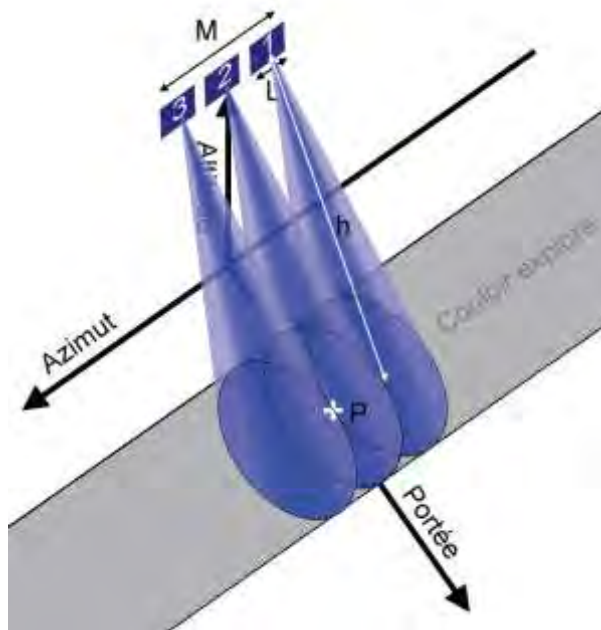


Figure 1.9 – Principe de l'imagerie Radar à Synthèse d'Ouverture (source : <https://commons.wikimedia.org/wiki/File:Sarresrp-fr.png>).

développée par Carl Wiley dans les années 50. Cette technique s'appuie sur le fait qu'un point (par exemple le point P sur la figure 1.9) est observé par plusieurs faisceaux radar lors du déplacement du satellite. En combinant les informations de chaque faisceau observant ce point, on peut obtenir une mesure radar (amplitude et phase) correspondant plus ou moins à la réponse de l'intersection de l'ensemble de ces faisceaux. Cette intersection a une superficie inférieure à la tâche au sol d'un seul faisceau radar. On obtient donc une image mieux résolue que l'image radar initiale, équivalente à une image prise par un instrument « synthétique » ayant une antenne plus grande.

Un des premiers imageurs RSO (SAR en anglais pour *Synthetic Aperture Radar*) satellitaire pour l'observation de la Terre a été embarqué sur la mission Seasat. Cette mission fut aussi un précurseur pour l'imagerie radar et initia une longue série de missions satellitaires, avec notamment les satellites européens ERS-1 et ERS-2 lancés respectivement en 1991 et 1995, le satellite ENVISAT, les missions canadiennes Radarsat-1 et Radarsat-2 lancées en 1995 et 2007, ou encore la mission japonaise JERS lancée en 1992. Le Tableau 1.3 récapitule les caractéristiques de ces différentes missions.

Mission	Durée de vie	Fréquence	Incidence	Fauchée	Rés. spatiale	Répétitivité
Seasat	1978 105 jours	L (1, 275 MHz)	$23^{\circ} \pm 3^{\circ}$	100 km	25 m	17 jours
ERS-1	1991-2000	L (1, 275 MHz)	$23^{\circ} \pm 3^{\circ}$	100 km	30 m	35 jours
JERS	1992-1998	C (5, 3 MHz)	$35^{\circ} \pm 3^{\circ}$	75 km	18 m	44 jours
ERS-2	1995-2011	C (5, 3 MHz)	$23^{\circ} \pm 3^{\circ}$	100 km	30 m	35 jours
Radarsat-1	1995-2000	C (5, 3 MHz)	$35^{\circ} \pm 15^{\circ}$	50 km-170 km	8 m-100 m	24 jours
SRTM	2000 (11 jours)	C (5, 3 MHz) X (9, 6 MHz)	$35^{\circ} \pm 20^{\circ}$ $17^{\circ} - 60^{\circ}$	225 km 50 km	30 m	-
ENVISAT	2002-2012	C (5, 3 MHz)	$30^{\circ} \pm 15^{\circ}$	100 km	30 m	35 jours
Radarsat-2	2007-	C (5, 405 MHz)	$35^{\circ} \pm 20^{\circ}$	8 km-500 km	3 m-100 m	24 jours
TerraSAR-X	2007-	X (9, 65 MHz)	$15^{\circ} - 60^{\circ}$	5 km-150 km	1 m-18 m	11 jours
TanDEM-X	2010-	X (9, 65 MHz)	$15^{\circ} - 60^{\circ}$	5 km-150 km	1 m-18 m	11 jours

Table 1.3 – Principales missions équipées d'un radar SAR et leur caractéristiques. La barre horizontale sépare les missions terminées (haut) des missions en cours (bas).

Une des applications de ces données consiste à l'estimation de masques d'eau (Smith and Alsdorf, 1998). Pour des angles d'observation entre 20° et 50° , les surfaces d'eau renvoient peu d'énergie contrairement au nadir et apparaissent très sombres sur les images et peuvent être délimitées. Il n'y a ainsi aucune donnée de hauteurs sur les zones en eau. Une autre application des radars imageurs a été l'élaboration de Modèles Numériques de Terrain (MNT ou DEM for Digital Elevation Model en anglais). La Shuttle Radar Topography Mission (SRTM; Farr et al., 2007), embarquée à bord de la navette spatiale américaine Endeavour en Février 2000, a utilisé deux antennes SAR observant la même zone sous des angles différents, pour en déduire un MNT des surfaces continentales entre 60° S et 60° N (cette technique de mesure est appelée interférométrie SAR).

Une méthode indirecte possible pour l'estimation des hauteurs d'eau dans les plaines d'inondations consiste à combiner des MNT et des images de masques d'eau. Par exemple, Matgen et al. (2007); Schumann et al. (2007, 2008) utilisent les images de l'instrument ASAR d'ENVISAT pour extraire un masque d'eau combinées à un MNT LiDAR; Schumann et al. (2008) y ajoutant le MNT SRTM. Plus récemment, Zwenzner and Voigt (2009); Schumann et al. (2011) se basent sur les données TerraSAR-X. L'avantage des ondes radar est qu'elles passent à travers les nuages (essentiel pour détecter les masques d'eau en cas d'inondations où il y a de fortes chances que des nuages recouvrent la surface). Cependant, ces mesures indirectes des hauteurs d'eau dépendent de plusieurs aspects qui engendrent de nombreuses sources d'erreurs. D'abord, cette méthode se base sur l'hypothèse que les étendues d'eau entre les rives sont planes et sans ondulations. Or, cette hypothèse n'est pas toujours vérifiée dans les rivières et très rarement dans les plaines inondées. De plus, les hauteurs d'eau estimées sont tributaires des incertitudes sur le MNT, la pente ainsi que de la résolution des images fournissant les masques d'eau.

LeFavour and Alsdorf (2005); Kiel et al. (2006) parviennent à extraire directement des mesures d'élévations de surface de l'eau via les images SAR SRTM. Les impulsions radars sont essentiellement réfléchies de manière spéculaire (c'est-à-dire dans une seule direction) par les surfaces d'eau libres. De ce fait, le coefficient de rétro-diffusion (traduisant l'énergie renvoyée à l'instrument radar) par ces surfaces est très faible. Malgré tout, les ondes en bande C et X peuvent être rétrodiffusées par ces mêmes surfaces à cause de la rugosité de surface induite par le vent. Partant de ce postulat, un produit de hauteurs d'eau SRTM a pu être testé sur le cours principal de l'Amazone et sur un affluent de la rivière Ohio aux États-Unis. Cependant, certains très faibles coefficients de rétro-diffusion font que certaines zones n'ont pas de données exploitables. De plus, même si des données sont disponibles, elles doivent être moyennées sur de très long tronçons (733 kilomètres sur l'Amazone).

1.3.3.4 Mesures par altimétrie LIDAR

La télédétection par laser ou LiDAR (anglais pour *LIght Detecting And Ranging*) est basée sur le même principe fondamental de l'altimétrie radar mais utilise les ondes de la lumière dans les domaines du visible, de l'infrarouge et de l'ultraviolet, issues d'un laser. Le LIDAR est généralement plus précis que l'altimétrie radar. Cependant il ne pénètre pas la couche nuageuse.

L'instrument LIDAR GLAS (Geoscience Laser Altimeter System) a été embarqué sur le satellite NASA ICESat (Ice, Cloud, and land Elevation Satellite) lancé en 2003. Malheureusement, des problèmes techniques sur l'instrument ont réduit sa durée de vie à 1 an seulement sur les 3 ou 4 ans initialement prévus. Cet imprévu a imposé un fonctionnement discontinu du satellite de 1 mois tous les 3 à 6 mois.

1.3.3.5 Mesures des débits via la télédétection

Comme pour les débits in situ, il n'y a pas de technique de télédétection directe pour mesurer le débit. Des méthodes d'estimation indirectes existent et sont basées sur la même approche qui est de simplement relier les hauteurs d'eau télédétectées (via l'altimétrie) ou les aires inondées (via l'imagerie), acquises près de stations in situ existantes, avec les données au sol correspondantes. Ceci revient à réutiliser des courbes de tarage ou des relations géomorphologiques (Kouraev et al., 2004; Coe and Birkett, 2004; Calmant and Seyler, 2006; Papa et al., 2007; Birkinshaw et al., 2014; Gleason et al., 2014). Une autre approche, introduite par Brakenridge et al. (2005), utilise les revisites fréquentes à haute résolution d'imageurs optiques comme MODIS (Tarpanelli et al., 2013, 2015). Le comportement différent de l'eau et du sol dans la bande proche infrarouge du spectre électromagnétique est exploité pour estimer le rapport de réflectance entre un pixel à l'intérieur de la rivière et un autre sur les bords. Les variations de cette fraction traduisent les variations de vitesses d'écoulement (relation linéaire) et donc de débits. Qu'elles soient appliquées à partir de données in situ ou satellite, ces méthodes souffrent de certaines limitations, telles que leur dépendance à une relation empirique qui lie des mesures occasionnelles du débit "vrai" à une autre variable (hauteur d'eau, aire inondée) plus facilement mesurable (Smith and Pavelsky, 2008).

Les mesures, bien que souvent indirectes, de cotes d'eau et de débits par la télédétection offrent d'énormes perspectives quant à l'observation du cycle de l'eau continental et en particulier, des fleuves et rivières. Cependant, les techniques utilisées sont entachées de nombreuses sources d'incertitudes qu'il est important de maîtriser pour pouvoir exploiter correctement les données.

Par ailleurs, bien que l'on vient de faire une présentation large de l'apport de la télédétection pour l'observation du cycle de l'eau, le travail de cette thèse se concentre sur l'étude des eaux de surfaces libres et des fleuves en particulier. Parmi toutes les données de télédétection décrites précédemment, seules des données altimétriques (nadir et interférométriques) seront utilisées dans ce travail de thèse. C'est pourquoi, dans la section suivante (Section 1.3.4), nous nous concentrerons seulement sur la description des incertitudes des altimètres nadir. L'altimètre interférométrique auquel je me suis intéressée au cours de cette étude sera présenté en détail à la Section 1.3.5.

1.3.4 Incertitudes, erreurs et limitations de l'altimétrie nadir

Un altimètre émet une série d'impulsion micro-ondes vers la surface et mesure le temps de retour entre l'impulsion émise et la réception de son écho. Le traitement des mesures de l'instrument permet de déterminer ensuite la distance entre l'antenne et la surface, en supposant que la vitesse de propagation de l'impulsion est constante (Rapley, 1990). Ainsi, la donnée brute, appelée forme d'onde, se présente comme la distribution dans le temps de la puissance de l'écho retournée vers l'antenne. Cet écho est finalement "traité" et corrigé, via des algorithmes dits de *retracking* (Frappart et al., 2006), pour être exprimé en hauteur [m] de la surface libre par rapport à une surface de référence (géoïde ou ellipsoïde).

Plusieurs sources d'incertitudes perturbent ces mesures. La mesure de la distance entre le satellite et la surface repose sur l'hypothèse que la vitesse de propagation est constante. Or, cette hypothèse est vérifiée seulement dans le vide et, concrètement, la vitesse diminue et varie lors du passage de l'onde dans les couches de l'atmosphère. Dans l'ionosphère, l'onde est ralentie lors du passage à travers les particules ionisées et dans la troposphère, la propagation de l'onde est impactée par les conditions d'humidité et de température. De plus, des incertitudes résident aussi dans la connaissance de la position exacte du satellite, sans omettre les incertitudes intrinsèques à l'appareil de mesure.

Par ailleurs, des incertitudes supplémentaires apparaissent lors de l'observation spécifique des eaux continentales. En effet, les formes d'ondes obtenues en observant les rivières et les plaines d'inondations sont influencées par la présence éventuelle de glace, de végétation très près des zones d'eau et elles sont aussi impactées par la topographie environnante. Aussi, les surfaces d'eau continentales sont plus lisses que celles des océans. Ceci favorise le phénomène de réflexion spéculaire et la forme des échos est plus irrégulière que celles obtenues sur les surfaces des océans - soit plus étroites, soit plus larges avec de nombreux pics (Fig. 1.10). En conséquence, ces formes d'onde sont plus délicates à interpréter. Une analyse formelle de l'incertitude sur la hauteur résultante est difficile à mener. Ainsi, l'incertitude totale sur l'estimation de la cote d'eau est le plus souvent quantifiée à l'aide des données in situ.

Les données d'altimétrie nadir souffrent d'une faible couverture spatiale et temporelle (Tableau 1.2). En espace, car les mesures sont essentiellement monodimensionnelles et aussi plusieurs centaines de kilomètres peuvent séparer deux traces adjacentes, voir Figure 1.8. Par exemple, pour la mission TOPEX/Poseidon, deux traces adjacentes ont environ 315 kilomètres d'écart à l'équateur et sont acquises à trois jours d'intervalle. Ces données ont aussi une faible couverture temporelle, car deux mesures consécutives sur une même station dépendent de l'orbite du satellite qui est de plusieurs jours (10 jours pour TOPEX/Poseidon et JASON-2, 35 jours pour ENVISAT).

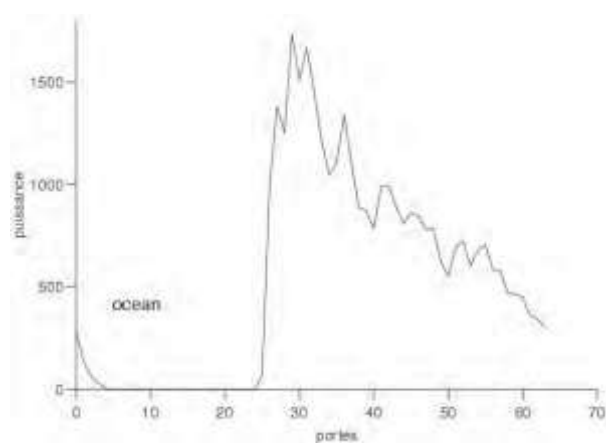
Il existe de nombreuses bases de données de cotes d'eau par altimétrie, comme par exemple la base de données Hydroweb (hydroweb.theia-land.fr). Cependant, les limitations spatio-temporelles des mesures et le fait qu'elles ne fournissent aucune information sur la largeur des fleuves ou leur pente, limite leur utilisation. Il est alors nécessaire d'avoir accès à des observations plus globales pouvant fournir des données de masques d'eau et de cotes d'eau en deux dimensions. C'est pour répondre à ce besoin qu'a été proposée la mission satellitaire *Surface Water and Ocean Topography* (SWOT ; Fu et al., 2012a).

1.3.5 La mission Surface Water and Ocean Topography (SWOT)

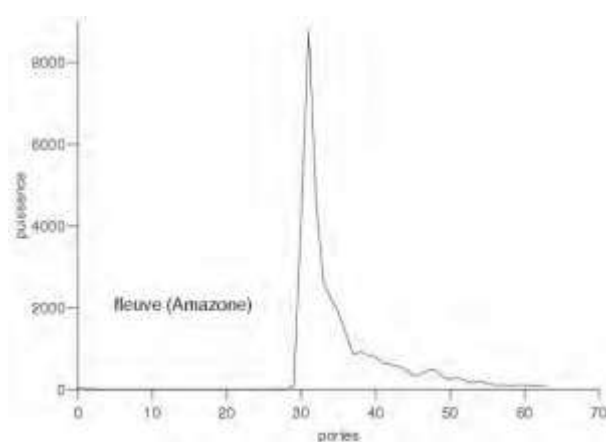
1.3.5.1 Genèse du projet et questions scientifiques

La mission est née de la nécessité de développer de nouveaux outils pour observer avec minutie les eaux de surfaces continentales ; les techniques actuelles ne permettant pas l'observation précise de ces eaux à l'échelle globale et surtout de leur variations spatio-temporelles (Alsdorf and Lettenmaier, 2003). Pour répondre aux besoins de la communauté scientifique hydrologique, une nouvelle mission spatiale appelée WATER (Water And Terrestrial Elevation Recovery ; Alsdorf et al., 2005) a été proposée pour étudier les surfaces d'eau continentale. La mission est basée sur l'héritage technologique des interféromètres SAR de la mission SRTM et de l'instrument WSOA (Wide-Swath Ocean Altimeter).

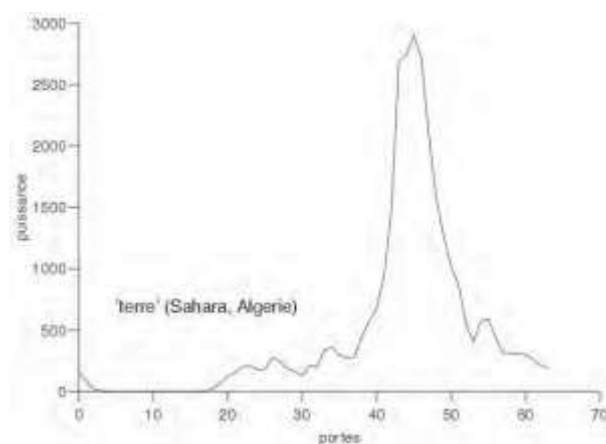
SRTM (Farr et al., 2007) utilisait deux antennes SAR, fonctionnant sur deux bandes de fréquence (bandes C et X), attachées aux extrémités d'un mât de 63 mètres et opérant avec des angles d'incidence entre 20° et 60° (Kiel et al., 2006), c.f. Fig. 1.11a. WSOA (Rodríguez and Pollard, 2001) était initialement prévu pour être embarqué sur la mission NASA/CNES Jason-2, mais le développement de cet instrument fut abandonné en 2004 et il n'a jamais été construit. WSOA possédait aussi deux antennes mais fonctionnant en bande Ku, autour d'un mât plus court de 6,4 mètres, se distinguait par ses angles de visée proche-nadir, c.f. Fig. 1.11b et devait avoir une résolution instrumentale d'environ 1 kilomètre. Quelques années plus tard, le potentiel de cette technique fut alors réutilisé pour les surfaces continentales. Alsdorf et al. (2007) proposa alors une nouvelle mission, WATER, en



(a)



(b)



(c)

Figure 1.10 - Exemples de formes d'onde réelles (puissance reçue en fonction de la distance à l'antenne) de l'altimètre Topex : (a) Forme d'onde sur l'océan, (b) Forme d'onde sur une rivière relativement "propre". Les échos sur les rivières peuvent prendre un très grand nombre de formes (c) Forme d'onde sur terre. (source : site Aviso+, www.aviso.altimetry.fr)

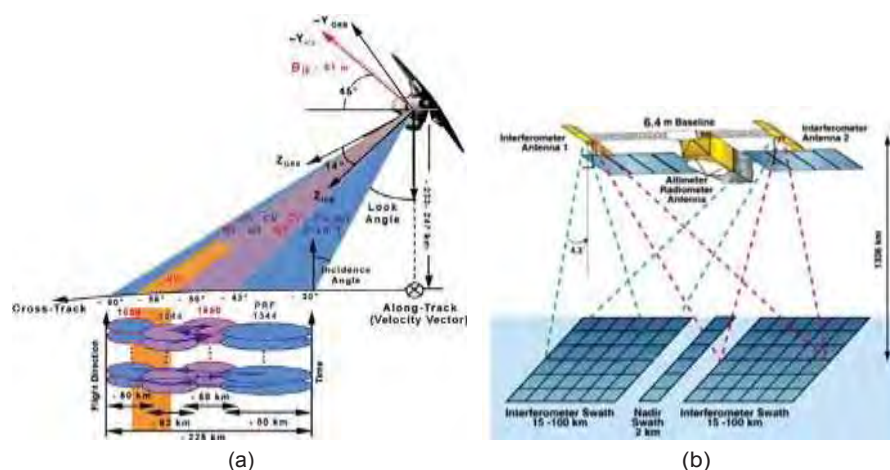


Figure 1.11 – Evolution des concepts des missions SAR interférométrique à large fauchée de SRTM à SWOT : (a) SRTM (crédits NASA), (b) WSOA (Fu and Rodríguez, 2004).

collaboration avec la communauté océanographique. Pour adapter le concept de WSOA aux besoins des surfaces continentales, la bande Ku fut remplacée par la bande Ka (les raisons de ce choix sont développées dans la section suivante). Recommandé à la NASA par le Decadal Survey, le projet fut alors rebaptisé SWOT (Surface Water and Ocean Topography ; Fu et al., 2009).

Depuis, SWOT est développé conjointement par la NASA, le CNES, la CSA (Canadian Space Agency, l'agence spatiale canadienne) et l'UKSA (United Kingdom Space Agency, l'agence spatiale anglaise) et est prévu pour un lancement en 2021 (swot.cnes.fr). Il sera alors capable d'observer l'ensemble du continuum eaux continentales-estuariers-océans.

Les objectifs principaux de la mission SWOT (Rodríguez, 2016) sont, pour l'océanographie, l'amélioration de la couverture spatio-temporelle des données par rapport à celles des altimètres utilisés actuellement en pouvant notamment observer la variabilité océanique à méso-échelle. Pour l'hydrologie, SWOT constituera un premier instrument de mesure dédié à l'observation des surfaces continentales. De plus, SWOT fournira des mesures sur une grille d'1 km² de la latitude 78° Sud à 78° Nord. Cette couverture s'étend aux réservoirs d'eau de surface continentale incluant les lacs et zones humides supérieurs à 250 × 250 m² et les cours d'eau ayant une largeur d'au moins 100 m (avec l'ambition de descendre jusqu'à 50 m). SWOT sera alors capable de détecter la plupart des phénomènes du cycle de l'eau à une échelle temporelle allant de la dizaine de jour à la saison et à une échelle spatiale allant de la centaine de mètres à plusieurs kilomètres, c.f. Fig. 1.12.

Comme indiqué par Rodríguez (2016), SWOT a pour objectif en hydrologie de répondre aux questions scientifiques suivantes :

- Quelles sont les échelles spatio-temporelles des processus hydrologiques contrôlant les eaux de surface et le transport de ces eaux sur les continents ?
- Quels sont les impacts (et leur répartition spatiale) de l'Homme sur les eaux de surface, par exemple par la retenue d'eau par les barrages, les prélèvements et largages dans les rivières et les lacs, des accords de partage des ressources en eaux entre pays transfrontaliers, les dérivations, les digues et toutes autres structures ?

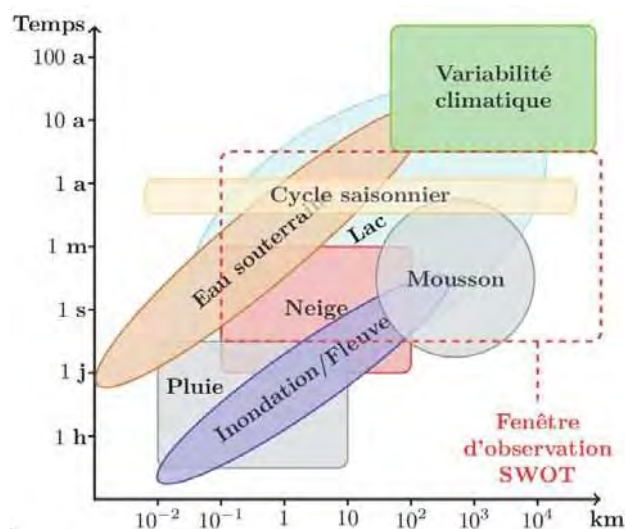


Figure 1.12 – Diagramme espace-temps des processus hydrologiques du cycle de l'eau continental en précisant la fenêtre d'observation de la mission SWOT (Blösch and Sivapalan, 1995; Skøien et al., 2003; Biancamaria et al., 2016).

- Quelle est la sensibilité, aux échelles régionale et globale, des réservoirs d'eau de surface au climat, aux précédents événements d'inondations, à la couverture végétale, aux fortes sécheresses et à la cryosphère ?
- Est-il possible de quantifier les étendues régionales et globales de plaines inondables, en combinant les mesures de niveaux d'eau, de largeurs, de pente télédétections avec la modélisation hydrologique des inondations ?
- Quelle est, à l'échelle globale, la géométrie hydraulique et la structure spatiale en trois dimensions des rivières ?

Pour pouvoir répondre à ces questions, les besoins scientifiques contraignant la conception de la mission ont été définis par Rodríguez (2016) (document aussi appelé SWOT Science Requirement Document). Ces besoins sont présentés dans le Tableau 1.4. C'est en se basant sur ces besoins que les paramètres et les instruments de la mission ont été conçus. Ceux-ci sont présentés à la section suivante.

1.3.5.2 La technologie SWOT

La charge utile principale de SWOT s'appelle KaRIn pour "Ka-band Radar Interferometer" (Fjörtoft et al., 2014). Il se compose de deux antennes SAR en bande Ka (fréquence à 35, 75 GHz pour une longueur d'onde à 8, 6 mm) séparées par un mât de 10 mètres observant les surfaces sous des angles d'incidence proche-nadir entre 0, 6° et 3, 9°. SWOT observe finalement les surfaces aux sols sur deux fauchées de 50 km, chacune distante de 20 km (Fig. 1.13).

La longueur d'onde plus courte de la bande Ka (8, 6 mm, ce qui correspond à 35, 75 GHz en fréquence) comparée aux autres systèmes SAR opérant en bandes L, C et X (pour l'observation de la Terre) a de nombreux avantages :

- Elle permet une meilleure résolution spatiale (qui dépend de la longueur d'onde électromagné-

Surfaces observées	Toutes surfaces d'eau détectées accessibles aux utilisateurs, mais : - erreurs <i>évaluées</i> pour étendues d'eau $\geq 250 \times 250 \text{ m}^2$ et pour rivières $\geq 100 \text{ m}$ de large sur des tronçons de $\sim 10 \text{ km}$ de long - erreurs <i>caractérisées</i> pour étendues d'eau entre $100 \times 100 \text{ m}^2$ et $250 \times 250 \text{ m}^2$ et pour rivières entre 50 m et 100 m de large, sur des tronçons $\sim 10 \text{ km}$ de long
Précision verticale	$< 10 \text{ cm}$ en moyennant sur des aires $> 1 \text{ km}^2$ $< 25 \text{ cm}$ en moyennant sur des aires entre $(250 \text{ m})^2$ et 1 km^2
Incertitude sur la pente	$1,7 \text{ cm/km}$ en moyennant sur des aires $> 1 \text{ km}^2$
Erreur relative sur les aires observées	$< 15\%$ sur les étendues d'eau et tronçons de rivières évalués $< 25\%$ sur le total d'étendues d'eau et de tronçons caractérisés
Durée de vie nominale	3 ans
Signalement eau gelée/pluie/layover	68% des données contaminées devront être correctement signalées
Données collectées	$> 90\%$ des données océan/continent pendant 90% du temps d'activité

Table 1.4 – Contraintes scientifiques de la mission SWOT et ses objectifs (Rodríguez, 2016).

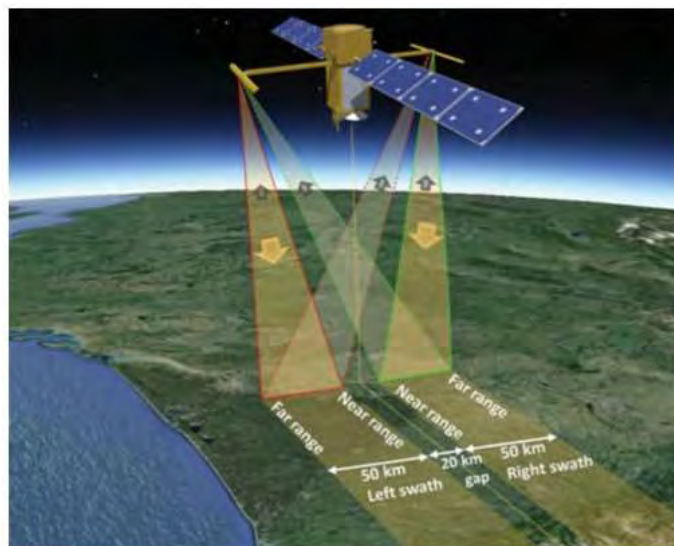


Figure 1.13 – Vue conceptuelle du satellite SWOT et de ses principaux instruments (Biancamaria et al., 2016).

tique) des images.

- Elle permet une distance réduite entre les deux antennes SAR, donc un mât plus court, pour atteindre la précision verticale souhaitée (par rapport aux autres bandes de fréquence), et par conséquent facilite la construction de l'instrument.
- Elle possède des propriétés de pénétration plus faible dans la végétation, le sol et la neige (Fjörtoft et al., 2014), qui permettra une meilleure estimation de l'élévation de l'eau des zones humides et des surfaces saturées.

En bande Ka, la rétrodiffusion sur les surfaces en eau est plus ou moins spéculaire alors que les surfaces continentales sont plus rugueuses. Les angles d'incidence proche-nadir sont particulièrement adaptés pour l'observation des étendues d'eau. En effet, l'eau rétrodiffusera la plupart de l'énergie émise vers les antennes de l'instrument alors que les surfaces terrestres rétrodiffuseront dans toutes les directions. Il en résulte que le coefficient de rétrodiffusion de l'eau est bien plus élevé (environ 15 dB supérieur) que celui du sol. Cependant, les angles d'incidence étant proches du nadir, la surface de l'eau doit avoir une rugosité suffisante pour rétrodiffuser un minimum d'énergie. De ce fait, si la surface de l'eau devient extrêmement lisse, par exemple si la vitesse du vent est très inférieure à 1 m.s^{-1} , une partie de la fauchée ne sera plus observée (Enjolras and Rodríguez, 2009; Moller and Esteban-Fernandez, 2015). Une autre limite importante de la bande Ka est sa forte sensibilité aux précipitations supérieures à 3 mm.h^{-1} (Alsdorf et al., 2007; Rodríguez, 2016).

Avant SWOT, très peu de missions satellites ont utilisé la bande Ka. La première a été la mission SARAL et son altimètre nadir ALtiKa, lancée en 2013. Même si les mesures obtenues avec ce nouvel instrument permettent une meilleure compréhension de la rétrodiffusion en bande Ka, la nature d'altimètre d'ALtiKa diffère de SWOT par ses angles de visée : ALtiKa intègre toute l'énergie rétrodiffusée sous un cône entre les angles $-0,3^\circ$ et $0,3^\circ$ au nadir (Steunou et al., 2015), alors que SWOT fonctionne avec des angles d'incidence entre $0,6^\circ$ et $3,9^\circ$. La seule autre mission utilisant la bande Ka est la mission GPM, lancée en 2014. Cette mission a à son bord un radar à précipitations fonctionnant sur deux bandes de fréquences, dont l'une est en bande Ka. L'instrument observe une fauchée de 125 km de large (angles d'incidence entre $\pm 8,5^\circ$). Bien que les données fournies par cet instrument peuvent certainement affiner notre compréhension de la bande Ka, ces données ont une résolution spatiale de 5 km, bien plus grossière que celle attendue sur SWOT. Le JPL (Moller and Esteban-Fernandez, 2015) et le CNES (Fjörtoft et al., 2014) ont alors organisé certaines campagnes de mesure aéroportées et de terrain pour mieux appréhender la rétrodiffusion en bande Ka sous les angles d'incidence proche nadir proches de ceux de SWOT. La campagne aéroportée AirSWOT (Rodríguez et al., 2010) en fait notamment partie.

KaRIn fonctionnera en mode bistatique. À un temps donné, une antenne émet un premier signal électromagnétique vers la fauchée la plus proche et les deux antennes reçoivent le signal rétrodiffusé dans leur direction respective. Chaque point de la fauchée est observé de deux points de vue différents (les deux antennes), ce qui permet une localisation précise de ces points. Plus précisément, la différence de phase entre les deux signaux rétrodiffusés (reçus par chaque antenne) ou interférogramme sera ensuite utilisée pour transformer les différences de phase en élévations d'eau. Plus de détails sur l'interférométrie SAR de SWOT peuvent être trouvés dans Fu et al. (2012b); Fjörtoft et al. (2014).

En plus de KaRIn, SWOT portera une charge utile supplémentaire dont un altimètre nadir fonctionnant sur deux fréquences (Bandes Ku et C) similaire à celui embarqué sur Jason-3, c.f. Tableau 1.2. Il fournira des mesures de cotes d'eau au nadir entre les deux fauchées de KaRIn. En plus de l'altimètre, un radiomètre facilitera, sur les océans, les corrections du biais engendré par les effets

de la troposphère humide. Cependant, il ne sera pas utilisé pour l'hydrologie car l'émissivité des surfaces continentales est bien plus importante que celle de l'atmosphère (contrairement aux océans). Les effets de la troposphère humide seront corrigés via l'utilisation d'un modèle atmosphérique. En conséquence, l'erreur résiduelle de la troposphère sera plus importante sur les continents que sur les océans.

1.3.5.3 Produits et erreurs SWOT en hydrologie

Pour l'hydrologie, la mission SWOT est développée pour observer toutes les rivières d'au moins 100 m de large et toutes les étendues d'eau (lacs, bassins, marécages permanents) avec une aire de plus de 250 m × 250 m. L'objectif est même de pouvoir arriver à observer les rivières d'au moins 50 m de large et les étendues ayant une surface d'au moins 100 m × 100 m. Si SWOT est capable de voir des réservoirs d'eau encore plus petits, les données seront aussi distribuées. KaRIn fournira ainsi des images radar bidimensionnelles de cotes des surfaces d'eau avec des pixels de dimension environ 6 m dans la direction azimuth et de 60 m (bord interne fauchée) à 10 m (bord externe fauchée) dans la direction distance (avant géolocalisation des pixels). Ces pixels constituent la résolution spatiale unitaire de mesures SWOT. Mais en pratique, la précision verticale des mesures réalisées à cette résolution par KaRIn sera métrique, voire décamétrique. Cette incertitude verticale s'accompagnera aussi d'une incertitude sur la géolocalisation des pixels (coordonnées longitude/latitude). Pour atteindre la précision spécifiée par le *Science Requirement Document* (Rodríguez, 2016), il sera nécessaire de moyenner les mesures sur un grand nombre de pixels, pour atteindre une précision verticale de 10 cm après moyenne sur une zone d'eau de 1km².

Il est alors important de préciser que les images sont initialement obtenues en projection radar et non en projection géolocalisée. En effet, l'instrument radar mesure la distance entre le point observé et l'antenne. Ainsi, en image radar, deux pixels consécutifs dans la direction distance correspondent à des points au sol ayant une distance similaire du satellite. Pour cette raison, quand les pixels sont géolocalisés après acquisition, ils sont éparpillés et ne forment pas une maille régulière. Par conséquent, le plus basique des produits géolocalisés SWOT sera certainement distribué sous forme de nuage de points.

SWOT produira principalement des mesures des cotes des surfaces d'eau observées, ainsi que des mesures de pentes de surface, largeurs et des cartes de masque d'eau. Il est important de rappeler que la *cote d'eau* correspond à la distance entre la surface d'eau libre et une surface de référence prédéfinie (un géoïde ou un ellipsoïde) ; au contraire de la *profondeur d'eau* qui correspond à la distance entre la surface d'eau et le fond du lit du cours d'eau. SWOT ne mesurera pas des profondeurs mais des cotes d'eau. Les produits SWOT, qui seront distribués par le CNES et la NASA, sont encore en cours de définition. Cependant, les principales caractéristiques de ces produits sont d'ores et déjà connues. Pour chaque trace et chaque cycle, ils seront constitués de :

- Un nuage de points géolocalisés contenant tous les pixels KaRIn identifiés comme « eau ». À chaque pixel sera alors associé la mesure de cote d'eau correspondante, une estimation de l'incertitude sur l'estimation de cette cote d'eau et une estimation de la fraction d'eau correspondante dans ce pixel.
- Un produit vectoriel du linéique des rivières/fleuves avec les cotes, les pentes, les longueurs, les débits moyennés sur les tronçons de longueurs variables (quelques km) et une estimation des erreurs associées à ces variables.
- Un produit vectoriel de noeuds situés le long du linéique avec les cotes, les pentes, les longueurs, les débits moyennés sur les noeuds (séparés d'une centaine de mètres) et toujours une estimation des erreurs sur ces variables.

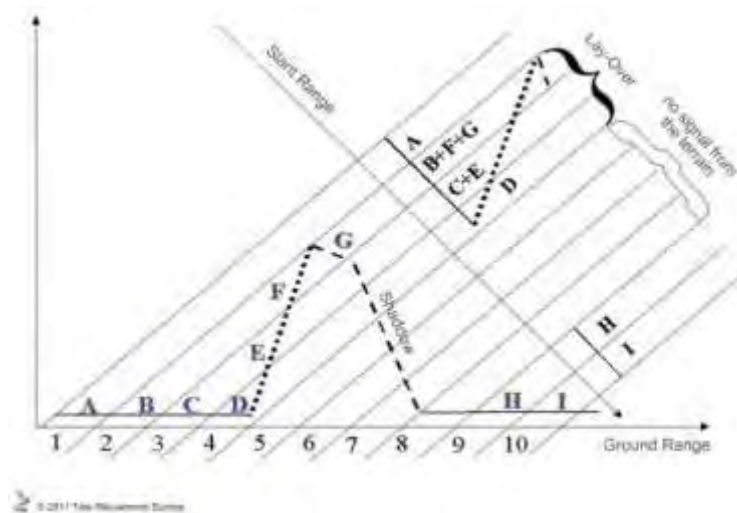
- Un produit vectoriel contenant le contour des lacs observés, auxquels sont associés pour chaque lac la cote moyenne, la superficie et la variation de stock d'eau par rapport à l'observation précédente et les erreurs associées.
- Un produit grillé à maille régulière de toutes les zones en eau observées (lacs, réservoirs, rivières, zones humides,...), en cours de définition.

Les principales sources d'erreurs qui affecteront les mesures de KaRIn sont le bruit thermique instrumental, le chatoiement (*speckle* en anglais), les erreurs de roulis, les effets de troposphère humide et sèche, les effets de l'ionosphère, le *layover* topographique et de végétation (voir Figure 1.14), les effets d'atténuation lors de la traversée de la végétation, la rugosité de la surface d'eau observée, le taux de précipitation (il risque d'y avoir des pertes de données pour des taux de précipitation supérieurs à 3 mm.h^{-1} ; Alsdorf et al., 2007; Rodríguez, 2016). Le bruit thermique (bruit blanc additif) et le chatoiement (bruit blanc multiplicatif) dominent l'erreur totale à l'échelle des pixels KaRIn ($10\text{-}60 \text{ m} \times 6 \text{ m}$) et engendrent des erreurs de l'ordre de plusieurs mètres dans la mesure de cote d'eau. Le bruit thermique se comporte comme un bruit blanc (erreur aléatoire indépendante sur chaque pixel de moyenne nulle) mais leur écart-type tend à augmenter vers le bord externe de la fauchée (Enjolras and Rodríguez, 2009). Heureusement, ces erreurs peuvent être réduites en moyennant les mesures sur plusieurs pixels d'eau. Les autres sources d'erreurs ne sont pas réduites par ce traitement. Sur 1 km^2 , concrètement un tronçon de 10 km de long pour une rivière de 100 m de large, SWOT aura une précision verticale de 10 cm. Le Tableau 1.5 détaille les différentes contributions des sources d'erreur dans ce bilan. Sur cette surface, les effets de la troposphère et les erreurs systématiques dans la direction distance sont les principales sources d'erreurs. Les effets de *layover* domineront cependant les zones de rives, en particulier étant donné les faibles angles d'incidence utilisés par KaRIn. Le *layover* apparaît quand, en projection radar, le sommet de la topographie (ou de la végétation) environnant la surface d'eau est à la même distance du satellite que la zone d'eau observée. Du point de vue du radar, l'énergie rétrodiffusée dans les pixels proviendra à la fois des surfaces en eau et de la topographie et induira une erreur sur la phase interférométrique, donc sur la cote d'eau et sur la position longitude/latitude du pixel. La Figure 1.14 illustre le problème de *layover*, en particulier pour les faibles angles d'incidence.

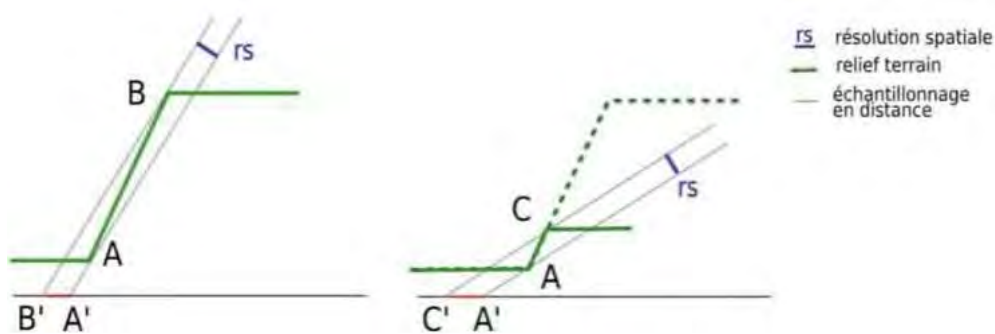
1.3.5.4 Déroulement de la mission

SWOT est prévu pour un lancement en 2021. Le Tableau 1.6 résume les différentes caractéristiques de la mission. Après le lancement, la mission commencera son activité par une phase initiale de calibration, sur les 3 premiers mois de la mission, sur une orbite d'échantillonnage rapide (répétitivité d'un jour). L'objectif de cette phase est d'effectuer des passages fréquents au-dessus de régions océaniques et continentales prédéfinies afin de calibrer les paramètres de l'instrument du satellite et aussi de calibrer les mesures. Ensuite, le satellite sera transféré sur son orbite nominale - orbite non-héliosynchrone à 890, 5 km d'altitude, inclinée de $77,6^\circ$ et ayant une répétitivité de 20, 86 jours - pour une période de 3 ans.

Les Figures 1.15-1.16 illustrent le nombre de revisites de SWOT par période de répétitivité. La couverture spatiale de la mission dépend des caractéristiques de son orbite, de la largeur de la fauchée ($2 \times 50 \text{ km}$), du trou au nadir de 20 km. Il advient que les régions tropicales seront moins observées que les zones situées dans les latitudes plus hautes : le nombre de revisites va de 2 à l'équateur à plus de 10 pour les régions au-delà de la latitude 70° N/S . Très peu de zones ne seront jamais vues par SWOT. Ces zones sont dues au trou au nadir de 20 km entre les deux fauchées et à la distance entre les traces qui ne permet pas toujours à deux fauchées adjacentes de se recouvrir dans les basses latitudes. La surface totale non-observée ne représente, au final, que 3, 55% de la surface



(a)



(b)

Figure 1.14 – Principe de fonctionnement du phénomène de layover et son comportement avec SWOT. (a) Principe général (image provenant de www.treuropa.com et <http://tre-altamira.com> ; Lion, 2012) : Les chiffres représentent les pixels et sont associés à des zones observées indexées en lettres majuscules. L'information observée par le satellite pour chaque pixel est délimitée par les traits obliques dont l'inclinaison dépend de l'angle d'incidence (la distance au satellite est constante le long de ces traits). Le layover apparaît dans le pixel 2 où les informations des zones B, F et G se superposent ainsi que dans le pixel 3 où les informations des zones C et E se superposent et aussi le pixel 4. (b) L'intensité du layover est fonction de l'angle de visée du satellite (Riccio, 2014) : à la même résolution spatiale (i.e. pour le même échantillonnage en distance), le phénomène de layover apparaît dans le cas de faibles angles d'incidence (le cas de SWOT, à droite) pour des faibles reliefs. Les angles d'incidence plus importants (à gauche) activent le layover pour des reliefs plus forts.

Sources	Erreur sur la cote d'eau [cm]	Erreur sur la pente [μ rad]
Ionosphère	0,8	0,1
Troposphère sèche	0,7	0,1
Troposphère humide	4,0	1,5
Orbite (erreur radiale)	1,62	0,5
Instrumental (KaRIn)	8,9	15,5
Aléatoire	(4,4)	15,3
Systématique (dir. distance)	(7,4)	(1,7)
Systématique (dir. azimuth)	(1,5)	(0,08)
Fréquence	(1,15)	(0,5)
Marge	1,23	1,75
Mouvement	0,8	1,6
Total	9,98	15,7
Marge	0,65	6,6
Erreur max exigée	10	17

Table 1.5 – Bilan des erreurs affectant les mesures de cotes d'eau et de pente SWOT après moyennage sur 1 km² (Fernandez and Stenou, 2016).

Orbite	
Altitude	890,5 km
Inclinaison	77,6°
Répétitivité	20,86 jours
KaRIn (charge utile principale)	
Largeur de fauchée (2 au total)	50 km
Distance entre les bords extérieurs des fauchées	120 km
Distance entre les bords intérieurs des fauchées (Trou au nadir)	20 km
Fréquence du radar	35,75 GHz/8,6 mm (bande Ka)
Distance entre les deux antennes	10 m
Résol. pix. distance (projection radar)	60 m en bord interne fauchée (0,6° d'incidence) 10 m en bord externe fauchée (3,9° d'incidence)
Résol. pix. azimuth (projection radar)	6 – 7 m
Charge utile supplémentaire	
Altimètre nadir	Même altimètre que Jason 3, c.f. Table (1.2)
Système de détermination d'orbite	GPS, instrument DORIS, rétro réflecteur laser
Radiomètre	3 fréquences (18, 23 et 34 GHz)

Table 1.6 – Caractéristiques de la mission SWOT.

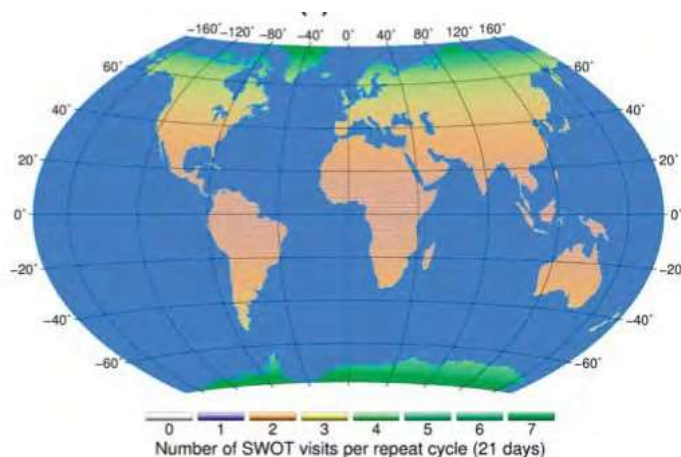


Figure 1.15 – Nombre de revisites de SWOT sur une période de répétitivité (21 jours) sur les surfaces continentales (les océans sont masqués en bleu) entre les latitudes 78° S et 78° N.

totale observable (Biancamaria et al., 2016).

1.3.5.5 Les attentes de SWOT en hydrologie

SWOT surveillera les variations spatio-temporelles des eaux de surface à l'échelle globale, et en particulier celles des fleuves et des rivières. Il produira ainsi les premières images cohérentes des variations de stock d'eau de surface et de débits des rivières. SWOT rendra accessible de nouvelles données très riches pour étudier les rivières, notamment celles peu suivies au sol, peu équipées en stations de mesures in situ ou dont les informations ne sont pas partagées avec la communauté internationale pour des raisons géopolitiques. De plus, SWOT fournira des informations cruciales sur l'impact des débits et de leurs variations sur les sociétés humaines. Ces informations incluent les caractéristiques des inondations et des sécheresses dans les bassins peu surveillés et aussi les variations des débits des fleuves transfrontaliers.

Ainsi, Pavelsky et al. (2014) s'intéressent à une question fondamentale : quelle proportion du réseau hydrographique global SWOT observera ? Quelles avancées cela représentent par rapport aux moyens actuels ? À partir des réseaux de rivières Hydro1K (Verdin and Greenlee, 1998) et HydroSHEDS (Lehner et al., 2008), le réseau global de stations in situ du GRDC et les relations de géométrie hydraulique admises, ils ont estimé que SWOT serait capable de voir plus de 60% des bassins ayant une aire d'au moins 50000 km². Si la technologie permet d'atteindre les rivières d'au moins 50 m de large, ce sera plus de 60% des bassins d'une surface de 10000 km² qui seront observés.

En plus de la mesure de la cote d'eau, de largeur et de pente de surface, de nombreuses études explorent la possibilité de produire une estimation du débit des rivières directement à partir de ces produits (Pavelsky and Durand, 2012). Actuellement, au moins cinq algorithmes de débits sont développés pour remplir cet objectif. Gleason and Smith (2014) et Gleason et al. (2014) se concentrent sur une approche qu'ils qualifient de "*at-many-stations hydraulic geometry*". Il s'agit d'une loi en puissance entre la largeur et le débit, les coefficients de cette loi étant liés le long d'un fleuve, ce qui permet de l'inverser et d'en déduire le débit. Bjerklie (2007) décrit une approche basée sur l'équation de Manning dont les puissances sont ajustées et dont le coefficient de Manning varie de manière sai-

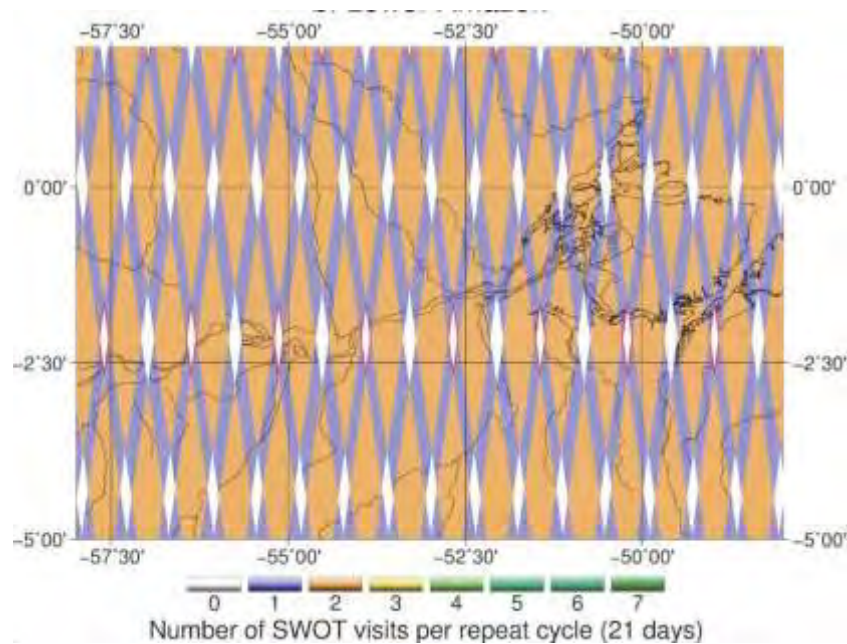


Figure 1.16 – Nombre de revisites de SWOT sur une période de répétitivité (21 jours) sur l'estuaire de l'Amazonie. Les losanges blancs (0 observations) avec les contours magenta sont dus au non-recouvrement des fauchées et les losanges blancs sans contours de couleurs sont dus aux trous au nadir.

sonnière. Garambois and Monnier (2015) proposent une méthode sur des approximations numériques et physiques des équations de Saint-Venant pour inverser la bathymétrie rectangulaire équivalente et le coefficient de frottements équivalents et en déduire le débit. Enfin, Durand et al. (2014) utilisent à la fois l'équation de Manning avec l'équation de continuité et un algorithme de Métropolis pour inverser bathymétrie, frottements et débit.

Les résultats de ces études sont encourageants et démontrent la faisabilité de l'extraction du débit à partir des seuls produits SWOT. Bien que ces algorithmes soient mis en oeuvre par des équipes différentes, leurs développements se font en collaboration car tous les auteurs sont membres du *SWOT Science Definition Team (SDT) Discharge Algorithms Working Group* (groupe de travail sur l'algorithme de débit au sein de l'équipe scientifique de SWOT). Des études d'intercomparaison ont aussi été menées sur différents types de rivières. Durand et al. (2016) ont notamment évalué ces algorithmes, ainsi que le débit médian sur ces algorithmes, sur un ensemble de 19 rivières aux conditions hydrauliques et géomorphiques différentes. Il a notamment été trouvé qu'au moins un des algorithmes est capable d'estimer le débit instantané avec une erreur relative de 35% pour 14 des 16 rivières non-tressées, et ce malgré les inondations, les effets de remous et l'écoulement dans plusieurs chenaux. De plus, il s'avère que parmi ces 35% d'erreurs, une part importante est due à un biais. En effet, l'écart-type médian de l'erreur pour les 16 rivières est de 12, 5%.

Dans ce cadre de cette thèse, le fait que SWOT soit une mission travaillant à l'échelle globale est très intéressante. SWOT va apporter une quantité considérable de nouvelles données pour l'hydrologie de surface, les lacs, les étendues d'eau, les fleuves et les rivières. Il fournira un ensemble de contraintes et de diagnostics qui permettront une meilleure représentation de certains processus encore mal appréhendés comme la dynamique des inondations. Ces données pourraient alors être combinées ou même intégrées à des modèles hydrologiques globaux afin de les corriger et d'améliorer leurs performances et leurs capacités de prédictions. La méthode qui consiste alors à combiner modèle et observations extérieures est appelée l'Assimilation de Données (AD). Dans la section qui va suivre, on va dresser une présentation très générale de l'AD et évoquer rapidement les principales familles de techniques disponibles à l'heure actuelle. Cependant, dans le cadre de cette thèse, on n'utilisera qu'une seule de ces méthodes, voir Chapitres 3 et 4.

1.4 Combiner modèle et observation avec l'assimilation de données

L'assimilation de données (AD) est l'ensemble des méthodes combinant un modèle numérique décrivant le comportement d'un système physique avec des observations directes ou indirectes de ce même système ; tout en tenant compte des incertitudes des deux sources d'information. L'objectif étant, in fine, d'obtenir une représentation de l'état du système étudié la plus proche possible de la réalité. L'AD recoupe les domaines de la modélisation physique, de l'analyse des incertitudes et de l'algorithmique dans le but de prédire avec précision le comportement d'un système. Héritée de la théorie du contrôle (Gelb, 1974), elle a été fortement développée dans le domaine des géosciences (notamment dans le cadre de la prévision numérique du temps) où ni le modèle, ni les observations ne peuvent fournir une description exacte et détaillée de l'état du système, quand ils sont utilisés séparément. Pour écrire cette section on s'est très largement appuyés sur Rochoux (2014) qui donne une description détaillée des méthodes d'assimilation de données.

1.4.1 L'assimilation de données comme un problème inverse probabiliste

Un modèle numérique décrivant un système physique est composé d'un ensemble d'équations simulant le comportement du-dit système, et plus particulièrement son évolution temporelle. Ces équations combinent les paramètres d'entrées, les conditions aux bords et la condition initiale du système pour les traduire en variables décrivant l'état du système, appelées variables d'état du modèle à un temps ultérieur à celui de la condition initiale. Le modèle numérique peut produire une description détaillée, à haute résolution temporelle et/ou spatiale, des variables d'intérêt. Il peut alors permettre d'améliorer notre compréhension des phénomènes physiques qu'il décrit - en mode ré-analyse (sur une période passée) - ou de prédire les comportements futurs du système - en mode opérationnel (prévisionnel).

Cependant, un modèle ne fournit qu'une approximation de la réelle évolution du système. Une première source d'incertitudes sont les approximations et les hypothèses faites pour dériver les équations. Ensuite, nous n'utilisons qu'une approximation de la majorité des paramètres et notre connaissance des conditions initiale et de bords est incomplète. Ces incertitudes dans les paramètres d'entrée se traduisent par des incertitudes dans les sorties du modèle. Ensuite, une autre source d'erreur non -négligeable provient des étapes de discrétisation spatiale et temporelle du système physique réel, indissociables de la construction des modèles numériques, ainsi que de l'utilisation de méthode numérique pour les résoudre. Ces incertitudes interdisent l'utilisation de modèles physiques comme étant entièrement déterministes et contraignent alors à quantifier ces incertitudes afin de pouvoir utiliser malgré tout les modèles incertains. Dans ce contexte, les valeurs incertaines des variables du modèle sont décrites sous la forme d'une densité de probabilité (PDF en anglais pour Probability Density Function) au lieu d'une unique valeur déterministe. On utilisera alors plus justement le terme de "modèles stochastiques" au lieu de "incertains". Ainsi, le modèle numérique fournit une première information a priori sur l'état du système, dite *ébauche*, nécessaire pour initialiser le schéma d'*analyse*, c'est-à-dire le schéma consistant à contruire une meilleure estimation des variables d'intérêts, plus proche de leur valeur *vraie*.

Parallèlement, les observations peuvent fournir une observation directe de l'état du système. Cependant, elles sont généralement éparées en espace et il n'est pas rare que les séries temporelles de mesures soient incomplètes. De plus, elles ne procurent qu'un cliché partiel et instantané du phénomène observé, sans aucune information sur son comportement physique. Quand des observations sont disponibles, les sources d'incertitudes sont majoritairement liées aux *erreurs de mesures instrumentales*. Par ailleurs, il est possible que la quantité mesurée ne soit pas exactement de la même nature que la quantité représentée par le modèle. Ainsi, ces observations doivent être traitées, soit en amont de l'assimilation, soit pendant une étape d'analyse, induisant nécessairement des *erreurs dites de représentation*.

L'AD consiste à corriger les caractéristiques du modèle numérique - ses paramètres d'entrées, son état - à partir des observations disponibles (même incomplètes et indirectes) des sorties de ce même modèle. En opposition à un problème numérique classique qui cherche à trouver les sorties modèles à partir des paramètres d'entrée (problème direct), l'AD correspond alors à un *problème inverse*. Etant donné les nombreuses sources d'incertitudes qui existent inévitablement dans le modèle numérique et les observations, l'AD combine toutes les sources d'informations disponibles pondérées par leur incertitudes. Le but étant d'ajuster les paramètres, les conditions aux bords/initiale ou l'état du modèle. Il en suit que l'AD est un problème inverse probabiliste, dû à la représentation des quantités inconnues par des variables aléatoires et leur PDFs. Dans ce sens, l'AD a pour objectif de déterminer la meilleure estimation des inconnues - l'*analyse* - à partir des erreurs statistiques de l'ébauche et

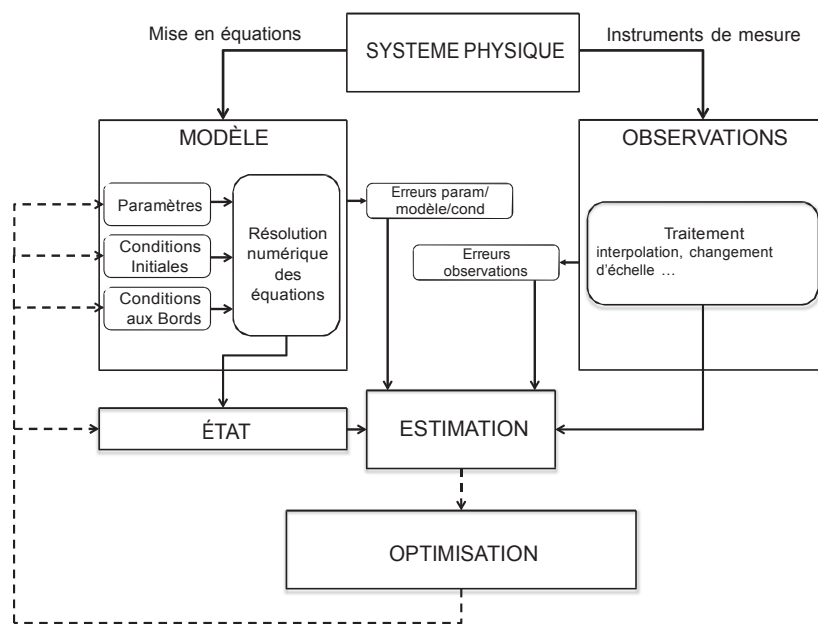


Figure 1.17 – Cadre général d’un problème d’assimilation de données. Le système physique étudié est d’une part, mis en équations dans un modèle numérique selon certaines hypothèses et approximations et d’autre part, observé par divers instruments dont la précision n’est pas exacte. L’AD combine ces deux sources d’information, étant donné les incertitudes de chacune, pour corriger soit les paramètres d’entrée du modèle (estimation de paramètre), soit son état directement (estimation d’état).

des observations. La Figure 1.17 résume les différentes entités impliquées dans un problème d’AD et comment elles interagissent. Enfin, l’un des points essentiels qui différencie les algorithmes d’AD est leur façon de représenter les erreurs et la PDF (par exemple PDF complète versus moments statistiques principaux).

1.4.2 Formalisme : variables d’assimilation de données

Les quantités mathématiques requises pour l’AD sont introduites dans cette section. Tout d’abord, on appelle *fenêtre d’assimilation* la période sur laquelle est réalisé un cycle complet d’assimilation. Elle est délimitée par deux temps d’observation, consécutifs ou non. La définition de la fenêtre d’assimilation dépend de la méthode d’AD choisie. Ici, $[k - 1, k]$ représente l’intervalle de temps entre deux temps d’observations successifs. Certaines méthodes d’assimilation peuvent cependant considérer des fenêtres d’assimilation plus longues incluant plusieurs temps d’observations. Les quantités mathématiques présentées dans cette section permettront alors de décrire le comportement du système physique le long d’une fenêtre d’assimilation, aussi appelée cycle d’assimilation.

Pour toutes les notations qui vont être introduites dans ce qui va suivre, la nomenclature suivante sera adoptée :

- Les éléments en lettre majuscule calligraphiée, tels que Z , sont des opérateurs mathématiques. Les opérateurs mathématiques traduisent une application liant deux espaces différents. En d’autres termes, ils peuvent être vus comme des fonctions qui transforment une quantité du premier espace (antécédent) en son équivalent (image) dans le deuxième espace.

- Les éléments en minuscule sans style particulier, tels que g , représentent des variables scalaires et des variables déterministes (des réalisations de variables aléatoires, des constantes).
- Les éléments en gras minuscule, tels que $\mathbf{g}$, représentent les vecteurs.
- Enfin, les éléments en gras majuscule, tels que $\mathbf{G}$, représentent les matrices.

1.4.2.1 Modèle de propagation du système physique

On note

$$\mathbf{w}_k = [w_{1,k}, w_{2,k}, \dots, w_{n_w,k}]^T \in \mathbb{R}^{n_w}$$

le vecteur contenant les variables d'état pronostiques d'un système physique à l'instant k . Les variables pronostiques sont les inconnues physiques dont l'évolution spatio-temporelle est décrite par des équations, souvent différentielles, qui définissent le modèle numérique. L'évolution spatiale est alors discrétisée sur une grille de n_w mailles. De ce fait, l'élément $w_{i,k}$ représente la valeur de l'élément à la maille i ($i = 1 \dots n_w$) à l'instant k . L'évolution spatiale et temporelle entre les instants $(k - 1)$ et k de la variable pronostique $\mathbf{w}$ est réalisée via un modèle numérique, certainement non-linéaire, noté $F_{[k-1,k]}$. Ce modèle dépend d'un nombre n_p de paramètres d'entrée et de forçages externes inclus dans le vecteur

$$\mathbf{p}_k = \begin{matrix} \mathbf{f} & & \mathbf{I}_T \\ p_{1,k} & p_{2,k} & \dots & p_{n_p,k} \end{matrix} \in \mathbb{R}^{n_p}.$$

Ces paramètres peuvent potentiellement évoluer dans le temps, via l'opérateur de paramètres $P_{[k-1,k]}$. Formellement, la dynamique du système physique de l'instant $(k - 1)$ à l'instant k s'exprime via le système mathématique :

$$\begin{aligned} \mathbf{w}_k &= F_{[k-1,k]}(\mathbf{w}_{k-1}, \mathbf{p}_{k-1}) \\ \mathbf{p}_k &= P_{[k-1,k]}(\mathbf{p}_{k-1}) \end{aligned} \quad (1.1)$$

Dans le cas où les paramètres d'entrée et les forçages sont constants dans le temps, l'opérateur $P_{[k-1,k]}$ se réduit alors à l'application identité $I_{[k-1,k]}$ et les paramètres/forçages externes sont invariants, ie. $\mathbf{p}_k = \mathbf{p}_{k-1}$.

Par ailleurs, en géosciences, les variables d'état pronostiques permettent de déterminer les variables d'état dites diagnostiques. Ces variables sont calculées séparément comme des variables dérivées des variables pronostiques. Elles sont généralement définies sur la même grille que les variables pronostiques, mais dans des soucis de généralités, on définit n_z la taille de la grille sur laquelle sont définies les variables diagnostiques. Les variables diagnostiques sont alors définies par :

$$\mathbf{z}_k = [z_{1,k}, z_{2,k}, \dots, z_{n_z,k}]^T \in \mathbb{R}^{n_z}$$

à partir de relations géométriques et/ou géomorphologiques pouvant faire intervenir les paramètres $\mathbf{p}_k$. L'ensemble de ces relations sont réunies dans l'opérateur Z , indépendant du temps, qui transforme alors les variables pronostiques en variables diagnostiques :

$$\mathbf{z}_k = Z(\mathbf{w}_k). \quad (1.2)$$

1.4.2.2 Modèle stochastique et vecteur de contrôle

Dans la suite, on appelle état *vrai* l'état réel ou exact du système étudié. Il est matérialisé par l'exposant " t ". L'intégration du système d'équations présenté en Eq. 1.1 est associée à des erreurs. Ces erreurs sont liées à la discrétisation, aux erreurs numériques apparaissant lors de la résolution des équations et aussi aux hypothèses simplificatrices faites lors de la formulation même des équations. Ces erreurs existent indépendamment de la connaissance de l'état *vrai* et seront désignées sous le

nom *d'erreurs modèle*. Elles seront notées E_k^F pour le modèle de propagation $F_{[k-1,k]}$ et E_k^P pour l'opérateur de paramètres $P_{[k-1,k]}$. Dans ce contexte, Eq. 1.1 peut être réécrite en fonction des termes *vrais* et de ces erreurs modèles :

$$\begin{aligned} \mathbf{w}_k^t &= F_{[k-1,k]}(\mathbf{w}_{k-1}^t, \mathbf{p}_{k-1}^t) + E_k^F \\ \mathbf{p}_k^t &= P_{[k-1,k]}(\mathbf{p}_{k-1}^t) + E_k^P. \end{aligned} \quad (1.3)$$

Pour la dérivation des algorithmes d'AD et aussi pour la simplification des notations, on introduit l'opérateur de modèle général $M_{[k-1,k]}$ avec l'erreur qui lui est associée E_k^M , regroupant les erreurs paramètres et les erreurs modèles E_k^F et E_k^P . Ainsi, le système en Eq. 1.1 se réduit à

$$\mathbf{x}_k = M(\mathbf{x}_{k-1}), \quad (1.4)$$

et en considérant l'erreur modèle, il se réduit à

$$\mathbf{x}_k = M(\mathbf{x}_{k-1}^t) + E_k^M. \quad (1.5)$$

Le vecteur $\mathbf{x}_k$ introduit dans les équations 1.4 et 1.5 est appelé *vecteur de contrôle* dans le cadre de l'AD, et se définit par :

$$\mathbf{x}_k = [x_{1,k}, x_{2,k}, \dots, x_{n_x,k}]^T \in \mathbb{R}^{n_x}.$$

Ce vecteur inclut les n_x variables qui doivent être estimées/corrigées dynamiquement par l'AD à l'instant k . Les variables de contrôle $x_{i,k}$ sont celles qui ont été identifiées comme ayant le plus d'incertitudes dans le système physique et auxquelles le modèle est le plus sensible. Les variables de contrôle peuvent inclure des paramètres d'entrées du modèle contenus dans le vecteur $\mathbf{p}_k$ et/ou des variables d'état pronostiques $\mathbf{w}_k$ et/ou diagnostiques $\mathbf{z}_k$ du modèle. La taille du vecteur de contrôle n_x peut alors varier drastiquement selon si l'on considère des variables spatialement distribuées ou des paramètres constants en temps et/ou en espace. L'espace dans lequel les variables de contrôle sont définies est couramment désigné comme *espace de contrôle*, et celui-ci n'est pas systématiquement de même nature que celui des variables d'état.

Pour la terminologie, si le vecteur de contrôle ne contient que des paramètres d'entrée (forçages externes, paramètres du modèle), on dit travailler en *estimation de paramètres* ; à l'opposé, on parlera d'*estimation d'état* si le vecteur de contrôle contient les variables d'état (pronostiques ou diagnostiques) du modèle. Dans le cadre de cette thèse, on utilisera les deux approches de manière disjointe : soit l'estimation de paramètres (Chapitre 3), soit l'estimation d'état (Chapitre 4).

Toutes les variables de contrôle sont définies avec des incertitudes. Leur valeur a priori à l'instant k provient d'une intégration du modèle préalable ou d'une connaissance a priori de ces variables. Cette information a priori, aussi appelée *ébauche*, est notée $\mathbf{x}_k^b$. Les incertitudes contiennent les erreurs dues aux changements d'échelle, à la résolution numérique, à la discrétisation, à la connaissance incomplète des phénomènes physiques représentés. Toutes ces incertitudes impliquent que la valeur exacte du vecteur de contrôle, notée $\mathbf{x}_k^t$ (avec l'exposant t pour "true"), est complètement inconnue. De ce fait, une représentation stochastique des erreurs est requise. L'ébauche $\mathbf{x}_k^b$ est alors traitée comme une variable aléatoire alors que $\mathbf{x}_k^t$ est déterministe. On définit l'erreur d'ébauche E_k^b comme la différence entre les vecteurs d'ébauche et vrai :

$$E_k^b = \mathbf{x}_k^b - \mathbf{x}_k^t. \quad (1.6)$$

1.4.2.3 Les observations et l'opérateur d'observation

Le vecteur d'observation $\mathbf{y}_k^o \in \mathbb{R}^{n_y}$ est le vecteur contenant les n_y observations disponibles pour l'assimilation sur l'intervalle de temps $[k-1, k]$, ie. :

$$\mathbf{y}_k^o = \begin{matrix} \mathbf{f} & & & & \mathbf{I}_T \\ y_{1,k}^o & y_{2,k}^o & \dots & y_{n_y,k}^o \end{matrix}.$$

L'espace de dimension n_y dans lequel vivent les observations est appelé *l'espace des observations*. Dans la plupart des problèmes d'estimation d'état, le vecteur d'observation $\mathbf{y}_k^o$ est de dimension inférieure à celle du vecteur de contrôle $\mathbf{x}_k$, alors que c'est généralement l'inverse dans les cas d'estimation de paramètres. Ces quantités mesurées proviennent de différentes sources d'observations (satellites, aéroportées, in situ), à différentes résolutions spatiales et temporelles. Elles sont aussi généralement de nature différente à celle des variables d'état et peuvent nécessiter un traitement préalable avant de pouvoir être assimilées.

Les observations apportent une information indirecte sur les variables de contrôle inconnues. Ceci sous-entend l'existence d'une relation non-linéaire entre les variables de contrôle $\mathbf{x}_k$ et les variables d'observations $\mathbf{y}_k^o$. On exprime généralement cette relation comme un opérateur H_k , appelé aussi *opérateur d'observation*. Il exprime les variables de contrôle en leur image dans l'espace des observations sur la fenêtre d'assimilation $[k-1, k]$. La finalité de l'opérateur d'observation H_k est de fournir les équivalents-modèle des observations $\mathbf{y}_k^o$ au temps k . Ces équivalents-modèle sont notés $\mathbf{y}_k$ et s'expriment comme

$$\mathbf{y}_k = H_k(\mathbf{x}_k).$$

$\mathbf{y}_k$ peut aussi être défini comme l'équivalent du vecteur de contrôle $\mathbf{x}_k$ dans l'espace des observations.

La définition de l'opérateur d'observation H_k dépend alors du choix des variables de contrôle, voir les Figures 1.18 et 1.19.

- En estimation d'état (Fig. 1.18), l'opérateur d'observation peut se réduire à l'opérateur identité si les variables observées sont directement les variables d'état du modèle. En pratique, il consistera en un opérateur de sélection S_k si les observations sont un sous-ensemble des variables d'état, voire en un opérateur d'interpolation si les observations et les variables d'état ne sont pas définies sur la même maille :

$$H_k = S_k.$$

- En estimation de paramètres (Fig. 1.19), l'opérateur d'observation consiste en la composition de l'opérateur de sélection/interpolation introduit pour l'estimation d'état, composé avec l'opérateur modèle $\mathbf{M}_{[k-1,k]}$. En effet, d'abord l'opérateur modèle transforme les paramètres en variables d'état du modèle équivalentes. Ensuite, ces variables d'état sont, à leur tour, exprimées dans l'espace des observations.

$$H_k = S_k \circ \mathbf{M}_{[k-1,k]}.$$

Dans la littérature, cet opérateur est appelé opérateur d'observation généralisée et est noté G . Cependant, dans ce manuscrit, on gardera la notation H_k à la fois pour l'estimation d'état et l'estimation de paramètres, par soucis de simplicité.

Des opérateurs supplémentaires peuvent encore être ajoutés si les variables d'état sont différentes des variables d'observations.

La définition de l'opérateur d'observation est sujette à des erreurs d'essentiellement deux types :

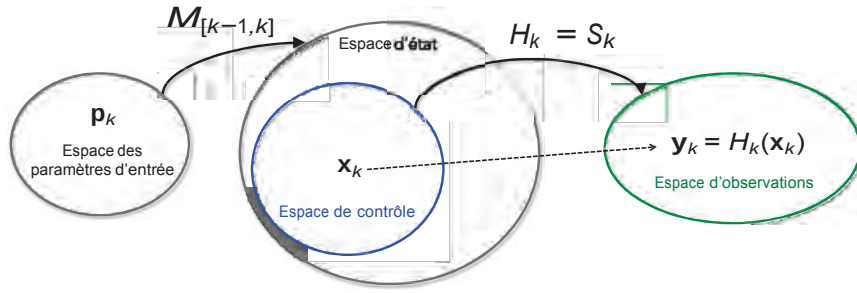


Figure 1.18 – Schéma des différents espaces dans lesquels vivent les variables impliquées dans l'AD en estimation d'état : l'espace de contrôle (bleu) dans lequel vivent les variables à corriger est inclus dans l'espace d'état du modèle (gris). L'opérateur H_k (opérateur de sélection et/ou d'interpolation) exprime ces variables dans l'espace des observations (vert).

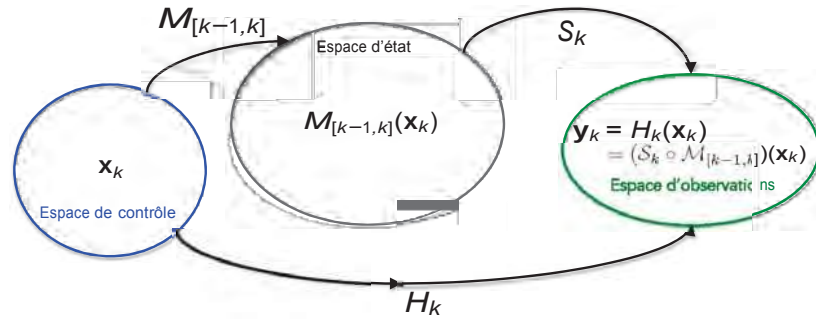


Figure 1.19 – Schéma des différents espaces dans lesquels vivent les variables impliquées dans l'AD en estimation de paramètres : l'espace de contrôle (bleu) dans lequel vivent les variables à corriger est inclus dans l'espace des paramètres du modèle. L'opérateur H_k est la composition de l'opérateur modèle $M_{[k-1,k]}$ (pour exprimer les variables de contrôle dans l'espace d'état en gris) et d'un opérateur de sélection/interpolation (pour exprimer les variables d'état dans l'espace des observations en vert).

- Les erreurs instrumentales, notées E_k^m et définies par $E_k^m = \mathbf{y}_k^o - \mathbf{y}_k^t$, qui apparaissent de manière systématique lorsque qu'une mesure est réalisée. Ces erreurs sont indépendantes de l'opérateur d'observation.
- Les erreurs de représentation, notées E_k^r et définies comme $E_k^r = \mathbf{y}_k^t - H_k(\mathbf{x}_k^t)$ et traduisent la définition imparfaite de l'opérateur d'observation malgré la connaissance de $\mathbf{y}_k^t$.

En supposant que ces deux sources d'erreurs sont non-corrélées, l'erreur d'observation totale E_k^o est définie comme la différence entre les observations $\mathbf{y}_k^o$ et le vecteur de contrôle vrai $\mathbf{x}_k^t$ projeté dans l'espace des observations $H_k(\mathbf{x}_k^t)$:

$$E_k^o = E_k^m + E_k^r = \mathbf{y}_k^o - H_k(\mathbf{x}_k^t). \quad (1.7)$$

1.4.2.4 Le vecteur d'innovation

Le vecteur d'innovation $\mathbf{d}_k^b$ mesure les écarts entre le vecteur d'observation et le vecteur de contrôle d'ébauche exprimé dans l'espace des observations soit :

$$\mathbf{d}_k^b = \mathbf{y}_k^o - \mathbf{y}_k^b = \mathbf{y}_k^o - H_k(\mathbf{x}_k^b). \quad (1.8)$$

À partir de cet écart dans l'espace des observations, l'AD cherche à obtenir une correction dans l'espace de contrôle. L'amplitude de la correction dépend des incertitudes respectives sur les observations et les prédictions du modèle. Si la confiance est très grande dans les observations, la correction va permettre de se rapprocher grandement des observations. Si la confiance est aussi grande entre les observations et les prédictions du modèle, la correction va donner une solution intermédiaire tirant parti des deux sources d'information. Si les observations sont très bruitées, la correction va être faible et permettre d'avoir une trajectoire restant sensiblement proche de l'ébauche.

1.4.3 Traitement stochastique des erreurs

1.4.3.1 Matrices des covariance d'erreur

Le vecteur de contrôle vrai $\mathbf{x}_k^t$ étant inconnu, les erreurs sur l'ébauche E_k^b , les observations E_k^o et le modèle E_k^m sont aussi inconnus. L'AD visant à combiner l'ébauche avec les observations en fonction de leurs erreurs, il est important de bien les caractériser. Ces erreurs sont alors modélisées comme des variables aléatoires définies par leur densité de probabilité (PDF). Ces PDF représentent l'ensemble des valeurs possibles pour ces erreurs et la probabilité d'occurrence de ces dites valeurs. Dans la plupart des algorithmes d'AD et en particulier pour celui utilisé dans le cadre de cette thèse, les erreurs sont décrites par des lois gaussiennes. Ainsi, les erreurs des variables aléatoires peuvent être complètement décrites par leurs deux premiers moments, à savoir leur espérance (notée $E[.]$ par la suite) et leur écart-type (on parlera plus systématiquement de covariance entre les différentes variables).

Pour rappel, les erreurs d'ébauche E_k^b sont liées à une mauvaise connaissance de la physique et/ou à des erreurs numériques. Partant de l'hypothèse que la dynamique du modèle est non-biaisée, ces erreurs d'ébauche satisfont l'égalité $E[E_k^b] = 0$. Ainsi, les statistiques de E_k^b sont décrites par une matrice carrée définie positive $\mathbf{P}_k^b$ de dimension $n_x \times n_x$, appelée *matrice de covariance des erreurs d'ébauche* et définie par :

$$\mathbf{P}_k^b = E(E_k^b)(E_k^b)^T \quad (1.9)$$

Les termes diagonaux de $\mathbf{P}_k^b$ représentent la variance de l'erreur pour chacun des éléments de E_k^b alors que les autres termes sont les covariances des erreurs des éléments deux à deux. Ces covariances représentent les corrélations entre les erreurs des éléments du vecteur de contrôle et permettent le transfert d'information d'une variable de contrôle à l'autre lors des étapes d'assimilation. Il advient que

$$E_k^b \sim N(0, \mathbf{P}_k^b) \quad (1.10)$$

où N représente la loi de probabilité gaussienne/normale.

Les erreurs d'observations E_k^o sont aussi supposées non-biaisées, c'est-à-dire de moyenne nulle ($E[E_k^o] = 0$). En général, un traitement doit préalablement leur être appliqué pour enlever le biais. Ainsi, similairement aux erreurs d'ébauche, le modèle d'erreur des observations est aussi représenté par une matrice de covariance d'erreur carrée symétrique définie positive, de dimension n_y , dénotée $\mathbf{R}_k$, et définie par :

$$\mathbf{R}_k = E(E_k^o)(E_k^o)^T \quad (1.11)$$

De la même façon que dans l'Eq. 1.10 :

$$E_k^o \sim N(0, \mathbf{R}_k) \quad (1.12)$$

Pour les méthodes présentées ici, on suppose que les observations ont des erreurs non-corrélées en espace et en temps (mais il existe d'autres méthodes, non-développées ici, qui ne sont pas basées

sur cette hypothèse). Ceci implique que la *matrice de covariance des erreurs d'observations* $\mathbf{R}_k$ est diagonale, chaque élément de la diagonale étant la variance de l'erreur associée à chaque observation.

L'opérateur modèle $M_{[k-1,k]}$ introduit inévitablement des erreurs. Ces erreurs sont indépendantes des erreurs d'observations et d'ébauche. Elles sont liées à l'incapacité du modèle à couvrir l'ensemble des échelles spatiales et temporelles, aux imprécisions dans la description de la physique, voir Eq. 1.3,1.5. Ces erreurs modèle E_k^M sont aussi prises en compte dans le schéma d'assimilation au moyen d'une *matrice de covariance d'erreurs modèle* notée $\mathbf{Q}_k^M$, telle que $E_k^M \sim N(0, \mathbf{Q}_k^M)$. Dans la plupart des configurations, le modèle est supposé parfait, i.e. $\mathbf{Q}_k^M = 0$.

Pour résumer, la résolution d'un problème inverse par l'AD supposant un modèle d'erreur gaussien requiert les composants suivantes :

- des observations du système physique $\mathbf{y}_k^o$ et leurs erreurs associées $(E_k^o, \mathbf{R}_k)$,
- une première estimation a priori du vecteur de contrôle ; l'ébauche $\mathbf{x}_k^b$ et ses erreurs associées $(E_k^b, \mathbf{P}_k^b)$,
- un opérateur modèle $M_{[k-1,k]}$ décrivant la dynamique spatio-temporelle du système physique et ses erreurs associées $(E_k^M, \mathbf{Q}_k^M)$,
- un opérateur d'observation H_k .

1.4.3.2 Formulation du problème d'estimation

A partir d'une estimation d'ébauche des variables de contrôle $\mathbf{x}_k^b$ (i.e. la meilleure estimation du vecteur de contrôle vrai $\mathbf{x}_k^t$ en amont de l'assimilation) et en utilisant les observations disponibles $\mathbf{y}_k^o$ l'AD identifie la meilleure estimation du vecteur de contrôle vrai $\mathbf{x}_k^t$ satisfaisant :

$$\begin{aligned} \mathbf{x}_k^b &= \mathbf{x}_k^t + E_k^b \\ \mathbf{y}_k^o &= H_k(\mathbf{x}_k^t) + E_k^o, \end{aligned} \quad (1.13)$$

et minimisant la distance entre l'estimation modèle et les observations, calculée via le vecteur d'innovation $\mathbf{d}_k^b = \mathbf{y}_k^o - H_k(\mathbf{x}_k^b)$. La rétroaction de l'AD consiste à retranscrire les écarts entre ébauche et observation dans l'espace des observations (i.e. le vecteur d'innovation) en une correction dans l'espace de contrôle, pondérée par les erreurs d'ébauche et d'observation. Ainsi, une estimation optimale du vecteur de contrôle, appelée *analyse* et notée $\mathbf{x}_k^a$ peut être formulée comme :

$$\mathbf{x}_k^a = \mathbf{x}_k^b + \delta \mathbf{x}_k^a. \quad (1.14)$$

Le vecteur de contrôle analyse $\mathbf{x}_k^a$ est aussi associé à une erreur $E_k^a = \mathbf{x}_k^a - \mathbf{x}_k^t$ et la *matrice de covariance des erreurs d'analyse* $\mathbf{P}_k^a$ telle que $E_k^a \sim N(0, \mathbf{P}_k^a)$.

La compréhension complète du système physique avec certitude est inaccessible. Cependant, l'analyse est suffisamment correcte pour être capable de fournir une bonne représentation du comportement du système sans être parfaite. L'AD ne peut pas directement réduire l'erreur de représentation mais elle peut cependant réduire les erreurs d'ébauche. Pour tout algorithme d'AD, l'optimalité de la méthode implique que les incertitudes du vecteur de contrôle soient réduites avec l'élaboration de l'analyse, signifiant que $E_k^a < E_k^b$. La distance aux observations doit aussi être réduite. En d'autres termes, la norme du vecteur d'innovation d'analyse $\mathbf{d}_k^a = \mathbf{y}_k^o - H_k(\mathbf{x}_k^a)$ doit être inférieure à celle du vecteur d'innovation d'ébauche $\mathbf{d}_k^b$. D'autres tests de performances peuvent être menés pour vérifier la consistance des matrices de covariance d'erreurs et, si nécessaire, pour les ajuster pour les cycles d'assimilation suivants (Desroziers et al., 2005; Chapnik et al., 2006).

1.4.4 Formule de Bayes et formulation générale d'un problème d'assimilation de données

L'objectif de cette section est de présenter le cadre d'estimation conceptuel général de l'AD servant de base à une grande majorité des techniques d'assimilation de données : approches variationnelles (Sasaki, 1970; Talagrand and Courtier, 1987), filtres de Kalman (Kalman, 1960) et ses extensions (Evensen, 1994b) ainsi que les filtres à particules (Doucet et al., 2000).

On suppose que les observations sont disponibles régulièrement, sur des pas de temps discrets d'un intervalle de temps $[0, K]$. La solution optimale du problème inverse est déduite de la situation du vecteur de contrôle $\mathbf{x}_k$ au temps k étant donné ses situations passées $\{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{k-1}\}$, ainsi que l'historique des observations $\{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{k-1}\}$. Cette configuration constitue la formulation de Bayes.

1.4.4.1 Formule de Bayes

On suppose que l'incertitude de vecteur de contrôle d'ébauche $\mathbf{x}_k^b$ s'exprime comme la PDF d'ébauche $p^b(\mathbf{x}_k)$ (Lorenç, 1986). Elle représente la probabilité que les variables de contrôle d'ébauche représentent le vecteur de contrôle vrai $\mathbf{x}_k^t$. Les observations $\mathbf{y}_k^o$ ont alors pour PDF associée $p(\mathbf{y}_k^o)$. Dans le contexte de la formule de Bayes, on suppose que la suite de variables aléatoires est une chaîne de Markov discrète vérifiant les propriétés suivantes :

- La PDF du vecteur de contrôle $\mathbf{x}_k$ au temps k ne dépend que de sa valeur à l'instant précédent $(k - 1)$, i.e. :

$$p(\mathbf{x}_k | \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{k-1}) = p(\mathbf{x}_k | \mathbf{x}_{k-1}). \quad (1.15)$$

- La PDF des observations $\mathbf{y}_k^o$ est un processus Markovien qui ne dépend que du contrôle au même pas de temps, i.e. :

$$p(\mathbf{y}_k^o | \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k) = p(\mathbf{y}_k^o | \mathbf{x}_k). \quad (1.16)$$

La PDF $p(\mathbf{x}_k | \mathbf{x}_{k-1})$ représente l'évolution temporelle probable du vecteur de contrôle de l'instant $(k - 1)$ à l'instant k et correspond ainsi à la PDF d'ébauche telle que :

$$p^b(\mathbf{x}_k) = p(\mathbf{x}_k | \mathbf{x}_{k-1}). \quad (1.17)$$

Pour déterminer un bon estimateur du vecteur de contrôle vrai, le théorème de Bayes donne une formule de la PDF de l'analyse $p^a(\mathbf{x}_k)$. Cette analyse est construite en utilisant toutes les informations disponibles (incluant l'ébauche et les observations) et vérifie :

$$p^a(\mathbf{x}_k) = p(\mathbf{x}_k | \mathbf{y}_k^o) = \frac{p^b(\mathbf{x}_k) p(\mathbf{y}_k^o | \mathbf{x}_k)}{p(\mathbf{y}_k^o)}. \quad (1.18)$$

Eq. 1.18 est obtenue en utilisant la théorie de Bayes, i.e.

$$p(\mathbf{x}_k \cap \mathbf{y}_k^o) = p^b(\mathbf{x}_k) p(\mathbf{y}_k^o | \mathbf{x}_k) = p(\mathbf{y}_k^o) p(\mathbf{x}_k | \mathbf{y}_k^o).$$

Elle représente la probabilité que l'estimation des variables de contrôle soit le vecteur de contrôle vrai $\mathbf{x}_k^t$ sachant que les observations de l'état du modèle sont $\mathbf{y}_k^o$ au temps k . Finalement, Eq. 1.18 peut être réécrite sous la formule générale :

$$p^a(\mathbf{x}_k) = p(\mathbf{x}_k | \mathbf{y}_k^o) \propto \frac{p^b(\mathbf{x}_k) p(\mathbf{y}_k^o | \mathbf{x}_k)}{p(\mathbf{y}_k^o)}. \quad (1.19)$$

1.4.4.2 Formulation de l'assimilation de données comme problème d'optimisation

Conceptuellement, le problème d'AD recherche la PDF $p^a(\mathbf{x}_k)$. Cependant, l'expression explicite de cette PDF s'avère généralement difficile à déterminer, à l'exception de quelques cas restrictifs comme avec des modèles linéaires ou des modèles d'erreurs gaussiens. Alors, cette PDF est généralement approximée par ses moments statistiques.

Maintenant, on suppose justement que le vecteur de contrôle d'ébauche $\mathbf{x}_k^b$ et les observations $\mathbf{y}_k^o$ suivent des lois gaussiennes ; la statistique des erreurs est donc décrite par des matrices de covariance. La PDF d'ébauche s'écrit comme

$$p^b(\mathbf{x}_k) \propto \exp -\frac{1}{2} (\mathbf{x}_k - \mathbf{x}_k^b)^T (\mathbf{P}_k^b)^{-1} (\mathbf{x}_k - \mathbf{x}_k^b) . \quad (1.20)$$

La PDF des observations conditionnées par le contrôle s'écrit similairement comme

$$p(\mathbf{y}_k^o | \mathbf{x}_k) \propto \exp -\frac{1}{2} (H_k(\mathbf{x}_k) - \mathbf{y}_k^o)^T (\mathbf{R}_k)^{-1} (H_k(\mathbf{x}_k) - \mathbf{y}_k^o) . \quad (1.21)$$

Sous l'hypothèse des erreurs gaussiennes, Eq. 1.19 devient

$$p^a(\mathbf{x}_k) \propto \exp -\frac{1}{2} (\mathbf{x}_k - \mathbf{x}_k^b)^T (\mathbf{P}_k^b)^{-1} (\mathbf{x}_k - \mathbf{x}_k^b) - \frac{1}{2} (H_k(\mathbf{x}_k) - \mathbf{y}_k^o)^T (\mathbf{R}_k)^{-1} (H_k(\mathbf{x}_k) - \mathbf{y}_k^o) . \quad (1.22)$$

Par la suite, on appelle $\mathcal{J}$ la fonction de $\mathbf{x}_k$ telle que

$$\mathcal{J}(\mathbf{x}_k) = -\ln [p^a(\mathbf{x}_k)] . \quad (1.23)$$

On recherche alors la valeur de $\mathbf{x}_k$ qui maximise la PDF $p^a(\mathbf{x}_k)$, ce qui équivaut au problème de minimisation suivant :

$$\max_{\mathbf{x}_k \in \mathbb{R}^{n_x}} p^a(\mathbf{x}_k) \longleftrightarrow \min_{\mathbf{x}_k \in \mathbb{R}^{n_x}} \{-\ln [p^a(\mathbf{x}_k)]\} = \min \mathcal{J}(\mathbf{x}_k) , \quad (1.24)$$

avec $\mathcal{J}$ la fonction coût du problème d'optimisation :

$$\mathcal{J}(\mathbf{x}_k) = \frac{1}{2} (\mathbf{x}_k - \mathbf{x}_k^b)^T (\mathbf{P}_k^b)^{-1} (\mathbf{x}_k - \mathbf{x}_k^b) + \frac{1}{2} (H_k(\mathbf{x}_k) - \mathbf{y}_k^o)^T (\mathbf{R}_k)^{-1} (H_k(\mathbf{x}_k) - \mathbf{y}_k^o) . \quad (1.25)$$

Cette fonction coût constitue la base d'une majorité d'algorithmes d'AD et sera plus détaillée en Section 1.4.7.

Pour résoudre le problème de minimisation, présenté en Eq. 1.24, deux types de méthodes peuvent être appliquées : les méthodes dites de *filtrage* ou les méthodes dites *variationnelles*. Il s'agit toujours d'estimer la PDF du vecteur de contrôle analyse $p^a(\mathbf{x}_k)$. Cependant, les méthodes estiment des moments caractéristiques de la PDF différents. Les méthodes variationnelles cherchent à déterminer le mode (ou valeur dominante) de la PDF, c'est-à-dire la valeur de $\mathbf{x}_k$ ayant la plus forte probabilité. Quant aux méthodes de filtrage, elles sont dédiées à l'estimation de la moyenne et de la variance (covariances) de la PDF. Il est cependant important de préciser que dans les cas de modèles linéaires et de modèles d'erreur gaussiennes, les solutions des deux méthodes sont identiques (Jazwinski, 1970; Lorenc, 1986; Cohn, 1997; Mirouze, 2010). Il est bon de mentionner qu'aucune technique particulière n'a été identifiée comme le choix idéal pour la résolution d'un problème inverse d'AD général. Ce

choix dépend énormément du système physique en jeu et ainsi l'AD ne peut pas être dissociée d'un rigoureux travail de compréhension du système et de ses sources d'incertitudes.

La minimisation directe de la fonction coût $\mathcal{J}$ définie en Eq. 1.25 est associée aux approches variationnelles d'AD. Plusieurs méthodes existent dans la littérature - 3DVAR (Parrish and Derber, 1992; Fisher, 1998; Gauthier et al., 1999), 4DVAR (Talagrand and Courtier, 1987), et leurs extensions incrémentales (Courtier et al., 1994) - mais elles ne seront pas utilisées dans ces travaux de thèse, qui se sont focalisés sur les méthodes de filtrage détaillées dans la suite.

1.4.5 Le Filtre de Kalman et ses extensions

1.4.5.1 Le Filtre de Kalman (KF) : une solution analytique au problème d'optimisation bayésien

Le problème bayésien d'optimisation présenté dans les Eq. 1.22 et 1.24 possède une solution analytique pour $\mathbf{x}_k^a$ sous les hypothèses que les opérateurs modèle $\mathbf{M}_{[k-1,k]}$ et d'observation H_k soient linéaires - notés alors $\mathbf{M}_{[k-1,k]}$ et H_k - et que les erreurs soient non-biaisées et gaussiennes. Sous ces hypothèses, la solution analytique telle que le gradient de la fonction coût $\mathcal{J}$ soit nul peut être formulée comme :

$$\mathbf{x}_k^a = \mathbf{x}_k^b + \mathbf{P}_k^b \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^b \mathbf{H}_k^T + \mathbf{R}_k)^{-1} (\mathbf{y}_k^o - H_k \mathbf{x}_k^b). \quad (1.26)$$

L'opérateur $\mathbf{H}_k$, et respectivement $\mathbf{M}_{[k-1,k]}$, est l'opérateur linéaire-tangent de H_k , $\mathbf{M}_{[k-1,k]}$ respectivement. Ces opérateurs sont formellement définis par :

$$\mathbf{H}_k = \frac{\partial H_k}{\partial \mathbf{x}_{k-1}} \Big|_{\mathbf{x}_{k-1} = \mathbf{x}_{k-1}^b}, \quad \mathbf{M}_{[k-1,k]} = \frac{\partial \mathbf{M}_{[k-1,k]}}{\partial \mathbf{x}_{k-1}} \Big|_{\mathbf{x}_{k-1} = \mathbf{x}_{k-1}^b}. \quad (1.27)$$

Ils représentent les variations des opérateurs H_k et $\mathbf{M}_{[k-1,k]}$ au voisinage de $\mathbf{x}_{k-1}$. Cependant, quand les opérateurs sont linéaires, il advient :

$$\mathbf{H}_k = H_k, \quad \mathbf{M}_{[k-1,k]} = \mathbf{M}_{[k-1,k]}.$$

En revenant sur l'Eq. 1.26, il apparaît que $\delta \mathbf{x}_k^a = \mathbf{K}_k (\mathbf{y}_k^o - H_k \mathbf{x}_k^b)$ avec $\mathbf{K}_k$ la matrice de gain de Kalman (explicitée en Eq. 1.33). La formule d'analyse Eq. 1.26 est la base du filtre de Kalman. Le KF permet aussi de calculer explicitement la matrice de covariance des erreurs d'analyse $\mathbf{P}_k^a$. En effet, sous les hypothèses de linéarité et de gaussianité des erreurs, il peut être montré que la PDF analyse $p^a(\mathbf{x}_k)$ est aussi gaussienne et s'exprime comme

$$p^a(\mathbf{x}_k) \propto \exp \left[-\frac{1}{2} (\mathbf{x}_k - \mathbf{x}_k^a)^T (\mathbf{P}_k^a)^{-1} (\mathbf{x}_k - \mathbf{x}_k^a) \right]. \quad (1.28)$$

Ceci implique que le KF agit directement sur la covariance des erreurs sur le vecteur de contrôle $\mathbf{x}_k$ pour produire une représentation exacte de la PDF d'analyse $p^a(\mathbf{x}_k)$; $\mathbf{P}_k^a$ est une estimation de l'incertitude dans l'analyse calculée par le KF. De plus, le KF propage l'information d'un temps d'observation au suivant; cette propagation est sujette aux possibles incertitudes de la dynamique du modèle mais mène à une actualisation du vecteur de contrôle quand de nouvelles observations sont disponibles. Ainsi, le KF se divise en deux étapes : une étape de prédiction et une étape d'analyse, voir Figure 1.20.

L'étape de prédiction consiste à intégrer le modèle, à partir de l'analyse du cycle d'assimilation précédent $\mathbf{x}_{k-1}^a$, et génère le vecteur contrôle d'ébauche du cycle suivant $\mathbf{x}_k^b$. La matrice de covariance

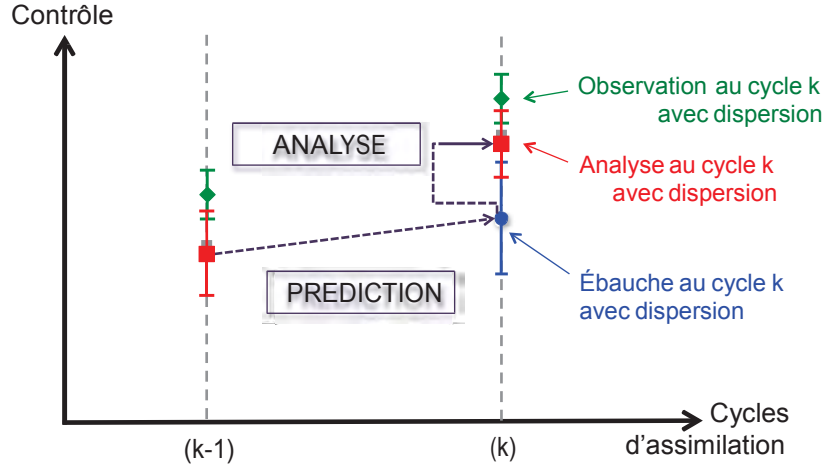


Figure 1.20 – Schéma de principe d'un cycle d'assimilation avec le Filtre de Kalman classique. L'étape de *prédiction* propage le contrôle jusqu'au temps d'observation et l'étape d'*analyse* corrige le contrôle en fonction de l'ébauche et de l'observation.

des erreurs d'ébauche $\mathbf{P}_k^b$ est aussi obtenue via la propagation de la matrice de covariance des erreurs d'analyse du cycle d'assimilation précédent $\mathbf{P}_{k-1}^a$. Ainsi, l'étape de prédiction peut être résumée au système d'équations suivant :

$$\mathbf{x}_k^b = \mathbf{M}_{[k-1,k]} \mathbf{x}_{k-1}^a, \quad (1.29)$$

$$\mathbf{P}_k^b = \mathbf{M}_{[k-1,k]} \mathbf{P}_{k-1}^a \mathbf{M}_{[k-1,k]}^T. \quad (1.30)$$

Si les erreurs modèle $\mathbf{E}_k^M$ sont prises en compte, l'étape de prédiction pour la matrice de covariance d'erreur est ajustée en tenant compte de la matrice de covariance des erreurs modèle $\mathbf{Q}_k$ tel que l'Eq. 1.30 devienne :

$$\mathbf{P}_k^b = \mathbf{M}_{[k-1,k]} \mathbf{P}_{k-1}^a \mathbf{M}_{[k-1,k]}^T + \mathbf{Q}_k. \quad (1.31)$$

A partir de l'information d'ébauche $(\mathbf{x}_k^b, \mathbf{P}_k^b)$ et l'information disponible dans les observations $\mathbf{y}_k^o$ au temps k , les équations d'analyse du KF se lisent :

$$\mathbf{x}_k^a = \mathbf{x}_k^b + \mathbf{K}_k (\mathbf{y}_k^o - \mathbf{H}_k \mathbf{x}_k^b), \quad (1.32)$$

$$\mathbf{K}_k = \mathbf{P}_k^b \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^b \mathbf{H}_k^T + \mathbf{R}_k)^{-1}, \quad (1.33)$$

$$\mathbf{P}_k^a = (\mathbf{I}_{n_x} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^b. \quad (1.34)$$

L'Eq. 1.32 montre que la correction du vecteur de contrôle $\mathbf{x}_k^b$ est proportionnelle au vecteur d'innovation $\mathbf{d}_k^b = \mathbf{y}_k^o - \mathbf{H}_k \mathbf{x}_k^b$ avec l'importance de la correction contrôlée par la matrice de gain $\mathbf{K}_k$. Cette matrice de gain, définie en Eq. 1.33 tient compte de la sensibilité des observations à des variations sur les variables de contrôle, ce qui permet de corriger le vecteur de contrôle à travers l'opérateur d'observations, tout en prenant en compte l'influence des erreurs d'ébauche et d'observations, respectivement $\mathbf{P}_k^b$ et $\mathbf{R}_k$. Enfin, l'Eq. 1.34 donne une estimation des statistiques des erreurs d'analyse avec $\mathbf{P}_k^a$ et assure que la variance d'erreur d'analyse est inférieure à la variance d'erreur d'ébauche, traduisant le fait qu'on devrait avoir une meilleure estimation.

Le KF reste la méthode d'assimilation de données séquentielle la plus utilisée, même si l'algorithme ne s'applique en théorie qu'aux modèles numériques linéaires et aux modèles d'erreurs gaussiens. Cependant, dans la plupart des situations, ces hypothèses ne sont pas vérifiées. La majorité des modèles présentent une dynamique non-linéaire et des incertitudes ne suivant pas des lois gaussiennes. Il est alors important d'étudier l'optimalité des solutions quand ces hypothèses ne sont plus valables. Dans la suite, on évoque les extensions du KF traitant les problèmes non-linéaires.

1.4.5.2 Extensions non-linéaires du filtre de Kalman

Le filtre de Kalman Étendu (EKF) est une première extension du KF adaptée pour les problèmes non-linéaires. L'analyse de l'EKF se base sur la linéarisation locale des opérateurs modèle $\mathbf{M}_{[k-1,k]}$ et d'observation $\mathbf{H}_k$ et dépend, de ce fait, de la dynamique de l'opérateur modèle $\mathbf{M}_{[k-1,k]}$ (Bouttier and Courtier, 1999). Les opérateurs linéaires-tangents résultants, $\mathbf{H}_k$ et $\mathbf{M}_{[k-1,k]}$ (Eq. 1.27), ne sont alors valides qu'au voisinage du vecteur de contrôle de référence (généralement défini comme le vecteur de contrôle d'ébauche $\mathbf{x}_k^b$ du cycle d'assimilation courant $[k-1, k]$). L'étape de prédiction de l'EKF est basée sur les équations suivantes :

$$\mathbf{x}_k^b = \mathbf{M}_{[k-1,k]}(\mathbf{x}_{k-1}^a), \quad (1.35)$$

$$\mathbf{P}_k^b = \mathbf{M}_{[k-1,k]} \mathbf{P}_k^a \mathbf{M}_{[k-1,k]}^T, \quad (1.36)$$

où l'Eq. 1.35 implique une intégration du modèle non-linéaire entre les instants $(k-1)$ et k , à partir du vecteur de contrôle analyse au temps d'assimilation $(k-1)$. Pour l'étape d'analyse, celle-ci se définit par :

$$\mathbf{x}_k^a = \mathbf{x}_k^b + \mathbf{K}_k (\mathbf{y}_k^o - \mathbf{H}(\mathbf{x}_k^b)) \quad (1.37)$$

$$\mathbf{K}_k = \mathbf{P}_k^b \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^b \mathbf{H}_k^T + \mathbf{R}_k)^{-1}, \quad (1.38)$$

$$\mathbf{P}_k^a = (\mathbf{I}_{n_x} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^b. \quad (1.39)$$

L'Eq. 1.38 montre que le linéaire tangent de l'opérateur d'observation $\mathbf{H}_k$ est toujours impliqué dans la définition de la matrice de gain $\mathbf{K}_k$ alors que le vecteur d'innovation $\mathbf{d}_k^b = \mathbf{y}_k^o - \mathbf{H}(\mathbf{x}_k^b)$ est directement évalué en utilisant l'opérateur d'observation non-linéaire $\mathbf{H}_k$ (Eq. 1.37). Dans ce système, une équation linéarisée (donc approximative) est utilisée lors de la détermination de l'innovation et pour le calcul de la matrice de gain de Kalman $\mathbf{K}_k$. Ceci implique que l'EKF est une approximation non-linéaire du filtre Bayésien. L'EKF ne produit une analyse cohérente que si la linéarisation locale est suffisante pour décrire correctement la dynamique du modèle. Il est alors bon de noter que les limitations de l'EKF peuvent être en partie surpassées en utilisant pendant l'étape d'analyse une boucle qui actualise les opérateurs linéarisés (Thirel et al., 2010; Harader et al., 2012).

Contrairement au KF, l'EKF n'est pas un estimateur optimal du vecteur de contrôle vrai $\mathbf{x}_k^f$ du fait de l'utilisation d'opérateurs linéaire-tangent approchés. L'EKF peut parfois linéariser de manière irréaliste les opérateurs modèle et d'observations. Ces linéarisations sont basées sur des dérivées locales qui sont souvent compliquées à calculer de manière fiable (Ros and Borga, 1997). De plus, l'EKF peut rapidement diverger lors de son exécution si l'estimation d'ébauche $\mathbf{x}_k^b$ est trop éloignée du vecteur de contrôle vrai $\mathbf{x}_k^f$. En effet, si la fonction coût à minimiser pour le problème n'est pas quadratique et non strictement monotone, l'EKF peut rencontrer des minima locaux et ne peut alors pas trouver l'estimateur optimal. Pour faire face à ces difficultés, les non-linéarités de l'opérateur d'observations et de l'opérateur modèle sont partiellement gérées en utilisant une approche basée sur une boucle externe (voir Fig. 1.21).

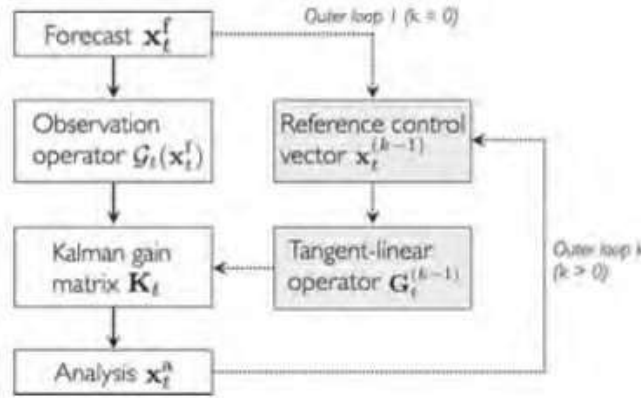


Figure 1.21 – Algorithme de fonctionnement de l’EKF utilisant l’approche de la boucle externe (Rochoux, 2014).

Le processus de boucle externe permet d’appliquer plusieurs fois l’étape d’analyse (Eq. 1.37-1.39) au sein d’un même cycle d’assimilation. A chaque nouvelle itération s , la matrice de gain $\mathbf{K}_k$ est actualisée, ce à partir d’une nouvelle estimation du linéaire-tangent d’observation $\mathbf{H}_k^{(s-1)}$. Ces différents linéaires tangents sont déterminés autour d’un vecteur de contrôle de référence différents d’une itération à l’autre : le vecteur de contrôle d’ébauche est utilisé comme vecteur de référence à la première itération - $\mathbf{x}_k^{(0)} = \mathbf{x}_k^b$ - et pour l’itération s , la référence est définie comme le vecteur d’analyse $\mathbf{x}_k^{(s-1)}$ de l’itération précédente ($s - 1$). Ce dernier ayant été obtenu via le linéaire-tangent $\mathbf{H}_k^{(s-1)}$ et le gain de Kalman $\mathbf{K}_k^{(s-1)}$ estimé à l’itération ($s - 1$), et ainsi de suite. La procédure est alors répétée jusqu’à ce que la convergence vers l’estimateur optimal $\mathbf{x}_k^a$ soit atteinte (le nombre d’itérations allant généralement de 1 à 10 ; Rochoux, 2014).

Mises à part les limites liées aux non-linéarités de l’EKF, des difficultés peuvent aussi être rencontrées lors de la construction et de la modélisation de la matrice de covariance des erreurs d’ébauche $\mathbf{P}_k^b$. Ces difficultés sont liées aux grandes tailles de ces matrices (qui sont alors lourdes à stocker au sein d’un code informatique) ou au simple fait que les erreurs ne suivent pas une loi gaussienne. De nombreuses références se concentrent sur la modélisation de $\mathbf{P}_k^b$ au début d’un cycle d’assimilation et se basent notamment sur des déformations, des transformations dans des bases d’ondelettes (Pannekoucke et al., 2008; Montmerle and Berre, 2010; Michel, 2013a,b) ou encore sur l’équation de diffusion (Weaver and Courtier, 2001); tout cela dans le but de décrire le plus justement possible des corrélations entre les variables de contrôle.

Les techniques d’assimilation de données basées sur l’utilisation d’ensembles, telles que le filtre de Kalman d’Ensemble (EnKF ; Evensen, 1994b; Houtemaker and Mitchell, 1998) proposent une alternative en déterminant stochastiquement les matrices de covariance d’erreurs. Ainsi, ces méthodes permettent de mieux tenir compte des non-linéarités de la physique, tout en évitant de stocker en mémoire les matrices d’un cycle d’assimilation à l’autre.

1.4.5.3 Le Filtre de Kalman d'Ensemble (EnKF)

Initialement proposé par Evensen (1994b) comme une alternative à l'EKF qui ne peut pas gérer les trop fortes non-linéarités des modèles dynamiques, l'EnKF est une méthode de Monte Carlo qui caractérise stochastiquement les matrices de covariance d'erreur d'ébauche $\mathbf{P}_{e,k}^b$ et de gain de Kalman $\mathbf{K}_{e,k}$ (l'indice e est pour qualifier les matrices ou vecteurs estimés stochastiquement). L'EnKF reste cependant basé sur des statistiques gaussiennes et fonctionne séquentiellement d'un temps d'observation à l'autre en appliquant successivement une étape de prédiction et une étape d'analyse comme dans le KF classique, voir Figure 1.22.

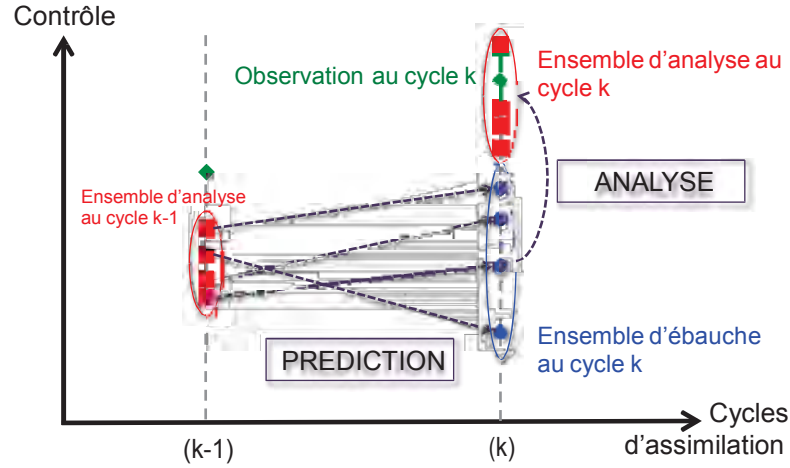


Figure 1.22 – Schéma de principe d'un cycle d'assimilation avec le Filtre de Kalman d'Ensemble (EnKF). La méthode fonctionne sur le même principe que le KF (voir Fig. (1.20)) mais propage et corrige un ensemble de vecteurs de contrôle plutôt qu'un seul. De plus, les erreurs sont estimées statistiquement à partir de l'ensemble.

1.4.5.3.a Propagation d'incertitudes

En opposition à l'EKF déterministe, l'EnKF ne requiert pas la linéarisation explicite des opérateurs $\mathbf{M}_{[k-1,k]}$ et $\mathbf{H}_k$ lors de l'étape de prédiction. La matrice de covariance des erreurs d'analyse $\mathbf{P}_{k-1}^a$ n'est pas propagée d'un cycle à l'autre pour estimer $\mathbf{P}_k^b$. L'EnKF estime plutôt la distribution des variables du vecteur de contrôle en utilisant une réunion de n_e simulations indépendantes, dont chaque simulation (appelée *membre*) est une réalisation du vecteur aléatoire $\mathbf{x}_k^b$. La réunion des n_e membres est appelée l'*ensemble d'ébauche* et s'écrit :

$$\mathbf{x}_{e,k}^b = \begin{bmatrix} \mathbf{x}_k^{b,[1]}, & \mathbf{x}_k^{b,[2]}, & \dots, & \mathbf{x}_k^{b,[n_e]} \end{bmatrix}. \quad (1.40)$$

Dans cet ensemble, le l -ème membre est un vecteur aléatoire multi-variés de dimension n_x tel que $\mathbf{x}_k^{b,[l]} \in \mathbb{R}^{n_x}$. Ainsi, au lieu d'intégrer les matrices de covariance d'erreur à partir de celles des cycles précédents, l'EnKF calcule $\mathbf{P}_{e,k}^b$ à partir de la distribution des vecteurs de contrôle de l'ensemble.

La propagation de la statistique de l'erreur du temps d'observation $(k-1)$ au temps k est simplement réalisée en propageant chaque membre l de l'ensemble en utilisant le modèle dynamique non-linéaire tel que

$$\mathbf{x}_k^{b,[l]} = \mathbf{M}_{[k-1,k]}(\mathbf{x}_{k-1}^{a,[l]}), \quad l = 1 \dots n_e. \quad (1.41)$$

Chaque membre de l'ensemble représente une réalisation particulière des trajectoires possibles du modèle en utilisant le modèle dynamique (potentiellement) non-linéaire. Ceci signifie que l'ensemble d'ébauche inclut certaines des non-linéarités de la dynamique qui sont négligées par l'EKF. Il faut rappeler que les perturbations aléatoires doivent d'abord être introduites dans les paramètres d'entrée ou les forçages externes pour tenir compte des incertitudes présentes dans le système. D'un point de vue stochastique, ceci est justifié par la nécessité de générer un échantillon représentatif de tous les comportements possibles du système. Idéalement, le vecteur de contrôle vrai $\mathbf{x}_k^t$ doit être contenu dans l'ensemble d'ébauche pour garantir le bon fonctionnement de l'EnKF. La génération de l'ensemble d'ébauche est donc une étape cruciale de l'EnKF.

La matrice de covariance des erreurs d'ébauche $\mathbf{P}_{e,k}^b$ est ensuite calculée à partir de l'ensemble, ce qui signifie qu'elle est remplacée par la matrice de covariance des erreurs d'ébauche empirique $\mathbf{P}_{e,k}^b$ définie par :

$$\mathbf{P}_{e,k}^b = \frac{1}{n_e - 1} \sum_{l=1}^{n_e} (\mathbf{x}_k^{b,[l]} - \overline{\mathbf{x}_k^b}) (\mathbf{x}_k^{b,[l]} - \overline{\mathbf{x}_k^b})^T, \quad (1.42)$$

où les barres horizontales au-dessus des vecteurs représente leur moyenne empirique sur l'ensemble défini comme

$$\overline{\mathbf{x}_k^b} = \frac{1}{n_e} \sum_{l=1}^{n_e} \mathbf{x}_k^{b,[l]}. \quad (1.43)$$

L'Eq. 1.42 implique que la moyenne de l'ensemble d'ébauche est utilisée, à la place du vecteur de contrôle vrai, comme le meilleur estimateur. De plus, la dispersion de l'ensemble est interprétée comme l'écart-type de l'erreur autour de ce meilleur estimateur. La qualité de l'estimation de $\mathbf{P}_k^b$ dépend alors de la taille de l'ensemble n_e et de la représentation de toutes les incertitudes dans l'ensemble.

L'EnKF a d'abord été testé avec succès sur des expériences jumelles (Evensen, 1994a), puis sur des expériences plus réalistes où Evensen and Leeuwen (1996) utilisent des données altimétriques Géosat. Lors des étapes d'analyse de ces expériences, chaque membre est actualisé indépendamment des autres membres selon l'Eq. 1.37 en utilisant l'opérateur modèle généralisé non-linéaire H_k , et en utilisant la matrice de gain de Kalman calculée stochastiquement à partir de l'ensemble d'ébauche. Chaque membre satisfait alors :

$$\mathbf{x}_k^{a,[l]} = \mathbf{x}_k^{b,[l]} + \mathbf{K}_{e,k} (\mathbf{y}_k^o - H_k(\mathbf{x}_k^{b,[l]})) \hat{\mathbf{I}}, \quad l = 1 \dots n_e, \quad (1.44)$$

avec

$$\mathbf{K}_{e,k} = \mathbf{H}_k^T \mathbf{P}_k^b (\mathbf{H}_k^T \mathbf{P}_k^b + \mathbf{R}_k)^{-1}. \quad (1.45)$$

1.4.5.3.b Ensemble d'observations

Par la suite, Burgers et al. (1998) a montré que pour que le schéma d'analyse soit consistant, il était nécessaire que les observations soient aussi considérées comme des variables aléatoires. Ceci est réalisé en ajoutant au vecteur d'observation original $\mathbf{y}_k^o$ une perturbation aléatoire, différente pour chaque membre, suivant une loi normale d'espérance nulle et de variance égale à la variance des erreurs d'observations. Ceci permet alors d'éviter une sous-estimation de la variance des erreurs sur les variables de contrôle et de prévenir l'appauvrissement de l'échantillon de contrôle qui arrive quand chaque membre est corrigé avec le même vecteur d'observation. Ce processus (appelé *randomization of data* en anglais) mène à la construction d'un ensemble de n_e réalisations d'observations avec une covariance égale à la matrice de covariance des erreurs d'observations $\mathbf{R}_k$, tel que :

$$\mathbf{y}_k^{o,[l]} = \mathbf{y}_k^o + \boldsymbol{\xi}_k^{o,[l]}, \quad \boldsymbol{\xi}_k^{o,[l]} \sim N(0, \mathbf{R}_k) \quad l = 1 \dots n_e, \quad (1.46)$$

où les perturbations des observations $\xi_k^{o,[l]}$ sont indépendantes des erreurs d'ébauche E_k^b . L'ensemble des observations est alors défini par :

$$\mathbf{y}_{e,k}^o = [\mathbf{y}_k^{o,[1]}, \mathbf{y}_k^{o,[2]}, \dots, \mathbf{y}_k^{o,[n_e]}] . \quad (1.47)$$

Dans le contexte de l'utilisation d'ensembles, la matrice de covariance des erreurs d'observation $\mathbf{R}_k$ peut être modélisée soit par sa version empirique, c'est-à-dire :

$$\mathbf{R}_{e,k} = \frac{1}{n_e - 1} \sum_{l=1}^{n_e} (\xi_k^{o,[l]} - \bar{\xi}_k^o) (\xi_k^{o,[l]} - \bar{\xi}_k^o)^T ,$$

soit en utilisant les écart-type des erreurs d'observations, c'est-à-dire Eq. 1.11. Cette matrice doit être inversible. Utiliser la définition empirique de $\mathbf{R}_k$ peut amener à une matrice non-diagonale, particulièrement quand la taille de l'ensemble n_e est petite, ce qui conduit à une matrice $\mathbf{R}_k$ coûteuse à inverser, voire non-inversible. Ainsi, la matrice de covariance des erreurs d'observations est généralement modélisée en considérant des erreurs non-corrélées, indépendantes de la taille de l'ensemble. Ceci assure une matrice diagonale, définie, positive, où chaque élément diagonal est égal à la variance de l'erreur d'observation.

1.4.5.3.c Formulation de l'ensemble d'analyse

On peut réécrire l'Eq. 1.44 de la façon suivante :

$$\mathbf{x}_k^{a,[l]} = \mathbf{x}_k^{b,[l]} + \mathbf{K}_{e,k} (\mathbf{y}_k^{o,[l]} - H_k(\mathbf{x}_k^{b,[l]})) , \quad l = 1 \dots n_e . \quad (1.48)$$

L'étape d'analyse construit alors un ensemble de vecteurs de contrôle d'analyse,

$$\mathbf{x}_{e,k}^a = [\mathbf{x}_k^{a,[1]}, \mathbf{x}_k^{a,[2]}, \dots, \mathbf{x}_k^{a,[n_e]}] , \quad (1.49)$$

avec lequel une matrice empirique de covariance d'erreur d'analyse $\mathbf{P}_{e,k}^a$ peut être reconstruite en appliquant :

$$\mathbf{P}_{e,k}^a = \frac{1}{n_e - 1} \sum_{l=1}^{n_e} (\mathbf{x}_k^{a,[l]} - \bar{\mathbf{x}}_k^a) (\mathbf{x}_k^{a,[l]} - \bar{\mathbf{x}}_k^a)^T . \quad (1.50)$$

Alternativement, il a été montré que $\mathbf{P}_{e,k}^a$ peut aussi être déterminée en appliquant :

$$\mathbf{P}_{e,k}^a = (\mathbf{I}_{n_x} - \mathbf{K}_{e,k} \mathbf{H}_k) \mathbf{P}_{e,k}^b , \quad (1.51)$$

par analogie avec l'Eq. 1.34. L'application de l'EnKF avec perturbation du vecteur d'observation a montré sa capacité à fournir des statistiques d'erreurs correctes pour l'analyse (Burgers et al., 1998; Evensen, 2003). Ainsi, l'ensemble de vecteurs de contrôle d'analyse $\mathbf{x}_{e,k}^a$ peut être propagé dans le temps via des intégrations du modèle jusqu'au temps d'assimilation suivant.

Contrairement à l'EKF, l'EnKF ne nécessite pas la propagation et le stockage des matrices de covariance d'erreurs. Elles sont reconstruites à chaque temps d'observation en se basant sur un ensemble de trajectoires modèle, fournies par un ensemble de taille n_e suffisamment large pour capter les corrélations les plus importantes entre toutes les variables de contrôle. Par ailleurs, Li and Xiu (2008) ont montré que les erreurs d'ensemble dues à l'échantillonnage de Monte Carlo dans l'EnKF peuvent être prédominantes sur les erreurs numériques et les erreurs modèles. L'algorithme est particulièrement adapté pour l'exécution parallèle des codes : les réalisations de chaque membre

peuvent être calculées indépendamment et, de ce fait, simultanément. Ainsi, le coût en temps de calcul est limité et est conditionné par le coût d'une intégration modèle.

Cependant, l'EnKF reste malgré tout un filtre gaussien comme le KF et l'EKF. Les statistiques des erreurs sont représentées à l'aide de modèles de covariance, signifiant que seulement deux moments de la PDF sont pris en compte - la moyenne et la variance - et les moments d'ordres supérieurs - l'asymétrie, l'aplatissement - sont négligés. Alors que les matrices de covariance sont des modèles parfaitement adaptés aux statistiques gaussiennes, les moments d'ordres supérieurs deviennent importants et non-négligeables si la dynamique du modèle et des observations est fortement non-linéaire.

Grâce à leur efficacité et leur flexibilité, l'EnKF et ses variantes sont devenus des méthodes d'AD très populaires, que ce soit pour l'estimation de paramètres ou pour l'estimation d'état. Cependant, même si ces algorithmes permettent de mieux gérer les non-linéarités, ils s'appuient sur des hypothèses gaussiennes pour décrire les statistiques des erreurs d'ébauche et d'observations. Les solutions issues de ces méthodes ne peuvent donc pas être considérées comme des solutions exactes du problème Bayésien. De plus, si les distributions sont fortement non-gaussiennes et si la relation entre les espaces de contrôle et d'observations est fortement non-linéaire, ces hypothèses peuvent introduire un biais systématique dans les analyses. Le Filtre à Particules (FP), ou filtre séquentiel de Monte Carlo dans la littérature (Gordon et al., 1993; Doucet et al., 2001), est une approche séquentielle et est, en théorie, parfaitement adaptée aux problèmes non-linéaires avec des statistiques non gaussiennes. L'un des avantages de cette méthode est qu'elle fournit une description complète de la PDF à partir d'un nuage de points (les *particules*), au lieu des moments statistiques habituels.

1.4.6 Filtres non-gaussiens et non-linéaires à particules

Le concept de base du FP est de décrire la PDF des variables de contrôle d'ébauche $p^b(\mathbf{x}_k)$ par un ensemble de n_e particules (équivalentes aux membres de l'EnKF). À chaque particule est associé un poids qui dépend de la distance pondérée aux observations $\mathbf{y}_k^o$, c'est-à-dire en fonction de l'incertitude sur les observations. Ensuite, la PDF d'analyse $p^a(\mathbf{x}_k)$ est construite à partir de ces poids : un poids plus important sera attribué aux particules suivant les trajectoires modèles les plus consistantes avec les observations. En augmentant le nombre de particules, on peut alors obtenir une représentation exacte de la PDF analyse, voir Figure 1.23. De nombreux algorithmes de FP ont été développés ces dix dernières années mais on se contentera ici de décrire l'algorithme de base et on présentera rapidement quelques variantes et évolutions de cet algorithme.

Les n_e particules, notées $\mathbf{x}_k^{[l]}$, $l = 1 \dots n_e$, à l'instant k , sont associées à un poids $w_k^{[l]}$ satisfaisant :

$$\sum_{l=1}^{n_e} w_k^{[l]} = 1$$

Ces particules caractérisent la PDF d'analyse sous la forme :

$$p^a(\mathbf{x}_k) = p(\mathbf{x}_k | \mathbf{y}_k^o) = \sum_{l=1}^{n_e} w_k^{[l]} \delta(\mathbf{x}_k - \mathbf{x}_k^{[l]}), \quad (1.52)$$

avec $\delta(\cdot)$ la fonction Dirac. Dans le cadre de l'AD, le FP consiste à estimer le poids $w_k^{[l]}$ de chaque particule compte tenu des observations disponibles et ce pour estimer la PDF analyse $p^a(\mathbf{x}_k)$.

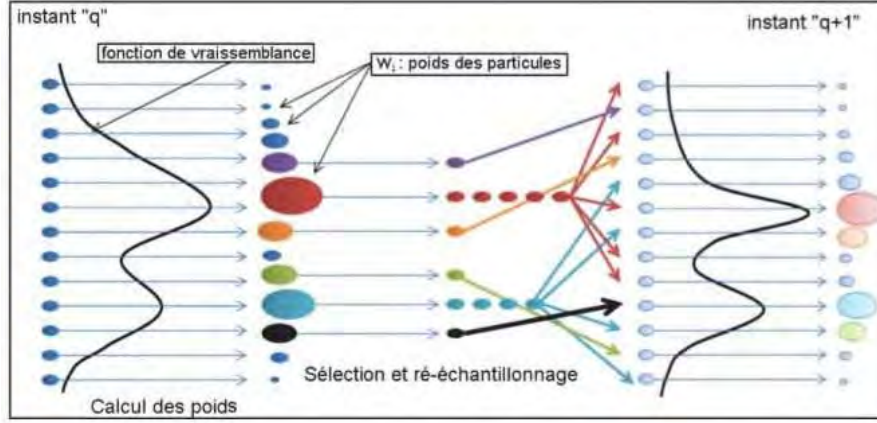


Figure 1.23 – Schéma de principe du Filtre à Particules (Mechri, 2014; Ottlé and Mahfouf, 2016). Un poids est attribué à chaque particule en fonction de leur vraisemblance avec les observations et ce poids détermine si la particule est supprimée de l'ensemble ou propagée voire dupliquée au cycle suivant.

Les premières implémentations du FP sont basées sur la méthode de l'*Échantillonnage Séquentiel par Importance* (SIS pour *Sequential Importance Sampling* en anglais ; Doucet et al., 2000). Les particules sont échantillonnées directement à partir de l'espace de contrôle et leur poids respectif est calculé selon les informations apportées par les observations. Les poids $w_k^{*,[l]}$ sont en effet calculés selon la formule :

$$w_k^{*,[l]} \propto w_{k-1}^{[l]} \frac{\left\{ p(y_k^o | x_k^{[l]}, \hat{y}_k) p(x_k^{[l]} | x_{k-1}^{[l]}, \hat{x}_{k-1}) \right\}}{\int q(x_k^{[l]} | x_{k-1}^{[l]}, y_k) dx_k^{[l]}}, \quad (1.53)$$

puis sont normalisés tels que :

$$w_k^{[l]} = \frac{w_k^{*,[l]}}{\sum_{l=1}^L w_k^{*,[l]}}. \quad (1.54)$$

La PDF $q(x_k^{[l]} | x_{k-1}^{[l]}, y_k^o, \hat{y}_k)$, appelée PDF préférentielle (importance PDF en anglais), permet de donner plus de poids aux valeurs des variables de contrôle qui ont la plus grande probabilité d'occurrence. Le choix de cette distribution est alors cruciale dans la mise en place de l'algorithme du FP. Pour plus de détails, Rochoux (2014) présente plusieurs distributions préférentielles.

Une fois les poids déterminés, une estimation de la PDF analyse peut être produite selon l'Eq. 1.52. Les poids $w_k^{[l]}$ et l'estimation de $p^a(x_k)$ sont ainsi actualisés à chaque fois qu'une nouvelle observation est disponible. En résumé, un algorithme de FP peut être décrit par les étapes suivantes (Gland, 2003) :

- **Mutation** : entre deux observations, les particules explorent l'espace d'état de manière indépendante en suivant le modèle a priori. En d'autres termes, à partir de la PDF analyse du cycle d'assimilation précédent $p^a(x_{k-1})$ et des poids corrigés des particules $(w_{k-1}^{[l]})_l$, on génère l'ensemble de particules $(x_k^{[l]})_l$ utilisé pour qualifier la nouvelle PDF analyse. Différentes méthodes existent pour générer l'ensemble des particules mais seront discutées dans le paragraphe suivant.

- **Pondération** : lorsque qu'une nouvelle observation est disponible, on évalue l'adéquation de chaque particule avec cette observation à partir d'une fonction de vraisemblance dépendant de l'écart entre l'observation réelle et l'observation attendue si le système occupait l'état d'ébauche, i.e l'innovation définie en Eq. 1.8. Un nouveau poids ($w_k^{[i]}$), pour chaque particule est alors déterminé.

Le FP basé sur la méthode SIS peut être vu comme une propagation séquentielle des poids et des particules d'un temps d'observation à l'autre.

Cette première méthode possède cependant plusieurs limitations. Premièrement, on constate très vite qu'un nombre réduit voire une seule trajectoire concentre tout le poids alors que les autres trajectoires affichent un poids quasi-nul : c'est ce qu'on appelle la *dégénérescence des poids*. La solution à ce problème consiste alors à redistribuer les particules. La méthode la plus commune de ré-échantillonnage est l'*Échantillonnage avec Rééchantillonnage par Importance* (SIR pour *Sequential Importance Resampling*) introduite par Gordon et al. (1993). Cette méthode ajoute une étape de **Redistribution** qui consiste à éliminer les particules avec un poids négligeable et multiplier celles avec un poids plus important (Kitagawa, 1996).

La redistribution n'est pas à effectuer à chaque itération mais lorsqu'on constate un déséquilibre trop fort dans la répartition des poids. Gland (2003) propose plusieurs critères de redistribution possibles. Aussi, la redistribution ne consiste pas à seulement tirer aléatoirement de nouvelles particules. Des étapes de régularisation permettent de proposer des particules dans un voisinage des particules déjà existantes. On parle ainsi de régularisation par noyau ou kernel filter (Hürzeler and Künsch, 1998) ou de filtre particulaire pré-régularisé (Musso et al., 2001). Idéalement, il faudrait effectuer une redistribution guidée par la nouvelle observation qui forcerait le mécanisme de mutation à proposer des particules dans des régions de forte vraisemblance. Parmi les techniques existantes, on trouve le *prior editing* (Gordon et al., 1993), l'*auxiliary particle filter* (Pitt and Shepard, 1999) ou encore l'*annealed importance sampling* (Neal, 2001).

Après l'étape de redistribution, les descendants d'une particule ayant été multipliée occupent la même position dans l'ensemble de contrôle. Si la transition entre le contrôle au temps ($k - 1$) et celui au temps k est déterministe - typiquement quand on utilise un modèle numérique considéré parfait - les particules n'explorent qu'une partie réduite de l'espace d'état : ce phénomène est appelé *dégénérescence des positions*. Une étape de *roughening* (rugosité en anglais) (Gordon et al., 1993) peut être ajoutée. Elle consiste à ajouter du bruit dans le modèle : le filtre particulaire post-régularisé (Musso et al., 2001) ou la méthode MCMC (Markov Chain Monte Carlo) (Gilks and Berzuini, 2001) sont issus de l'ajout de cette étape.

Pour conclure rapidement sur les FP, ces algorithmes permettent ainsi une grande flexibilité quant à la forme des PDF des variables de contrôle et des observations. Ils génèrent, en théorie, des solutions plus générales au problème d'optimisation Bayésien mais restent des algorithmes coûteux en ressources et temps de calcul, notamment à cause des méthodes d'échantillonnage (de type Monte Carlo) qui peuvent devenir très lourdes avec l'augmentation de la complexité du système physique. Des algorithmes plus avancés que le SIR existent et permettent notamment de réduire le coût de calcul en réduisant le nombre de particules. Une description plus détaillée des FP peut être trouvée dans Moradkhani et al. (2005).

1.4.7 Les méthodes variationnelles

L'autre grande catégorie de méthodes d'AD sont regroupées sous l'appellation des méthodes variationnelles. Elles consistent en la minimisation directe de la fonction coût introduite dans l'Eq. 1.25.

1.4.7.1 Formulation du problème de minimisation

On rappelle l'expression de la fonction coût ici :

$$\mathcal{J}(\mathbf{x}_k) = \frac{1}{2} \mathbf{x}_k - \mathbf{x}_k^b \text{ }^T (\mathbf{P}_k^b)^{-1} \mathbf{x}_k - \mathbf{x}_k^b + \frac{1}{2} H_k(\mathbf{x}_k) - \mathbf{y}_k^o \text{ }^T (\mathbf{R}_k)^{-1} H_k(\mathbf{x}_k) - \mathbf{y}_k^o . \quad (1.55)$$

$\mathcal{J}$ mesure la différence pondérée entre un vecteur de contrôle quelconque $\mathbf{x}_k$ et le vecteur de contrôle d'ébauche d'un côté, entre les observations $\mathbf{y}_k^o$ et le vecteur de contrôle $\mathbf{x}_k$ exprimée dans l'espace des observations via l'opérateur d'observations $H_k \equiv H_k(\mathbf{x}_k)$ de l'autre. Chaque terme est pondéré par la confiance (ou certitude) associée que l'on a dans ces termes, modélisée par l'inverse des matrices de covariance d'erreurs $(\mathbf{P}_k^b)^{-1}$ et $(\mathbf{R}_k)^{-1}$. L'approche variationnelle correspond alors au problème de minimisation suivant :

$$\min_{\mathbf{x}_k \in \mathbb{R}^{n_x}} \left[\mathbf{x}_k - \mathbf{x}_k^b \right]_{(\mathbf{P}_k^b)^{-1}}^2 + \left[\mathbf{y}_k^o - H_k(\mathbf{x}_k) \right]_{(\mathbf{R}_k)^{-1}}^2 , \quad (1.56)$$

où $\|\cdot\|_{\mathbf{N}}$ représente la norme au sens de la métrique $\mathbf{N}$ soit $\|\cdot\|_{\mathbf{N}}^2 = (\cdot)^T \mathbf{N}(\cdot)$.

Le problème tel qu'il est défini dans l'Eq. 1.56 implique que l'analyse doit être à la fois consistante avec l'ébauche et avec l'observation, étant donné leurs incertitudes respectives. La minimisation de la fonction coût $\mathcal{J}$ mène à un vecteur de contrôle d'analyse $\mathbf{x}_k^a$ satisfaisant :

$$\nabla \mathcal{J}(\mathbf{x}_k^a) = 0. \quad (1.57)$$

Ainsi, la solution minimisant $\mathcal{J}$ maximise la PDF analyse $p^a(\cdot)$.

Dans cette approche, comme pour les méthodes de type filtrage, les statistiques des erreurs sont décrites uniquement par leurs covariance. En statistiques gaussienne, ces moments sont suffisants pour représenter exactement les PDFs. De plus, il n'est pas nécessaire d'avoir des opérateurs linéaires : les opérateurs généralisés d'observation H_k , et implicitement de modèle $\mathbf{M}_{[k-1,k]}$, sont directement utilisés dans la formulation du problème.

1.4.7.2 Méthodes séquentielles 3D-VAR et 3D-VAR incrémentale

Les méthodes variationnelles regroupent toute une variété d'algorithmes, les plus connus sont les méthodes 3D-VAR, 4D-VAR et leur homologue incrémental.

En supposant que les erreurs modèle E_k^M (Eq. 1.5) sont négligeables sur la période d'intégration $[k-1, k]$, la fonction coût est exactement celle présentée en Eq. 1.25, i.e :

$$\mathcal{J}_{3D}(\mathbf{x}_k) = \frac{1}{2} \mathbf{x}_k - \mathbf{x}_k^b \text{ }^T (\mathbf{P}_k^b)^{-1} \mathbf{x}_k - \mathbf{x}_k^b + \frac{1}{2} H_k(\mathbf{x}_k) - \mathbf{y}_k^o \text{ }^T (\mathbf{R}_k)^{-1} H_k(\mathbf{x}_k) - \mathbf{y}_k^o . \quad (1.58)$$

La physique du modèle, impliquée via le terme $H_k(\mathbf{x}_k)$, constitue une forte contrainte du problème de minimisation. En effet, le fait de considérer le modèle comme parfait (E_k^M négligeable ou $\mathbf{Q}_k = 0$)

impose que les trajectoires du modèle physique soient complètement déterministes et limite le champ des trajectoires possibles. Il existe ainsi une version de J_{3D} pour laquelle les erreurs modèle sont prises en compte. La nouvelle fonction-coût, $J_{3D,w}$ constitue le problème faiblement-constraint ("w" pour *weak*, faible en anglais) et s'écrit :

$$J_{3D,w}(\mathbf{x}_k) = J_{3D}(\mathbf{x}_k) + \frac{1}{2} \xi^T (\mathbf{Q}_k)^{-1} \xi, \quad (1.59)$$

où ξ est l'estimation de l'erreur modèle au temps k . Cette implémentation est intensément étudiée actuellement par la communauté de la prévision météorologique et ne sera pas plus développée ici. Dans le cadre de cette présentation générale des techniques d'AD, on considérera toujours le modèle d'évolution comme parfait.

La minimisation de la fonction J_{3D} passe par l'utilisation de son gradient ∇J_{3D} :

$$\nabla J_{3D}(\mathbf{x}_k) = (\mathbf{P}_k^b)^{-1} \mathbf{x}_k - \mathbf{x}_k^b - \mathbf{H}_k^T (\mathbf{R}_k)^{-1} \mathbf{y}_k^o - H_k(\mathbf{x}_k), \quad (1.60)$$

avec $\mathbf{H}_k$ l'opérateur linéaire tangent défini à la Section (1.4.5.2), Eq. 1.27 Le vecteur de contrôle analyse $\mathbf{x}_k^a$ est ensuite déterminé en résolvant $\nabla J_{3D}(\mathbf{x}_k^a) = 0$.

On constate que J_{3D} n'est pas quadratique par rapport au vecteur de contrôle $\mathbf{x}_k$ à cause notamment de la non-linéarité de l'opérateur d'observation $\mathbf{H}_k$. De ce fait, J_{3D} peut contenir plusieurs minima locaux, sa minimisation devient une tâche beaucoup plus délicate et surtout très coûteuse en temps et ressources de calcul. Courtier et al. (1994) ont réussi à réécrire le problème de minimisation non-quadratique de J_{3D} en une suite de minimisations de fonctions intermédiaires quadratiques : c'est la méthode 3D-VAR incrémentale.

La méthode consiste à reformuler J_{3D} en fonction de l'incrément $\delta \mathbf{x}_k = \mathbf{x}_k - \mathbf{x}_k^b$:

$$J_{3D-inc}(\delta \mathbf{x}_k) = \frac{1}{2} \delta \mathbf{x}_k^T (\mathbf{P}_k^b)^{-1} \delta \mathbf{x}_k + \frac{1}{2} \mathbf{d}_k^b - \mathbf{H}_k \delta \mathbf{x}_k^T (\mathbf{R}_k)^{-1} \mathbf{d}_k^b - \mathbf{H}_k \delta \mathbf{x}_k, \quad (1.61)$$

où $\mathbf{d}_k^b$ est le vecteur d'innovation comme défini en Eq. 1.8 et où le terme $\mathbf{H}_k \delta \mathbf{x}_k$ est obtenu en linéarisant $H_k(\mathbf{x}_k)$. Il s'en suit que J_{3D-inc} est bien quadratique (linéaire en $\delta \mathbf{x}_k$). Le problème de minimisation revient à trouver l'incrément $\delta \mathbf{x}_k^*$ qui annule le gradient de J_{3D-inc} , i.e. :

$$\begin{aligned} \nabla J_{3D-inc}(\delta \mathbf{x}_k^*) = 0 &\iff (\mathbf{P}_k^b)^{-1} \delta \mathbf{x}_k^* - \mathbf{H}_k^T (\mathbf{R}_k)^{-1} \mathbf{d}_k^b - \mathbf{H}_k \delta \mathbf{x}_k^* = 0 \\ &\iff \delta \mathbf{x}_k^* = (\mathbf{P}_k^b)^{-1} - \mathbf{H}_k^T (\mathbf{R}_k)^{-1} \mathbf{H}_k^{-1} \mathbf{H}_k^T (\mathbf{R}_k)^{-1} \mathbf{y}_k^o - H_k(\mathbf{x}_k^b). \end{aligned} \quad (1.62)$$

Les deux fonctions coût 1.60 et 1.62 étant équivalentes, il advient que :

$$\begin{aligned} \delta \mathbf{x}_k^* &= \mathbf{x}_k^a - \mathbf{x}_k^b \\ \iff \mathbf{x}_k^a &= \mathbf{x}_k^b + \delta \mathbf{x}_k^* \\ \iff \mathbf{x}_k^a &= \mathbf{x}_k^b + (\mathbf{P}_k^b)^{-1} - \mathbf{H}_k^T (\mathbf{R}_k)^{-1} \mathbf{H}_k^{-1} \mathbf{H}_k^T (\mathbf{R}_k)^{-1} \mathbf{y}_k^o - H_k(\mathbf{x}_k^b). \end{aligned} \quad (1.63)$$

Aussi, en utilisant les propriétés matricielles, on obtient que

$$(\mathbf{P}_k^b)^{-1} - \mathbf{H}_k^T (\mathbf{R}_k)^{-1} \mathbf{H}_k^{-1} \mathbf{H}_k^T (\mathbf{R}_k)^{-1} = \mathbf{P}_k^b \mathbf{H}_k^T \mathbf{H}_k \mathbf{P}_k^b \mathbf{H}_k^T + \mathbf{R}_k^{-1}. \quad (1.64)$$

L'Eq. 1.64 montre ainsi la similitude avec le filtre de Kalman.

Dans la pratique, la minimisation de J_{3D-inc} est réalisée via des algorithmes itératifs comme le gradient conjugué ou de type quasi-Newton, en prenant le vecteur de contrôle d'ébauche comme initialisation. Chaque itération nécessite l'évaluation de J_{3D-inc} et de son gradient ∇J_{3D-inc} au vecteur de contrôle courant. Il n'y a cependant pas de garantie que la méthode converge et en général, un nombre maximal d'itérations est fixé à l'avance pour générer $\mathbf{x}_k^a$.

Pour passer au cycle d'assimilation suivant, le modèle dynamique est intégré à partir de l'analyse du cycle précédent $\mathbf{x}_k^a$ pour générer $\mathbf{x}_{k+1}^b$. Les méthodes 3D-VAR et 3D-VAR incrémentale se comporte de manière séquentielle comme les méthodes de filtrage. En effet, l'étape de propagation du modèle part de l'analyse du cycle précédent et évolue jusqu'au prochain temps d'observation où une nouvelle étape de minimisation sera effectuée, et ainsi de suite. Cependant, cette méthode engendre de nombreuses discontinuités (à chaque assimilation) dans la trajectoire analyse du modèle. De ce fait, l'assimilation séquentielle des observations apparaît problématique pour obtenir une trajectoire analyse lisse et régulière. Pour compenser ce problème, les méthodes 4D-VAR ont été introduites comme une généralisation du 3D-VAR à un groupe d'observations réparties sur une plus grande fenêtre d'assimilation $[0, K]$. On ne détaillera pas ces méthodes ici. Voir Rochoux (2014); Talagrand and Courtier (1987) pour plus de détails sur ces algorithmes.

Pour terminer sur les méthodes variationnelles, un point qui n'a pas été discuté auparavant est que le succès ou non de ces méthodes dépend, entre autre, de notre capacité à définir le plus justement possible l'opérateur $\mathbf{H}_k^T$ que l'on définit comme l'opérateur adjoint à l'opérateur H_k (utilisé en Eq. 1.60-1.62). À l'opposé de l'opérateur d'observation, cet opérateur permet d'exprimer les observations dans l'espace de contrôle. Une approche basique pour estimer cet opérateur consiste à estimer l'opérateur linéaire-tangent (dérivées partielles) par des différences finies puis à transposer l'opérateur. Cependant, cette technique devient très vite limitée lorsque les opérateurs généralisés deviennent fortement non-linéaires. Quand les modèles le permettent, l'adjoint peut être explicitement écrit à partir des équations différentielles décrivant le comportement du système. Dans la cadre de cette thèse, la complexité du LSM et le grand nombre de phénomènes qu'il décrit (voir Chapitre 2) ne permet pas le développement explicite de l'opérateur adjoint et donc l'utilisation d'une méthode variationnelle. De plus, les LSMs ont un coût de calcul limité par rapport aux modèles atmosphériques 3-D pour lesquels l'assimilation de données variationnelle s'est développée initialement. Les approches de type filtrage de Kalman sont ainsi largement privilégiées dans la littérature pour la modélisation de surface (voir par exemple de Rosnay et al., 2013).

L'AD a été appliquée avec succès pour de nombreux systèmes météorologiques et océanographiques (Houtemaker and Mitchell, 1998; Bennet, 2002). L'utilisation de l'AD en hydrologie de surface a, elle, commencé bien après et fait l'objet de recherches actives depuis les dernières décennies (Lahoz and Lannoy, 2014). La section suivante présente quelques applications de l'AD en hydrologie.

1.5 Application de l'assimilation de données en hydrologie de surface

En pratique, toutes les méthodes actuellement utilisées pour l'assimilation de données en hydrologie sont inspirées ou adaptées des systèmes atmosphériques et océanographiques. Comme précisé par Reichle (2008), ces modèles d'écoulements sont de nature chaotique, c'est-à-dire que de faibles

erreurs dans la condition initiale peuvent mener à des erreurs bien plus importantes au fur et à mesure des intégrations modèle. De ce fait, les algorithmes d'AD utilisés avec ces modèles cherchent principalement à corriger la condition initiale. À l'opposé, la dynamique des modèles d'hydrologie est, par nature, dissipative. Les perturbations de la condition initiale tendent à rapidement disparaître avec la propagation du modèle dans le temps. Autrement dit, la condition initiale corrigée n'a que peu de persistance en temps. Ainsi, l'AD appliquée sur ces modèles se concentrera plus, de manière générale, sur la correction des paramètres d'entrée et des forçages atmosphériques.

L'assimilation d'observations dans les modèles de surfaces continentales (LSM) considère à la fois des données in situ et des données télédéctées. Souvent, les données télédéctées sont assimilées et les données in situ sont utilisées pour valider les résultats après assimilation. Les observations regroupent des mesures d'humidité du sol (Reichle et al., 2002), d'étendue du manteau neigeux (Andreadis and Lettenmaier, 2006), de température de brillance de surface (Crow and Wood, 2003). Les observations sont alors assimilées pour contraindre la paramétrisation des modèles, mais peuvent aussi permettre de mieux initialiser des états critiques pour la prédiction saisonnière à interannuelle :

- Les observations de neige et d'humidité peuvent être assimilées pour contraindre les différents réservoirs d'eau et éviter des stocks irréalistes (Houser et al., 2010; Reichle et al., 2013),
- L'AD peut permettre d'ajuster les paramètres du modèle (Pauwels et al., 2009; Vrugt et al., 2013).
- Enfin, la structure du modèle peut être directement améliorée via l'AD, la confrontation continue des sorties modèle avec les observations permettant d'identifier les insuffisances de la modélisation.

Par contre, il y a encore eu peu d'études d'AD dans des RRM. On peut attribuer le retard de développement des méthodes d'AD combinant des données télédéctées aux sorties des RRM à :

- l'absence d'instruments de télédétection dédiés à l'étude des surfaces continentales et plus particulièrement du volume d'eau à surface libre,
- des algorithmes de *retracking* pas toujours adaptés pour retirer l'information des mesures d'altimétrie nadir dans certains cas,
- un manque de maturité des techniques pour mieux améliorer et contraindre les modèles avec des données télédéctées.

L'état de l'art des méthodes d'assimilation appliquées à ces modèles sera plus amplement développé dans les Chapitres 3 et 4 lorsque l'on appliquera ces méthodes à notre modèle hydrologique, introduit au Chapitre 2.

Conclusions

Ce chapitre donne un aperçu de l'ensemble des techniques disponibles pour étudier le cycle de l'eau continental :

- des modèles hydrologiques qui représentent avec plus ou moins de détails l'ensemble des processus hydrologiques impliqués dans le cycle de l'eau avec un accent sur les modèles globaux de surfaces continentales (LSM) et leur modèle de routage associé (RRM),
- des techniques de mesure in situ et télédéctées qui observent avec plus ou moins de précisions et de manière plus ou moins directe ces différents processus hydrologiques, avec un accent sur les méthodes d'observation et de suivi des fleuves et des rivières.

L'assimilation de données permet ainsi de combiner ces deux sources d'informations. Dans ce chapitre, on a alors présenté les méthodes d'assimilation de données qui ont été développées pour un

large panel d'applications dans les géosciences et l'ingénierie. Le Filtre de Kalman est la technique d'assimilation de données séquentielle la plus répandue. Cependant, ce filtre suppose un modèle dynamique linéaire ainsi que des distributions statistiques gaussiennes pour le modèle et les observations. Des extensions du Filtre de Kalman qui compensent en partie ces limitations ont été développées, parmi elles, le Filtre de Kalman Etendu qui se base sur des linéarisations locales et le Filtre de Kalman d'Ensemble basé sur une description stochastique du comportement du modèle et des statistiques des erreurs.

Ce travail de thèse s'inscrit dans le cadre de l'estimation des flux d'eau à la surface des continents, de l'échelle saisonnière à interannuelle et d'en estimer les incertitudes, puis de les réduire. Comme cela a été présenté dans ce chapitre, le couplage LSM/RRM permet d'avoir une vision spatialisée des flux d'eau à la surface des continents et en particulier dans le réseau de routage. Les évolutions constantes de ces modèles permettent de prendre en compte de plus en plus de processus physiques, bien que leur résolution spatiale soit encore de l'ordre de 50 km. Ces modèles pouvant tourner à l'échelle globale, il est nécessaire de disposer de jeux de données à grande échelle (voire globaux) pour pouvoir les valider et les corriger. Les réseaux de données in situ sont évidemment très importants pour ce type d'étude, mais comme ils sont très inhomogènes spatialement et temporellement, les données satellitaires sont de plus en plus utilisées pour les compléter. Mais à l'heure actuelle, peu d'études ont utilisées des données satellitaires pour corriger les hauteurs et les débits distribués simulés par ces modèles et quantifier les erreurs résultantes.

L'objectif de ce travail de thèse est alors d'utiliser l'assimilation de données de type filtrage (EnKF) pour combiner les informations de base apportées par un modèle de surfaces continentales avec des observations de ces surfaces apportées par l'altimétrie satellitaire en général et la mission SWOT en particulier (voir Fig. 1.24). Dans le chapitre qui suit, on présentera en détail le modèle hydrologique qui, dans les chapitres 3 et 4 suivants, sera combiné à des données satellites via l'AD pour corriger ses paramètres et ses sorties.

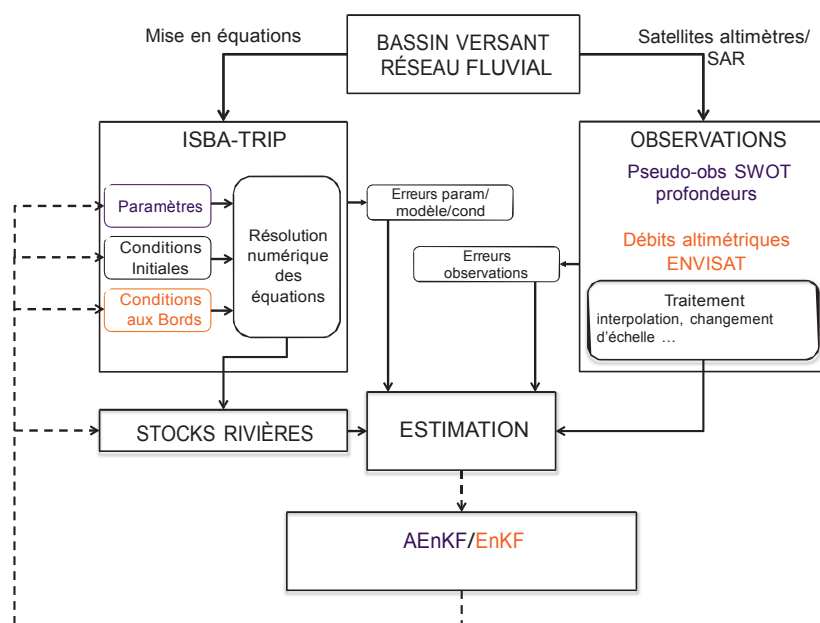


Figure 1.24 – Cadre général du problème d’assimilation de données dans le cadre de cette thèse. Le système physique étudié ici est la dynamique d’un grand bassin versant et plus particulièrement celle du réseau fluvial modélisée par le modèle ISBA-TRIP. Les observations proviennent d’instruments radar embarqués sur des missions altimétriques (SWOT, ENVISAT). L’AD, via l’EnKF (au Chapitre 4) et l’AEnKF (au Chapitre 3) combine ces deux sources d’information, étant donné les incertitudes de chacune, pour corriger soit les paramètres d’entrée du modèle (estimation de paramètre en orange au Chapitre 3), soit son état directement (estimation d’état en violet au Chapitre 4).

Chapitre 2

Description et étude de sensibilité du modèle hydrologique grande échelle ISBA-TRIP

Sommaire

Introduction	77
2.1 Le modèle de surface continentale ISBA-TRIP	79
2.1.1 Historique et différentes versions d'ISBA	79
2.1.2 Description du modèle ISBA-3L	80
2.1.3 Version originale et développements successifs du modèle de routage TRIP	90
2.1.4 Paramètres spécifiques de TRIP	93
2.2 Le domaine d'étude : le bassin de l'Amazone	95
2.3 Analyse de Sensibilité (AS) de la composante "routage de rivière" du modèle ISBA-TRIP	96
2.3.1 Choix de la méthode d'Analyse de Sensibilité	97
2.3.2 Formulation de la méthode ANOVA	98
2.3.3 Application de l'ANOVA au modèle TRIP	101
2.3.4 Résultats	105
Conclusions et perspectives	123

Introduction

Lors du chapitre précédent, une présentation générale de l'ensemble des méthodes (données et modèles) existantes pour étudier le cycle de l'eau continental a été menée. Parmi toutes les méthodes présentées, nous allons nous concentrer sur l'étude du cycle de l'eau continental à grande échelle, c'est-à-dire à l'échelle des grands bassins fluviaux mondiaux - tels que l'Amazone, le Mississippi ou encore le Niger - à l'échelle continentale et globale. On se basera sur l'utilisation d'un modèle hydrologique grande échelle : le modèle ISBA-TRIP (Decharme et al., 2012).

ISBA (Interactions Sol-Biosphère-Atmosphère) constitue la composante "surfaces continentales

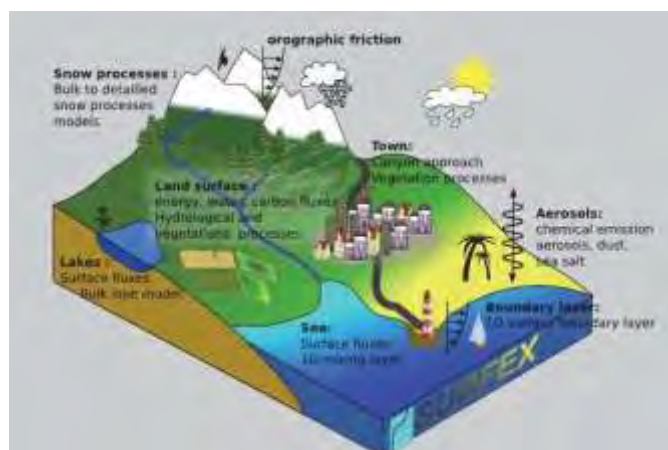


Figure 2.1 – Illustration de l'ensemble des surfaces modélisées par la plateforme SurfEx.

naturelles" (LSM) et TRIP (Total Runoff Integrating Pathways) le modèle "de routage global des fleuves" (RRM) de la plateforme globale de modélisation de surface SurfEx (pour Surface Externalisée) développée et utilisée par Météo-France (<http://www.umr-cnrm.fr/surfex/>). SurfEx, illustrée en Figure 2.1 est composée de 4 tuiles (ou *tiles* en anglais), qui permettent de calculer les bilans de masse et d'énergie et dont 3 sont liées aux surfaces continentales :

- la tuile « lac », avec le modèle FLake,
- la tuile « surfaces naturelles », avec le modèle ISBA (qui peut être couplé avec TRIP au sein de SurfEx, en activant ou non les plaines d'inondation de TRIP),
- la tuile « surfaces urbanisées », avec le modèle TEB
- la tuile « mer et océan », qui permet de calculer les flux de surface sur les océans et mers, en utilisant, au choix, trois paramétrisations plus ou moins complexes (paramétrisation simple, calcul itératif des flux de surface ou modèle de couche de mélange océanique 1D).

SurfEx, et a fortiori ISBA-TRIP, peut être couplé à des modèles atmosphériques et océaniques pour modéliser l'ensemble du système climatique ou être utilisé en mode découplé, pour des études plus locales et/ou de réanalyses dans lesquelles des forçages atmosphériques externes sont utilisés. À chaque pas de temps, les forçages atmosphériques sont constitués de champs de température, pression, humidité de l'air ainsi que des flux radiatifs et de vents, et des champs de précipitations liquides et solides. Dans le cadre de cette thèse, nous utiliserons exclusivement le modèle ISBA-TRIP en mode découplé de l'atmosphère et de l'océan.

L'objectif in fine est de corriger le modèle ISBA-TRIP, soit en corrigeant ses paramètres d'entrée (voir Chapitre 3), soit ses sorties (voir Chapitre 4), grâce à l'assimilation de données satellites. Mais avant de pouvoir faire cela, il est nécessaire de comprendre le fonctionnement du modèle et en particulier la sensibilité de ses sorties à ses paramètres d'entrée. C'est l'objectif de ce chapitre, qui s'organise de la sorte : on décrira d'abord en détail le modèle ISBA-TRIP (Section 2.1), puis on présentera le bassin de l'Amazonie, qui est la zone d'étude utilisée tout au long de ces travaux (Section 2.2) et enfin on mènera une étude de sensibilité globale du modèle ISBA-TRIP sur la zone d'étude (2.3). Cette étude de sensibilité a fait l'objet d'un article actuellement sous presse dans Journal Of Hydrometeorology (disponible en Section C).

2.1 Le modèle de surface continentale ISBA-TRIP

Le modèle ISBA-TRIP correspond au couplage du modèle de surface ISBA (Interactions Sol, Biosphère, Atmosphère) avec le modèle de routage TRIP (Total Runoff Integrating Pathways). Le premier est chargé d'évaluer les flux et les bilans d'énergie et de masse d'eau à la surface et dans la première dizaine de mètre du sol, et le deuxième dirige ces flux des têtes de bassins vers l'exutoire en suivant la topographie. Dans cette section, on commence par décrire le modèle ISBA en commençant par un bref historique des développements et différentes versions du modèle (Section 2.1.1). Puis, on décrira plus en détail la version ISBA-3L utilisée dans le cadre de cette thèse (Section 2.1.2).

2.1.1 Historique et différentes versions d'ISBA

Dans le cadre de cette thèse, j'ai utilisé exclusivement le modèle de surface naturelle ISBA, développé initialement par Noilhan and Planton (1989) et Noilhan and Mahfouf (1996). Le modèle de surface ISBA calcule les échanges d'énergie et d'eau entre le continuum sol-végétation-neige et la couche basse de l'atmosphère. Le schéma est relativement simple car il a été conçu pour des applications climatiques, mais considère malgré tout les composantes les plus importantes des processus de surface c'est-à-dire le contenu en eau et énergie du sol et de la neige, la quantité d'eau interceptée par la canopée et la neige, l'évapotranspiration, le drainage et le ruissellement. Initialement, le sol d'ISBA était basé sur un schéma de forçage-relaxation (Blackadar, 1976; Deardoff, 1977) pour l'évolution de sa température et était décrit par deux couches de sol pour évaluer son contenu en eau, formant ISBA-2L. La méthode α (Mahfouf and Noilhan, 1991) est utilisée pour décrire l'évaporation et le traitement du drainage gravitationnel est aussi pris en compte (Mahfouf and Noilhan, 1996).

Une version multi-couches d'ISBA basée sur les transferts diffusifs a été ensuite développée (ISBA-DF; Boone, 2000; Decharme et al., 2011). Dans cette version, seuls les transferts verticaux, activés par les gradients hydrique et thermique sont pris en compte. Parallèlement, les limitations de l'approche à deux couches (ISBA-2L) ont été mises en évidence et ont imposé la distinction entre une couche du sol où les racines de la végétation extraient de l'eau et une couche sous-racines. Boone et al. (1999); Boone (2000) ont alors travaillé à l'inclusion d'une troisième couche dans le sol, toujours en utilisant le schéma de forçage-relaxation : le modèle ISBA-3L.

La version initiale d'ISBA n'incluait pas le traitement de la photosynthèse pour la végétation. La version ISBA-A-gs (Calvet et al., 1998) incorpore un modèle de surface capable de simuler le cycle diurne du carbone et les flux de vapeurs d'eau. Il modélise les différentes réponses aux changements de concentrations du CO₂, améliore la photosynthèse et réduit la transpiration. Plus récemment, Gibelin et al. (2008) ont développé ISBA-CC : une amélioration d'ISBA-A-gs proposant une représentation plus détaillée de la respiration des écosystèmes et du comportement de la biomasse.

Le traitement de la glace dans le sol a été introduit par Boone et al. (2000) et est modélisé par un équivalent volumétrique en eau dans chaque couche du sol. Par ailleurs, ISBA proposait un premier modèle de neige basé sur Douville et al. (1995) qui fit office de précurseur au schéma ISBA-ES obtenu en simplifiant le modèle CROCUS (Brun et al., 1989) et validé sur le bassin du Rhône (Boone and Etchevers, 2001). Dans ce modèle, la couverture neigeuse est représentée sur 3 couches (au lieu des 50 proposées par CROCUS). Récemment, plus de couches et une physique plus détaillée, provenant de CROCUS, ont été ajoutées (Decharme et al., 2016).

De nombreuses améliorations de la physique ont été apportées au cours des années. Un modèle global tel qu'ISBA fonctionne à basse résolution mais la qualité des simulations hydrologiques est très

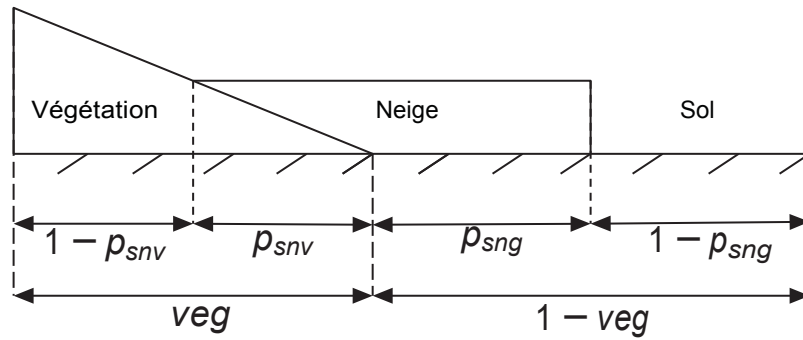


Figure 2.2 – Paramétrisation sous-maille de la surface dans le modèle ISBA : une partie de la surface de la cellule peut être recouverte de végétation (veg), l'autre fraction étant considérée comme du sol nu ($1 - veg$). Le manteau neigeux peut recouvrir une partie de la végétation (p_{snv}) et une partie du sol nu (p_{sng}).

dépendante de nombreux processus ayant lieu à des échelles spatiales non-résolues. Ainsi, Decharme and Douville (2005) ont paramétrisé des processus de ruissellement sous-maille dans ISBA et Habets et al. (2008) ont introduit la modélisation du ruissellement de surface sous-maille. Aussi, les transferts d'eau entre les couches du sol ont été améliorés en ajoutant des profils de conductivité hydraulique à saturation suivant des lois exponentielles (Decharme et al., 2006).

Dans le cadre de cette thèse, on utilise la version ISBA-3L en mode découplée, c'est-à-dire que les conditions atmosphériques qui forcent le modèle proviennent d'une simulation atmosphérique préalable indépendante.

2.1.2 Description du modèle ISBA-3L

ISBA est construit en se basant sur le principe de conservation de la masse et de l'énergie. Les équations d'équilibre sont ensuite résolues indépendamment les unes des autres. Pour représenter les flux d'eau et d'énergie, ISBA-3L utilise 8 variables pronostiques :

- T_g [K], la température à la surface du sol,
- T_2 [K], la température moyenne de la couche de sol,
- w_1 [$m^3.m^{-3}$], l'équivalent volumétrique en eau de la couche superficielle du sol (que l'on dénotera ensuite par w_1),
- w_2 [$m^3.m^{-3}$], l'équivalent volumétrique en eau de la couche racinaire du sol,
- w_3 [$m^3.m^{-3}$], l'équivalent volumétrique en eau de la couche sous-racinaire du sol,
- w_r [mm] (ou équivalent à [$kg.m^2$]), la masse d'eau dans le réservoir d'interception de la canopée.

À ces variables, peuvent être ajoutées les variables pronostiques associées à la glace du sol :

- $w_{1,f}$ [$m^3.m^{-3}$], l'humidité du sol équivalente à la glace contenue dans w_1 (" f " est utilisé ici pour l'anglais *frozen*),
- $w_{2,f}$ [$m^3.m^{-3}$], l'humidité du sol équivalente à la glace contenue dans w_2 ,

ainsi qu'à la neige :

- w_s [mm] (ou équivalent à [$kg.m^2$]), la masse d'eau sous forme de neige sur la surface,
- ρ_s [$kg.m^{-3}$], la densité de la neige,
- α_s [-], l'albédo de la neige.

La prise en compte entre les différentes couvertures de la surface - c'est-à-dire sol nu/végétation couvert(e) de neige ou non - est permise par l'utilisation de valeurs fractionnaires (variables dans le temps ou non). La Figure 2.2 illustre cette division sous-maille de la surface telles que :

- veg est la fraction de sol couvert de végétation,
- p_{sng} est la fraction de sol nu couverte de neige,
- p_{snv} est la fraction de végétation couverte de neige,
- p_{sn} est la fraction totale de surface couverte de neige telle que $p_{sn} = (1 - veg).p_{sng} + veg.p_{snv}$.

Les entrées d'eau sur la surface sont les précipitations liquides (R) et/ou solides (S) tombant dans chaque maille (et fournies par les forçages atmosphériques). Ces forçages pilotent les mouvements des masses d'eau et les différents mécanismes physiques sur et à travers la surface, illustrés en Figure 2.3. Une partie des précipitations sont interceptées par la canopée alors que le reste atteint le sol nu, gonflées par les gouttes d'eau ruisselant de la végétation. Selon la capacité d'infiltration du sol et sa saturation en eau, l'eau ruisselle ou s'infiltre. Chaque couche du sol échange de l'eau par diffusion verticale. Simultanément, les plantes puisent de l'eau dans la couche racinaire qui peut être émise dans l'atmosphère via la transpiration. Enfin, l'eau de la surface peut aussi retourner à l'atmosphère par évaporation ou sublimation. L'eau s'évapore du sol nu, de la canopée et par l'évapotranspiration des plantes.

Les échanges d'énergie entre la surface et l'atmosphère sont contrôlés par les rayonnements solaires et atmosphériques. La partie des rayonnements absorbées par la surface contribue à son réchauffement et permet le phénomène d'évaporation.

2.1.2.1 Traitement de l'eau liquide

La surface du sol se divise en sol nu et sol couvert de végétation, chaque fraction du sol est liée à des processus physiques particuliers.

Une fraction des précipitations R ($[kg.m^{-2}.s^{-1}]$) est interceptée par la canopée (w_r , [mm]) alors que le reste des précipitations tombe directement sur le sol nu. De plus, si le feuillage est saturé, un phénomène d'égouttement R_r ($[kg.m^{-2}.s^{-1}]$) a lieu et le sol nu reçoit plus d'eau. L'évolution temporelle de la masse d'eau dans le réservoir d'interception (canopée) est obtenue en résolvant l'Eq. 2.1 (Deardoff, 1977) :

$$\frac{dw_r}{dt} = veg.R - E_r - R_r, \quad w_r > w_{r,max} \text{ [mm]}. \quad (2.1)$$

veg est la fraction de la cellule ISBA couverte par la végétation et détermine la fraction de précipitations interceptées par la canopée. L'égouttement de la canopée apparaît quand la masse d'eau dans w_r excède une valeur maximale $w_{r,max}$ qui dépend de la densité du feuillage. Pour déterminer $w_{r,max}$, ISBA utilise la formule proposée par Dickinson et al. (1998) :

$$w_{r,max} = 0.2veg.LAI \quad (2.2)$$

où le LAI est l'indice de surface foliaire (Leaf Area Index en anglais), une quantité adimensionnelle caractérisant la densité de la végétation. Il contrôle le taux de transpiration et d'interception de la canopée. Ce paramètre ainsi que d'autres paramètres physiographiques sont fournis par la base de données ECOCLIMAP (Faroux et al., 2013).

Le bilan d'eau sur la couche superficielle du sol w_1 (c'est-à-dire la première des trois première couche du sol d'ISBA-3L) comprend alors les précipitations non-interceptées ainsi que l'égouttement

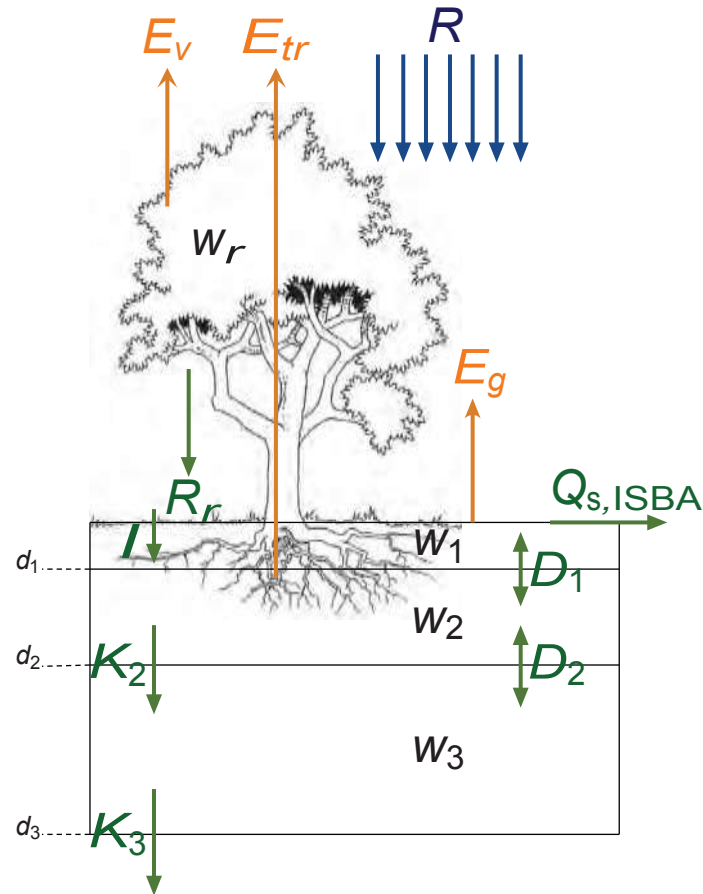


Figure 2.3 – Illustration des mécanismes simulés dans le schéma ISBA-3L (sans neige) : les précipitations R (et S non-représentées ici) tombant sur la surface se partagent entre la canopée (dont le contenu en eau est noté w_r) et le sol nu. Une partie des précipitations interceptées par la canopée peut ruisseler vers la surface sous forme de gouttes R_r . Le sol se divise en trois couches : superficielle (w_1) jusqu'à la profondeur d_1 , racinaire (w_2) jusqu'à la profondeur d_2 et profonde (w_3) jusqu'à la profondeur d_3 . L'eau disponible à la surface peut ruisseler ($Q_{s,ISBA}$), s'infiltrer (I) ou directement s'évaporer du sol (E_g) et de la végétation (E_v). La canopée puise l'eau de la couche racinaire et la rejette dans l'atmosphère via l'évapotranspiration (E_{tr}). Les couches du sol s'échange de l'eau par diffusion verticale (D_1 , D_2) ou l'eau peut être drainée vers les couches plus profondes du sol (K_2 , K_3).

de la canopée. Les trois couches du sol échangent de l'eau via des transferts verticaux, dépendant de l'humidité de chaque couche et la texture du sol. Le système complet d'équation pour les variables pronostiques du sol est donné par :

$$\frac{dw_1}{dt} = \frac{C_1}{\rho_w d_1} [(1 - veg)R + R_r - Q_{s,ISBA} - E_g] - D_1, \quad (2.3)$$

$$\frac{dw_2}{dt} = \frac{1}{\rho_w d_2} [(1 - veg)R + R_r - Q_{s,ISBA} - E_g - E_{tr}] - D_2 - K_2, \quad (2.4)$$

$$\frac{dw_3}{dt} = \frac{d_2}{d_3 - d_2} [D_2 + K_2] - \frac{1}{d_3 - d_2} K_3 \quad (2.5)$$

avec

- ρ_w [kg.m⁻³], la masse volumique de l'eau,
- d_1, d_2, d_3 [m], les profondeurs de chaque couche,
- C_1 [], le coefficient de relaxation du sol,
- D_1, D_2 [s⁻¹], les termes de diffusion,
- K_2, K_3 [s⁻¹], les termes de drainage,
- $Q_{s,ISBA}$ [kg.m⁻².s⁻¹], le ruissellement de surface.

E_g et E_{tr} sont respectivement l'évaporation d'eau liquide du sol nu (différent de la sublimation d'eau solide E_s qui peut aussi apparaître) et la transpiration des plantes. Enfin, le terme $Q_{s,ISBA}$ représente le flux d'eau quittant la cellule par le ruissellement de surface.

2.1.2.1.a Ruissellement de surface $Q_{s,ISBA}$ et infiltration /

Dans la version initiale d'ISBA, le ruissellement de surface n'apparaissait que quand w_2 excédait la valeur de saturation w_{sat} . C'est ce qui est appelé, en hydrologie, le ruissellement de Dunne (c'est-à-dire, le ruissellement apparaissant quand le sol est saturé en eau). Cependant, quand l'échelle de variabilité du ruissellement est plus petite que l'échelle de la maille du modèle, le sol n'est jamais saturé et la production de ruissellement est très faible. Pour pallier ce problème, une fraction de cellule toujours saturée a été introduite dans le modèle, afin de produire du ruissellement. Dans ISBA, deux approches sont disponibles pour tenir compte du ruissellement sous-maille.

Dans la première option, la paramétrisation sous-maille est introduite dans Habets et al. (1999) selon l'approche développée dans le modèle VIC (Variable Infiltration Capacité, Dumenil and Todini, 1992; Zhao, 1992; Wood et al., 1998). On suppose que chaque cellule est composée d'une infinité de réservoirs élémentaires dont la capacité d'infiltration varie continuellement, d'un bassin à l'autre, de 0 à une valeur maximale i_m donnée. Pour une teneur en eau $w_2 < w_{sat}$, certains réservoirs élémentaires seront pleins et d'autres ne le seront pas. Si i est la teneur en eau des réservoirs non-saturés, tous les réservoirs ayant une capacité inférieure à i sont saturés ; ce qui permet de déterminer la fraction de cellule saturée $A(i)$. Dans le cas d'un événement pluvieux, tous les réservoirs ayant une capacité inférieure à $i + R$ seront remplis et produiront du ruissellement :

$$A(i) = 1 - \left(1 - \frac{i}{i_m}\right)^b \Rightarrow Q_{s,VIC}^D = \int_i^{i+R} A(i) di. \quad (2.6)$$

Ainsi, pour une humidité du sol de w , la fraction de cellule saturée est donnée par :

$$A(w_2) = 1 - \left(1 - \frac{w_2}{w_{sat}}\right)^{\frac{b}{b+1}}. \quad (2.7)$$

Finalement, suivant les tests préliminaires dans Habets et al. (1999), une valeur seuil a été introduite pour éviter la production de ruissellement lors de la saison sèche (w_{crit}). Habets et al. (1999) ont fixé ce seuil égal au point de flétrissement w_{wilt} (*wilting point* en anglais), qui est la valeur seuil en-dessous de laquelle il n'y a plus de transpiration de la végétation. Ainsi, on a : $w_{crit} = w_{wilt}$.

La deuxième option repose sur l'approche du couplage ISBA-TOPMODEL. Elle a été initialement introduite par Habets and Saulnier (2001) et généralisée par Decharme et al. (2006). La méthode utilise le *déficit de stock d'eau* D_t [m] pour déterminer la fraction saturée de la cellule f_{sat} . Le déficit de stock d'eau prend des valeurs entre d_2 - la profondeur de la couche de sol w_2 - et $d_0 = (w_{sat} - w_{wilt})d_2$ - le déficit correspondant à $w_2 = w_{wilt}$. Quand $D_t = d_0$, la fraction de cellule saturée est égale à $f_{sat} = 0$ et quand $D_t = 0$, $f_{sat} = 1$. Enfin, le ruissellement de surface est simplement déterminé par :

$$Q_{s, TOP}^D = f_{sat} \cdot (1 - veg) \cdot R \quad (2.8)$$

Dans un second temps, il est important de considérer aussi, en complément du ruissellement de Dunne, le ruissellement d'Horton. Ce phénomène de ruissellement apparaît lorsque l'intensité des précipitations dépasse le taux d'infiltration effectif de la surface. Ce phénomène tend à dominer dans les zones désertiques ou semi-arides où les courts épisodes pluvieux peuvent être très intenses. Le ruissellement d'Horton est simplement calculé selon Decharme et al. (2006) :

$$Q_s^H = (1 - \delta_f) \max(0, S_m + (1 - veg)R - I_{unf}) + \delta_f \max(0, S_m + (1 - veg)R - I_f), \quad (2.9)$$

où δ_f est la fraction de sol gelé, S_m la fonte de la neige et I_{unf}/I_f les fonctions d'infiltration sur un sol non-gelé/gelé.

Finalement, la production totale de ruissellement est donnée par :

$$Q_{s, ISBA} = Q_s^D + (1 - f_{sat})Q_s^H. \quad (2.10)$$

Dans le cadre de cette thèse, le ruissellement de surface de Dunne est modélisé en appliquant l'approche du modèle TOPMODEL, i.e. $Q_{s, TOP}^D$ Eq. 2.8.

Par déduction, l'infiltration est définie comme la différence entre la masse totale d'eau liquide atteignant le sol et l'eau qui ruisselle, soit :

$$I = (1 - veg) \cdot R + R_r - Q_{s, ISBA} \quad (2.11)$$

2.1.2.1.b Termes de diffusion D_1 et D_2

La diffusion verticale de l'humidité est un mouvement ascendant ou descendant de l'eau à travers les couches du sol et dépendant du gradient d'humidité. La diffusion (ou transfert d'eau) entre la couche superficielle et la couche racinaire (notée D_1) dépend du volume d'eau dans la couche superficielle (w_1) et du volume d'eau d'équilibre entre les forces de gravité et de capillarité ($w_{1,eq}$). Elle est calculée en utilisant l'Eq. 2.12 (Noilhan and Planton, 1989). Pour le terme de diffusion D_2 , le gradient d'humidité entre la couche racinaire et sous-racinaire est directement utilisé (Eq. 2.13; Boone et al., 1999). Les équations permettant le calcul de D_1 et D_2 sont les suivantes :

$$D_1 = \frac{C_2}{T} (w_1 - w_{1,eq}), \quad (2.12)$$

$$D_2 = \frac{C_4}{\tau} (w_2 - w_3) . \quad (2.13)$$

avec τ la constante de relaxation, égale à 1 jour, alors que C_2 et C_4 sont les paramètres hydrologiques de forçage-relaxation caractérisant la vitesse à laquelle le profil d'eau est rétabli à l'équilibre.

2.1.2.1.c Termes de drainage K_2 et K_3

Le drainage gravitationnel apparaît quand w excède la capacité du terrain w_{fc} . Les équations permettant le calcul de K_2 (Mahfouf and Noilhan, 1996) et K_3 (Boone et al., 1999) sont données ci-dessous :

$$K_2 = \frac{C_3 d_3}{\tau d_2} \max(0, w_2 - w_{fc}) , \quad (2.14)$$

$$K_3 = \frac{C_3 d_3}{\tau d_3 - d_2} \max(0, w_3 - w_{fc}) . \quad (2.15)$$

avec C_3 le paramètre de forçage-relaxation qui représente la vitesse à laquelle le profil d'humidité dans le sol retourne à l'équilibre.

2.1.2.1.d Propriétés hydriques du sol

Un sol quelconque va être caractérisé par sa texture (proportion d'argile, de sable et de limon). À partir de la texture du sol et de son humidité, sa conductivité hydraulique k [$m.s^{-1}$] et son potentiel hydrique φ [m] peuvent être déduits à partir de formules empiriques issues des travaux de Brooks and Corey (1966); Clapp and Hornbeger (1978). La conductivité hydraulique, homogène à une vitesse, exprime l'aptitude d'un milieu poreux à laisser passer un fluide. D'autre part, le potentiel hydrique, toujours négatif, mesure la capacité du sol à retenir l'eau. Plus le potentiel est bas, plus le sol va retenir l'eau. Les travaux de Clapp and Hornbeger (1978) ont permis d'exprimer la conductivité hydraulique et le potentiel hydrique de chaque couche du sol w_i comme une fonction du type du sol (Brooks and Corey, 1966) :

$$\varphi(w_i) = \varphi_{sat} \left(\frac{w_i}{w_{sat}} \right)^{-b} , \quad (2.16)$$

$$k(w_i) = k_{sat,c} \left(\frac{w_i}{w_{sat}} \right)^{2b+3} , \quad (2.17)$$

$$(2.18)$$

avec w_{sat} l'humidité du sol à saturation, φ_{sat} le potentiel hydrique à saturation, $k_{sat,c}$ la valeur minimale (dite valeur "compactée") de la conductivité hydraulique à saturation et b la pente de la courbe de rétention hydrique, tous dépendant de la texture du sol.

Suite aux travaux de Montaldo and Albertson (2001), Decharme et al. (2006) a redéfini la conductivité hydraulique à saturation k_{sat} comme une fonction exponentielle de la profondeur du sol z telle que :

$$k_{sat}(z) = k_{sat,c} \times \exp(-f(z - d_c)) , \quad (2.19)$$

où d_c [m] est la profondeur compactée et f [m^{-1}] est le coefficient de décroissance. En conséquence, la conductivité hydraulique k a été réécrite comme :

$$k(w_i, z) = k_{sat}(z) \left(\frac{w_i}{w_{sat}} \right)^{2b+3} . \quad (2.20)$$

Enfin, les coefficients adimensionnels de forçage-relaxation C_1 , C_2 , C_3 and C_4 sont aussi décrits par des formules empiriques dépendant de la texture du sol. Nous ne détaillerons pas l'expression de ces coefficients car Boone et al. (1999) les fournit explicitement en Annexe A de l'article. Decharme et al. (2006) a alors réécrit ces coefficients afin qu'ils tiennent compte du profil exponentiel de la conductivité hydraulique à saturation.

Maintenant que les flux dans les couches du sol ont été présentés, nous allons nous concentrer sur les flux de surfaces.

2.1.2.2 Traitement des flux de surface

2.1.2.2.a Flux d'énergie

La totalité des rayonnements émis par le soleil R_G [W.m^{-2}] dans le domaine des courtes longueurs d'onde n'atteint pas la surface. Elles peuvent être modifiées par plusieurs facteurs, notamment atmosphériques (nuages, humidité, aérosols) et humains (bâtiments). La capacité de la surface à réfléchir les rayonnements de faibles longueurs d'onde est modélisée par l'albédo α dont les valeurs, entre 0 et 1, sont fortement dépendantes des caractéristiques de la surface. En général, l'albédo du sol est compris entre 0,1 et 0,9, celui de la végétation varie entre 0,1 et 0,2 et celui de la neige se situe entre 0,5 et 0,9. Étant donné la définition de l'albédo, la fraction de rayonnements absorbés par la surface correspond à $(1 - \alpha)R_G$. De plus, l'atmosphère émet aussi des rayonnements, notées R_A [W.m^{-2}], dans le domaine des grandes longueurs d'onde vers la surface (ainsi que vers l'espace); une partie étant aussi réfléchi par la surface. La surface se comporte comme un corps gris avec une émissivité E . Selon la loi de Kirchoff, avec une transmissivité de surface de 0 - caractéristique du comportement des corps gris -, une émissivité de E , la réflectivité de la surface est donnée par $(1 - E)$: les rayonnements grandes longueurs d'onde absorbés par la surface sont alors de ER_A . La surface elle-même émet des rayonnements de grandes longueurs d'onde vers l'atmosphère. La quantité de rayonnements perdues est déterminées en utilisant la température de la surface T_g et la loi de Boltzmann - $E\sigma T_g^4$. Le bilan des rayonnements reçus par la surface R_n , voir Figure 2.4 est alors donné par

$$R_n = (1 - \alpha)R_G + E(R_A - \sigma T_g^4). \quad (2.21)$$

R_n constitue une importante part de l'énergie disponible à la surface et est utilisée dans les équations décrivant l'évolution temporelle de la température du sol, toujours basées sur la méthode de forçage-relaxation :

$$\frac{dT_g}{dt} = C_T (R_n - H - LE) - \frac{2\pi}{T} (T_g - T_2), \quad (2.22)$$

$$\frac{dT_2}{dt} = \frac{1}{T} (T_g - T_2). \quad (2.23)$$

$$(2.24)$$

La température à la surface T_g évolue avec le forçage diurne selon le flux de chaleur $G = R_n - H - LE$ et un terme de relaxation contrôlé par la température moyenne du sol profond T_2 . C_T [$\text{J.kg}^{-1}.\text{K}^{-1}$] est la capacité calorifique du sol tenant compte des différentes fractions de couverture du sol : veg pour la végétation, p_{sng} pour le sol nu neigeux et p_{snv} pour la végétation couverte de neige.

$$(C_T)^{-1} = \frac{(1 - veg)(1 - p_{sng})}{C_g} + \frac{veg(1 - p_{snv})}{C_v} + \frac{(1 - veg)p_{sng} + veg.p_{snv}}{C_s}, \quad (2.25)$$

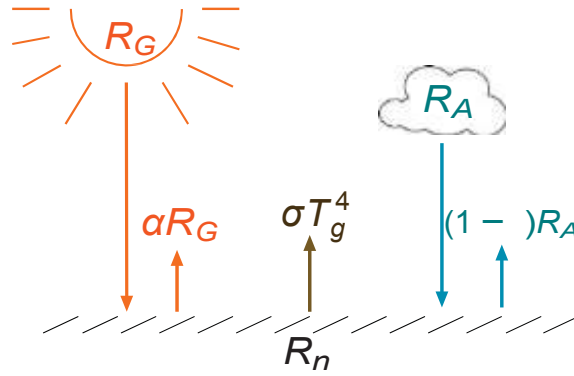


Figure 2.4 – Illustration des différentes composantes du bilan radiatifs de la surface : le soleil émet la surface des rayonnements de courtes longueurs d'onde R_G et l'atmosphère émet des rayonnements de grandes longueurs d'onde R_A . Une partie des rayonnements sont réfléchies en fonction de l'albédo α (rayonnements de courtes longueurs d'onde) et de la réflectivité $1 - \epsilon$ (rayonnements de longues longueurs d'onde) de la surface. La surface elle-même émet des rayonnements vers l'atmosphère/l'espace en fonction de son émissivité ϵ et de sa température T_g .

C_g, C_v, C_s étant respectivement les capacités calorifiques du sol nu, de la végétation et de la neige.

La surface reçoit un flux radiatif total de R_n . Cette chaleur entrante est amputée des flux turbulents ou de chaleur sensible H ainsi que des flux d'évapotranspiration LE . Le flux de chaleur sensible H consiste en l'ensemble des échanges d'énergie dus aux mouvements des fluides (ici le vent). Elle est directement proportionnelle au gradient de température entre la surface et l'atmosphère :

$$H = \rho_a c_p C_H v_a (T_g - T_a), \text{ [W.m}^{-2}\text{]} \quad (2.26)$$

où

- ρ_a [kg.m⁻³] est la masse volumique de l'air,
- c_p [J.kg⁻¹.K⁻¹] est la capacité calorifique de l'air,
- C_H [] est le coefficient de traînée dépendant de la stabilité thermique et de la rugosité de surface,
- v_a [m.s⁻¹] est la vitesse du vent,
- T_g, T_a [K] sont respectivement la température de la surface et de l'air.

2.1.2.2.b Flux de vapeur d'eau

Les flux de vapeur d'eau interagissent évidemment avec les flux d'énergie. E [kg.m⁻².s⁻¹] est la vapeur d'eau totale émise par la surface. Elle se compose de l'évaporation d'eau liquide sur le sol nu E_g , sur la végétation E_r et la transpiration E_{tr} ainsi que de la sublimation de la neige E_s et de la glace du sol $E_{g,f}$. Ces quantités dépendent du gradient de l'humidité spécifique entre la surface et l'atmosphère et les fractions correspondantes différents types de sol concernés dans la cellule. L'évaporation sur le sol nu est donnée par :

$$E_g = (1 - veg)(1 - p_{sng})(1 - \delta_i)\rho_a C_H v_a (h_u q_{sat}(T_g) - q_a), \text{ [kg.m}^{-2}\text{.s}^{-1}\text{]} \quad (2.27)$$

où

- δ_i [] est la fraction de surface gelée utilisée pour partitionner l'évaporation et la sublimation sur le sol nu,

- $q_{sat}(T_g)$ [] est l'humidité spécifique de saturation à la température T_g ,
- q_a [] est l'humidité spécifique de l'atmosphère dans la plus basse couche de l'atmosphère,
- h_u [] est l'humidité relative de la surface, liée à l'humidité de la couche superficielle du sol w_1 .

Similairement :

$$E_{g,f} = (1 - veg)(1 - p_{sng})\delta_i\rho_a C_H v_a (h_{ui}q_{sat}(T_g) - q_a), \text{ [kg.m}^{-2}\text{.s}^{-1}\text{]} \quad (2.28)$$

avec h_{ui} l'humidité relative de la glace. De plus, la sublimation de la neige s'écrit :

$$E_s = p_{sn}\rho_a C_H v_a (h_{ui}q_{sat}(T_g) - q_a). \text{ [kg.m}^{-2}\text{.s}^{-1}\text{]} \quad (2.29)$$

L'évapotranspiration de la végétation E_v est donnée par

$$E_v = veg(1 - p_{snv})\rho_a C_H v_a h_v (q_{sat}(T_g) - q_a) \text{ [kg.m}^{-2}\text{.s}^{-1}\text{]}. \quad (2.30)$$

La constante adimensionnelle h_v est le coefficient d'Halstead, défini dans (Noilhan and Planton, 1989) par :

$$h_v = (1 - \delta) \frac{R_a}{R_a + R_s} + \delta. \quad (2.31)$$

Elle dépend ainsi :

- de la fraction δ [-] de feuillage couverte par l'eau interceptée :

$$\delta = \frac{w_r}{w_{r,max}}^{\frac{2}{3}}, \quad (2.32)$$

- de la résistance de l'air R_a [s.m⁻¹] :

$$R_a = (C_H v_a)^{-1}, \quad (2.33)$$

- et de la résistance de la surface du sol R_s [s.m⁻¹]. Ce terme est construit de telle manière que la transpiration soit stoppée quand le sol est très sec c'est-à-dire quand l'humidité du sol dans la couche racinaire w_2 atteint le point de flétrissement w_{wilt} .

Il s'avère que l'évapotranspiration totale soit la somme de l'évaporation directe de la canopée E_r et la transpiration des plantes E_{tr} :

$$E_r = veg(1 - p_{snv})\rho_a \frac{\delta}{R_a} (q_{sat}(T_g) - q_a), \text{ [kg.m}^{-2}\text{.s}^{-1}\text{]} \quad (2.34)$$

$$E_{tr} = veg(1 - p_{snv})\rho_a \frac{1 - \delta}{R_a + R_s} (q_{sat}(T_g) - q_a), \text{ [kg.m}^{-2}\text{.s}^{-1}\text{]}. \quad (2.35)$$

Les flux de chaleur correspondant à ces flux de vapeur sont déterminés en utilisant la capacité calorifique de vaporisation L [J.kg⁻¹] et de sublimation L_i [J.kg⁻¹] donnant le flux total de chaleur :

$$LE = LE_g + LE_v + L_i E_s + L_i E_{g,f} \quad (2.36)$$

On note finalement que le terme résiduel $G = R_n - H - LE$ correspond aux flux de chaleur par conduction à travers le sol.

2.1.2.3 Traitement de la neige

ISBA propose plusieurs schémas pour gérer le stock d'eau sous forme de neige à la surface. Ils vont d'un schéma monocouche très simplifié à des schémas multicouches qui décrivent l'intégralité du comportement du manteau neigeux (schémas ES Boone and Etchevers, 2001, EBA Bazile et al., 2002 et CROCUS Vionnet et al., 2012). Le bassin d'étude de cette thèse se situe dans des zones géographiques où l'impact de la neige n'est pas dominant, aussi le schéma de neige monocouche développé par Douville et al. (1995) est utilisé.

L'évolution de l'équivalent en masse d'eau du manteau neigeux w_s [mm] est donnée par :

$$\frac{\partial w_s}{\partial t} = S - E_s - melt, \quad (2.37)$$

où

- S [$\text{kg.m}^{-2}.\text{s}^{-1}$] sont les précipitations de neige,
- E_s [$\text{kg.m}^{-2}.\text{s}^{-1}$] est le flux de sublimation (Eq. 2.29),
- $melt$ [$\text{kg.m}^{-2}.\text{s}^{-1}$] est la fonte de la neige.

Du fait de l'importante variabilité des propriétés thermiques liées à la compacité de la neige, la densité ρ_s de celle-ci est considérée comme une nouvelle variable pronostique qui décroît exponentiellement à un taux donné τ_f par jour (Verseghy, 1991). De plus, la présence de neige sur la surface peut fortement influencer les transferts de masse et d'énergie entre le sol et l'atmosphère. En effet, la couche de neige modifie l'albédo de la surface et donc son bilan radiatif R_n - voir Eq. 2.21. Ainsi, l'albédo de la neige α_s est aussi considéré comme une variable pronostique supplémentaire qui varie selon la fonte ou non de la neige. L'albédo moyen α_t sur une cellule est alors donné par :

$$\alpha_t = (1 - p_{sn})\alpha + p_{sn}\alpha_s, \quad (2.38)$$

avec p_{sn} la fraction de sol neigeux (végétation et sol confondu). Similairement, l'émissivité moyenne de la surface est aussi impactée et devient

$$E_t = (1 - p_{sn})E + p_{sn}E_s, \quad E_s = 1. \quad (2.39)$$

Plus de détails sont disponibles sur Douville et al. (1995).

2.1.2.4 Traitement du sol gelé

La prise en compte du sol gelé nécessite l'addition d'un terme de changement de phase dans les équations thermiques 2.24 et hydrologiques 2.5. Dans le modèle, seulement les couches superficielle et racinaire du sol peuvent geler. Le mécanisme de gel/dégel est modélisé similairement au mécanisme d'assèchement/humidification (Boone et al., 2000). Ainsi les coefficients de forçage-relaxation sont redéfinis pour ces équations. L'évolution temporelle des variables pronostiques liées au gel/dégel sont alors décrites par :

$$\frac{dw_{1,f}}{dt} = \frac{1}{d_1\rho_w} (F_{g,w} - E_{g,f}), \quad (0 \leq w_{1,f} \leq w_{sat} - w_{min}), \quad (2.40)$$

$$\frac{dw_{2,f}}{dt} = \frac{1}{(d_2 - d_1)\rho_w} F_{2,w}, \quad (0 \leq w_{2,f} \leq w_{sat} - w_{min}), \quad (2.41)$$

où $w_{1,f}$ et $w_{2,f}$ sont les humidités de sol équivalentes en glace dans w_1 et w_2 . L'indice "g" est relatif à la surface, "2" au sol profond et "f" au gel (f pour "freezing"). On trouve la ré-écriture des Eq. 2.24 et Eq. 2.5 dans Boone et al. (2000).

Les termes de changement de phase - F [$\text{kg.m}^{-2}.\text{s}^{-1}$] - représentent les flux de masse dus à la formation de la glace où à la fonte et sont exprimés à la fois pour la couche superficielle du sol et pour la couche racinaire. Pour plus de détails sur la définition des termes de changement de phase, voir Boone (2000).

Tout ces modules permettent d'établir le bilan d'eau et d'énergie dans chaque cellule une à une et plus particulièrement, d'estimer le ruissellement de surface et le drainage gravitationnel. Au cours du cycle de l'eau, l'eau qui ruisselle alimente les rivières et fleuves et contribuent au débit. Cependant, comme ce bilan de masse et d'énergie est établi sur chaque cellule indépendamment des autres, ISBA seul n'est pas suffisant pour estimer ensuite les débits des rivières. De plus, l'eau infiltrée peut remonter à la surface et aussi alimenter les rivières et fleuves. Les modèles de routage tels que TRIP (Oki and Sud, 1998) - sont basés sur un réseau hydrologique suivant la topographie des bassins afin de lier les cellules les unes autres selon le sens d'écoulement de l'eau. Ces modèles permettent de déplacer l'eau de pixel en pixel, jusqu'à l'exutoire (d'où leur nom) et indirectement convertissent ainsi le ruissellement de surface en débit équivalent.

2.1.3 Version originale et développements successifs du modèle de routage TRIP

La version initiale développée par Oki and Sud (1998) proposait un réseau d'écoulement hydrologique global à une résolution de $1^\circ \times 1^\circ$ construit de telle sorte qu'une cellule peut recevoir les flux sortants de plusieurs cellules mais se déverse dans une seule et unique cellule. Pour chaque cellule du bassin, TRIP considère le débit entrant dans la cellule via les cellules amont ainsi que le ruissellement de surface dans la cellule même. Le modèle en déduit alors le débit sortant de la cellule. Les cellules amonts (donnant le débit entrant) et avalés (indiquant dans quelle cellule le flux sortant est envoyé) sont définies à partir du réseau de routage de TRIP, illustré pour l'Amazonie à $1^\circ \times 1^\circ$ en Figure 2.5.

Dans la version originale du modèle (Oki and Sud, 1998), TRIP fonctionnait alors en mode découplé, c'est-à-dire que les données de ruissellement pour chaque cellule provenaient des sorties d'un modèle de surface tel qu'ISBA (Oki et al., 1999). La variable pronostique de TRIP est la masse d'eau dans le tronçon de rivière se trouvant dans la cellule. Plusieurs tronçons de rivières pouvant exister dans une même cellule de TRIP (du fait de sa grande échelle), ces tronçons sont regroupés sous un seul tronçon équivalent de forme régulière (rectangulaire), Figure 2.6 compare alors le réseau fluvial réel avec le réseau de routage TRIP sur le bassin de l'Amazonie. Enfin, à partir des variations de masse d'eau et sous l'hypothèse d'une vitesse d'écoulement constante de 0.5 m.s^{-1} , le débit sortant de la cellule peut être estimé.

Depuis, de nombreuses modifications ont été apportées à TRIP, essentiellement par les travaux de Decharme et al. (2008, 2010) : l'unique réservoir de surface initial a été complété par un réservoir profond pour les eaux souterraines (Decharme et al., 2010) et un réservoir d'inondations pour les plaines inondées (Decharme et al., 2008). Enfin, une vitesse d'écoulement variable basée sur Manning (1891); Arora and Boer (1999) a été introduite. ISBA et TRIP sont pour la première fois couplés par Decharme and Douville (2006) dans le cadre d'une étude d'intercomparaison de modèles et de produits de précipitations : le ruissellement de surface ISBA $Q_{\text{in, ISBA}}^S$ alimente le réservoir de surface de TRIP ainsi que le drainage profond Q_{sub} . Par la suite, plusieurs études se sont aussi intéressées aux performances de ce système couplé à l'échelle régionale (Decharme et al., 2008) et globale (Alkama et al., 2010; Decharme et al., 2010, 2012).



Figure 2.5 – Réseau de routage du modèle TRIP à $1^{\circ} \times 1^{\circ}$ sur le bassin de l'Amazon (image provenant de <http://hydro.iis.u-tokyo.ac.jp/~taikan/TRIPDATA/TRIPDATA.html>).

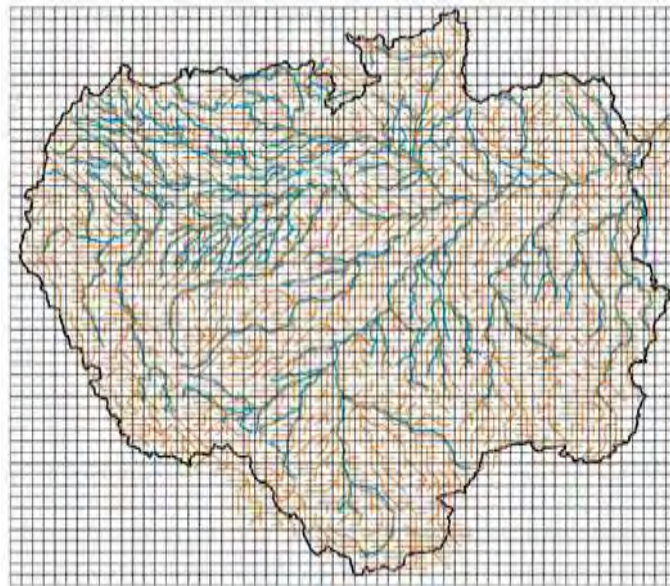


Figure 2.6 – Sur le bassin de l'Amazon, délimité par un polygone noir, le réseau fluvial est représenté par le chevelu bleu. En comparaison, la maille régulière grise représente la maille du modèle TRIP à $0,5^{\circ} \times 0,5^{\circ}$ et les liens en lignes droites oranges représentent le réseau hydrologique TRIP à cette résolution.

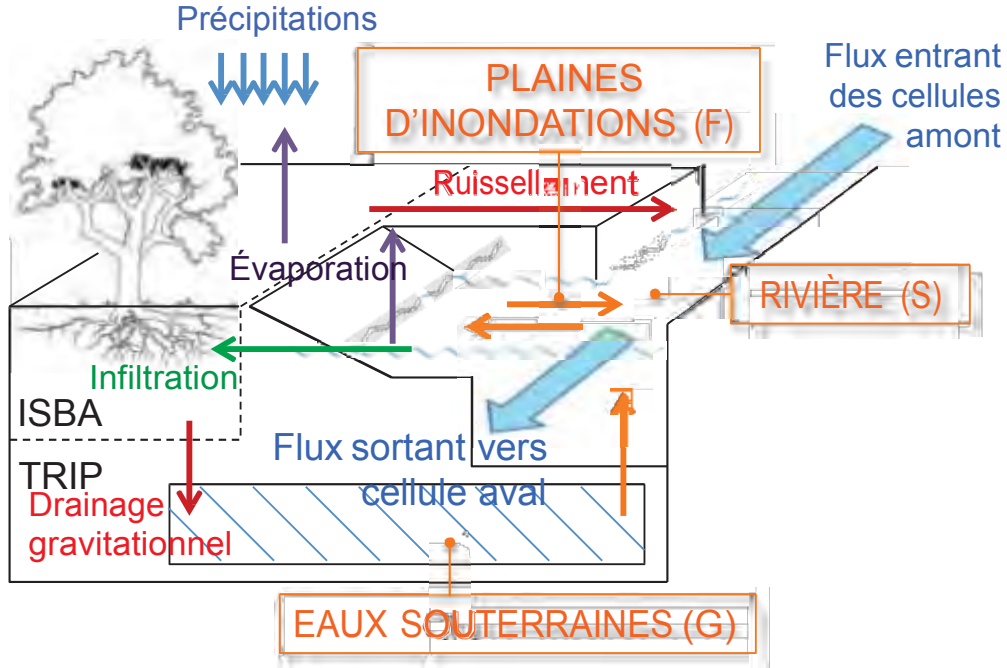


Figure 2.7 – Le système ISBA-TRIP dans une cellule donnée : Le ruissellement de surface ISBA ($Q_{in,ISBA}^S$) est intercepté par le réservoir de surface (S) représentant la rivière et le drainage gravitationnel (Q_{sub}) alimente le réservoir profond (G). Le réservoir de surface S peut aussi se déverser ou recevoir de l'eau du réservoir des plaines d'inondations F. Enfin, l'eau de surface est transférée d'une cellule à l'autre selon le réseau de routage TRIP.

La version actuelle de TRIP, couplée avec ISBA, utilisée dans le cadre de cette thèse, consiste en un système de trois réservoirs : le réservoir de surface S modélisant la rivière, le réservoir profond G pour les eaux souterraines et le réservoir des plaines d'inondations F. Ce système est illustrée en Figure 2.7. Le système d'équations de TRIP permet d'estimer l'évolution temporelle des masses/stocks d'eau [kg] dans chacun de ces réservoirs basée sur le modèle d'une onde cinématique :

$$\frac{dS(t)}{dt} = Q_{in,ISBA}^S(t) + Q_{in,TRIP}^S(t) + Q_{out}^G(t) + Q_{out}^F(t) - Q_{in}^F(t) - Q_{out}^S(t), \quad (2.42)$$

$$\frac{dF(t)}{dt} = Q_{in}^F(t) - Q_{out}^F(t) + (P_F(t) - I_F(t) - E_F(t)), \quad (2.43)$$

$$\frac{dG(t)}{dt} = Q_{sub}(t) - Q_{out}^G(t). \quad (2.44)$$

Seul le réservoir de surface S envoie de l'eau d'une cellule à l'autre selon le réseau de routage TRIP : une cellule peut recevoir l'eau de plusieurs cellules amont et n'est liée qu'à une seule cellule en aval. Pour une cellule donnée, les entrées d'eau sont le flux sortant des cellules amont $Q_{in,TRIP}^S(t)$ [kg.s⁻¹] et le ruissellement de surface propre à la cellule fourni par ISBA $Q_{in,ISBA}^S$ [kg.s⁻¹]. De plus, S reçoit de l'eau en provenance du sol, et plus particulièrement de G, $Q_{out}^G(t)$ [kg.s⁻¹], et peut échanger des masses d'eau avec la plaine d'inondation adjacente F $Q_{out}^F(t) - Q_{in}^F(t)$ [kg.s⁻¹]. Enfin, le débit sortant $Q_{out}^S(t)$ [kg.s⁻¹] est envoyé vers la cellule aval selon le réseau de routage TRIP.

Le drainage gravitationnel ISBA $Q_{sub}(t)$ [kg.s⁻¹] alimente le réservoir profond G qui se déverse à son tour dans la rivière S . Cette modélisation simple ne représente pas la réelle dynamique des eaux souterraines mais seulement le retard avec lequel l'eau drainée contribue au débit de la rivière. Le flux sortant de la rivière est estimé par :

$$Q_{out}^G(t) = \frac{G(t)}{\tau_G} \text{ [kg.s}^{-1}\text{]} \quad (2.45)$$

avec τ_G [s] le temps de résidence. Une version ultérieure de TRIP développée dans Vergnes et al. (2012); Vergnes (2012); Vergnes et al. (2014), non exploitée dans cette thèse, tient compte de la dynamique complexe des eaux souterraines.

Le schéma des plaines d'inondations ne s'active que si la hauteur d'eau dans la rivière h_S [m] excède une certaine hauteur critique H_c [m]. Le flux dans et hors des plaines d'inondations $Q_{in}^F(t)/Q_{out}^F(t)$ [kg.s⁻¹] sont estimés à partir de la formule de Manning et de la pente hydrologique entre la rivière et la plaine d'inondations

$$s_{in} = \frac{\max(0, h_S(t) - H_c - h_F(t))}{(W + W_F(t))/2}, \quad s_{out} = \frac{\max(0, h_F(t) + H_c - h_S(t))}{(W + W_F(t))/2}, \quad (2.46)$$

où $h_F(t)$ [m] est la hauteur d'eau dans F . À partir du stock d'eau dans le réservoir F , la hauteur d'eau $h_F(t)$ est estimée à partir d'une distribution sous-maille, définie à la résolution de 1km × 1km, qui associe le stock d'eau de F à la hauteur correspondante. Il en est de même pour la fraction de cellule inondée f_F . La géométrie du réservoir est supposée rectangulaire avec une largeur $W_F(t)$ qui évolue avec son remplissage et vidage. De plus, avec l'implémentation du schéma d'inondation, TRIP a maintenant une rétro-action sur ISBA. En effet, une partie des précipitations est désormais interceptée par la plaine d'inondation (P_F [kg.s⁻¹]) et l'eau contenue dans F peut infiltrer le sol (I_F [kg.s⁻¹]) et être stockée dans les couches du sol gérées par ISBA. Enfin, une fraction du volume d'eau de F peut aussi s'évaporer (E_F [kg.s⁻¹]).

2.1.4 Paramètres spécifiques de TRIP

S correspond à un tronçon de rivière équivalent à l'intérieur de chaque cellule définie à grande échelle. En effet, dans la réalité, plusieurs tronçons de rivières peuvent co-exister dans la zone géographique de la cellule (voir Figure 2.6). La section de la rivière est rectangulaire et est caractérisée par sa pente s [-], sa largeur W [m], sa profondeur critique H_c [m], sa longueur L [m] et un coefficient de Manning n [-] qui quantifie les frottements et la résistance du fond de la rivière. Ces différents paramètres vont être maintenant présentés mais pour plus de détails, se référer à Decharme et al. (2012).

La longueur de tronçon L [m] est la longueur d'arc entre le centre de la cellule considérée et le centre de la cellule directement en aval de celle-ci, multipliée par un coefficient de méandres pour tenir compte des méandres de la rivière. Dans la définition de TRIP, ce coefficient de méandres a pour valeur, par défaut, $\mu = 1,4$.

La pente s [-] dans une cellule donnée est déduite de la différence d'élévation E de la cellule avec l'élévation de la cellule aval E_{next} . Ces élévations sont elles-même basées sur le Modèle

Numérique de Terrain (MNT) STN-30p défini à une résolution de $0.5^\circ \times 0.5^\circ$ fourni par la base de données ISLSCP2 (http://daac.ornl.gov/ISLSCP_II/islscpii.shtml). STN-30p est lui-même issu du MNT HYDRO1K (<https://lta.cr.usgs.gov/HYDRO1K>) défini globalement à la résolution de 1 km :

$$s = \min \left(10^{-5}, \frac{E - E_{\text{next}}}{L} \right). \quad (2.47)$$

La valeur de 10^{-5} est alors la pente minimum exigée pour garantir l'écoulement.

La largeur W [m] est spatialement distribuée et constante dans le temps. Elle est calculée globalement à partir d'une relation géomorphologique la liant au débit annuel moyen Q_{yr} [$\text{m}^3 \cdot \text{s}^{-1}$] (Knighton, 1998; Arora et al., 1999; Moody and Troutman, 2002) :

$$W = \max \left(20, \beta \times Q_{yr}^{\frac{1}{2}} \right), \quad (2.48)$$

avec $\beta = 32 [\text{m}^{-\frac{1}{2}} \cdot \text{s}^{\frac{1}{2}}]$ pour des bassins équatoriaux tels que l'Amazonie ou le Congo, $\beta = 20$ pour des bassins tropicaux comme le Niger, le Gange et $\beta = 15$ pour des bassins de zones tempérées comme le Mississippi ou l'Amour.

La profondeur de plein bord H_c [m] est déduite de W via la formule non-linéaire

$$H_c = W^{\frac{1}{3}}. \quad (2.49)$$

Le coefficient de Manning n [-] peut être défini comme la résistance ou frottement du lit de la rivière à l'écoulement et est généralement compliqué à estimer. Il prend des valeurs basses (0.025-0.03) pour les cours d'eau naturels profonds et sablonneux et des valeurs plus élevées pour les petits cours d'eau montagneux et rocheux et les plaines d'inondations (0.75-0.1) (Maidment, 1993). De précédentes études globales attribuent une valeur globale au Manning de 0.035 mais il est reconnu que le coefficient de Manning doit être spatialement distribué sur l'ensemble du bassin. Dans TRIP, le coefficient de Manning est distribué selon une simple hypothèse considérant que les cellules en tête de bassin, contenant des cours d'eau peu larges et potentiellement en zone montagneuse (donc ayant un lit rocheux), ont des coefficients de Manning élevés et que cette valeur diminue quand on se rapproche de l'embouchure. Arbitrairement, les valeurs minimales et maximales ont été choisies comme $n_{min} = 0.04$ et $n_{max} = 0.06$:

$$n = n_{min} + (n_{max} - n_{min}) \frac{SO_{max} - SO}{SO_{max} - SO_{min}}, \quad (2.50)$$

avec SO étant le "niveau du cours d'eau" dans le réseau hydrographique (stream order en anglais).

Tout ces paramètres sont critiques pour déterminer la vitesse d'écoulement dans la rivière à partir de la formule de Manning. Ici, l'équation de Manning est appliqué pour un écoulement dans un tronçon ayant une section rectangulaire :

$$v(t) = \frac{s^{\frac{1}{2}}}{n} \left(\frac{W \cdot h_s(t)}{W + 2h_s(t)} \right)^{\frac{2}{3}}, \quad (2.51)$$

où h_s [m] est la hauteur d'eau dans le réservoir de surface estimée à partir du stock d'eau S et de la géométrie rectangulaire du réservoir :

$$h_s(t) = \frac{S(t)}{\rho_w \times LW}, \quad (2.52)$$

avec ρ_w [kg.m⁻³] la masse volumique de l'eau. La vitesse permet ensuite d'estimer le débit de sortie de la cellule Q_{out}^S :

$$Q_{out}^S(t) = \frac{v(t)}{L} S(t). \quad (2.53)$$

Le réseau de routage TRIP est défini globalement et le système ISBA-TRIP peut tourner pour n'importe quel bassin du monde. L'étude de Decharme et al. (2012) a évalué les performances du système ISBA-TRIP sur une période de 12 ans (1993-2004) et sur un ensemble de 12 grands bassins. L'étude compare les débits simulés à l'exutoire du bassin à des débits in situ. L'étude donne des résultats comme 25% de RMSE (erreur quadratique) à Óbidos (Amazonie), 46% de RMSE à Kusur (Lena), 17% de RMSE à Vicksburg (Mississippi) et 68% de RMSE à Pakse (Mekong). La section suivante présente le bassin étudié dans le cadre de ces travaux de thèse.

2.2 Le domaine d'étude : le bassin de l'Amazonie

L'objectif de cette thèse étant de s'intéresser aux grands bassins, on s'est focalisé sur le plus grand bassin versant du monde : le bassin de l'Amazonie (qui sera utilisé pour l'ensemble des travaux présentés dans cette thèse).

En termes de débit moyen (2×10^5 m³.s⁻¹) et d'aire drainée ($6,15 \times 10^6$ km²), l'Amazonie est le plus grand fleuve du monde. Son bassin versant recouvre à peu près 40% de l'Amérique du Sud et le débit sortant à son embouchure représente 30% de la quantité totale d'eau douce déversée dans l'océan Atlantique (Wisser et al., 2010).

Comme illustré en Figure 2.8, l'Amazonie prend sa source dans les Andes Péruviennes sur le sommet du Nevado Mismi. De ce pic à sa confluence avec le Marañon près de Nauta au Pérou, la rivière est généralement appelée la rivière Ucayali. Son débit moyen est alors de $1,6 \times 10^4$ m³.s⁻¹. Le fleuve s'écoule à travers la forêt tropicale brésilienne dans une vaste plaine d'inondation et est alimenté par de nombreux affluents importants : le Japurá, le Purus et le Negro (16% du débit total) à Manaus. À cette confluence, la rivière a atteint 56% de son débit total. Cette portion de la rivière est couramment appelée le Solimões alors que le dernier tronçon, de Manaus à l'embouchure prend le nom d'Amazonie. Cette dernière section reçoit l'eau des principaux affluents du bassin à savoir : le Madeira (17% du débit total), la Tapajós et le Xingu (11% du débit total) (Molinier et al., 1993).

D'un point de vue géologique, le bassin de l'Amazonie est divisé en trois unités morphostructurelles majeures : les boucliers sur la partie Est du bassin (le bouclier Guyannais au Nord et le bouclier Brésilien au Sud), la cuvette amazonienne centrale et la Cordillère des Andes à l'Ouest (Stallart and Edmont, 1983). En ce qui concerne la végétation, le bassin est recouvert essentiellement de forêts tropicales humides (71%) et de savanes (29%) (Sioli, 1984).

L'ensemble du bassin de l'Amazonie est situé dans une zone climatique tropicale humide. Le bassin central est sous une zone de climat équatorial, impliquant de fortes températures de surface, un fort

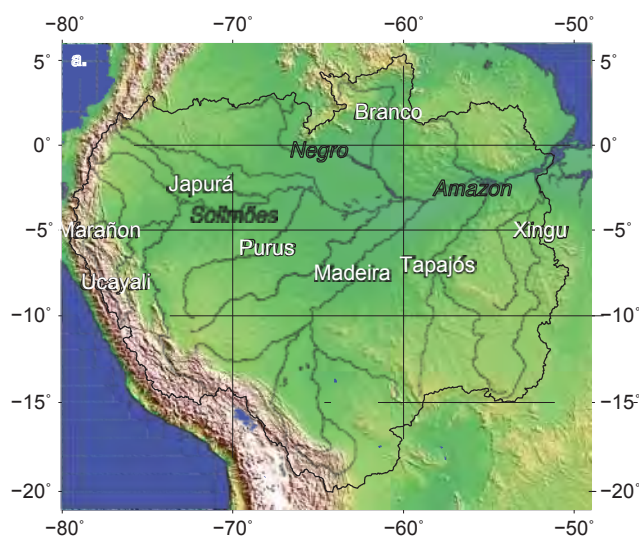


Figure 2.8 – Le bassin de l'Amazon : le cours principal et ses principaux affluents.

taux d'humidité de l'air et d'importantes précipitations liquides. Ainsi, une large zone autour et tout le long du cours principal est inondée chaque année, entraînant un amortissement des pics de débits. Les régions nord et sud du bassin sont sous un climat tropical engendrant une alternance entre une saison sèche et une saison humide. Cependant, à cause de l'inversion des saisons autour de la zone de convergence intertropicale, les saisons de maximum de précipitations ont lieu à des périodes différentes de l'année au nord et au sud du bassin (Meade et al., 1991). Il en résulte que le pic de débit dans les affluents droits arrive quelques mois plus tôt que celui des affluents gauches.

2.3 Analyse de Sensibilité (AS) de la composante "routage de rivière" du modèle ISBA-TRIP

Le système ISBA-TRIP, présenté précédemment, permet ainsi de modéliser et simuler la partie continentale du cycle de l'eau à l'échelle de grands bassins fluviaux, voir aux échelles continentales et globales. Les versions initiales d'ISBA et de TRIP ont profité de nombreuses améliorations pour mieux représenter la réalité comme la prise en compte de plus de processus physiques ou encore la spatialisation des paramètres. Il apparaît évident que pour mieux reproduire les processus observés avec les modèles hydrologiques, la complexité du modèle est souvent augmentée (Grayson and Blöschl, 2001). Dans ce contexte, Sieber and Uhlenbrook (2005) a indiqué que l'Analyse de Sensibilité (AS) peut être un outil puissant pour à la fois identifier les paramètres les plus sensibles et comprendre la structure du modèle hydrologique et ses comportements. Ainsi, nous allons à présent mener une AS du système ISBA-TRIP. De plus, cette thèse s'inscrit dans le cadre de la mission SWOT qui observera plus particulièrement les rivières et les plaines d'inondations et qui sont, de fait, modélisées par le modèle TRIP. On se concentrera alors sur l'AS du modèle TRIP exclusivement.

2.3.1 Choix de la méthode d'Analyse de Sensibilité

Les AS peuvent être utilisées pour deux types d'objectifs : soit elles permettent d'explorer l'espace multidimensionnel des paramètres d'entrée afin de comprendre les sources d'incertitudes modèle (Hornberger and Spear, 1981), soit elles permettent d'identifier les paramètres clés et leur rôle dans la réponse du modèle (Saltelli et al., 2008).

La première option a été largement étudiée, et ce principalement avec la méthode GLUE (Generalizes Likelihood Uncertainty Estimation Beven and Binley, 1992) à la fois en modélisation de bassins versants (Freer et al., 1996) et pour les modèles pluie-débit (Blazkova and Beven, 2004). La méthode du GLUE suppose que, pour un bassin donné, modélisé numériquement par un modèle donné, plusieurs combinaisons de paramètres d'entrée peuvent être indifféremment acceptés comme pouvant reproduire la réponse réelle du modèle. En comparant un ensemble de réponses modèle simulées à des observations de même nature, une valeur de vraisemblance est affectée à chaque combinaison de paramètres d'entrée de l'ensemble. Toutes les simulations classées comme acceptable (selon un critère de vraisemblance choisi par l'utilisateur) sont ré-échantillonnées pour construire la fonction densité de cet ensemble. Les valeurs de vraisemblances sont finalement utilisées pour estimer les incertitudes associées aux prédictions du modèle. Une description détaillée de la méthode GLUE est disponible dans Roux (2004); Garambois (2012). D'autres méthodes sont couramment utilisées comme la RSA (Regional Sensitivity Analysis, Wagener et al., 2003) ou encore la méthode BATEA (BAYesian Total Error Analysis method, Kavetski et al., 2006) qui permettent de calibrer et d'estimer l'incertitude des modèles.

Alternativement, pour pouvoir identifier les paramètres clés du modèle, l'AS peut être considérée comme une étude qui distribue l'incertitude totale de la sortie modèle en différentes contributions liées à l'incertitude des paramètres d'entrée (Saltelli et al., 2008). Les méthodes basées sur l'analyse de la variance (ANOVA pour ANalysis Of VAriance) sont particulièrement adaptées pour ce type d'étude (Francos et al., 2003; Hall et al., 2005; van Griensven et al., 2006). Contrairement au GLUE, l'ANOVA ne nécessite pas l'utilisation d'observations externes. La variance de la sortie du modèle est artificiellement générée en considérant les paramètres d'entrée comme des variables aléatoires. À partir d'un ensemble de combinaison de paramètres d'entrée, l'ANOVA détermine la contribution de chaque paramètre à la variance inconditionnelle de la sortie modèle.

Enfin, il est important de distinguer les études de sensibilités locales des études de sensibilité globales. Ici, les termes "local" et "global" ne sont pas associés à l'échelle du domaine géographique d'étude mais à la partie de l'espace multidimensionnel des paramètres d'entrée qui est considéré. Les études locales se concentrent sur une partie réduite de l'espace des paramètres d'entrée, généralement, les paramètres d'entrée sont perturbés un par un pendant que les autres restent fixes. Les études globales vont étudier l'impact de la totalité des paramètres d'entrée. Les paramètres sont tous perturbés simultanément et des statistiques sont réalisées.

De nombreuses études de sensibilités ont déjà été appliquées à des modèles de routage de rivières (Aronica et al., 1998; Apel et al., 2004; Bates et al., 2004; Pappenberger et al., 2004; Werner et al., 2005). Ces études, essentiellement basées sur la méthode GLUE, sont généralement menées sur des événements courts (inondations) à l'échelle des bassins. Plus particulièrement, Pappenberger et al. (2010) a mené une étude de sensibilité globale du modèle TRIP2 (Ngo-Duc et al., 2007), couplé au modèle HTESSEL (Hydrological component of the Tiled ECMWF Scheme for Surface Exchange over Land, Balsamo et al., 2009), un modèle de surface équivalent à ISBA. L'étude est basée sur une réanalyse de 10 ans (1986-1995) du modèle ECMWF. Cette réanalyse a alors été comparée à

un ensemble de mesures de débits in situ de plus de 400 stations à travers le globe au moyen d'une mesure de performances développée par Pappenberger and Beven (2004). L'ensemble de ces mesures de performances moyennées dans le temps a ensuite permis l'application d'à la fois la méthode GLUE et la méthode ANOVA. La sensibilité de plusieurs paramètres critiques tels que la pente, la longueur du tronçon ou d'autres paramètres impliqués dans la formule de Manning a pu être étudiée.

Au cours de cette thèse, on va chercher à combiner le modèle ISBA-TRIP avec des observations satellites dans un cadre d'assimilation de données ensembliste (voir Chapitre 1, section 1.4.5.3). Ces méthodes nécessitent la génération d'un ensemble dans lequel les paramètres d'entrée sont perturbés selon leur incertitude. La variance induite dans les sorties du modèle est associée à l'erreur modèle et celle-ci doit être assez importante pour permettre à l'assimilation de données d'apporter une correction (sur les paramètres et/ou sur les sorties) à partir des observations. De ce fait, avant de mettre en place l'algorithme d'assimilation de données, il est nécessaire d'identifier les paramètres les plus sensibles dans les sorties du modèle TRIP. Ces paramètres seront alors ceux que l'assimilation de données sera en mesure de corriger. On choisit alors d'appliquer une méthode globale de type ANOVA pour l'AS du modèle TRIP car sa configuration est très similaire à la méthode d'assimilation de données que l'on souhaite appliquer par la suite.

2.3.2 Formulation de la méthode ANOVA

La formulation de la décomposition ANOVA est basée sur Sobol (2001). Tout d'abord, d'un point de vue purement déterministe, la sortie modèle Y peut être écrite comme une fonction de n paramètres indépendants et incertains $p_1, p_2, \dots, p_n$:

$$Y = \mathcal{M}(p_1, p_2, \dots, p_n), \quad (2.54)$$

où $\mathcal{M}$ est l'opérateur modèle et les paramètres $p_1, p_2, \dots, p_n$ sont définis sur l'ensemble à n dimension Ω^n , l'ensemble de toutes les valeurs possibles des paramètres. On s'assure au préalable que l'intervalle de variation de ces paramètres correspond à la supposée incertitude dans la valeur du paramètre. De ce fait, dans l'AS qui va suivre, les paramètres $p_1, p_2, \dots, p_n$ ne prendront que des valeurs réalistes.

Maintenant, soit $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un jeu de paramètres normalisés, c'est-à-dire un jeu de paramètre $(p_1, p_2, \dots, p_n)$ normalisés entre 0 et 1. L'application qui lie les paramètres normalisés à la sortie Y est définie par :

$$Y = \mathbf{f}(\mathbf{X}) = \mathbf{f}(X_1, X_2, \dots, X_n), \quad (2.55)$$

où $\mathbf{f}$ est aussi l'opérateur modèle mais prenant les paramètres normalisés en entrée.

La décomposition de Sobol suppose que Y peut être écrit comme une somme de fonctionnelles à valeurs réelles dont le nombre de paramètres d'entrée augmente de 1 à n telle que :

$$Y = \mathbf{f}_0 + \sum_{i=1}^n \mathbf{f}_i(X_i) + \sum_{i=1}^n \sum_{j>i}^n \mathbf{f}_{i,j}(X_i, X_j) + \dots + \mathbf{f}_{1,2,\dots,n}(X_1, X_2, \dots, X_n). \quad (2.56)$$

Sous l'hypothèse que $\mathbf{f}$ est intégrable sur l'hypercube unité de dimension n , on suppose que l'hypothèse ci-dessous est vérifiée :

$$\forall 1 \leq s \leq n, \forall 1 \leq i_1 \leq \dots \leq i_s \leq n, \quad \mathbf{f}_{i_1,\dots,i_s}(X_{i_1}, \dots, X_{i_s}) dX_{i_1} \dots dX_{i_s} = 0. \quad (2.57)$$

La fonction $f_{i_1, \dots, i_s}$ peut prendre n'importe quelle valeur réelle. Donc, l'intégrale dans l'Eq. 2.57 signifie que la fonction $f_{i_1, \dots, i_s}$ est à valeur moyenne nulle. On déduit de l'Eq. 2.57 que les fonctionnelles sont orthogonales et peuvent être exprimées comme une intégrale de $f(\mathbf{X})$. En effet :

$$f(\mathbf{X})d\mathbf{X} = f_0, \quad (2.58)$$

$$f(\mathbf{X})d\mathbf{X}_{\sim i} = f_0 + f_i(X_i), \quad (2.59)$$

où $d\mathbf{X}_{\sim i}$ signifie que l'intégration est faite sur tous les paramètres sauf X_i . Consécutivement :

$$f(\mathbf{X})d\mathbf{X}_{\sim i, j} = f_0 + f_i(X_i) + f_j(X_j) + f_{i, j}(X_i, X_j). \quad (2.60)$$

Enfin, si f est aussi à carré intégrable sur l'hypercube unité de dimension n , il en est de même pour toutes les fonctionnelles et :

$$f^2(\mathbf{X})d\mathbf{X} - f_0^2 = \sum_{i=1}^n \sum_{i_1 < \dots < i_s} f_{i_1, \dots, i_s}^2 dX_{i_1} \dots dX_{i_s}. \quad (2.61)$$

Maintenant, on sort du cadre déterministe pour tenir compte concrètement des incertitudes des paramètres. Pour cela, on considère alors les paramètres d'entrée $\mathbf{X}$ comme des variables aléatoires uniformément distribuées sur l'hypercube unité de dimension n . Ceci signifie que chaque paramètre suit une loi uniforme continue entre 0 et 1. Les paramètres d'entrée étant des variables aléatoires, le sortie modèle $Y = f(\mathbf{X})$ et toutes les fonctionnelles $f_{i_1, \dots, i_s}$ sont aussi des variables aléatoires. Comme pour toutes variables aléatoires continues, il en découle que les intégrales de f et $f_{i_1, \dots, i_s}$ représentent des espérances et les intégrales de f^2 et $f_{i_1, \dots, i_s}^2$ représentent des variances. Par exemple,

- L'espérance de la sortie modèle est donnée par :

$$E(Y) = \int f(\mathbf{X})d\mathbf{X} = f_0, \quad (2.62)$$

- L'espérance conditionnelle de la sortie modèle Y étant donné les facteurs de premier ordre est définie comme :

$$E(Y|X_i) = \int f(\mathbf{X})d\mathbf{X}_{\sim i} = f_0 + f_i(X_i). \quad (2.63)$$

L'Eq. 2.63 signifie que l'espérance de Y est calculée sur toutes les valeurs possibles de $\mathbf{X}_{\sim i}$ tout en gardant X_i fixé.

- Et une définition similaire peut être appliquée pour les facteurs d'ordres supérieurs :

$$E(Y|X_i, X_j) = f_0 + f_i(X_i) + f_j(X_j) + f_{i, j}(X_i, X_j). \quad (2.64)$$

Les méthodes d'AS basées sur l'analyse de la variance ont pour objectif d'analyser la variance inconditionnelle de la sortie d'un modèle en fonction d'un ensemble de combinaisons de paramètres d'entrée où tous les paramètres sont perturbés simultanément (Efron and Stein, 1981; Sobol, 1993). Ces méthodes ont l'avantage d'être efficaces sur tout type de modèle même nonlinéaires et non monotones (Saltelli and Bolado, 1996). La variance totale de la sortie modèle Y est alors formulée par :

$$\text{Var}[Y] = \int_{[0, 1]^n} f^2(\mathbf{X})d\mathbf{X} - f_0^2 = \sum_{i=1}^n \sum_{i_1 < \dots < i_s} f_{i_1, \dots, i_s}^2 dX_{i_1} \dots dX_{i_s}. \quad (2.65)$$

On peut réécrire Eq. 2.65 en se basant sur Eq. 2.56 et en utilisant Eq. 2.57. La variance totale (ou inconditionnelle) est alors écrite sous sa forme de décomposition de l'ANOVA :

$$\text{Var}[Y] = \sum_{i=1}^n V_i + \sum_{i=1}^n \sum_{j>i}^n V_{i,j} + \dots V_{1,2,\dots,n}, \quad (2.66)$$

où

$$V_i = \text{Var}[f_i] = \int_0^1 f_i^2(X_i) dX_i \quad (2.67)$$

est appelé l'effet principal du paramètre X_i sur Y . L'effet principal représente la réduction de la variance totale à laquelle on s'attendrait si le paramètre X_i pouvait être fixé. Et,

$$V_{i,j} = \text{Var}[f_{i,j}] = \int_0^1 \int_0^1 f_{i,j}^2(X_i, X_j) dX_i dX_j \quad (2.68)$$

est la covariance engendrée par les effets combinés des paramètres X_i et X_j sur Y . On dit que deux paramètres interagissent quand leurs effets sur Y ne peuvent pas être uniquement expliqués par V_i et V_j . Il en est de même pour les termes d'ordre supérieur.

Finalement, l'objectif de l'ANOVA est d'estimer une groupe d'Indices de Sensibilité (SI pour Sensitivity Indices) - un pour chaque paramètre - qui représente la contribution du paramètre associé dans la variance totale de la sortie modèle. Les SI sont en réalité déduits des variances partielles définies en Eq. 2.67-2.68. Les indices de sensibilités $S_{i_1, \dots, i_s}$ sont déterminés en divisant la variance partielle associée $V_{i_1, \dots, i_s}$ par la variance totale :

$$S_{i_1, \dots, i_s} = \frac{V_{i_1, \dots, i_s}}{\text{Var}[Y]}. \quad (2.69)$$

Il apparaît évident que :

$$\sum_{i=1}^n S_i + \sum_{i=1}^n \sum_{j>i}^n S_{i,j} + \dots + S_{1,2,\dots,k} = 1. \quad (2.70)$$

L'essentiel de l'application de la méthode ANOVA réside dans le calcul des variances partielles définies en Eq. 2.67-2.68. Il existe alors plusieurs méthodes pour les calculer. Par exemple la méthode de Sobol (Sobol, 1993), ou l'(E)FAST (simple ou Extended Fourier Amplitude Sensitivity Test) (Cukier et al., 1973; Francos et al., 2003). La différence principale entre la méthode de Sobol et FAST repose sur l'approche choisie pour calculer les intégrales multi-dimensionnelles définissant les variances. Alors que Sobol utilise une procédure d'intégration basée des méthodes de Monte Carlo, la méthode FAST recherche des fonctions de type sinusoïdale. Reusser et al. (2011) a testé les trois approches d'échantillonnage (FAST, EFAST et Sobol) et a trouvé des résultats comparables entre les trois méthodes. Même si ces méthodes sont populaires, elles requièrent un grand nombre de runs modèle et ont un coût de calcul élevé pour estimer correctement les SI. En effet, pour n'importe quelle approche choisie, l'évaluation des intégrales en Eq. 2.67-2.68 est réalisée stochastiquement à partir d'un ensemble de réalisations de sortie modèle. D'abord, un ensemble de jeux de paramètres d'entrée $\mathbf{X}_e = \{\mathbf{X}^{[1]}, \mathbf{X}^{[2]}, \dots, \mathbf{X}^{[n_e]}\}$, avec n_e la taille de l'ensemble, est généré à partir d'un générateur de suites pseudo-aléatoires. Ensuite, l'ensemble des sorties modèle est constitué en faisant tourner le modèle f n_e fois avec, pour chaque run modèle, un jeu de paramètres différent.

L'étape suivante consiste à évaluer les SIs à partir de l'ensemble des sorties modèle. Actuellement, certaines études se concentrent sur le calcul direct des double-intégrales multidimensionnelles définissant les SIs (voir Eq. 2.65, 2.67, 2.68) tout en essayant de réduire le coût de calcul (Jansen, 1999; Saltelli et al., 2010). D'autres alternatives ont cependant réussi à se séparer du problème de dimension du problème (liée au nombre de paramètres) en utilisant des représentations simplifiées des modèles (HDMR - High Dimensional Model Representation) (Rabitz et al., 1999; Ratto et al., 2007).

Pour l'étude de sensibilité mise en place pour le modèle TRIP, on utilisera l'approche développée par Ratto et al. (2007) appelée *SDP-metamodeling* (SDP pour State-Dependent-Parameter). Elle est conçue pour estimer les fonctionnelles f_i jusqu'à l'ordre 2, c'est-à-dire les f_i s et les f_{ij} s de l'Eq. 2.56. Cette étape est réalisée via la routine *ss-anova* (<https://ec.europa.eu/jrc/en/econometric-statistical-software#SS-ANOVA-R>). Une fois que les termes $f_i(X_i)$ sont estimés pour l'ensemble tout entier, les S_i associés sont directement déduits à partir de l'estimateur construit par Doksum and Samarov (1995). L'avantage de cette méthode est son adaptabilité à n'importe quelle méthode d'échantillonnage pour l'ensemble des paramètres d'entrée tout en fournissant des résultats rapides, justes et non-biaisés. La routine fournit, en plus des SIs, un coefficient de corrélation R^2 mesurant la similitude entre le modèle original et le modèle émulé.

Par ailleurs, Sieber and Uhlenbrook (2005); Reusser et al. (2011) ont introduit deux extensions temporelles des méthodes d'analyse de variance, à savoir : les TEDPAS (TEmporal Dynamics of PArameter Sensitivity) et les TIGER (Timebseries of Grouped Error) dans le but d'effectuer une analyse de sensibilité temporelle de n'importe quelle sortie modèle ayant une variabilité temporelle comme le débit des fleuves. Ce type de méthode vise à identifier les paramètres dominants pendant des phénomènes hydrologiques ponctuels et à mettre en évidence l'interaction des paramètres. Par exemple, Garambois et al. (2013) a mené une étude de sensibilité temporelle du modèle hydrologique distribué MARINE pendant des événements de crues éclair à partir d'une ANOVA basée sur la méthode de Sobol. Autre exemple, Guse et al. (2014) a utilisé une ANOVA temporelle basée sur la méthode FAST pour identifier les paramètres responsables des mauvaises performances du modèle SWAT pendant les périodes de sécheresses.

Pour cette étude, on utilisera l'approche de Ratto et al. (2007) pour estimer exclusivement les SIs de premier ordre. Si les SIs de premier ordre suffisent à expliquer la variance totale de la sortie modèle, cela signifie que le modèle est purement additif. Ce résultat sera alors confirmé par un coefficient de corrélation R^2 proche de 1 et des effets d'interactions (calculés via $1 - \sum_i S_i$) très bas. Dans le cas contraire, cela indiquera que les effets d'interactions ne sont pas négligeables. Pour plus de détails, se référer à Ratto et al. (2007). De plus, on réalisera une AS temporelle comme celles proposées par Garambois et al. (2013) et Reusser and Zehe (2011). Concrètement, au lieu de générer un ensemble de sorties scalaires de modèle, on génère un ensemble de séries temporelles de sorties modèle et on applique la méthode ANOVA à chaque pas de temps de sortie modèle pour générer une série temporelle de SI.

2.3.3 Application de l'ANOVA au modèle TRIP

Nous allons maintenant détailler l'application de la méthode ANOVA au cas du modèle TRIP.

2.3.3.1 Configuration générale

Pour cette étude, l'AS est basée sur un échantillonnage de l'espace des paramètres de $n_e = 1024$ tirages (cette valeur a été choisie selon Garambois et al. (2013) et Ratto et al. (2007), pour avoir un

coût en calcul raisonnable). Pour générer l'ensemble des jeux de paramètres d'entrée normalisés $\mathbf{X}_e$, on utilise le générateur de suites pseudo-aléatoire LP_r développé par Sobol (Sobol, 1967). Les suites de Sobol utilisent la base 2 pour former des partitions de plus en plus fines de l'intervalle unité - [0, 1] et les ré-ordonnent ensuite selon chaque dimension. La routine LP_r visite l'intégralité de l'hypercube unité (elle peut générer jusqu'à $2^{30} - 1$ points de l'hypercube unité jusqu'à la dimension 52) et ainsi, l'intervalle de définition de chaque paramètre est exploré.

Un élément de l'ensemble des sorties modèle $Y^{[I]}$ consiste en une série temporelle journalière sur la période de 2008 à 2010 d'une sortie de TRIP sur l'ensemble du bassin de l'Amazone. On étudiera plusieurs sorties principales de TRIP pour la modélisation des fleuves et rivières, à savoir : la profondeur d'eau dans la rivière (Eq. 2.52), le débit (Eq. 2.53) et enfin l'anomalie de hauteur d'eau (Eq. 2.73), dérivée de la profondeur. L'AS temporelle est ainsi menée sur une période de trois ans. Les années 2006 et 2007 sont utilisées comme *spin up* pour éviter les effets de la condition initiale et des états transitoires dans la variance de la sortie modèle et l'AS n'est pas effectuée sur ces deux années. L'AS est menée sur la période de 2008 à 2010. 2009 a été une année particulièrement humide pour l'Amazone. En effet, les inondations de mi-Juin ont atteint leur plus haut niveau depuis 50 ans. Au contraire, 2010 a été une année plutôt sèche et le bassin a subi une très forte sécheresse. Enfin, l'année 2008 est considérée comme une année moyenne. De ce fait, mener l'AS sur cette période 2008 à 2010 permettra d'étudier tous les types d'années hydrologiques.

Les nombreuses simulations élémentaires d'ISBA-TRIP nécessaires pour générer l'ensemble des sorties ont été menées sur la plateforme de calcul hautes performances CALMIP (<https://www.calmip.univ-toulouse.fr/spip/>) et sur son super ordinateur EOS. Par ailleurs, ISBA-TRIP est exécuté en mode découplé et est forcé par les données atmosphériques du Global Soil Wetness Project 3 (GSWP3, <http://hydro.iis.u-tokyo.ac.jp/GSWP3>). Ce projet se compose de trois simulations à l'échelle globale ayant pour objectif d'étudier les effets des changements climatiques à long terme (sur 100 ans) sur le cycle énergie-eau-carbone et sur leurs interactions. On utilise alors les conditions atmosphériques à 3 heures de ces simulations. Ces données atmosphériques sont issues d'une réduction d'échelle dynamique de la Réanalyse Globale du 20^{ème} siècle à 2° (Compo et al., 2011) qui assimile différentes observations atmosphériques dans le modèle opérationnel de l'NCEP (National Centers for Environment Prediction) : le Climate Forecast System (CFS).

2.3.3.2 Choix des paramètres

Pour cette étude, la totalité des paramètres TRIP est considérée : le coefficient de Manning de la rivière n et de la plaine d'inondation n_F , le temps de résidence des eaux souterraines τ_G , la pente du lit de la rivière s , la largeur de la rivière W , la profondeur critique H_c et le coefficient de méandrement de la longueur du tronçon μ . Le Tableau 2.1 donne un récapitulatif des valeurs nominales de ces paramètres.

L'ANOVA requiert que l'on perturbe ces paramètres tous simultanément pour générer un ensemble de jeux de paramètres possibles et générer la variance de la sortie modèle. Or, mis à part le temps de résidence τ_G et le coefficient de méandrement μ qui sont des paramètres constants, tous les autres paramètres sont spatialement distribués sur le bassin d'étude. Pour éviter la sur-paramétrisation en perturbant chaque cellule une à une, on perturbe l'ensemble du champ de paramètres en lui appliquant un facteur multiplicatif constant spatialement. De plus, une attention particulière a été portée à la perturbation du champ de pente s . Selon Pappenberger et al. (2010), le champ de pente s est d'abord perturbé en lui appliquant un facteur en puissance s_{pow} puis, un facteur multiplicatif s_{mult} . Les différents facteurs multiplicatifs et de puissance ont une valeur nominale de 1 et sont définis sur

Paramètres TRIP	Gamme/Valeur nominale
τ_G	30 jours
W	20-15000 m
H_c	3-17 m
n	0,04-0,06
s	$2 \cdot 10^{-4} - 5 \cdot 10^{-2}$
μ	1,4
n_F	0,03-0,09

Table 2.1 – Nom et valeur/gamme nominale des paramètres TRIP.

Param. SA	Dir/Mult/Pow	Val. nom.	Pert. $\pm E\%$	Param. TRIP	Gamme de variations SA
τ_G	Dir	30 jours	-	τ_G	15 - 60 jours
W_{mult}	Mult	1	$\pm 50\%$	W	10-22500
$H_{c,mult}$	Mult	1	$\pm 50\%$	H_c	1,5-25,5
n_{mult}	Mult	1	$\pm 75\%$	n	0,01-0,105
s_{mult}	Mult	1	$\pm 20\%$	s	$3 \cdot 10^{-5} - 7 \cdot 10^{-2}$
s_{pow}	Pow	1	$\pm 20\%$	s	gamme obtenue avec simult. s_{mult} et s_{pow}
μ	Dir	1,4	$\pm 30\%$	L	1,0-1,8
$n_{F,mult}$	Mult	1	$\pm 75\%$	n_F	0,007-0,15

Table 2.2 – Intervalles de définitions des paramètres de perturbations utilisés pour l'AS : la 2^{ème} colonne donne la nature du paramètre : "Dir" pour valeur directe (ces champs ne sont pas spatialisés), "Mult" pour un facteur multiplicatif et "Pow" pour un facteur en puissance. La 3^{ème} colonne donne la valeur nominale du paramètre. La 4^{ème} colonne donne la gamme de perturbations de chaque paramètre, précisés dans la 5^{ème} colonne. Et la 6^{ème} colonne présente la gamme complète de variation du paramètre physique d'entrée TRIP après perturbations.

un intervalle de $\pm E\%$ autour de la valeur nominale. La valeur de E est choisie de telle sorte que la gamme des valeurs perturbées soit représentative de l'incertitude du paramètre Galtelli et al. (2009). On s'est alors basé sur des tables de valeurs existantes pour les coefficients de Manning (Maidment, 1993), des études antérieures pour la pente et le temps de résidence (Decharme et al., 2012; Paris et al., 2016) et des bases de données issues d'images satellites optiques (Yamazaki et al., 2014) pour la largeur, pour fixer la valeur de E pour chaque paramètre et obtenir la gamme de variations des paramètres donnée en Tableau 2.2.

Le Tableau 2.3 résume l'ensemble des expériences d'AS menées sur le bassin de l'Amazonie. La première étude de sensibilité est nommée AMA8 - "AMA" pour Amazonie et "8" pour huit paramètres perturbés (voir Tableau 2.2). On remarque que, mis à part la pente du lit de la rivière s à laquelle deux coefficients de perturbations sont associés, tout les autres paramètres TRIP ne sont associés qu'à un seul paramètre de perturbation multiplicatif. Pour ces paramètres en particulier et pour simplifier les écritures, on fera référence à la sensibilité à un paramètre multiplicatif donné directement avec le paramètre physique TRIP associé (colonne 5 dans la Tableau 2.2). Cette première AS nous donnera un premier aperçu de la sensibilité de TRIP à ses paramètres.

À partir de ces résultats, on effectuera une deuxième AS, nommée AMA45, dont l'objectif sera d'étudier la sensibilité régionalisée des paramètres (ignorée dans la première étude car les paramètres

Simulation	Bassin	Nbre de paramètres	Particularités
AMA8	Amazone	8, voir Tableau 2.2	-
AMA45	Amazone	45	Étude régionalisée
AMA8_ H_c	Amazone	8, voir Tableau 2.2	$H_{c,mult} \in [0, 3, 1, 2]$

Table 2.3 – Liste des différentes simulations d’AS.

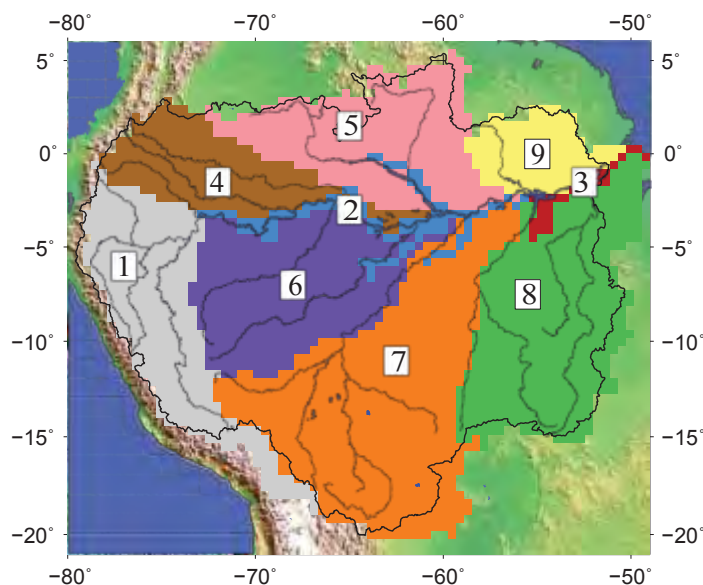


Figure 2.9 – Régions hydro-géomorphologiques définies sur le bassin de l’Amazonie.

étaient identiquement perturbés sur l’ensemble du bassin). Un sous-ensemble (5 sur les 8 initiaux) des paramètres d’entrée, considérés comme les plus critiques pour la compréhension du comportement de TRIP, est alors choisi et les paramètres sont perturbés distinctement sur différentes régions géographiques du bassin. Dans ce contexte, le bassin de l’Amazonie a été divisé en 9 zones hydro-géomorphologiques. L’intérêt d’une telle subdivision est d’étudier l’impact séparé des paramètres de chaque zone. Ces zones ont été construites selon : (1) des arguments hydrologiques où le cours principal est séparé des principaux affluents et les bassins versants des principaux affluents eux-même ont leur propre zone, (2) des arguments géologiques où l’on distingue les 3 principales unités morphostructurales du bassin. Ces zones, illustrées en Figure 2.9, sont alors :

- La partie Andine du bassin, en amont, jusqu’à la ville d’Iquitos au Pérou (zone 1),
- Le cours principal d’Iquitos à Óbidos (le Solimões) (zone 2),
- Le cours principal d’Óbidos à l’exutoire (zone 3),
- Les affluents gauches du Napo au Japurá inclus (zone 4),
- Les affluents gauches de la confluence Japurá à Óbidos, incluant le Negro (zone 5),
- Les affluents droits d’Iquitos à la confluence du Purus à la ville d’Anamã (zone 6),
- Les affluents droits d’Anamã à Óbidos, incluant le Madeira (zone 7),
- Les affluents droits se déversant dans le cours principal en zone 3, incluant le Tapajós et le Xingu (zone 8),
- Les affluents gauches se déversant dans le cours principal en zone 3 (zone 9).

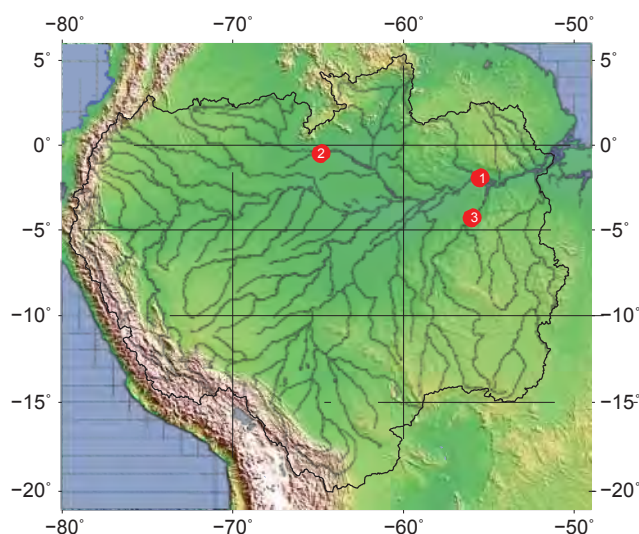


Figure 2.10 – Localisation sur le bassin de l’Amazonie des stations in situ illustrant les résultats de l’analyse de sensibilité : (1) Óbidos (zone 3), (2) Serrinha (zone 2), (3) Itaituba (zone 3).

Enfin, la troisième AS, dénommée $AMA8_{H_c}$, est basée sur la même configuration que $AMA8$. Seulement, la gamme de variation du paramètre $H_{c,mult}$ est différente, de telle sorte que les valeurs minimales et maximales de la hauteur critique H_c soient plus basses et engendrent plus d’inondations. L’intérêt de cette expérience est de voir plus clairement l’impact des fortes inondations dans l’AS.

La configuration des AS est la même pour les trois analyses en termes de nombre de membres dans les ensembles ($n_e = 1024$), de forçages atmosphériques (GSWP3) et runs ISBA-TRIP (trois ans de simulation de 2008 à 2010). Par ailleurs, ISBA-TRIP estime les hauteurs d’eau, les débits et les fractions inondées à chacune des 2028 cellules TRIP formant le bassin de l’Amazonie. Les résultats de l’AS ne seront détaillés que pour un nombre réduit de cellules correspondant à des stations in situ connues et localisées à des cellules stratégiques pour comprendre les résultats. Les stations sélectionnées sont :

1. La station d’Óbidos, située sur le cours principal, à l’entrée de la zone hydro-géomorphologique 3 (station 1 dans la Figure 2.10),
2. La station de Serrinha, sur le Negro, à l’entrée de la zone 2 et recevant des flux entrants de la zone 5 (station 2 dans la Figure 2.10),
3. La station d’Itaituba, sur le Tapajós, à l’entrée de la zone 3 et recevant des flux entrants de la zone 8 (station 3 dans la Figure 2.10).

Plus de pixels ont été analysés dans l’article donné en Section C. Les trois pixels cités ci-dessus sont cependant très représentatifs des différents cas analysés dans le papier.

2.3.4 Résultats

Nous allons à présent présenter les résultats de ces analyses de sensibilité.

Station	AMA8		AMA45		AMA8- H_c	
	R^2	$1 - \sum_i S_i$	R^2	$1 - \sum_i S_i$	R^2	$1 - \sum_i S_i$
Óbidos	0,95	0,06	0,97	0,04	0,93	0,06
Serrinha	0,95	0,05	0,96	0,05	0,95	0,06
Itaituba	0,94	0,07	0,95	0,06	0,93	0,07

Table 2.4 – Statistiques des AS sur les hauteurs d'eau. Pour chaque station et chaque simulation : Coefficients de corrélation du métamodèle moyenné sur la période d'étude (R^2) et effets d'interactions moyennés sur la période d'étude ($1 - \sum_i S_i$).

2.3.4.1 Résultats sur les hauteurs d'eau

On rappelle d'abord la définition de la hauteur d'eau h_S dans le réservoir de la rivière S de section rectangulaire $W \times H_c$ et de longueur L :

$$h_S(t) = \frac{S(t)}{\rho L W}. \quad (2.71)$$

Le Tableau 2.4 présente, pour chaque station d'étude et chaque expérience, le coefficient de corrélation du métamodèle R^2 ainsi que les effets d'interaction, estimés par $1 - \sum_i S_i$. Le premier quantifie la ressemblance entre les sorties-modèle réelles $\mathbf{Y}_e$ et les sorties-modèle émulées par la routine ss-anova (Ratto et al., 2007) qui estime les SIs. Le second donne l'importance des effets d'interaction de plusieurs paramètres dans la variance totale par rapport aux effets principaux étudiés. Dans le cas des hauteurs d'eau TRIP, le coefficient R^2 proche de 1 et les effets d'interactions proche de 0 indique que la hauteur d'eau simulée par TRIP a un comportement quasi-additif.

Les Figures 2.11-2.14 illustrent les indices de sensibilité de premier ordre à chacune des stations et chacune des simulations. Pour les sensibilités aux hauteurs d'eau, des caractéristiques communes aux trois simulations peuvent être observées.

Les ensembles des hauteurs d'eau simulées présentent une large dispersion et le 1er et le 9ème décile sont clairement distincts. Il peut alors être remarqué que les hauteurs d'eau ont une importante sensibilité aux paramètres de TRIP. De manière générale, les séries temporelles de SI ne montrent pas de sensibilités inter-annuelles et 4 paramètres sont systématiquement activés : en premier, le coefficient de Manning de la rivière n (courbe violette), suivi de près par le coefficient en puissance de la pente s_{pow} (courbe verte) ; ensuite, la largeur de la rivière W (ligne orange) et le temps de résidence des eaux souterraines τ_G (ligne bleue). Enfin, ponctuellement, le coefficient de méandrement de la longueur μ s'active aussi (ligne rouge en Fig. 2.11c et Fig. 2.14c). En résumé, S_n , $S_{s_{pow}}$ et S_W expliquent respectivement 40%, 35% et 20% de la variance de la sortie modèle et S_{τ_G} et S_μ expliquent aussi un faible pourcentage de la sensibilité pendant la saison sèche. Pour résumer les paramètres les plus sensibles par ordre d'importance :

$$n > s_{pow} > W > \tau_G >> \mu \quad (2.72)$$

Si cet ordre d'importance entre les paramètres se retrouvent pour toutes les trois sections, leurs variations saisonnières varient d'une station à l'autre, voir pour AMA8 en Figure 2.11. Les sensibilités à Óbidos - et sur l'ensemble du cours principal en général (pour plus de figures voir l'article en Section C)

- montrent très peu variations intersaisonnières : seulement 3 paramètres - n , s_{pow} et W - sont actifs et la sensibilité à ces paramètres est quasi-constantes dans le temps. Ces 3 paramètres sont directement impliqués dans l'équation de Manning, voir Eq. 2.51, qui estime la vitesse d'écoulement dans le tronçon de rivière. Cette vitesse d'écoulement est utilisée dans l'équation différentielle décrivant le comportement du réservoir S et influe sur le stock d'eau dans le réservoir S . Ainsi, ces trois paramètres représentent la contribution de S dans la détermination de h_S . Les sensibilités à Serrinha - et sur l'ensemble des cellules des affluents gauche (pour plus de figures voir l'article en Section C) - présentent un comportement similaire mais plus bruité qui peut être associé aux événements pluvieux. Il s'avère que les cellules dans les affluents gauches ont une aire de drainage plus petites que celles des cellules du cours principal et sont alors plus sensibles aux événements locaux de pluies. De plus, on observe de nombreux pics de sensibilité de W (ligne orange) synchronisés avec des élargissements soudains de dispersion dans l'ensemble des hauteurs d'eau simulées. Ceci montre que lors de changements de régime soudain, la géométrie du réservoir contrôle la dynamique des hauteurs d'eau. Enfin, les sensibilités à Itaituba ont une certaine variabilité intersaisonnière : en saison de hautes eaux, les paramètres n et s_{pow} dominent comme pour les autres stations mais en saison de basses eaux, la sensibilité au paramètre τ_G augmente et surpasse celles des autres paramètres. Ce paramètre représente la contribution au débit de base des eaux souterraines vers la rivière. Pour cette station, la différence entre les hautes eaux et basses eaux est plus marquée dans l'ensemble (panneaux du haut dans la Figure 2.14c que pour les autres stations, ce qui semble confirmer que le débit de base est plus dépendant des échanges avec le sol et donc plus sensible à τ_G . La méthode ANOVA utilisée ici détermine les sensibilités comme des fractions de la variance totale. Ainsi, la contribution du réservoir souterrain prend des proportions plus importantes en saison de basses eaux quand les autres processus (précipitations, ruissellement) sont moins importants. Ce schéma temporel des sensibilités est bien plus observable pour les affluents droits que pour les affluents gauches et ceci est expliqué par des aires drainées plus grandes et aussi de plus grandes amplitudes entre le maximum et le minimum des ensembles de hauteurs d'eau simulées.

À présent, la simulation AMA45 permet d'étudier l'impact régionalisé des 5 paramètres les plus importants que sont : le coefficient de Manning de la rivière n , le coefficient en puissance de la pente s_{pow} , la largeur de la rivière W , le temps de résidence des eaux souterraines τ_G et le coefficient de méandrement μ . Les résultats sont illustrés en Figure 2.12-2.13. Les sensibilités des paramètres d'étude, organisées par zone d'appartenance des paramètres montrent clairement que les paramètres géométriques (W et μ) et morphologiques (n et s_{pow}) locaux - au sens des paramètres de la même zone que la cellule d'étude - sont activés ; et que pour le paramètre τ_G , le paramètre des zones amont à la cellule d'étude est activé. Avec cette nouvelle perspective des paramètres régionalisés, il apparaît que la géomorphologie locale du réservoir gouverne la hauteur d'eau en contrôlant la quantité d'eau sortant de la cellule. Entre-temps, en période de basses eaux, la hauteur d'eau est aussi sensible aux entrées d'eau de l'amont.

Enfin, une courte remarque sur l'expérience AMA8_ H_c . Cette expérience est identique à l'expérience AMA8 à ceci près que le paramètre H_c , la profondeur du lit de la rivière, est perturbée de telle sorte que les valeurs minimales et maximales soient plus basses que celles de l'expérience AMA8, voir Tableau 2.3. Comme indiqué précédemment, ces valeurs plus basses de la hauteur critique (voir Section 2.1.3 et Eq. 2.49) vont permettre d'activer plus fréquemment et plus durablement les plaines d'inondations. Malgré cela, il n'y a pas de différences notoires entre les résultats des deux simulations. En regardant dans le détail les sensibilités à Óbidos (Fig. 2.14a), on remarque une très légère activation du paramètre H_c (ligne jaune) précédant celle du paramètre τ_G . Cette activation a lieu alors que les hauteurs d'eau sont en train de baisser suite à la saison de hautes eaux et peut être associée au vidage des plaines d'inondations (réservoir F) dans la rivière (réservoir S). En effet, on rappelle

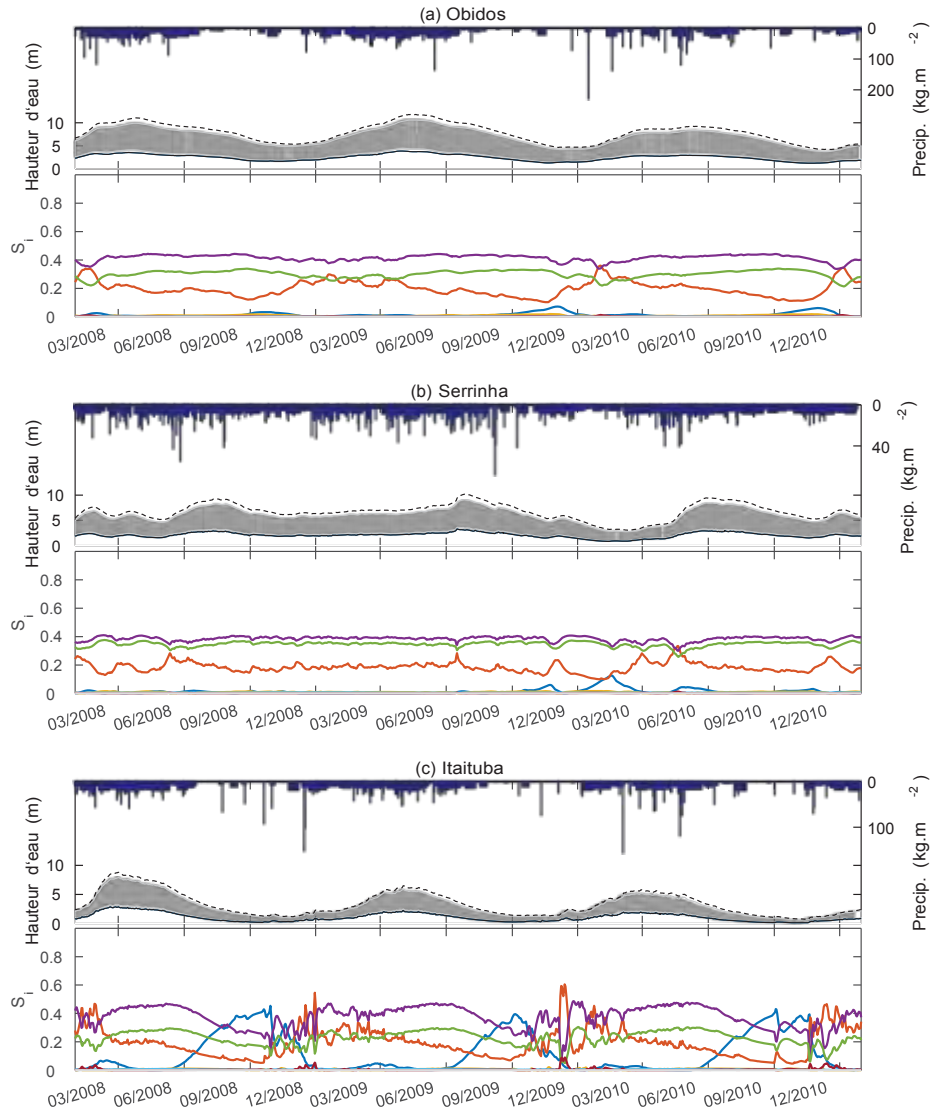


Figure 2.11 – Résultats de l'expérience AMA8 sur les hauteurs d'eau à Óbidos (a), Serrinha (b) et Itaituba (c). Pour chaque panneau, le graphique du haut représente, en fonction du temps, et par rapport à l'axe des ordonnées ascendant (gauche), la dispersion de l'ensemble des hauteurs d'eau simulées (zone grise) avec le 1^{er} (courbe pleine noire) et 9^{ème} (courbe tirets noire) décile de l'ensemble et par rapport à l'axe des ordonnées descendant (droite) les précipitations journalières dans la cellule observée. Pour chaque panneau, le graphique du bas représente, en fonction du temps, les séries temporelles des SIs de premier ordre : S_{r_e} (ligne bleue foncée), S_W (ligne orange), S_{H_e} (ligne jaune), S_n (ligne violette), $S_{s_{mult}}$ (ligne bleue claire non visible), $S_{s_{pow}}$ (ligne verte), S_μ (ligne rouge foncée), S_{n_F} (ligne grise non visible).

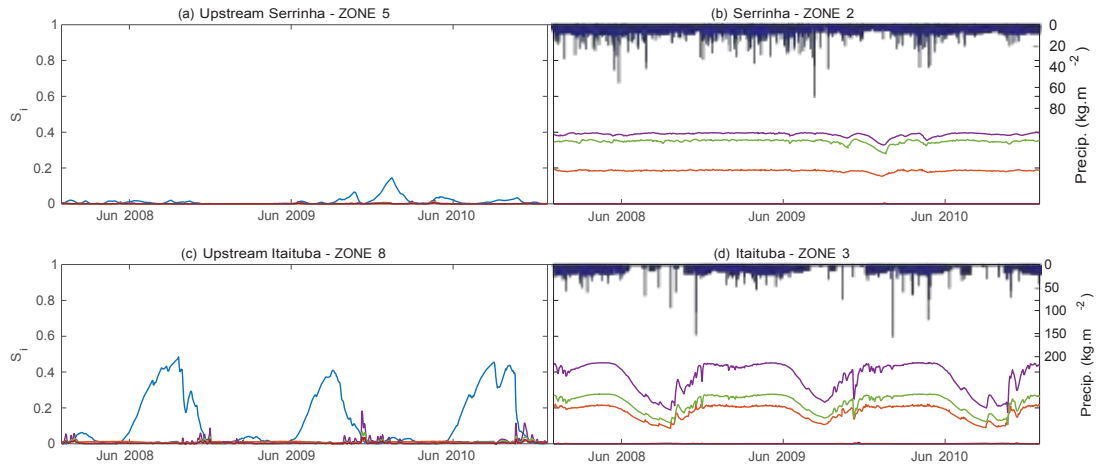


Figure 2.12 – Résultats de l'expérience AMA45 sur les hauteurs d'eau à Serrinha (a-b) et Itaituba (c-d). Sont représentées, par rapport à l'axe des ordonnées ascendant (gauche), les séries temporelles des SIs de premier ordre dans chaque zone activées : S_{TG} (ligne bleue foncée), S_W (ligne orange), S_n (ligne violette), S_{pow} (ligne verte), S_μ (ligne rouge foncée). Sont représentées, par rapport à l'axe des ordonnées descendant (droit), les précipitations locales. Toutes les courbes dans les panneaux (a) et (b) (respectivement (c) et (d)) représentent les SIs à Serrinha (respectivement à Itaituba). Le panneau (a) (respectivement (c)) représentent la contribution des paramètres des zones amont à Serrinha (respectivement à Itaituba). Le panneau (b) (respectivement (d)) représentent la contribution des paramètres locaux à Serrinha (respectivement à Itaituba).

que le schéma d'inondation de TRIP s'active quand la hauteur d'eau dans la rivière h_S dépasse la profondeur de la rivière H_c . De ce fait, comme le paramètre r_G qui est associé au réservoir des eaux souterraines G , le paramètre H_c est associé au réservoir des plaines d'inondations F .

Assimiler des profondeurs d'eau aurait le plus d'impact pour corriger à la fois les profondeurs d'eau directement et aussi les débits modélisés. Cependant, les missions satellites actuelles fournissent des mesures de cotes des surfaces d'eau référencées par rapport à un géoïde ou un ellipsoïde difficile à comparer à la profondeur d'eau de la rivière dans le modèle. Ainsi, il est aussi intéressant de regarder l'impact des paramètres sur les anomalies de hauteurs d'eau qui peut permettre de s'affranchir des incertitudes de bathymétrie.

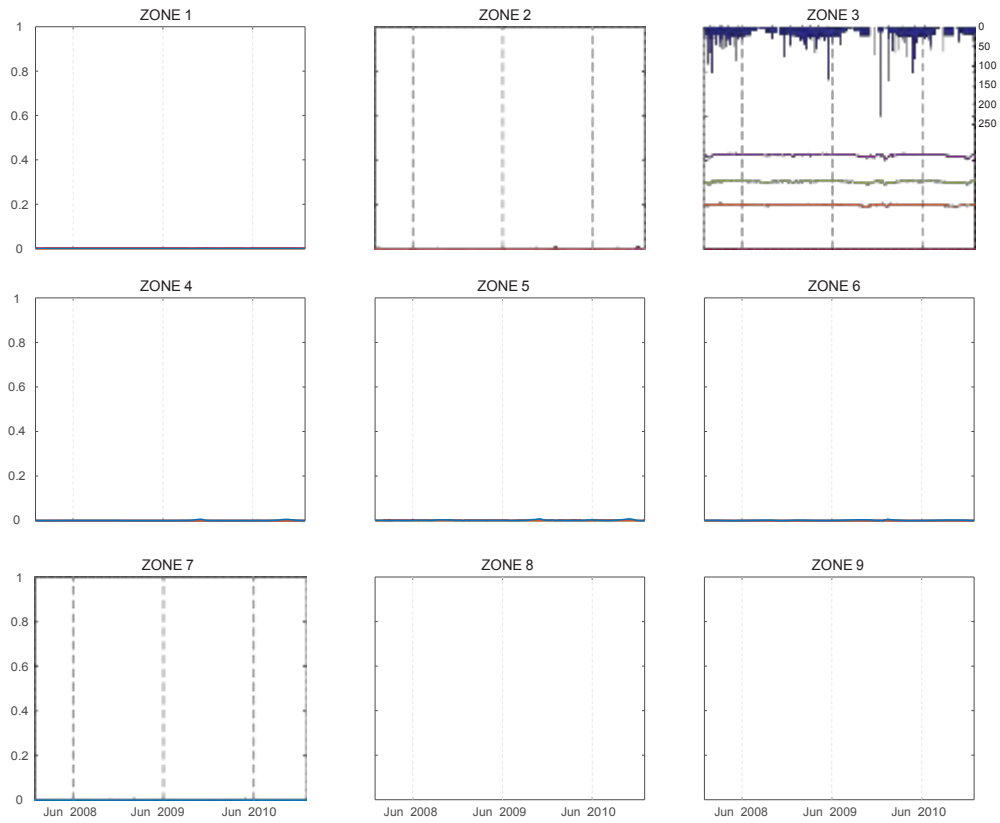


Figure 2.13 – Résultats de l'expérience AMA45 sur les hauteurs d'eau à Óbidos. Sont représentées, par rapport à l'axe des ordonnées ascendant (gauche), les séries temporelles des SIs de premier ordre dans chaque zone : S_{τ_G} (ligne bleue foncée), S_W (ligne orange), S_n (ligne violette), $S_{s_{pow}}$ (ligne verte), S_{μ} (ligne rouge foncée). Sont représentés, par rapport à l'axe des ordonnées descendant (droit), les précipitations locales.

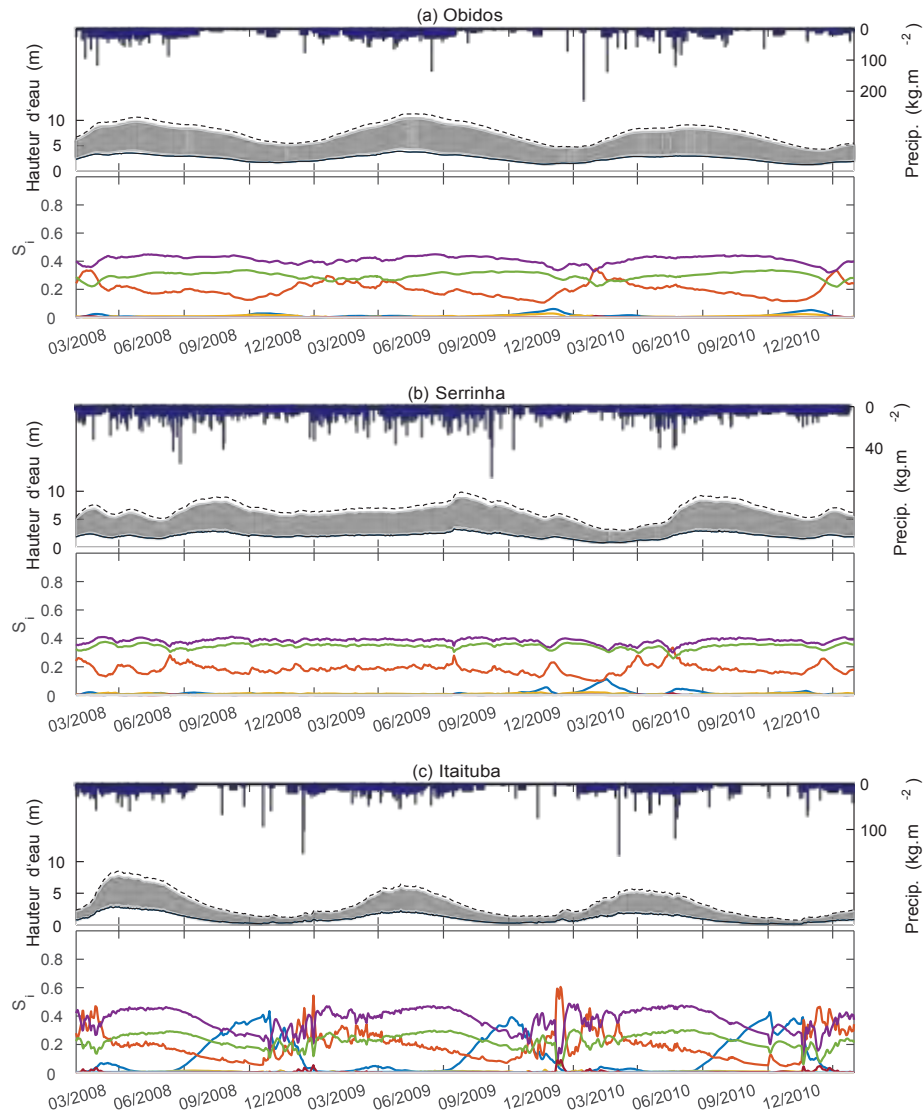


Figure 2.14 – Résultats de l'expérience AMA8_ H_c sur les hauteurs d'eau à Óbidos (a), Serrinha (b) et Itaituba (c). Pour chaque panneau, le graphique du haut représente, en fonction du temps, et par rapport à l'axe des ordonnées ascendant (gauche), la dispersion de l'ensemble des hauteurs d'eau simulées (zone grise) avec le 1^{er} (courbe pleine noire) et 9^{ème} (courbe tirets noire) décile de l'ensemble et par rapport à l'axe des ordonnées descendant (droite) les précipitations journalières dans la cellule observée. Pour chaque panneau, le graphique du bas représente, en fonction du temps, les séries temporelles des SIs de premier ordre : S_{TG} (ligne bleue foncée), S_W (ligne orange), S_{H_c} (ligne jaune), S_n (ligne violette), S_{mult} (ligne bleue claire non visible), S_{spow} (ligne verte), S_μ (ligne rouge foncée), S_{nF} (ligne grise non visible).

2.3.4.2 Résultats sur les anomalies

On commence d'abord par définir les anomalies de hauteurs d'eau a_S telles qu'elles ont été définies pour l'AS. Pour chaque membre de l'ensemble des hauteurs d'eau simulées $h_S^{[I]}$ $I = 1 \dots n_e$ on détermine la hauteur d'eau moyenne $\overline{h_S^{[I]}}$ sur la série temporelle (sur les années 2008 à 2010) et on la soustrait à la hauteur d'eau à chaque pas de temps, telle que :

$$\forall t, \forall I = 1 \dots n_e, a_S^{[I]}(t) = h_S^{[I]}(t) - \overline{h_S^{[I]}}, \quad \overline{h_S^{[I]}} = \frac{1}{T} \sum_{t=1}^T h_S^{[I]}(t). \quad (2.73)$$

Maintenant, le Tableau 2.5 montre, pour chaque station d'étude et chaque expérience, des coefficients de corrélation du métamodèle R^2 qui sont le plus souvent supérieurs à 0,9 (et toujours supérieurs à 0,8) ainsi que des effets d'interaction vers 10% pour AMA8 et AMA45 et d'environ 15% pour AMA8_ H_c . Les interactions plus importantes pour AMA8_ H_c peuvent s'expliquer par les non-linéarités provoquées par l'activation plus fréquentes des plaines d'inondations. Ceci confirme une bonne convergence de la méthode pour les anomalies de hauteurs d'eau.

Figures 2.15-2.18 montrent alors les indices de sensibilité de premier ordre pour les stations résultat et chacune des simulations comparés à l'ensemble des anomalies simulées. Parallèlement, on a aussi calculé les SI pour la hauteur d'eau moyenne $\overline{h_S^{[I]}}$ et les valeurs obtenues pour les paramètres dominant n , s_{pow} et W est donnée en encadré dans les Figures.

Pour toutes les simulations, le comportement général des sensibilités sur une majeure partie du temps d'analyse est très similaire à celui des expériences sur la hauteur d'eau, en particulier pour le coefficient de Manning, la pente du lit de la rivière ainsi que pour la largeur. Aussi, les comportements particuliers observés sur une seule expérience, par exemple la régionalisation des sensibilités avec AMA45 et l'activation de la profondeur critique H_c avec AMA8_ H_c , sont aussi reproduits sur l'étude des anomalies. Cependant, quand l'ensemble des anomalies est proche de 0, le paramètre τ_G devient un paramètre très sensible sur de courte période et, parallèlement, les sensibilités aux autres paramètres chutent fortement.

La rapide étude de sensibilité de la hauteur d'eau moyenne $\overline{h_S}$ a montré que la variance de cette dernière est expliquée à 95% par les effets principaux de n , s_{pow} et W . De plus, lorsque les anomalies sont proches de 0, leur variance est aussi très faible et toute la variance des hauteurs d'eau brutes est issue de la variance de la hauteur d'eau moyenne. Ainsi, la variance des paramètres n , s_{pow} et W se retrouvent essentiellement dans la hauteur d'eau moyenne et sont quasiment sans effet sur

Station	AMA8		AMA45		AMA8_ H_c	
	R^2	$1 - \sum_i S_i$	R^2	$1 - \sum_i S_i$	R^2	$1 - \sum_i S_i$
Óbidos	0,90	0,11	0,93	0,08	0,84	0,17
Serrinha	0,87	0,13	0,90	0,11	0,81	0,20
Itaituba	0,91	0,09	0,93	0,08	0,91	0,10

Table 2.5 – Statistiques des AS sur les anomalies. Pour chaque station et chaque simulation : Coefficients de corrélations du métamodèle moyenné sur la période d'étude (R^2) et effets d'interactions moyennés sur la période d'étude ($1 - \sum_i S_i$).

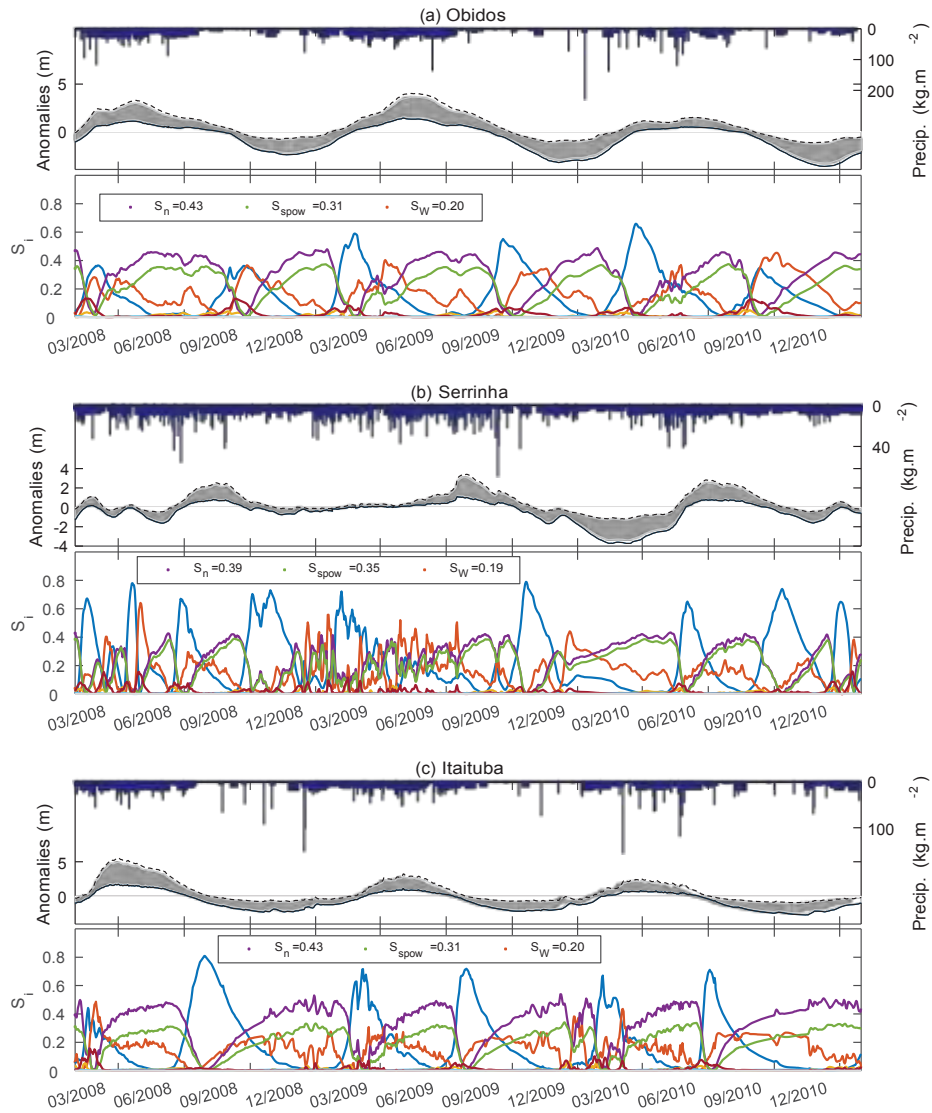


Figure 2.15 – Résultats de l'expérience AMA8 sur les anomalies à Óbidos (a), Serrinha (b) et Itaituba (c). Pour chaque panneau, le graphique du haut représente, en fonction du temps, et par rapport à l'axe des ordonnées ascendant (gauche), la dispersion de l'ensemble des hauteurs d'eau simulées (zone grise) avec le 1^{er} (courbe pleine noire) et 9^{ème} (courbe tirets noire) décile de l'ensemble et par rapport à l'axe des ordonnées descendant (droite) les précipitations journalières dans la cellule observée. Pour chaque panneau, le graphique du bas représente, en fonction du temps, les séries temporelles des SIS de premier ordre : S_{T_G} (ligne bleue foncée), S_W (ligne orange), S_{H_c} (ligne jaune), S_n (ligne violette), S_{smult} (ligne bleue claire non visible), S_{spow} (ligne verte), S_μ (ligne rouge foncée), S_{nF} (ligne grise non visible). À titre d'information, en encadré, sont indiqués les SIS de la hauteurs d'eau moyenne pour chaque graphique du bas sur chaque panneau.

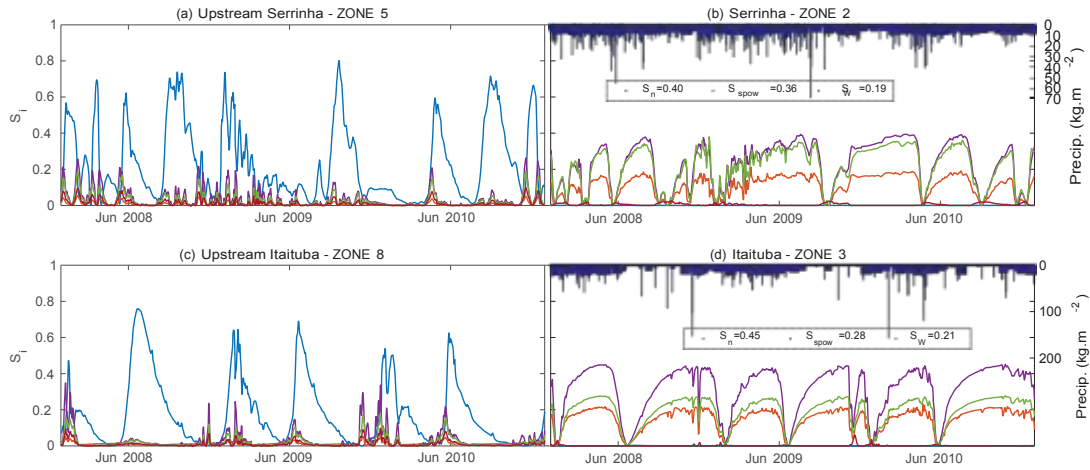


Figure 2.16 – Résultats de l'expérience AMA45 sur les anomalies à Serrinha (a-b) et Itaituba (c-d). Sont représentées, par rapport à l'axe des ordonnées ascendant (gauche), les séries temporelles des SIs de premier ordre dans chaque zone activées : S_{r_G} (ligne bleue foncée), S_W (ligne orange), S_n (ligne violette), S_{spow} (ligne verte), S_μ (ligne rouge foncée). Sont représentées, par rapport à l'axe des ordonnées descendant (droit), les précipitations locales. Toutes les courbes dans les panneaux (a) et (b) (respectivement (c) et (d)) représentent les SIs à Serrinha (respectivement à Itaituba). Le panneau (a) (respectivement (c)) représentent la contribution des paramètres des zones amont à Serrinha (respectivement à Itaituba). Le panneau (b) (respectivement (d)) représentent la contribution des paramètres locaux à Serrinha (respectivement à Itaituba). En encadré, sont indiqués les SIs de hauteurs d'eau moyenne.

les anomalies. De plus, pour les anomalies proches de 0, la variance est faible. Même si elle est principalement expliquée par r_G en pourcentage, elle reste très faible en absolu. On rappelle que les SIs calculés par la méthode ANOVA sont définis comme une fraction de la variance totale. Les effets des paramètres n , s_{pow} et W étant très faibles, les autres paramètres de l'étude, en particulier r_G ici, apparaissent comme beaucoup plus sensibles. En d'autres termes, les fortes sensibilités aux paramètres r_G et aussi H_c (pour la simulation AMA8_ H_c) pour les anomalies ne sont pas caractéristiques de processus physiques qui n'ont pas déjà été observés lors de l'étude des hauteurs d'eau. Les fortes sensibilités sont alors liées à la méthode employée qui définit les sensibilités comme une fraction de la variance totale. L'étude des anomalies permet alors de mieux mettre en évidence les autres processus impactant la hauteur d'eau mais qui restent petits en proportion par rapport à ceux mis en évidence en Section 2.3.4.1 pour les profondeurs d'eau.

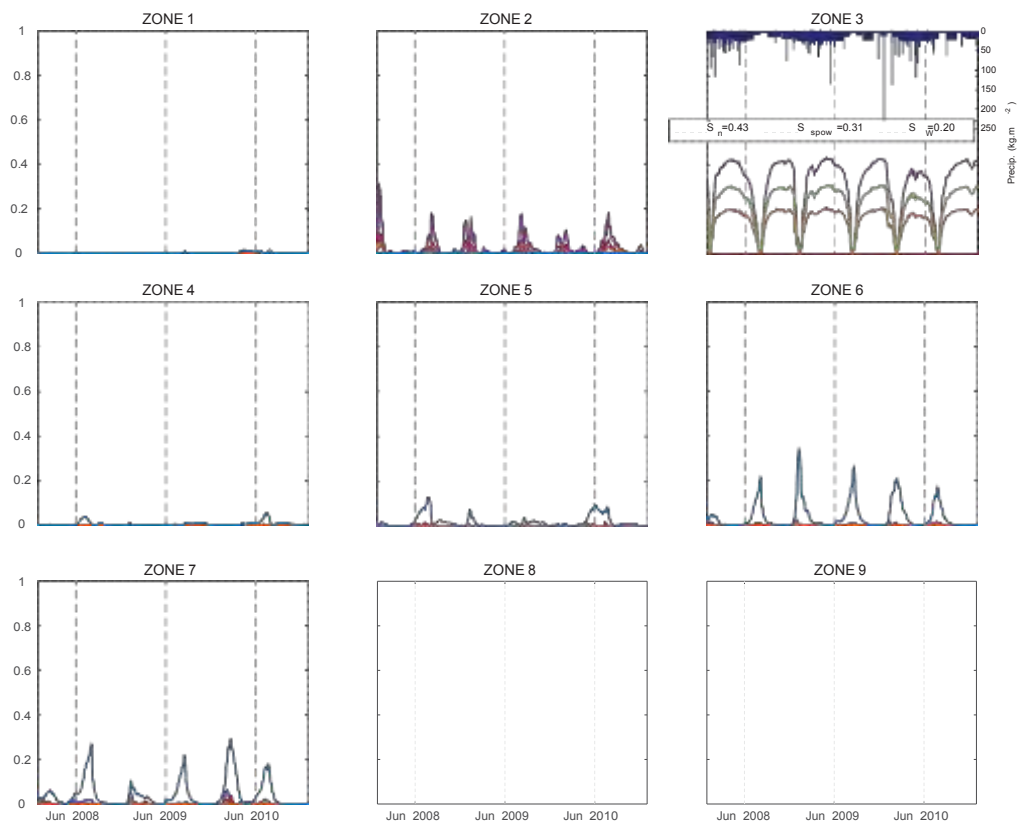


Figure 2.17 – Résultats de l'expérience AMA45 sur les anomalies à Óbidos. Sont représentées, par rapport à l'axe des ordonnées ascendant (gauche), les séries temporelles des SIs de premier ordre dans chaque zone : S_{TG} (ligne bleue foncée), S_W (ligne orange), S_n (ligne violette), S_{spow} (ligne verte), S_μ (ligne rouge foncée). Sont représentés, par rapport à l'axe des ordonnées descendant (droit), les précipitations locales. En encadré, sont indiqués les SIs de la hauteurs d'eau moyenne.

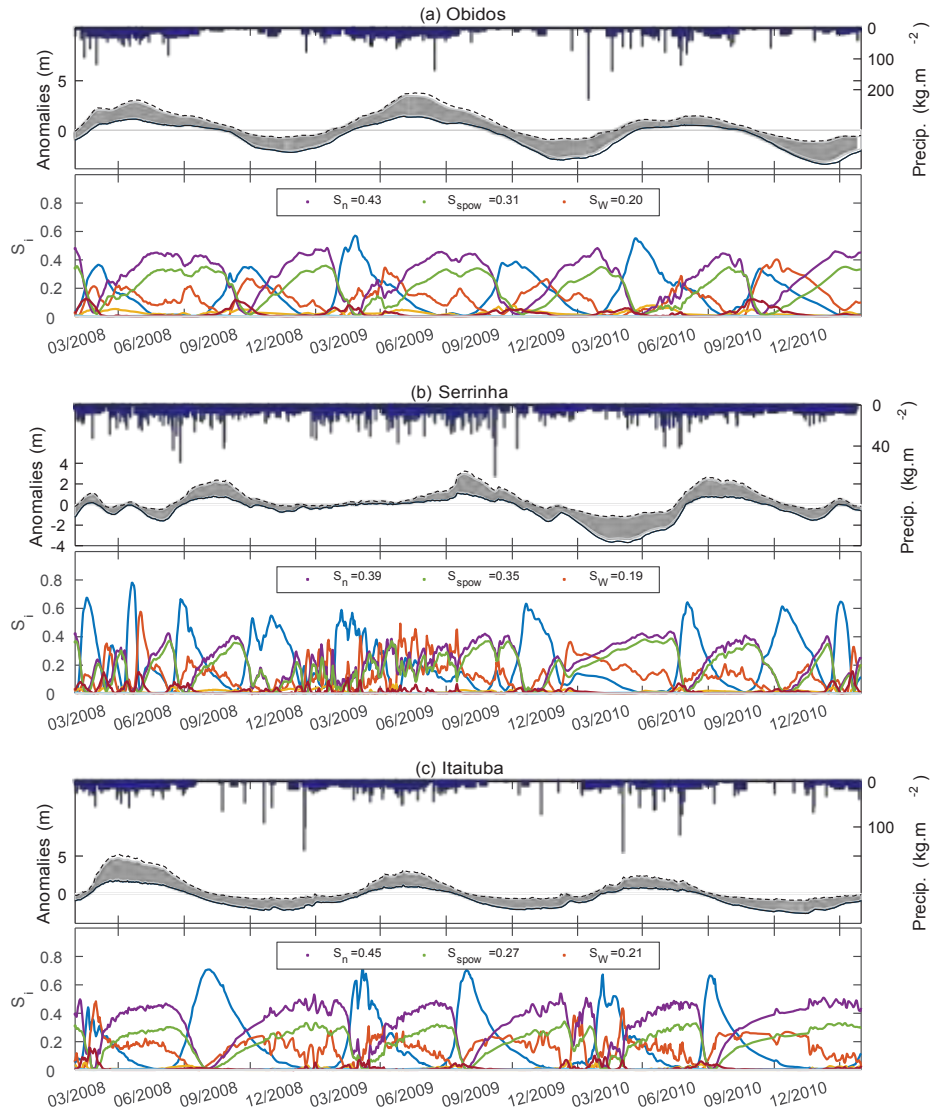


Figure 2.18 – Résultats de l'expérience AMA8_ H_c sur les anomalies à Óbidos (a), Serrinha (b) et Itaituba (c). Pour chaque panneau, le graphique du haut représente, en fonction du temps, et par rapport à l'axe des ordonnées ascendant (gauche), la dispersion de l'ensemble des hauteurs d'eau simulées (zone grise) avec le 1^{er} (courbe pleine noire) et 9^{ème} (courbe tirets noire) décile de l'ensemble et par rapport à l'axe des ordonnées descendant (droite) les précipitations journalières dans la cellule observée. Pour chaque panneau, le graphique du bas représente, en fonction du temps, les séries temporelles des SIs de premier ordre : S_{T_G} (ligne bleue foncée), S_W (ligne orange), S_{H_c} (ligne jaune), S_n (ligne violette), S_{smult} (ligne bleue claire non visible), S_{spow} (ligne verte), S_μ (ligne rouge foncée), S_{n_F} (ligne grise non visible). En encadré, sont indiqués les SIs de la hauteurs d'eau moyenne.

2.3.4.3 Résultats sur les débits

Dans le modèle ISBA-TRIP, la variable de débit est associée au flux de masse d'eau sortant du réservoir de surface dans chaque cellule et est définie par :

$$Q(t) = \frac{1}{L} v(t) S(t), \quad (2.74)$$

où $v(t)$ est la vitesse d'écoulement estimée par l'équation de Manning (Eq. 2.51), L est la longueur du tronçon et $S(t)$ le stock d'eau dans le réservoir.

Le Tableau 2.6 présente, pour chaque station d'étude et chaque expérience, le coefficient de corrélation du métamodèle R^2 ainsi que les effets d'interaction, estimés par $1 - \frac{L}{S_i}$. Le coefficient R^2 est légèrement plus bas que ceux obtenus pour les hauteurs d'eau mais reste généralement au-dessus de 0,85. Cependant, les effets d'interactions représentent alors 10 à 15 % de la variance inconditionnelle (en d'autres termes, les effets principaux n'expliquent que 85 à 90 % de la variance totale).

Contrairement aux hauteurs d'eau et aux anomalies, l'ensemble des débits simulés, voir la zone grisée en Figures 2.19 et 2.22, est très peu dispersé et les 1er et 9ème décile sont très proches. Une première observation sur l'étude des débits est que, contrairement aux hauteurs d'eau, les débits sont beaucoup moins sensibles aux paramètres étudiés. Les trois expériences affichent les mêmes résultats suivants : les sensibilités ne montrent pas de différences d'une année à l'autre. De plus, les paramètres dominants sont similaires à ceux ressortis lors de l'étude des hauteurs d'eau et des anomalies mais avec un ordre d'importance différent :

$$\tau_G > n > s_{\text{pow}} > \mu > W. \quad (2.75)$$

Mais la différence notable est que le paramètre dominant est le temps de résidence des eaux souterraines τ_G et S_{τ_G} peut prendre des valeurs supérieures à 0,9, en particulier en pleines saisons de hautes et basses eaux. Le paramètre τ_G est le paramètre d'exfiltration du réseau de drainage (Roux et al., 2011) et représente le transfert de masse d'eau des eaux souterraines vers la rivière. La forte sensibilité du débit à ce paramètre est en accord avec la littérature sur les modèles pluie-débit et aussi avec les conclusions de Pappenberger et al. (2010). En régime stable, le débit est principalement actionné par les flux de masses en provenance de l'amont et des eaux souterraines. Néanmoins, en période transitoire entre les saisons de hautes et basses eaux et aussi lors d'événements pluvieux, la sensibilité au paramètre τ_G chute fortement au profit des paramètres géomorphologiques - n , s_{pow} et W . Sur ces périodes transitoires, le débit est dirigé par les changements de régime, en d'autres termes, les changements de vitesses. Ainsi, les paramètres impliqués dans l'équation de Manning influencent

Station	AMA8 $\frac{L}{S_i}$		AMA45 $\frac{L}{S_i}$		AMA8 $\frac{H_c}{S_i}$	
	R^2	$1 - \frac{L}{S_i}$	R^2	$1 - \frac{L}{S_i}$	R^2	$1 - \frac{H_c}{S_i}$
Óbidos	0,89	0,11	0,93	0,09	0,80	0,21
Serrinha	0,91	0,09	0,95	0,07	0,83	0,17
Itaituba	0,91	0,09	0,94	0,06	0,90	0,11

Table 2.6 – Statistiques des AS sur les débits. Pour chaque station et chaque simulation : Coefficients de corrélations du métamodèle moyenné sur la période d'étude (R^2) et effets d'interactions moyennés sur la période d'étude ($1 - \frac{L}{S_i}$).

de manière non-négligeable le débit sur ces périodes. En outre, la longueur du tronçon L , représentée ici par le coefficient de méandrement μ , a aussi un effet important sur le débit et ce pour deux raisons : (1) ce paramètre est directement impliqué dans l'équation pour estimer le débit (Eq. 2.74) et (2) c'est un paramètre géométrique important lors des phases de changement de régime.

Les résultats de l'expérience régionalisée AMA45 illustre parfaitement le caractère intégrateur du débit, c'est-à-dire que le débit est la somme de tous les processus hydrologiques ayant eu lieu en amont et localement d'un point de vue spatial (processus ayant eu lieu dans l'ensemble de l'aire drainée et à la station même) et temporel (événement passé qui parcourt le réseau de drainage avant d'atteindre la cellule étudiée). On remarque alors que seules les zones amont à la cellule d'étude sont activées par des paramètres sensibles. Ce phénomène s'observe très bien pour Serrinha (Fig. 2.20a-b) et pour Itaituba (Fig. 2.20c-d) où une unique zone amont alimente la cellule. C'est aussi le cas pour Óbidos, (Fig. 2.21), qui se situe en zone 3 et où les zones amont 1, 2, 4, 5, 6 et 7 sont activées. Néanmoins, un comportement particulier est observé à Óbidos. Toute l'eau provenant de l'aire drainée s'écoule nécessairement par la zone 2 avant d'atteindre la station d'Óbidos. Ainsi, toutes les zones les plus en amont (1, 4, 5, 6 et 7) sont activées via le temps de résidence τ_G alors que ce sont les paramètres géomorphologiques de la zone la plus proche, ici la zone 2, qui sont actifs. Dans cette même zone, τ_G est légèrement sensible aussi mais comme l'aire de cette zone est petite par rapport aux autres zones amonts, l'impact de ce paramètre est moins important. Dans les autres zones où τ_G est actif, deux types de pics sont observés : ces pics correspondent aux saisons de basses eaux (pics les plus hauts) et aux saisons de hautes eaux (pics les plus bas) dans leur zone respective. Ces pics sont alors légèrement décalés par rapport aux saisons de hautes/basses eaux dans leur zone respective à cause du temps de transfert d'une zone à l'autre.

Enfin, l'expérience AMA8_ H_c sur les débits illustrent plus clairement l'échange de masses d'eau entre la rivière et les plaines d'inondations observés lors de l'étude des hauteurs d'eau et des anomalies. Plus particulièrement à Óbidos, la profondeur critique H_c (ligne jaune dans les séries temporelles des SI) devient un paramètre sensible en pleine saison de hautes eaux. On associe cette activation à l'activation du schéma des plaines d'inondations engendrant des échanges de masses entre la rivière (réservoir TRIP S) et les plaines inondées (réservoir TRIP F). En contrepartie, la sensibilité au paramètre τ_G est moins importante mais ceci est complètement lié au choix de la méthode ANOVA qui estime les sensibilités comme une fraction de la variance totale.

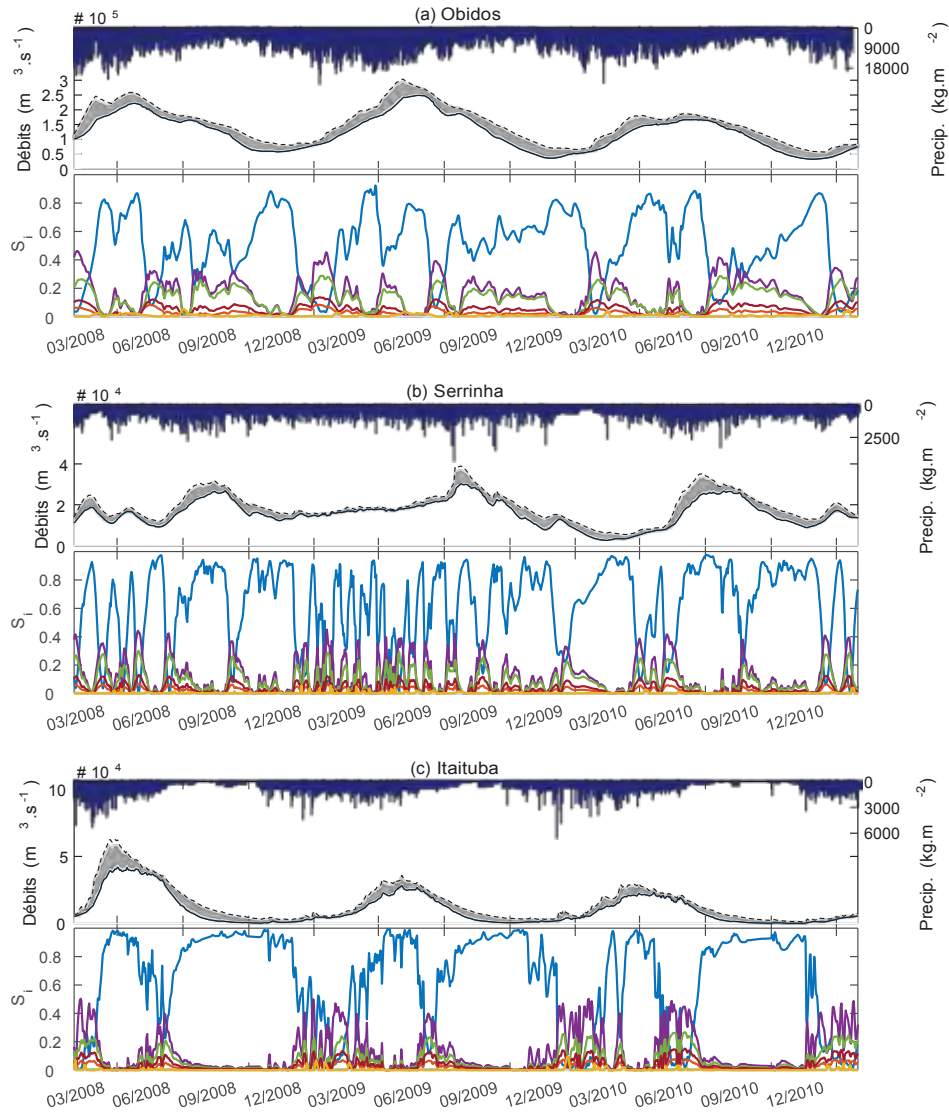


Figure 2.19 – Résultats de l'expérience AMA8 sur les débits à Óbidos (a), Serrinha (b) et Itaituba (c). Pour chaque panneau, le graphique du haut représente, en fonction du temps, et par rapport à l'axe des ordonnées ascendant (gauche), la dispersion de l'ensemble des hauteurs d'eau simulées (zone grise) avec le 1^{er} (courbe pleine noire) et 9^{ème} (courbe tirets noire) décile de l'ensemble et par rapport à l'axe des ordonnées descendant (droite) les précipitations journalières dans la cellule observée. Pour chaque panneau, le graphique du bas représente, en fonction du temps, les séries temporelles des SIs de premier ordre : S_{T_G} (ligne bleue foncée), S_W (ligne orange), S_{H_c} (ligne jaune), S_n (ligne violette), S_{mult} (ligne bleue claire non visible), S_{spow} (ligne verte), S_μ (ligne rouge foncée), S_{n_F} (ligne grise non visible).

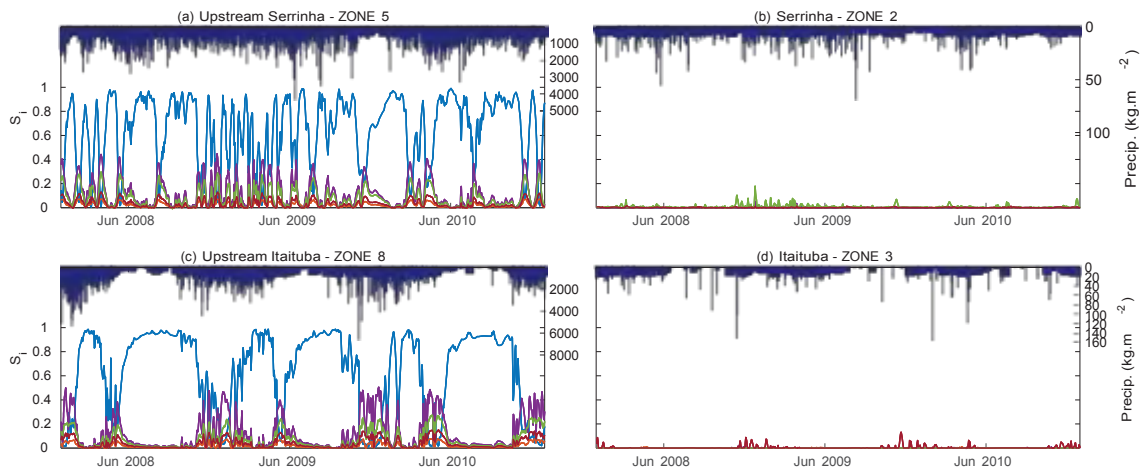


Figure 2.20 – Résultats de l'expérience AMA45 sur les débits à Serrinha (a-b) et Itaituba (c-d). Sont représentées, par rapport à l'axe des ordonnées ascendant (gauche), les séries temporelles des SIs de premier ordre dans chaque zone activées : S_{TG} (ligne bleue foncée), S_W (ligne orange), S_N (ligne violette), $S_{S_{pow}}$ (ligne verte), S_μ (ligne rouge foncée). Sont représentés, par rapport à l'axe des ordonnées descendant (droit), les précipitations de toute l'aire drainée. Toutes les courbes dans les panneaux (a) et (b) (respectivement (c) et (d)) représentent les SIs à Serrinha (respectivement à Itaituba). Le panneau (a) (respectivement (c)) représentent la contribution des paramètres des zones amont à Serrinha (repectivement à Itaituba). Le panneau (b) (respectivement (d)) représentent la contribution des paramètres locaux à Serrinha (respectivement à Itaituba).

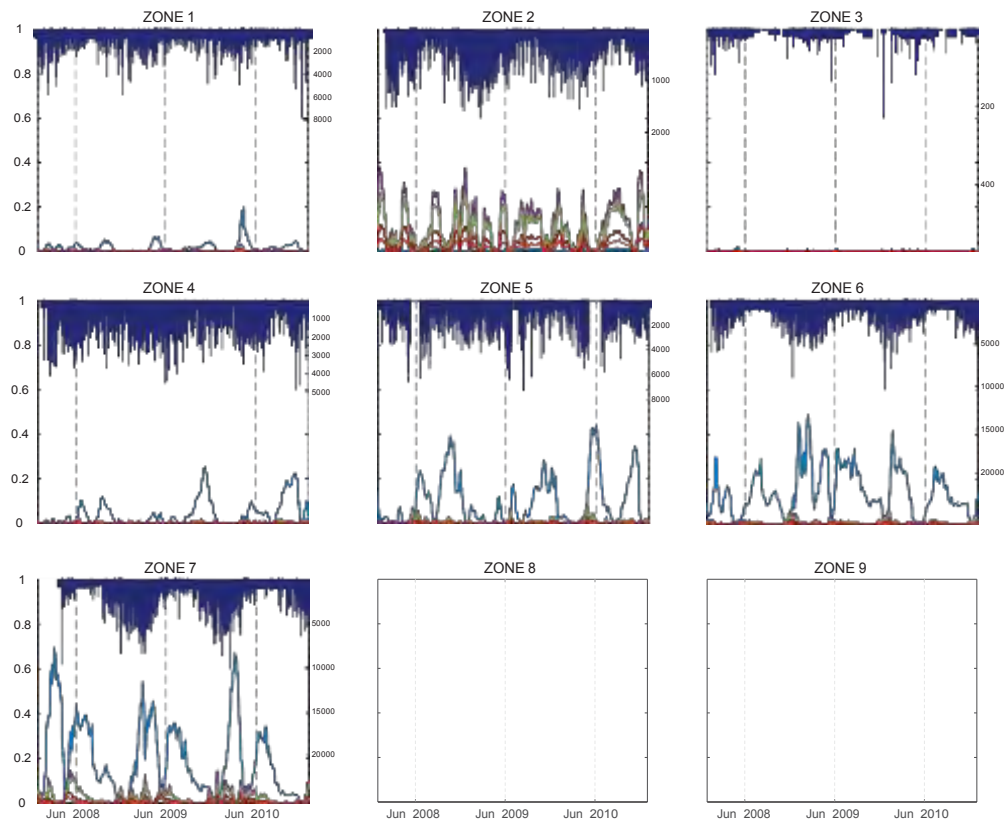


Figure 2.21 – Résultats de l'expérience AMA45 sur les débits à Óbidos. Sont représentées, par rapport à l'axe des ordonnées ascendant (gauche), les séries temporelles des SIs de premier ordre dans chaque zone : S_{TG} (ligne bleue foncée), S_W (ligne orange), S_n (ligne violette), S_{spow} (ligne verte), S_μ (ligne rouge foncée). Sont représentés, par rapport à l'axe des ordonnées descendant (droit), les précipitations de toute l'aire drainée.

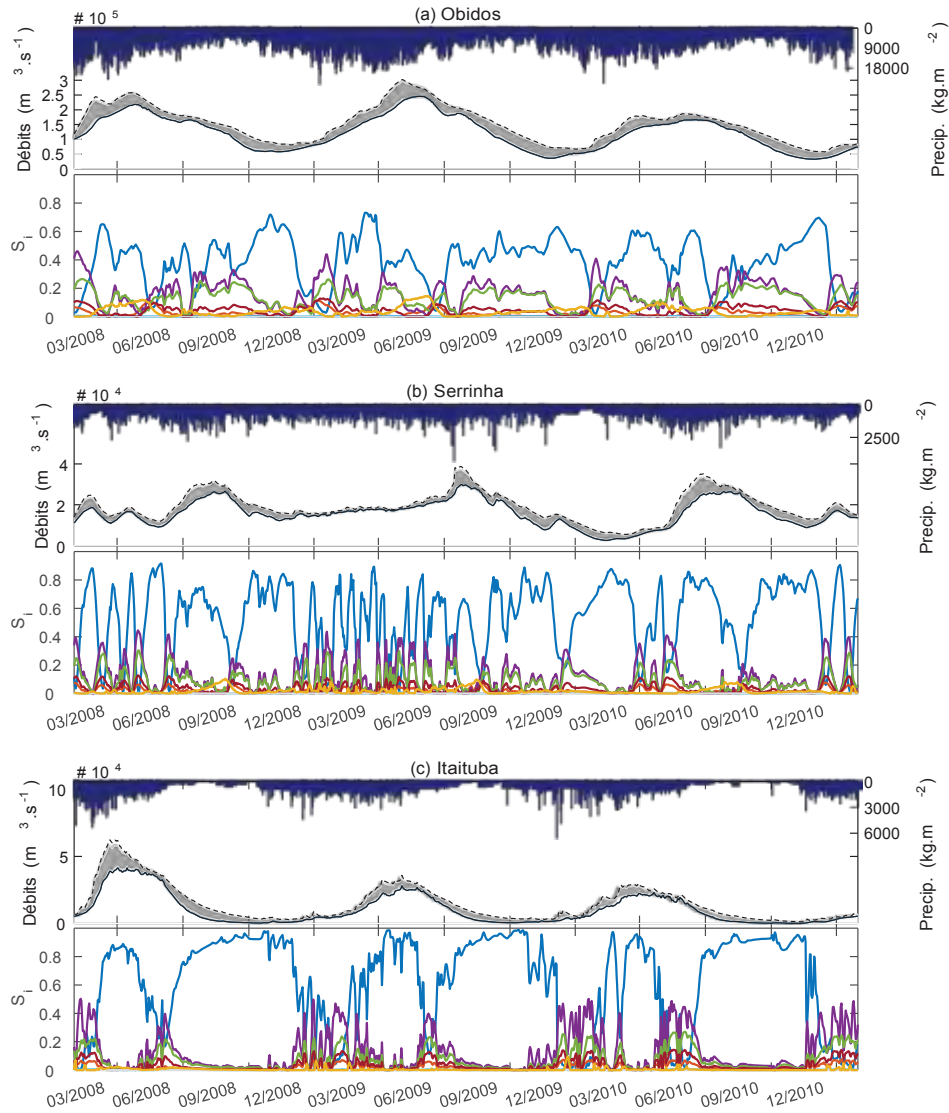


Figure 2.22 - Résultats de l'expérience AMA8_Hc sur les débits à Óbidos (a), Serrinha (b) et Itaituba (c). Pour chaque panneau, le graphique du haut représente, en fonction du temps, et par rapport à l'axe des ordonnées ascendant (gauche), la dispersion de l'ensemble des hauteurs d'eau simulées (zone grise) avec le 1^{er} (courbe pleine noire) et 9^{ème} (courbe tirets noire) décile de l'ensemble et par rapport à l'axe des ordonnées descendant (droite) les précipitations journalières dans la cellule observée. Pour chaque panneau, le graphique du bas représente, en fonction du temps, les séries temporelles des SI de premier ordre : S_{TG} (ligne bleue foncée), S_W (ligne orange), S_{Hc} (ligne jaune), S_n (ligne violette), S_{mult} (ligne bleue claire non visible), S_{spow} (ligne verte), S_μ (ligne rouge foncée), S_{nF} (ligne grise non visible).

Conclusions et perspectives

L'étude menée dans ce chapitre avait pour objectif d'analyser la sensibilité du modèle hydrologique grande échelle ISBA-TRIP sur le bassin de l'Amazonie. On a utilisé pour cela une méthode de décomposition de la variance pour identifier les paramètres clés parmi les paramètres de la composante routage de rivière du modèle sur une période de trois ans. Trois analyses différentes ont été menées pour évaluer la sensibilité des sorties modèle (profondeur d'eau, anomalies de hauteurs d'eau et débits) et ce à différentes échelles spatiales. La première étude, AMA8, considère des paramètres dont l'incertitude est définie identiquement sur l'ensemble du bassin. La deuxième étude, AMA45, utilise les mêmes paramètres mais avec une incertitude régionalisée selon une division hydrologique et géologique du bassin de l'Amazonie. Le but étant de séparer les contributions des paramètres locaux et des paramètres amont sur à la fois la profondeur d'eau, les anomalies de hauteurs d'eau et les débits. Enfin, l'analyse AMA8_ H_c reprend la configuration de l'expérience AMA8 en perturbant différemment la hauteur critique H_c de la rivière pour étudier la sensibilité aux paramètres en présence d'inondations.

Les trois études n'ont pas montré de sensibilités inter-annuelle. Pour l'ensemble des expériences, l'importance du coefficient de Manning et de la pente a été mise en évidence. Ces affirmations sont en accord avec l'utilisation de la formule de Manning pour estimer la vitesse d'écoulement dans la rivière. De plus, la largeur et la longueur du tronçon de la rivière ont une influence non-négligeable. Contrairement aux hauteurs d'eau, le temps de résidence a un impact considérable sur les débits et aussi sur les anomalies de hauteurs d'eau (sur des périodes de temps plus courtes et dans une moindre mesure).

La sensibilité des profondeurs d'eau est relativement constante dans le temps avec le coefficient de Manning et la pente de la rivière (via un coefficient en puissance) expliquant 40% et 35% respectivement de la variance inconditionnelle. La largeur de la rivière, elle, prend alors 20% de la variance. Aussi, quand la profondeur d'eau de l'ensemble est très basse, on observe un impact non négligeable du temps de résidence des eaux souterraines. Dans l'ensemble, les profondeurs d'eau sont sensibles aux paramètres géomorphologiques. L'étude régionalisée indique que les paramètres géomorphologiques locaux (c'est-à-dire dans la même zone que la cellule étudiée) pilotent les hauteurs d'eau une majeure partie de l'année. Enfin, la dernière expérience met en évidence les échanges de masses d'eau entre la rivière et les plaines inondées en période de hautes eaux via le paramètre H_c . Finalement, les effets principaux (SIs de premier ordre modélisant l'action d'un seul paramètre à la fois) suffisent pour expliquer plus de 90% de la variance totale des profondeurs d'eau. Ceci signifie que la profondeur d'eau dans TRIP a un comportement quasi-additif et peut s'exprimer comme une somme des contributions de chaque paramètre :

$$h_S(t) = h_{S,0}(t) + h_{S,n}(n, t) + h_{S,s_{pow}}(s_{pow}, t) + h_{S,W}(W, t) + h_{S,\tau_G}(\tau_G, t) + h_{S,\mu}(\mu, t). \quad (2.76)$$

L'étude de sensibilité des anomalies de hauteurs d'eau montrent des sensibilités dans l'ensemble très similaires à celles observées pour les profondeurs d'eau. Seulement, lorsque l'ensemble des anomalies est proche de zéro, les sensibilités aux paramètres géomorphologiques (n , s_{pow} et W) chutent complètement et les autres paramètres deviennent plus sensibles. Il apparaît que la profondeur d'eau moyenne (utilisée pour générer les anomalies par rapport aux profondeurs d'eau) prend la totalité de la variance dans ces périodes particulières. Ces résultats sont représentatifs de la méthode par décomposition de variance. Les résultats sur les anomalies mettent alors en évidence les processus physiques secondaires influençant la profondeur d'eau.

Les sensibilités des débits présentent de nombreuses structures temporelles : des structures saisonnières liées à l'alternance des saisons de hautes et basses eaux et des structures de courtes durées associées aux événements pluvieux en amont. Le temps de résidence des eaux souterraines est le paramètre le plus sensible et explique parfois plus de 90% de la variance totale. En régime permanent, ce paramètre représente le principal apport d'eau dans la rivière. Les paramètres géomorphologiques sont sensibles sur les périodes de transition entre la saison des hautes et basses eaux. Il est important de constater que, du fait de la faible dispersion de l'ensemble des débits simulés, la variable de débit est en fait peu sensible aux paramètres TRIP. Une hypothèse raisonnable serait de considérer les précipitations et la configuration d'ISBA (couverture végétale, caractéristiques du sol) comme les principaux vecteurs de débits.

À partir de ces résultats, il apparaît nécessaire d'avoir une meilleure connaissance des paramètres géomorphologiques et de leur valeur afin d'améliorer les diagnostics de profondeurs d'eau et de débits dans ISBA-TRIP. Pour le moment, la largeur de la rivière est définie à partir de relations empiriques mais elle pourrait être directement prise de bases de données récentes telles que NARWidth (Allen and Pavelsky, 2015) ou GWD-LR (Global Width Database for Large River, Yamazaki et al., 2014). Similairement, d'autres Modèles Numériques de Terrain (MNT) peuvent remplacer celui utilisé actuellement dans ISBA-TRIP (GTOPO30) pour obtenir une meilleure estimation de la pente comme le MNT SRTM pour les latitudes inférieures à 60° édité par HydroSHEDS (Lehner and Grill, 2013). De plus, SWOT fournira des informations sur la largeur et la pente de la rivière qui peuvent être réutilisées pour affiner la paramétrisation de TRIP. Des images optiques peuvent aussi être utilisées pour redéfinir et spatialiser le coefficient de méandrement. Une meilleure définition de ces paramètres limiterait alors les incertitudes du modèle et le paramètre le plus incertain resterait le coefficient de Manning, le temps de résidence des eaux souterraines qui restent des paramètres difficiles à estimer ainsi que le paramètre H_c qui contrôle les inondations dans le modèle.

L'étude présentée dans ce chapitre se concentre seulement sur les paramètres du modèle de routage de rivière et a été appliquée sur un seul bassin hydrologique mais elle offre de très nombreuses perspectives. La plateforme actuelle peut être facilement étendue à d'autres grands bassins tels que le Mississippi ou le Niger. Aussi, il serait intéressant d'étudier l'impact du modèle de surface (ici le modèle ISBA), des forçages atmosphériques et en particulier des précipitations sur les sorties TRIP. Un autre point important est que les résultats d'une telle AS sont assez dépendant des paramètres choisis, de leur gamme de perturbations mais aussi du modèle choisi pour l'étude. En effet, le modèle TRIP fonctionne sur un modèle d'onde cinématique. D'autres modèles incluent un modèle de propagation d'onde diffusive (Yamazaki et al., 2011; Winsemius et al., 2013) dans leur modèle de routage combiné à une description plus détaillée de la topographie et de la dynamique des plaines d'inondations (Sampson et al., 2015b). Le formalisme de la méthode ANOVA peut facilement être exporté à ces modèles et donnera certainement des résultats différents dus à la physique différentes des modèles (onde cinématique contre onde diffusive) et des schémas spatio-temporels des sensibilités différents.

Pour terminer, cette Analyse de Sensibilité était un travail préliminaire nécessaire afin de préparer des travaux de réanalyses et d'estimation de paramètres via l'Assimilation de Données (Moradkhani, 2008; Reichle, 2008). Ces méthodes utilisent des données de débits et de cotes d'eau issues de la télédétection ou de mesures in situ pour corriger l'état et/ou les paramètres d'un modèle physique. Parmi ces données, SWOT présente un fort potentiel pour améliorer la modélisation hydrologique à grande échelle. L'idée est d'utiliser les produits SWOT dans une maquette d'AD afin de corriger les principaux paramètres mis en évidence par l'AS. Le produit principal de SWOT s'affilie à des données de hauteurs d'eau. Au vu des résultats de l'AS sur la sensibilité des hauteurs d'eau et des anomalies, un

tel produit apporterait assez d'information pour corriger les paramètres critiques de TRIP tels que les coefficients de Manning. Par ailleurs, des algorithmes pour calculer les débits dérivés des observables SWOT est actuellement en développement (voir Chapitre 1 Section 1.3.5.5). Le produit de débit SWOT permettrait alors de corriger le temps de résidence des eaux souterraines car ce dernier s'est avéré être le plus sensible aux débits. Le Chapitre suivant va alors se concentrer sur la mise en place d'une maquette d'assimilation de données utilisant des produits de type SWOT afin de corriger les paramètres du modèle TRIP.

Chapitre 3

Correction des paramètres de TRIP via l'assimilation de pseudo-observations SWOT

Sommaire

Introduction	127
3.1 Description des variables observées et des variables à corriger . .	130
3.1.1 Description des pseudo-observations SWOT	131
3.1.2 Description des variables de contrôle	133
3.2 Choix de la méthode d'assimilation	135
3.2.1 Définition de la fenêtre d'assimilation et conséquences	135
3.2.2 Le Filtre de Kalman d'Ensemble Asynchrone en estimation de paramètres	136
3.2.3 Configuration des simulations d'estimation de paramètres	140
3.3 Résultats pour l'assimilation des profondeurs d'eau	146
3.3.1 Résultats des expériences EP1F et EP1NF	146
3.3.2 Résultats de l'expérience EP2F	155
3.4 Résultats pour l'assimilation des anomalies de hauteurs d'eau . .	159
3.4.1 Résultats des expériences EP3-1 et EP3-2	159
3.4.2 Résultats des expériences EP4F et EP4NF	166
Conclusions et perspectives	175

Introduction

Dans ce chapitre, on va maintenant chercher à étudier les potentielles contributions de la future mission large fauchée SWOT pour la modélisation hydrologique à grande échelle. Le satellite SWOT et son instrument KaRIn (Rodríguez, 2016; Fu et al., 2012a; Fjörtoft et al., 2014) observera toutes les rivières et fleuves d'au moins 100 mètres de large situées entre les latitudes 78° Sud et 78° Nord. Le produit principal de SWOT consistera en des mesures en 2D (contrairement aux altimètres nadirs actuels) des cotes d'eau, pentes et largeurs des rivières observées avec leur incertitude, sous deux fauchées de 50 km (séparées par un trou de 20 km au nadir). Se référer au Chapitre 1, Section 1.3.5 pour la description de la mission SWOT.

L'analyse de sensibilité d'ISBA-TRIP, menée au Chapitre 2, Section 2.3, a montré que les hauteurs d'eau simulées par ISBA-TRIP étaient particulièrement sensibles aux coefficients de Manning (frottements) ainsi qu'à la pente du lit de la rivière et à la largeur. Ainsi, l'assimilation des observations de cotes d'eau SWOT dans le modèle ISBA-TRIP permettrait de corriger la valeur de ces paramètres du modèle afin d'améliorer ses performances.

La combinaison du modèle hydrologique grande échelle ISBA-TRIP avec les observations SWOT, via l'assimilation de données en estimation de paramètres, est une contribution directe des produits SWOT pour l'hydrologie à grande échelle. Cependant, le lancement de la mission SWOT n'est pas prévu avant 2021. D'ici là, on ne peut pas utiliser, pour les expériences d'assimilation de données (AD) impliquant SWOT, des données réelles issues de KaRIn. Ainsi, l'ensemble des études d'AD se basant sur les produits SWOT sont principalement basées sur des expériences dites jumelles. En expériences jumelles, on choisit au préalable une configuration des paramètres, des forçages et du modèle que l'on considère comme la « vérité ». On en déduit directement le run « vrai » en intégrant le modèle à partir de cette configuration vraie. Ensuite, d'un côté, les observations synthétiques utilisées en expériences jumelles sont obtenues en perturbant les observables, ici les cotes d'eau, issues du run vrai avec un modèle d'erreur représentatif des erreurs d'observation. D'un autre côté, les variables de contrôle (paramètres et/ou variables d'état du modèle) d'ébauche au premier cycle d'assimilation sont obtenus en perturbant directement les variables de contrôle vraies, afin d'obtenir des erreurs qui doivent être une bonne approximation des erreurs de modélisation. L'algorithme d'AD doit alors être capable de retrouver la configuration vraie des variables de contrôle à partir de la configuration d'ébauche erronée et des observations virtuelles (ou pseudo-observations). La Figure 3.1 présente le schéma de principe de fonctionnement d'expériences jumelles dans le cadre de SWOT. Les expériences jumelles permettent de complètement maîtriser les différentes sources d'incertitudes impliquées dans l'expérience d'AD. Ainsi, le contexte idéalisé de ces expériences constitue une étape nécessaire lors du développement de maquettes d'AD, notamment celles impliquant les données SWOT.

L'idée de base de l'estimation de paramètres est qu'une part importante des incertitudes dans les sorties/prévisions du modèle est issue d'erreurs dans les paramètres d'entrée de celui-ci. De ce fait, de nombreuses études se sont consacrées à la calibration des paramètres des modèles pour améliorer leurs sorties (Sorooshian and Dracup, 1980; Duan et al., 1992; Gupta et al., 1998). Aussi, Getirana et al. (2013) a développé une méthode automatique de calibration des paramètres géomorphologiques du système ISBA-HyMAP (Getirana et al., 2012; HyMAP étant un RRM pouvant être forcé avec des sorties d'un LSM) à partir d'observations de débits in situ et de cotes d'eau altimétriques. L'utilisation de l'AD pour l'estimation de paramètres s'est développée suite aux limites des méthodes de calibration qui nécessitent généralement un nombre considérable d'observations pour être appliquées (Thiemann et al., 2001). L'AD permet d'initialiser la recherche des valeurs des paramètres à partir d'une connaissance a priori de ces valeurs et de la description des incertitudes à l'aide de densités de probabilités (Misirli et al., 2003). Chertok and Lardner (1996), par exemple, ont assimilé des pseudo-observations de hauteurs d'eau pour corriger les paramètres hydrauliques d'un modèle d'écoulement unidirectionnel en utilisant une méthode variationnelle. D'autres applications de l'estimation de paramètres dans des domaines similaires ont été réalisées. Montzka et al. (2011) utilise un filtre à particules pour corriger les propriétés hydrauliques du sol à partir d'observations satellite d'humidité de surface; ou encore Alzraiee et al. (2014) se base sur un Filtre de Kalman d'Ensemble (EnKF) pour corriger les caractéristiques de l'aquifère à partir de données de tomographie hydraulique.

Dans le cas de la mission SWOT, des études se basant sur l'estimation de paramètres à partir de données de type SWOT ont été menées. Durand et al. (2008), par exemple, cherchent à corriger

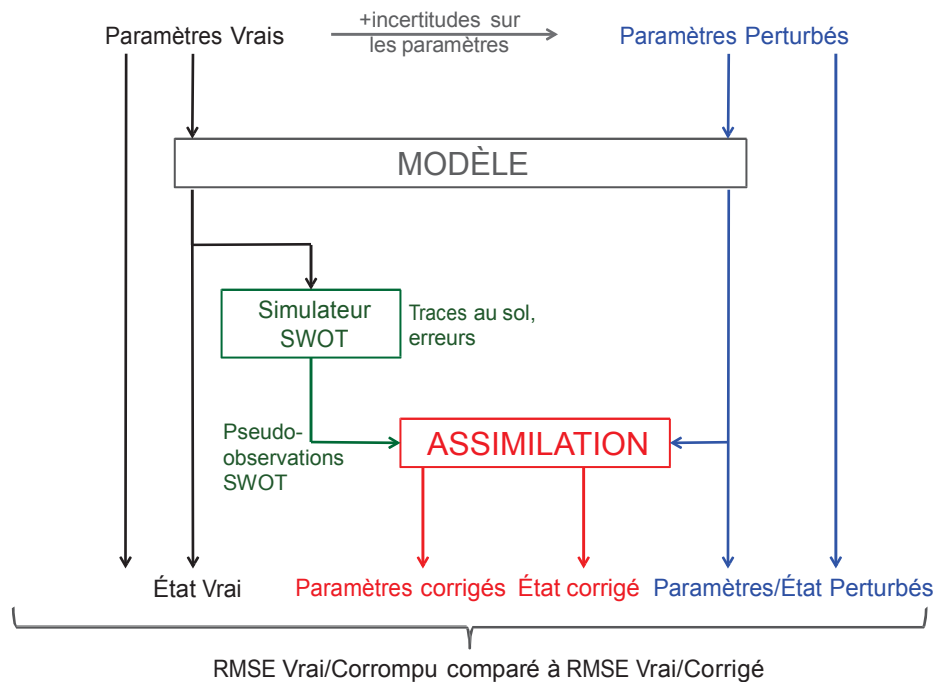


Figure 3.1 – Schéma de principe d’une expérience jumelle : Un jeu de paramètres vrais est pré-défini à l’avance et est perturbé pour générer le jeu de paramètres d’ébauche initial. Les deux jeux de paramètres sont utilisés par une intégration modèle pour générer les variables d’état d’ébauche et les variables d’état vraies. Ces dernières sont perturbées pour générer les observations. L’étape d’assimilation combine le vecteur de contrôle d’ébauche et les observations pour générer un vecteur de contrôle corrigé. L’apport de l’assimilation est déterminé en calculant les erreurs quadratiques moyennes (RMSE) entre l’état vrai, l’état d’ébauche et l’état corrigé.

la bathymétrie du modèle hydrodynamique LISFLOOD-FP en assimilant dans un EnKF des mesures d'élévations de surface SWOT. Similairement, Yoon et al. (2012) utilisent des pseudo-observations SWOT pour corriger la bathymétrie du modèle LISFLOOD-FP à partir d'une extension de l'EnKF appelée *Ensemble Batch Smoother* (EnBS; Dunne and Entekhabi, 2005). Cependant, il n'existe encore que très peu d'études utilisant l'AD pour corriger les paramètres de RRM globaux tels que TRIP. Récemment, Pedinotti et al. (2014) a réalisé une des premières expériences d'estimation de paramètres sur un modèle grande échelle et a utilisé un Filtre de Kalman Étendu (EKF) en expérience jumelle pour corriger la distribution des coefficients de Manning du modèle ISBA-TRIP, à partir de pseudo-observations de hauteurs d'eau SWOT.

Le choix du coefficient de Manning comme paramètre le plus sensible et qui nécessite d'être corrigé a été confirmé grâce à l'analyse de sensibilité menée au Chapitre 2. Dans l'expérience menée par Pedinotti et al. (2014), les coefficients de Manning des réservoirs de surface constituant le cours principal du Niger, plus précisément, les cellules du réseau TRIP ayant une largeur W d'au moins 100 mètres, sont perturbées et corrigées une à une. Dans le cadre de ces simulations orientées vers SWOT, un simulateur SWOT très simple, développé par Biancamaria et al. (2011), a été utilisé. À partir d'une orbite donnée (à l'époque l'orbite qui était envisagée pour SWOT avait une répétitivité de 22 jours) le simulateur génère, pour une date donnée, la trace SWOT correspondante et le masque des cellules TRIP sous la fauchée. Les hauteurs d'eau issues du run vrai dans les cellules se trouvant dans le masque de la fauchée sont ensuite perturbées et forment les observations à la date en cours. Bien que l'erreur sur les Manning soient réduites de 40%, ainsi que les erreurs sur les hauteurs d'eau (30%) et sur les débits (7%), la méthode souffre des limitations de l'EKF face aux nonlinéarités du modèle ISBA-TRIP. Pedinotti et al. (2014) mentionne, entre autres, dans ses conclusions, l'intérêt de passer à une méthode d'assimilation ensembliste - de type EnKF - pour mieux tenir compte des autres sources d'incertitudes inhérentes au modèle et mieux gérer les nonlinéarités du modèle.

L'objectif de ce chapitre est alors de développer une maquette d'AD basée sur l'algorithme du Filtre de Kalman d'Ensemble afin de corriger les paramètres du modèle TRIP à partir de pseudo-observations de type SWOT. On commencera par décrire la méthode d'assimilation dans le cadre d'expériences jumelles, à savoir les pseudo-observations SWOT et l'algorithme choisi. Une série d'expériences jumelles ont ensuite été menées, afin d'étudier les performances de la maquette et comment celle-ci gère des sources d'erreurs fréquentes, telles que les erreurs de bathymétrie et la présence ou non de plaines inondées qui n'ont pas été étudiées par Pedinotti et al. (2014). La dernière section est alors consacrée aux conclusions et perspectives.

3.1 Description des variables observées et des variables à corriger

On rappelle d'abord les notations générales introduites au Chapitre 1, Section 1.4. En estimation de paramètres, le vecteur de contrôle au cycle k , $\mathbf{x}_k \in \mathbb{R}^{n_x}$, est constitué de n_x paramètres d'entrée du modèle que l'on considère comme les plus incertains ou les plus délicats à estimer (car on ne peut pas les observer/mesurer directement). Plus particulièrement en expériences jumelles, on va être amené à manipuler plusieurs types de vecteurs de contrôle :

- le vecteur de contrôle vrai $\mathbf{x}^t$, invariant dans le temps. Ce dernier contient les n_x valeurs des paramètres d'entrée de contrôle que l'on choisit préalablement comme les valeurs vraies et qu'on doit retrouver via l'AD,
- le vecteur de contrôle d'ébauche au cycle k , $\mathbf{x}_k^b$,
- et le vecteur de contrôle analyse au même cycle, $\mathbf{x}_k^a$.

L'opérateur modèle $M_{[k-1,k]}$ permet de transformer les paramètres de contrôle $\mathbf{x}_k$ en variables pronostiques du modèle $\mathbf{z}_k \in \mathbf{R}^{n_z}$ et éventuellement en variables diagnostiques $\mathbf{w}_k \in \mathbf{R}^{n_w}$ via l'opérateur Z_k :

$$\mathbf{z}_k = M_{[k-1,k]}(\mathbf{x}_k) \in \mathbf{R}^{n_z} \quad \mathbf{w}_k = Z_k(\mathbf{z}_k) = Z_k \circ M_{[k-1,k]}(\mathbf{x}_k) \in \mathbf{R}^{n_w}. \quad (3.1)$$

Dans le cadre de cette thèse, ces opérateurs sont définis dans le modèle ISBA-TRIP.

Le vecteur d'observations $\mathbf{y}_k^o$ est composé des n_y observations disponibles au cycle d'assimilation courant k et l'opérateur d'observation H_k permet d'exprimer les variables de contrôle $\mathbf{x}_k$ dans l'espace des observations :

$$\mathbf{y}_k = H_k(\mathbf{x}_k). \quad (3.2)$$

Dans le cadre particulier de l'estimation de paramètres, l'opérateur d'observation est nécessairement la composition de l'opérateur modèle $M_{[k-1,k]}$ (et éventuellement de l'opérateur diagnostiques Z_k), avec un autre opérateur de sélection ou d'interpolation (selon la nature des observations et de leur résolution).

Enfin, en expériences jumelles, les valeurs vraies des variables observées $\mathbf{y}_k^f$ sont aussi connues à l'avance.

3.1.1 Description des pseudo-observations SWOT

On cherche à générer des observations synthétiques de type SWOT. SWOT fournira des cartes bidimensionnelles de cotes d'eau des rivières (d'au moins 100 mètres de large). Les variables de sorties d'ISBA-TRIP se rapprochant le plus de ce produit SWOT sont les variables diagnostiques de hauteurs d'eau des rivières. Pour simplifier les écritures, à partir de maintenant, on redéfinit l'opérateur Z_k comme l'opérateur transformant les variables pronostiques d'ISBA-TRIP $\mathbf{z}_k$ (stocks d'eau dans le réservoir de la rivière dans TRIP) en hauteurs d'eau $\mathbf{w}_k$ des réservoirs de rivières TRIP.

Il nous faut maintenant définir plus en détails l'opérateur d'observation H_k , et plus précisément la nature de l'opérateur qui permet de transformer l'ensemble des variables diagnostiques de hauteurs d'eau $\mathbf{w}_k^f$ en l'ensemble des variables observées $\mathbf{y}_k^f$. L'orbite actuelle de SWOT est fixée à une altitude de 890,5 km pour une répétitivité de 21 jours et l'instrument KaRIn observera les surfaces sous deux fauchées de 50 km (séparées par un trou de 20 km au nadir). Pour simuler la couverture spatio-temporelle des observations SWOT, on réutilise le simulateur utilisé par Biancamaria et al. (2011); Pedinotti et al. (2014). Pour une date donnée, étant donné la définition de l'orbite, on est en mesure de générer les traces au sol des fauchées SWOT. Cette trace au sol est ensuite mise à l'échelle du modèle ISBA-TRIP ($0,5^\circ \times 0,5^\circ$) en sélectionnant les cellules TRIP dont au moins 50% de l'aire est comprise dans le polygone de la fauchée. De ce fait, dans les traces SWOT à la résolution du modèle, on ne peut pas voir le trou au nadir de 20 km, trop petit par rapport aux cellules (environ $50 \text{ km} \times 50 \text{ km}$). La Figure 3.2 illustre alors la sélection des cellules ISBA-TRIP sous la fauchée réelle sur le bassin de l'Amazonie. Ainsi, l'opérateur d'observation H_k est la composition de l'opérateur $Z_k \circ M_{[k-1,k]}$ avec un opérateur de sélection S_k qui consiste à sélectionner les cellules ISBA-TRIP sous la fauchée SWOT selon les jours d'observations :

$$\mathbf{y}_k = S_k \circ Z_k \circ M_{[k-1,k]}(\mathbf{x}_k) = H_k(\mathbf{x}_k). \quad (3.3)$$

Finalement, pour générer le vecteur d'observation $\mathbf{y}_k^o$ à partir du vecteur des hauteurs d'eau vraies sous le masque d'observation $\mathbf{y}_k^f$, on perturbe simplement chacune des n_y hauteurs d'eau vraies

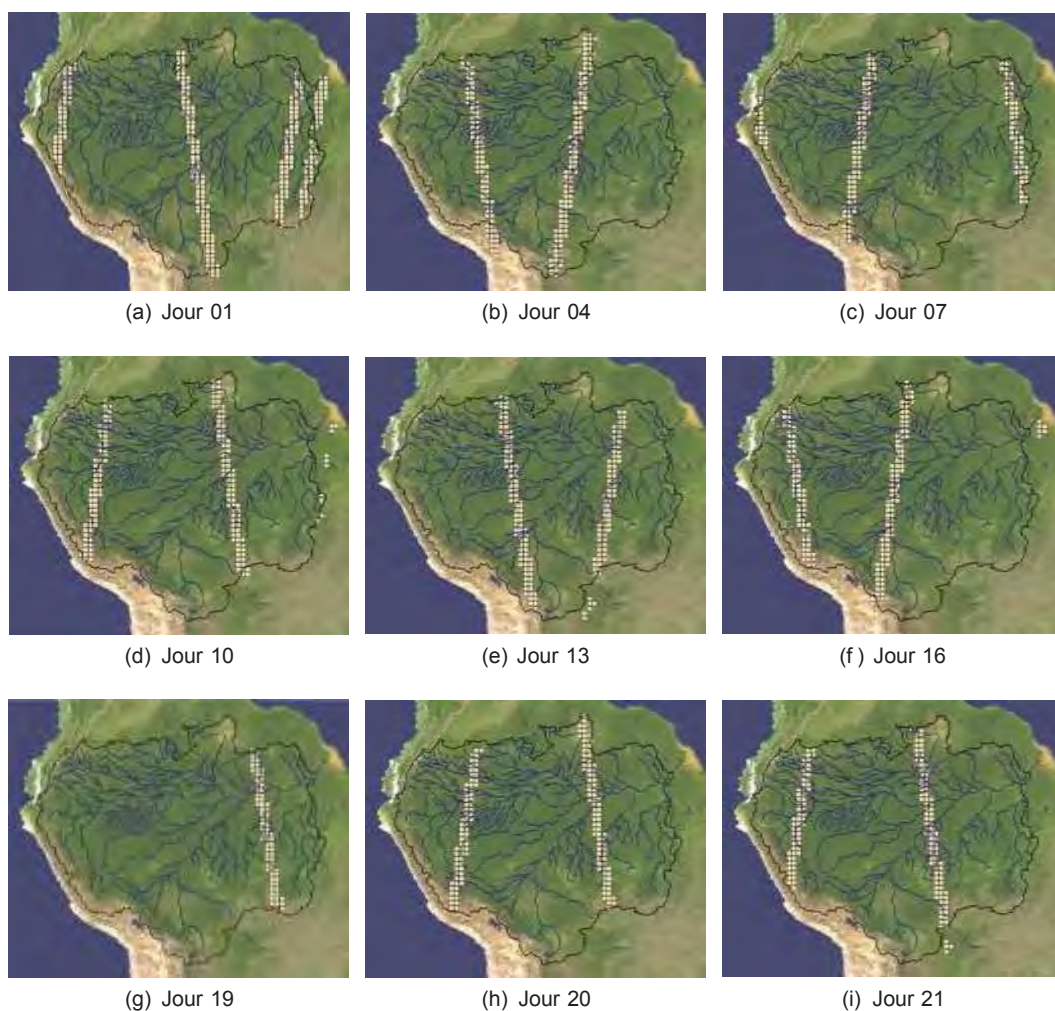


Figure 3.2 – Traces SWOT au-dessus du bassin de l’Amazone à la résolution d’ISBA-TRIP. Le contour du bassin est représenté par un polygone noir et les principaux cours d’eau du bassin sont représentés par le chevelu bleu marine. Les pixels TRIP compris dans le masque d’observation sont mis en évidence en beige.

$\mathbf{y}_{k,j}^t$, $j = 1 \dots n_y$, en leur ajoutant un bruit gaussien $\mathcal{E}_{k,j}$ de moyenne nulle et d'écart-type σ^o . Conformément au bilan d'erreurs défini dans le Science Requirements Document (Rodríguez, 2016) (voir le Tableau 1.5, Chapitre 1), on utilisera comme erreur d'observation un bruit blanc de $\sigma^o = 10$ cm :

$$\forall j = 1 \dots n_y, \mathbf{y}_{k,j}^o = \mathbf{y}_{k,j}^t + \mathcal{E}_{k,j}^o, \mathcal{E}_{k,j}^o \sim N(0, \sigma^o). \quad (3.4)$$

Pour information, la construction de H_k est basée sur les hypothèses simplificatrices suivantes :

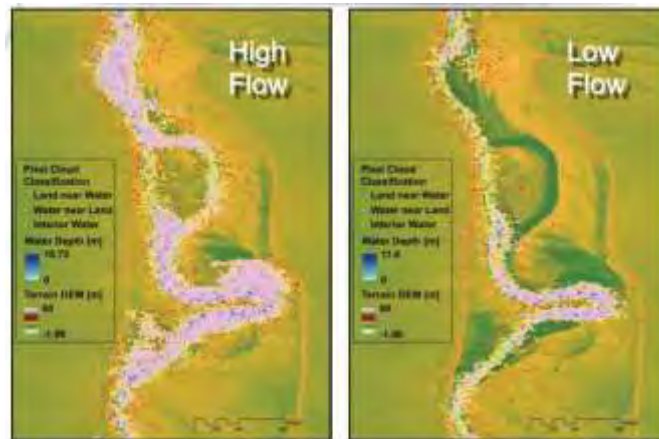
- Les mesures SWOT sont originalement des cotes d'eau référencées par rapport à un géoïde, alors que les hauteurs d'eau ISBA-TRIP sont référencées par rapport à la bathymétrie intrinsèque du modèle. Dans le cadre des expériences jumelles, les hauteurs d'eau observées et simulées sont référencées par rapport au même géoïde. Dans un premier temps, la bathymétrie de la rivière équivalente est connue et est la même pour les observations SWOT (ainsi, les cotes d'eau = les hauteurs d'eau).
- Le produit de cotes d'eau SWOT, tel qu'il est défini dans le Science Requirements Document (Rodríguez, 2016), se présente sous deux formes : (1) un nuage de points géoréférencés et de fait à une résolution irrégulière (en opposition à la maille régulière de TRIP) et bien plus fine que celle de TRIP, voir Figure 3.3a et (2) un produit contenant le linéique du réseau fluvial, découpé en tronçons de quelques kilomètres avec des mesures moyennées sur ces tronçons qui restera tout de même à une résolution plus fine que celle de TRIP, voir Figure 3.3b. Il est clair qu'à l'échelle du bassin de l'Amazonie et donc pour une modélisation à grande échelle telle que celle d'ISBA-TRIP, c'est ce deuxième produit qui est le plus intéressant. Cependant, même en utilisant ce produit, un traitement de ces observations - par « moyennage » par exemple - ou une mise à l'échelle des sorties ISBA-TRIP sera nécessaire pour pouvoir comparer des deux sources d'informations. Vu l'état actuel du développement de la maquette d'assimilation, on émet l'hypothèse simplificatrice que les points d'observations SWOT sont à la même résolution que les cellules du modèle.
- Enfin, on suppose que les erreurs d'observations sont modélisées par un bruit blanc. On ne tient pas compte ici des corrélations dans les erreurs qui existent le long de la fauchée (bruit instrumental, layover, mouvement du satellite, traversée de la troposphère, pluie ...).

Maintenant que la nature des observations que l'on va assimiler a été clairement défini, on va définir plus en détails la nature et la forme des variables de contrôle.

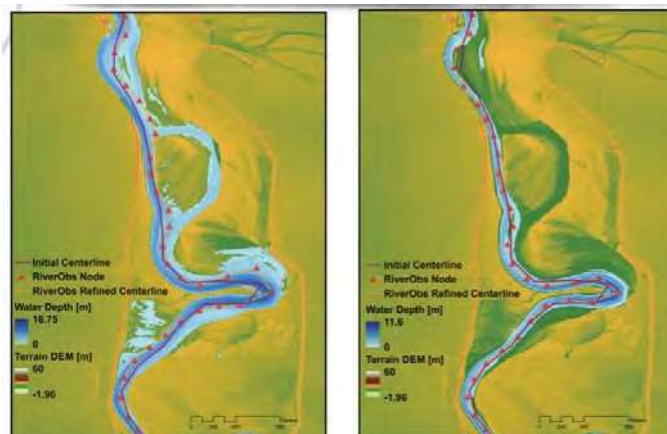
3.1.2 Description des variables de contrôle

Les résultats de l'analyse de sensibilité ont montré que les hauteurs d'eau étaient sensibles à essentiellement trois paramètres géomorphologiques, par ordre décroissant d'importance : le coefficient de Manning, la pente et la largeur du tronçon. Potentiellement, ces trois paramètres peuvent être corrigés via l'assimilation. Cependant, dans cette étude, on se concentrera sur la correction des coefficients de Manning seuls.

La distribution d'origine des coefficients de Manning dans TRIP est spatialement distribuées sur les bassins, c'est-à-dire que les tronçons dans chaque cellule TRIP ont chacun leur propre valeur de coefficient de Manning. Pedinotti et al. (2014) a choisi de corriger les valeurs de Manning de chaque cellule indépendamment. Cette configuration génère cependant de l'équifinalité et l'AD ne converge pas nécessairement vers les valeurs vraies. De plus, le coefficient de Manning est un paramètre difficile à estimer et il est encore plus compliqué de vérifier les valeurs des coefficients à l'échelle de la cellule. Malgré tout, la distribution originelle suppose que la valeur du coefficient de Manning décroît plus



(a) Nuage de points



(b) Linéique

Figure 3.3 – Exemple de (a) nuages de points et de (b) linéique que pourra fournir SWOT (ici issu du simulateur SWOT) sur un tronçon de la rivière Sacramento (Figures générées par l'équipe de Mike Durand de l'Ohio State University, Williams et al., 2016).

on se rapproche de l'embouchure et à l'inverse, augmente quand on remonte vers les têtes de bassins (voir Chapitre 2).

Pour définir les variables de contrôle, on conserve cette distribution d'origine des paramètres et on va chercher à l'ajuster non pas cellules par cellules mais régions par régions. Pour cela, on reconsidère la division du bassin en zones hydro-géomorphologiques (voir Figure 3.5). Pour chaque zone, comme pour la perturbation faite pour l'analyse de sensibilité, l'AD cherchera à optimiser un coefficient multiplicatif de la distribution des coefficients de Manning. Finalement, le vecteur de contrôle se compose des n_x (= nombre de zones hydro-géomorphologiques) coefficients correcteurs multiplicatifs de la distribution des coefficients de Manning :

$$\mathbf{x}_k = [n_{\text{mult},1} \dots n_{\text{mult},n_x}]^T. \quad (3.5)$$

Pour corriger la distribution des coefficients de Manning via l'AD, on décide d'utiliser une méthode ensembliste séquentielle de type Filtre de Kalman d'Ensemble (EnKF, Evensen, 2003). Cette méthode est décrite dans la section ci-après.

3.2 Choix de la méthode d'assimilation

Avant de choisir un algorithme d'AD, il nous faut d'abord commencer par définir la fenêtre d'assimilation sur laquelle on va travailler.

3.2.1 Définition de la fenêtre d'assimilation et conséquences

Des pseudo-observations SWOT seront disponibles tous les jours sur le bassin de l'Amazonie. Comme on se place en expériences jumelles, il n'y a pas de risques d'observations manquantes sur la série temporelle des observations à une cellule donnée. Par ailleurs, la répétitivité de SWOT étant de 21 jours, les cellules observées sur une période inférieure à 21 jours changent d'un jour à l'autre avec la position des traces SWOT. Pedinotti et al. (2014) a choisi une fenêtre d'assimilation de 2 jours pour reproduire un comportement opérationnel où les données sont disponibles en quasi-temps réel ou avec un léger retard. Les séries temporelles de coefficient de Manning après analyse affichent, autour d'une tendance générale qui converge plus ou moins vers la valeur vraie, une succession régulière de variations brutales. Ce comportement a alors été associé au fait qu'à certains instants, le coefficient est correctement corrigé dû à une observation directe de la cellule et, aux autres instants, la correction est due à l'observation de cellules adjacentes. Pour limiter ces comportements, on décide alors d'utiliser une fenêtre d'assimilation de 21 jours. Au cours d'une même fenêtre d'assimilation, les observations ont alors lieu, pour la plupart, à des instants différents du temps d'analyse.

Suite aux conclusions de Pedinotti et al. (2014), on décide de baser l'AD sur un algorithme ensembliste de type EnKF (Evensen, 2003, 2009). La définition classique de l'EnKF implique qu'une analyse soit générée dès qu'une observation est disponible, dans le cas présent, tous les jours. Or, dans le cadre de notre étude, une fenêtre d'assimilation de 21 jours va être utilisée. Par conséquent, le problème actuel consiste à assimiler des observations désynchronisées avec le temps d'analyse de l'EnKF.

Un tel problème a déjà été considéré par plusieurs études : le Lisseur de Kalman d'Ensemble ou EnKS (Ensemble Kalman Smoother en anglais), introduit par Evensen and Leeuwen (2000), est

une extension directe de l'EnKF consistant, dès qu'une nouvelle observation est disponible, à tenir compte de la totalité des observations passées, en plus de la nouvelle, pour générer l'analyse. Les aspects pratiques de l'EnKS sont aussi discutés dans Evensen (2003, 2009). L'EnKS est en fait une version séquentielle, plus pratique à implémenter, du Lisseur d'Ensemble ou ES (Ensemble Smoother en anglais). Initialement, Leeuwen and Evensen (1996) proposèrent ce lisseur qui détermine une estimation du contrôle minimisant la variance de la densité spatio-temporelle décrivant l'évolution du modèle. Cette méthode tenait alors compte de toutes les observations passées et futures. Cependant, elle s'est montrée bien moins efficace que l'EnKF et l'EnKS lors d'études comparatives (Evensen, 1997; Evensen and Leeuwen, 2000). Parallèlement, Hunt et al. (2004) développa le 4D-EnKF (« 4D » en écho aux méthodes d'assimilation variationnelles 4D-Var ; Talagrand and Courtier, 1987) qui assimilent des observations ayant lieu à des instants différents). Dans le 4D-EnKF, les observations modèles -

$\mathbf{y}_k^{b,[l]} = H_k(\mathbf{x}_k^{b,[l]})$ - au temps d'observation $k^i \in [k-1, k]$, différent du temps d'analyse k , sont réécrites comme une combinaison linéaire de l'ensemble des observations modèle au temps d'analyse

$$\mathbf{Y}_{e,k}^b = \mathbf{y}_k^{b,[l]} \begin{matrix} n_e \\ l=1 \end{matrix} :$$

$$\forall l = 1 \dots n_e : \mathbf{y}_k^{b,[l]} = H_k(\mathbf{x}_k^{b,[l]}) = \sum_{l=1}^n w_l \mathbf{y}_k^{b,[l]}.$$

Cette étape constitue une réécriture de l'opérateur d'observation comme une fonction de $\mathbf{Y}_{e,k}^b$ seul. Ainsi, le problème est ramené à un EnKF classique. Plus tard, Hunt et al. (2007) proposera une version asynchrone (c'est-à-dire avec des observations à des temps différents du temps d'analyse) du LETKF (Local Ensemble Transform Kalman Filter ; Bishop et al., 2001; Ott et al., 2004).

Dans le cadre de l'estimation de paramètres mise en place dans ce chapitre, on se basera sur le Filtre de Kalman d'Ensemble Asynchrone ou AEnKF (Asynchronous Ensemble Kalman Filter) décrit par Sakov et al. (2010) et Rakovec et al. (2015). Comme précisé dans Sakov et al. (2010) et Rakovec et al. (2015), l'AEnKF n'est pas une nouvelle méthode à part entière mais une modification simple de l'EnKF. Le principe est d'augmenter la dimension de l'état afin de considérer à la fois les observations aux temps d'analyse et celles aux temps passés. Ceci entraîne l'augmentation des dimensions des matrices de covariance d'erreurs mais celles-ci contiennent alors les covariances entre des observations ayant lieu à des temps différents. Au final, l'analyse classique de l'EnKF, voir Eq. 1.48 au Chapitre 1, est appliquée. La Figure 3.4 compare les schémas d'analyse de l'EnKF, de l'EnKS et de l'AEnKF. À notre connaissance, l'AEnKF n'a pas été utilisé pour de l'estimation de paramètres.

Sakov et al. (2010) et Rakovec et al. (2015) décrivent l'AEnKF dans le cadre d'expériences d'estimation d'état. Nous allons à présent décrire l'AEnKF dans le cadre de l'estimation de paramètres.

3.2.2 Le Filtre de Kalman d'Ensemble Asynchrone en estimation de paramètres

Pour commencer, si l'indice k représente toujours l'index du cycle d'assimilation, il doit se distinguer de l'indice de jour (car 1 cycle d'assimilation dure 21 jours) qui est l'unité de temps des observations. On notera alors en indice :

- $k(ii)$, $ii = 1 \dots 2$, pour dénoter le ii -ème jour du cycle d'assimilation k ,
- k^i , les indices de jours en absolu, c'est-à-dire indépendamment du cycle d'assimilation.

Un vecteur de contrôle au cycle d'assimilation k , $\mathbf{x}_k$, est composé des n_x (=nombre de zones

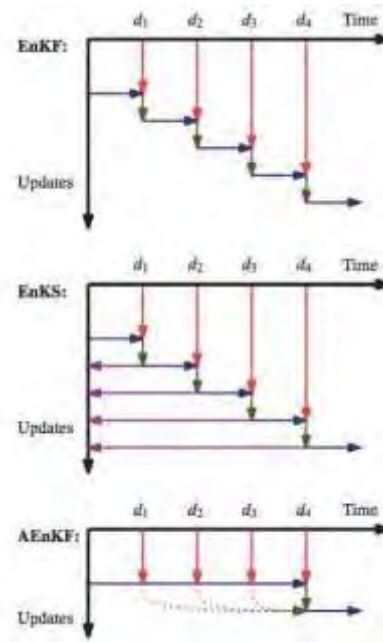


Figure 3.4 – Rakovec et al. (2015) : Illustration du schéma d'analyse de l'EnKF, de l'EnKS et de l'AEnKF. L'axe des abscisses représente le temps et les observations (d_1, d_2, d_3, d_4) sont ici données à des intervalles de temps réguliers. Les flèches bleues représentent la propagation du modèle, les flèches rouges représentent l'introduction des observations et les flèches vertes indiquent les étapes d'analyse. Les flèches violettes pour l'EnKS montrent l'étendue de l'analyse (et prend en compte toutes les observations passées). Les flèches pointillées vertes pour l'AEnKF montrent que les observations antérieures au temps d'analyse mais dans la période de propagation du modèle sont prises en compte pour l'analyse.

hydro-géomorphologiques du bassin étudié) coefficients multiplicateurs de la distribution des coefficients de Manning, $n_{\text{mult},i}, i = 1 \dots n_x$:

$$\mathbf{x}_k = [x_{k,1} \dots x_{k,n_x}]^T = [n_{\text{mult},1} \dots n_{\text{mult},n_x}]^T. \quad (3.6)$$

Dans le cadre de l'AEnKF, on génère un ensemble de vecteur de contrôle d'ébauche $\mathbf{x}_{e,k}^b$ de dimension $n_x \times n_e$:

$$\mathbf{x}_{e,k}^b = [\mathbf{x}_k^{b,[1]}, \dots, \mathbf{x}_k^{b,[n_e]}], \quad (3.7)$$

où $\mathbf{x}_k^{b,[l]}, l = 1 \dots n_e$ suit une loi normale de moyenne $\bar{\mathbf{x}}_k^b$ et de matrice de covariance $(\sigma_{\mathbf{x}_k^b})^2 \cdot \mathbf{I}_{n_x}$, avec $\mathbf{I}_{n_x}$ la matrice identité de dimension $n_x \times n_x$.

Au ii -ème jour du cycle k , les $n_{y,k(ii)}$ observations sont regroupées dans le vecteur $\mathbf{y}_{k(ii)}^o$ telles que :

$$\mathbf{y}_{k(ii)}^o = [y_{k(ii),1}^o \dots y_{k(ii),n_{y,k(ii)}}^o]^T \in \mathbb{R}^{n_{y,k(ii)}}. \quad (3.8)$$

Le vecteur global d'observations au cycle k , $\mathbf{y}_k^o$ concatène alors les 21 vecteurs d'observations journaliers $\mathbf{y}_{k(ii)}^o, ii = 1 \dots 21$ tel que :

$$\mathbf{y}_k^o = [\mathbf{y}_{k(1)}^{o,T} \dots \mathbf{y}_{k(21)}^{o,T}]^T \in \mathbb{R}^{n_{y,k}}, \quad n_{y,k} = \sum_{ii=1}^{21} n_{y,k(ii)}. \quad (3.9)$$

De plus, conformément à Burgers et al. (1998), on génère un ensemble d'observations $\mathbf{Y}_{e,k}^o$ de taille $n_{y,k} \times n_e$ en ajoutant n_e fois un bruit blanc E_k^o de covariance $\mathbf{R}_k$ au vecteur d'observations $\mathbf{y}_k^o$:

$$\mathbf{Y}_{e,k}^o = [\mathbf{y}_k^{o,[1]}, \dots, \mathbf{y}_k^{o,[n_e]}], \quad (3.10)$$

avec

$$\forall l = 1 \dots n_e, \mathbf{y}_k^{o,[l]} = \mathbf{y}_k^o + E_k^{o,[l]}, \quad E_k^{o,[l]} \sim \mathcal{N}(0, \mathbf{R}_k).$$

Identiquement au vecteur d'observations, l'opérateur d'observation global au cycle k , H_k , est la concaténation des opérateurs d'observations journaliers $H_{k(ii)}$, définis à partir de l'Eq. 3.3, en considérant l'opérateur $\mathbf{M}$ intégrant le modèle entre les instants $k-1$ et $k(ii)$ ainsi que les opérateurs diagnostique et sélection à l'instant $k(ii)$, $Z_{k(ii)}$ et $S_{k(ii)}$. Il en résulte que :

$$H_k = \begin{bmatrix} H_{k(1)} \\ H_{k(2)} \\ \vdots \\ H_{k(21)} \end{bmatrix} = \begin{bmatrix} S_{k(1)} \circ Z_{k(1)} \circ \mathbf{M}_{[k-1,k(1)]} \\ S_{k(2)} \circ Z_{k(2)} \circ \mathbf{M}_{[k-1,k(2)]} \\ \vdots \\ S_{k(21)} \circ Z_{k(21)} \circ \mathbf{M}_{[k-1,k(21)]} \end{bmatrix}. \quad (3.11)$$

Selon les mêmes équations que l'EnKF, l'AEnKF génère, pour chaque membre $l = 1 \dots n_e$ de l'ensemble de contrôle, un vecteur de contrôle analyse $\mathbf{x}_k^{a,[l]}$ tel que :

$$\forall l = 1 \dots n_e, \mathbf{x}_k^{a,[l]} = \mathbf{x}_k^{b,[l]} + \mathbf{K}_{e,k} (\mathbf{y}_k^{o,[l]} - H_k(\mathbf{x}_k^{b,[l]})), \quad (3.12)$$

où $\mathbf{K}_{e,k}$ est le gain de Kalman défini par : $\mathbf{K}_{e,k} = \mathbf{P}_{e,k}^b \mathbf{H}_{e,k}^T (\mathbf{H}_{e,k} \mathbf{P}_{e,k}^b \mathbf{H}_{e,k}^T + \mathbf{R}_k)^{-1}$.

La matrice de covariance $\mathbf{P}_{e,k}^b \mathbf{H}_{e,k}^T$ est estimée stochastiquement par la matrice des covariance entre les coefficients de Manning (variables de contrôle) et les hauteurs d'eau simulées par le modèle à partir des coefficients de Manning de contrôle. On l'estime stochastiquement - $\mathbf{P}^b \mathbf{H}^T_{e,k}$ - par :

$$\mathbf{P}^b \mathbf{H}^T_{e,k} = \frac{1}{n_e - 1} \sum_{l=1}^{n_e} (\mathbf{x}_k^{b,[l]} - \overline{\mathbf{x}_{\cdot,k}^b}) (\mathbf{H}_k(\mathbf{x}_k^{b,[l]}) - \overline{\mathbf{H}_k(\mathbf{x}_{\cdot,k}^b)})^T, \quad (3.13)$$

où $\overline{\mathbf{x}_{\cdot,k}^b}$ représente la moyenne sur l'ensemble des vecteurs de contrôle d'ébauche $\mathbf{x}_{e,k}^b$ (idem pour $\overline{\mathbf{H}_k(\mathbf{x}_{\cdot,k}^b)}$). De même, la matrice de covariance $\mathbf{H}_{e,k} \mathbf{P}_{e,k}^b \mathbf{H}_{e,k}^T$ est aussi estimée stochastiquement par la matrice de covariance entre les hauteurs d'eau simulées aux cellules d'observations :

$$\mathbf{H} \mathbf{P}^b \mathbf{H}^T_{e,k} = \frac{1}{n_e - 1} \sum_{l=1}^{n_e} (\mathbf{H}_k(\mathbf{x}_k^{b,[l]}) - \overline{\mathbf{H}_k(\mathbf{x}_{\cdot,k}^b)}) (\mathbf{H}_k(\mathbf{x}_k^{b,[l]}) - \overline{\mathbf{H}_k(\mathbf{x}_{\cdot,k}^b)})^T. \quad (3.14)$$

Pour limiter les effets de covariances irréalistes dûs à l'échantillonnage (utilisation d'un ensemble pour estimer les densités de probabilités), on traitera la matrice $\mathbf{P}^b \mathbf{H}^T_{e,k}$ une fois estimée stochastiquement et avant de l'utiliser lors de l'étape d'analyse. Ce traitement consiste à annuler les covariances reliant des zones avec des cellules qui n'ont théoriquement pas de lien entre elles. Ce « lien » est basé sur la définition de la matrice de covariance d'erreurs $\mathbf{P}^b \mathbf{H}^T_{e,k}$ et sur le réseau de routage de TRIP.

La matrice de covariance des erreurs d'observations $\mathbf{R}_k$ est la concaténation de l'ensemble des matrices de covariance d'erreurs d'observations élémentaires au jour $k(ii)$, $i = 1 \dots n_e$:

$$\mathbf{R}_k = \begin{pmatrix} \mathbf{R}_{k(1)} & & & \\ & \mathbf{R}_{k(2)} & 0 & \\ & 0 & \ddots & \\ & & & \mathbf{R}_{k(21)} \end{pmatrix}. \quad (3.15)$$

Les erreurs d'observations étant construites comme des bruits blancs d'écart-type $\sigma^o = 10$ cm, toutes les matrices $\mathbf{R}_{k(ii)}$ sont diagonales définies positives telles que :

$$\mathbf{R}_{k(ii)} = (\sigma^o)^2 \cdot \mathbf{I}_{n_{y,k(ii)}}, \quad (3.16)$$

où $\mathbf{I}_{n_{y,k(ii)}}$ est la matrice identité de dimension $n_{y,k(ii)}$. Il en advient que :

$$\mathbf{R}_k = (\sigma^o)^2 \cdot \mathbf{I}_{n_{y,k}}. \quad (3.17)$$

Enfin, une fois que l'ensemble d'analyse $\mathbf{X}_{e,k}^a = [\mathbf{x}_k^{a,[l]}]_{l=1}^{n_e}$ est déterminé, on est en mesure de lancer le cycle d'assimilation suivant. En estimation d'état, on corrige les sorties du modèle qui sont ensuite directement ré-utilisées comme condition initiale pour le cycle suivant. Cependant, en estimation de paramètres, le cyclage de l'analyse n'est pas si direct. Comme on corrige les paramètres d'entrée, les sorties du modèle après l'étape d'analyse ne sont pas impactées par les corrections et on ne peut pas les utiliser directement pour initialiser le prochain cycle d'assimilation. En estimation de paramètres, il est alors nécessaire de ré-effectuer les runs d'ensemble modèle sur la fenêtre d'assimilation mais en prenant les paramètres analyse en entrée. Ainsi, l'apport des corrections des paramètres est introduit dans la condition initiale du modèle pour le cycle d'assimilation suivant.

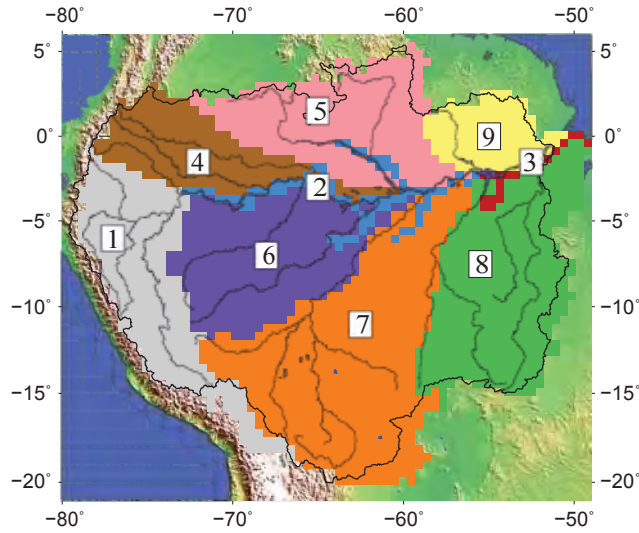


Figure 3.5 – Division du bassin de l’Amazonie en $n_x = 9$ zones hydro-géomorphologiques.

3.2.3 Configuration des simulations d’estimation de paramètres

On va maintenant détailler les simulations d’expériences jumelles réalisées dans le cadre de l’estimation de paramètres et la configuration générale des expériences.

On mènera les expériences sur le bassin de l’Amazonie car l’analyse de sensibilité d’ISBA-TRIP a été menée sur ce même bassin. Toujours en suivant l’analyse de sensibilité, on réutilisera les données atmosphériques GSWP3 pour forcer le modèle.

3.2.3.1 Description générale des expériences d’assimilation

Comme vu au Chapitre 2, le bassin de l’Amazonie a été découpé en 9 zones hydro-géomorphologiques (rappelée en Figure 3.5). Il en découle que $n_x = 9$.

Pour la simulation vraie, on choisit le vecteur de contrôle suivant :

$$\mathbf{x}^f = 1.65 \ 0.85 \ 0.85 \ 0.95 \ 0.9 \ 0.95 \ 0.9 \ 1.3 \ 1.4 . \quad (3.18)$$

On augmente volontairement les coefficients de Manning sur la zone montagneuse de la Cordillère des Andes (zone 1) et sur les boucliers brésiliens et guyannais (zone 8 et 9); et on les diminue sur les autres zones, en particulier sur le cours principal (zone 2 et 3).

Le vecteur de contrôle pour le run libre est défini par :

$$\mathbf{x}^f = 1.50 \ 0.50 \ 0.50 \ 0.50 \ 0.50 \ 0.50 \ 0.50 \ 1.50 \ 1.50 . \quad (3.19)$$

On conserve la logique dans les augmentations/réductions des coefficients mais on prend des valeurs moins précises, plus générales.

Au premier cycle d’assimilation, les valeurs de $\overline{x_{1,i}^b}$, $i = 1 \dots n_x$ (donc les neuf coefficients multiplicatifs des cartes de coefficients de Manning), sont prises égales à celles du run libre avec une erreur

Zone ...	influence les cellules de la zone ...
1	1, 2, 3
2	2, 3
3	3
4	4, 2, 3
5	5, 2, 3
6	6, 2, 3
7	7, 2, 3
8	8, 3
9	9, 3

Table 3.1 – Effets des zones hydrogéomorphologiques (associées aux variables de contrôle) sur l'ensemble des cellules du bassin hydrologique (associées aux variables observées).

de 25%, i.e. $\sigma_{x_{1,i}^b} = 0,25$, $i = 1 \dots n_x$. Aux cycles d'assimilation suivants, la valeur moyenne des variables de contrôle est égale à la valeur moyenne d'analyse estimée au cycle précédent :

$$\forall k > 1, \forall i = 1 \dots n_x, \bar{x}_k^b = \bar{x}_{k-1}^a, \quad (3.20)$$

et l'erreur sur ces valeurs est maintenue à 25% pour conserver assez de dispersion dans l'ensemble d'ébauche. Enfin, le Tableau 3.1 détaille les liens entre les différentes zones du bassin utilisés pour traiter la matrice $\mathbf{P}^b \mathbf{H}^T_{e,k}$. Concrètement, on est parti du principe que l'élément situé à la ligne n et à la colonne m de la matrice $\mathbf{P}^b \mathbf{H}^T_{e,k}$ représente l'impact de l'erreur sur la variable de contrôle n sur l'erreur de la variable modèle (exprimée dans l'espace des observations) m . En d'autres termes, cet élément caractérise l'effet de l'erreur dans le coefficient de Manning de la zone hydrogéomorphologique n sur l'observation modèle m . Ainsi, la zone n a un impact sur la cellule observée m que si cette zone est en amont de la cellule observée.

La Figure 3.2 illustre les traces SWOT à la résolution du modèle ISBA-TRIP sur le bassin de l'Amazonie au fil des jours. Pour chaque cycle de 21 jours, le nombre total d'observations est de $n_{y,k} = 3628$. Il est possible que certaines cellules soient observées plusieurs fois sur un même cycle, la Figure 3.6 détaille le nombre d'observations par cellule sur une fenêtre d'assimilation complète. Par ailleurs, lors de la construction des masques d'observation, ne devait être sélectionnées que les cellules dont la largeur TRIP étaient supérieures ou égale à 100 mètres. Or, la résolution du modèle fait que la totalité des cellules du bassin ont une largeur de plus de 100 mètres et sont sélectionnées dans le masque d'observations.

Enfin, on prend une taille d'ensemble de $n_e = 101$ comme une taille à la fois raisonnable pour le temps de calcul et assez grande pour correctement estimer les matrices de covariance d'erreurs. Le Tableau 3.2 résume l'ensemble des paramètres de configuration des expériences.

Le Tableau 3.3 détaille maintenant les différentes expériences d'estimation de paramètres menées dans ce chapitre. Le cadre idéalisé des expériences jumelles nous permet de forcer certains caractères des simulations hydrologiques (inondations, erreurs sur la bathymétrie) afin de voir comment la maquette d'assimilation se comporte face aux problèmes engendrés.

On va ainsi chercher à étudier le comportement de la maquette en présence d'inondations. Pour cela, on génère deux runs vrais. Les deux utilisent la distribution de coefficients de Manning vraie



Figure 3.6 – Couverture spatio-temporelle des pseudo-observations SWOT (à la résolution du modèle ISBA-TRIP) sur un cycle de 21 jours : (orange) 0 observation, (jaune) 1 observation, (vert) 2 observations sur le cycle.

Paramètres	Valeurs									
n_x	9									
n_y	3628									
n_e	101									
$\mathbf{x}^t$	1.65	0.85	0.85	0.95	0.9	0.95	0.9	1.3	1.4	
$\mathbf{x}^f$	1.50	0.50	0.50	0.50	0.50	0.50	0.50	1.50	1.50	
$\mathbf{x}_i^b$	$\mathbf{x}^f$									
$\sigma_{\mathbf{x}^b}$	0, 25 [%]									
σ^o	0, 10 [m]									

Table 3.2 – Récapitulatif des paramètres de configuration pour les expériences jumelles d'estimation de paramètres.

Nom de l'expérience	Inondations	Erreurs sur la bathymétrie	Nature des observations	Date de début	Nbre de cycles
EP1F	oui	non	Hauteurs d'eau	01 Janvier 2008	17
EP1NF	non	non	Hauteurs d'eau	01 Janvier 2008	17
EP2F	oui	oui	Hauteurs d'eau	01 Janvier 2008	17
EP3-1	oui	non	Anomalies de hauteurs	04 Mars 2008	3
EP3-2	oui	non	Anomalies de hauteurs	09 Septembre 2008	3
EP4F	oui	oui	Anomalies de hauteurs	04 Mars 2008	3
EP4NF	non	oui	Anomalies de hauteurs	04 Mars 2008	3

Table 3.3 – Description des expériences d'estimation de paramètres.

définie en Eq. 3.18 mais ont une distribution de la hauteur critique H_c de la rivière équivalente différente. La distribution d'origine de la hauteur critique de TRIP, H_c dans le Chapitre 2, est :

- soit, uniformément réduite de 60% - pour les expériences EP1F, EP2F, EP3-1, EP3-2 et EP4F - pour forcer des inondations pour le run vrai (H_c^t) et les runs pour l'assimilation,
- soit uniformément augmentée de 50% - pour les expériences EP1NF et EP4NF - pour forcer un run vrai sans aucune inondation tout au long de l'année.

3.2.3.2 Expériences avec inondations

Pour rappel, les trois expériences EP1F, EP1NF, EP3-1 et EP3-2, les paramètres du modèle TRIP, en particulier la largeur, la pente et la profondeur du réservoir de la rivière, sont exactement les mêmes entre le run vrai et les runs d'ensemble, il n'y a pas d'erreur sur la bathymétrie pour ces expériences. Les modèles utilisent un MNT (HYDRO1K pour TRIP) pour définir les pentes mais ce dernier peut avoir des biais. De plus pour TRIP, une relation géomorphologique est utilisée pour définir la distribution des largeurs et des hauteurs critiques. Pour étudier l'impact de l'erreur sur la bathymétrie, on mène les expériences EP2F, EP4F, EP4NF où l'on introduit manuellement une différence entre la bathymétrie vraie (et donc celle des observations) et celles des runs d'ensemble. Pour ce faire, on conserve la même pente et la même largeur et on ne perturbe que la distribution de hauteur critique vraie H_c^t en insérant un biais telle que :

$$H_c^b/H_c^a = H_c^t \times 0,75. \quad (3.21)$$

Dans la suite, on dénotera plus généralement par « bathymétrie » tout ce qui est en rapport avec la distribution de hauteur critique. Ainsi, l'Eq. 3.21 indique que la bathymétrie vraie est réduite de 25% pour les runs d'ensemble de l'AEnKF.

3.2.3.3 Expériences assimilant les anomalies de hauteurs d'eau

Une autre erreur potentielle lors de l'assimilation de cote d'eau serait d'avoir un biais constant entre les cotes d'eau TRIP et les cotes d'eau SWOT : la profondeur d'eau SWOT est correcte par rapport à la profondeur d'eau TRIP, mais par contre il y a un biais constant soit dans le MNT utilisé par TRIP, soit dans l'estimation des cotes d'eau altimétrique. Un exemple simplifié de cette situation est présenté sur la Figure 3.7. Dans ce cas, SWOT capture la bonne dynamique des variations de cote d'eau (car les profondeurs d'eau SWOT et TRIP, respectivement h_s^{SWOT} et h_s^{TRIP} , sont égales). Par contre, l'assimilation directe des cotes d'eau induirait un biais (puisque les cotes du lit du fleuve sont différentes, $H_{\text{alti}}^{\text{SWOT}} \neq H_{\text{alti}}^{\text{TRIP}}$), qui fausserait l'estimation du Manning. Une solution pour gérer ce problème est d'assimiler des anomalies de cotes d'eau, plutôt que les profondeurs d'eau directes. La

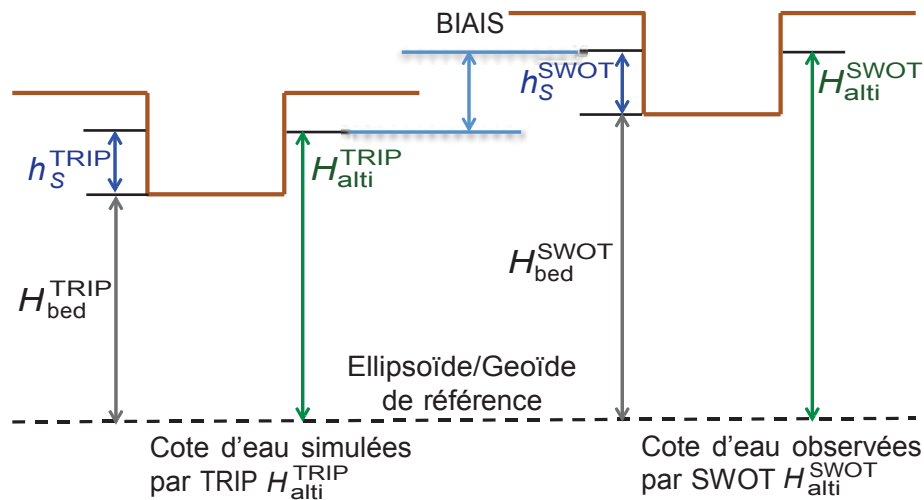


Figure 3.7 – Image de principe illustrant le cas d'un biais entre la cote d'eau observée SWOT (H_{alti}^{SWOT}) et la cote d'eau simulée TRIP (H_{alti}^{TRIP}), dû à un biais dans la cote du lit du fleuve.

maquette d'assimilation sera donc testée aussi avec des anomalies de cote d'eau, afin de déterminer si ces anomalies sont suffisantes pour retrouver le Manning vrai, ce qui permettrait de s'affranchir d'erreurs potentielles dues à un biais constant entre les cotes d'eau SWOT et les cotes d'eau TRIP.

L'étude de sensibilité a montré que, contrairement aux hauteurs d'eau, les anomalies ne sont pas identiquement sensibles aux coefficients de Manning tout au long de l'année mais sur des périodes plus courtes correspondant aux pleines saisons de hautes et basses eaux (voir Fig. 2.15-2.18). De ce fait, les expériences assimilant des hauteurs d'eau se feront sur une période totale de 1 an (environ 17 cycles de 21 jours) et les expériences assimilant des anomalies se feront sur des périodes plus courtes (3 cycles de 21 jours), correspondant aux périodes sur lesquelles les anomalies sont bien sensibles aux coefficients de Manning.

Pour générer les anomalies de hauteurs d'eau, le principe est de soustraire à la hauteur d'eau une hauteur d'eau de référence. La hauteur d'eau de référence est générée de la même façon pour les observations et l'ébauche en effectuant une intégration du modèle sur l'année précédant le début de la fenêtre d'assimilation et se terminant exactement à la date de début de la fenêtre. La hauteur d'eau de référence à une cellule donnée est alors la hauteur d'eau moyenne sur l'année à cette cellule. Cette méthode de définition des anomalies a plusieurs avantages :

- Lorsque l'on assimile les anomalies, la hauteur d'eau de référence est estimée pour chaque membre un à un. Ainsi, la hauteur de référence de chaque membre de l'ensemble d'ébauche est différente de celle des observations et son incertitude est prise en compte dans la génération de l'ensemble des observations modèle d'ébauche $\mathbf{Y}_{e,k}^b = H_k(\mathbf{X}_{e,k}^b)$.
- Aussi, le fait d'intégrer le modèle sur une certaine période avant la fenêtre d'assimilation permet de supprimer les effets de la condition initiale. Ainsi, la variance de l'ensemble $\mathbf{Y}_{e,k}^b$ ne proviendra que des incertitudes manuellement introduites à savoir : celles sur les coefficients de Manning (celle sur les hauteurs de référence lors de l'assimilation des anomalies).

La Figure 3.8 résume alors le fonctionnement temporel d'un cycle d'assimilation de données pour ces expériences d'estimation de paramètres. La variable K_{ref} représente la durée sur laquelle le modèle

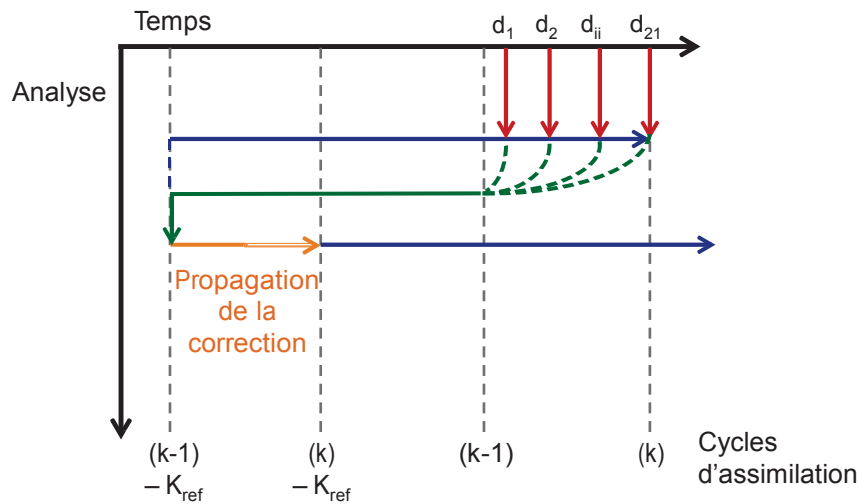


Figure 3.8 – Illustration du schéma d’analyse de l’AEnKF dans le cadre de l’estimation de paramètres. L’axe des abscisses représente le temps, le début des cycles d’assimilation sont représentés par les axes verticaux en tirets gris : on illustre ici le cycle k . Les flèches bleues représentent les périodes de propagation du modèle, les flèches rouges représentent l’introduction des observations au cycle courant et les flèches vertes indiquent les étapes d’analyse. En estimation de paramètres, le modèle est d’abord propagé sur la même période d’assimilation pour propager les effets de la correction des paramètres, représentée par la flèche orange. K_{ref} est égale à 4×21 jours pour les expériences EP1* et EP2* et 18×21 jours pour les expériences EP3* et EP4*.

est propagé avant la date de début du cycle d’assimilation pour supprimer les effets de la condition initiale : $K_{ref} = 4 \times 21$ jours pour les expériences EP1* et EP2* et $K_{ref} = 18 \times 21$ jours pour les expériences EP3* et EP4*.

Enfin, numériquement, les instances de l’AEnKF sont couplées au modèle ISBA-TRIP via le coupleur OpenPALM (www.cerfacs.fr/globc/PALM_WEB, Lagarde et al., 2001; Buis et al., 2006) et les runs d’ensemble sont menés en parallèle via la fonctionnalité Parasol du coupleur. L’Annexe A présente la maquette d’assimilation et décrit plus en détails le coupleur OpenPALM et Parasol.

Les sections qui suivent montrent les différents résultats obtenus pour les expériences détaillées en Tableau 3.3. Pour illustrer les résultats, on pourra détailler les profondeurs et/ou anomalies d’eau à certaines stations in situ, chacune situées dans une zone hydrogéomorphologique différente. La Figure 3.9 détaille la localisation des 9 stations in situ tests utilisées.

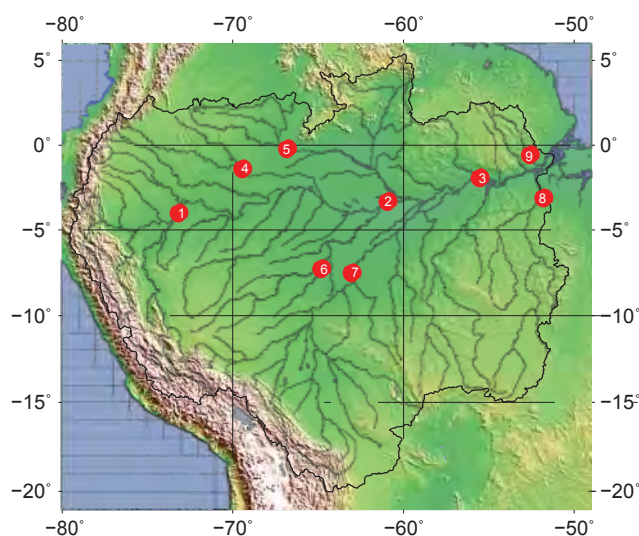


Figure 3.9 – Localisation sur le bassin de l’Amazonie des stations in situ illustrant les résultats de l’estimation de paramètres : (1) Tamashiyacu (zone 1), (2) Manacapuru (zone 2), (3) Óbidos (zone 3), (4) Villa Bittencourt (zone 4), (5) Curicuriari (zone 5), (6) Lábrea (zone 6), (7) Humaita (zone 7), (8) Uhe Belo Monte (zone 8), (9) São Francisco (zone 9).

3.3 Résultats pour l’assimilation des profondeurs d’eau

Dans cette section, on étudie le comportement de la maquette d’assimilation lorsque l’on assimile des profondeurs d’eau. Ceci revient donc à supposer que la bathymétrie est parfaitement connue et que l’on peut calculer des profondeurs d’eau à partir des mesures de cotes d’eau SWOT.

3.3.1 Résultats des expériences EP1F et EP1NF

Les expériences EP1F et EP1NF consistent à assimiler des observations de profondeurs d’eau pour corriger la distribution de coefficients de Manning de TRIP tout en ne considérant aucune erreur sur la bathymétrie (pour rappel, c’est-à-dire pas d’erreur sur la hauteur critique H_c) entre les simulations d’ensemble et la simulation vraie servant à générer les observations. L’expérience EP1F est menée à partir du run vrai avec inondations et l’expérience EP1NF est menée à partir du run vrai sans inondations. Les expériences d’AD sont réalisées sur une période de 17 cycles d’assimilation de 21 jours à partir du 1er Janvier 2008 jusqu’au 22 Décembre 2008 (ce qui revient à une année quasiment complète d’expérience). On rappelle alors que l’objectif de la maquette est de corriger la valeur de 9 facteurs multiplicatifs de la distribution des coefficients de Manning, chacun associé à une zone hydrogéomorphologique (rappelée en Figure 3.5). Enfin, l’AD est basée sur un Filtre de Kalman d’Ensemble Asynchrone, ce qui implique que les matrices de covariance d’erreur d’ébauche et d’analyse sont calculées à partir d’un ensemble. Dans la suite, on va caractériser cet ensemble par sa moyenne (d’ensemble) et sa dispersion.

Pour commencer, la Figure 3.10 donne, pour chaque zone, la valeur moyenne de l’ensemble d’analyse (après correction) au dernier cycle d’assimilation (cercles rouges) avec la dispersion de l’ensemble (barres verticales rouges). Ces valeurs sont ensuite comparées aux valeurs moyennes au premier cycle d’assimilation (carrés bleus) et aux valeurs vraies (losanges noirs). Pour l’expérience

EP1F (voir Fig. 3.10a), les moyennes d'ensemble analyse au dernier cycle sont rapprochées de la valeur vraie pour toutes les zones sauf pour les zones 1 et 9 où la valeur finale est à égale distance du vrai que la valeur au premier cycle. Dans tous les cas, la dispersion des ensembles analyse est réduite par rapport à la dispersion initiale. La 4^{ème} colonne du Tableau 3.4 donne explicitement les valeurs illustrées en Figure 3.10a. Les observations sont similaires pour l'expérience EP1NF (Figure 3.10b et 5^{ème} colonne du Tableau 3.4). Ces premiers résultats montrent le bon fonctionnement de la maquette d'AD dans les configurations de EP1F et EP1NF. On va alors regarder plus en détails le comportement de la maquette au fil des cycles d'assimilation.

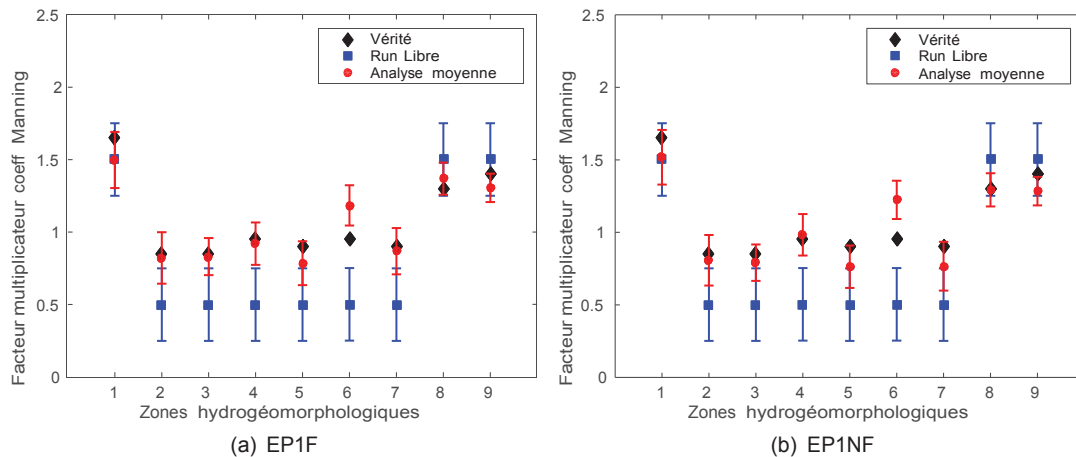


Figure 3.10 – Expériences EP1F (a) et EP1NF (b) : Comparaison des valeurs de facteurs multiplicatifs de la distribution de coefficients de Manning par zones : valeur vraie (losanges noirs), valeur a priori au 1er cycle d'assimilation avec erreur (carrés bleus) et valeur d'analyse moyenne au dernier cycle d'assimilation avec dispersion de l'ensemble (cercles rouges).

La Figure 3.11 détaille, pour l'expérience EP1F et pour chaque zone, l'évolution au cours des cycles de la valeur moyenne de l'ensemble des facteurs multiplicatifs de la distribution des Mannings après l'étape d'analyse (courbe rouge). Ces valeurs analyse sont toujours comparées à la valeur vraie (courbe constante noire) et à la valeur d'ébauche au premier cycle (courbe constante bleue). De plus, on matérialise, dans chaque graphe individuel, la présence d'inondations. Pour le run vrai (noir) et le run libre (bleu), on sélectionne dans chaque zone les cellules TRIP situées sur le réseau hydrologique principal c'est-à-dire les cellules ayant un *stream order* (coefficient SO introduit au Chapitre 2) supérieur à 6 (cette valeur a été choisie comme étant une valeur seuil minimale du SO pour que le réservoir de surface de TRIP correspondent à un tronçon réel de la rivière). On compte alors le nombre de jour où le schéma d'inondation de TRIP est effectivement activé. Les cycles dans lesquelles les inondations sont prises en compte sont ceux où plus de la moitié des cellules sélectionnées sont inondées au moins un jour du cycle. Les cycles « inondés » sont représentés par des courbes en tirets. Pour les ensembles d'ébauche et d'analyse, on matérialise les inondations sur la Figure 3.11 si plus de la moitié de l'ensemble est inondé selon les mêmes critères que la simulation vraie et libre : elles sont représentées par une zone grisée pour l'ensemble d'ébauche et par une courbe tirets pour l'ensemble d'analyse. Ce traitement permet de mettre en évidence les importantes inondations présentes sur les zones du cours principal (zones 2 et 3 essentiellement et aussi en zone 1), le long des affluents droits (zones 7 et 8) et dans une moindre mesure en zones 5 et 9. On peut noter que, dès le premier cycle,

Zones	Vrai	A priori (Sans assim)	Analyse (Après assim)	
			EP1F	EP1NF
1	1,65	1,5 (0,25)	1,497 (0,193)	1,516 (0,188)
2	0,85	0,5 (0,25)	0,822 (0,177)	0,806 (0,174)
3	0,85	0,5 (0,25)	0,831 (0,128)	0,789 (0,126)
4	0,95	0,5 (0,25)	0,920 (0,146)	0,981 (0,143)
5	0,9	0,5 (0,25)	0,785 (0,151)	0,761 (0,146)
6	0,95	0,5 (0,25)	1,184 (0,139)	1,222 (0,132)
7	0,9	0,5 (0,25)	0,868 (0,159)	0,764 (0,168)
8	1,3	1,5 (0,25)	1,368 (0,111)	1,291 (0,115)
9	1,4	1,5 (0,25)	1,305 (0,098)	1,282 (0,098)

Table 3.4 – Comparaison, pour chaque zone, des valeurs vraies (colonne 2), des valeurs a priori (colonne 3) avec dispersion (entre parenthèses) avec les valeurs analyse moyenne finale de l'expérience EP1F avec dispersion (colonne 4) et les valeurs analyse moyenne finale de l'expérience EP1NF avec dispersion (colonne 5). Les valeurs analyses en vert montrent qu'il y a eu amélioration par rapport à la valeur a priori. Celles en noir n'ont été ni améliorées, ni dégradées.

toutes les valeurs analyses sont proches de la valeur vraie. Puis, sur les cycles suivants, ces valeurs oscillent de part et d'autre de la valeur vraie. On observe une convergence plus nette de l'analyse vers la valeur vraie pour les zones 2, 3, 4, 7 et 8. La convergence est plus modérée pour les zones 1, 5, 6 et 9 où l'amplitude des oscillations est plus importante. Ces oscillations de part et d'autre de la valeur vraie confirment le bon fonctionnement de la maquette d'assimilation. En effet, à chaque cycle d'assimilation, la correction apportée par la maquette aux facteurs multiplicatifs est « judicieuse », c'est-à-dire que quand la valeur d'analyse au cycle précédent (équivalente à la valeur d'ébauche au cycle courant) est en-dessus (respectivement en-dessous) de la valeur vraie, la valeur analyse au cycle courant est inférieure (respectivement supérieure) à la valeur d'ébauche. Toutefois, il apparaît que les corrections apportées sont souvent trop fortes, en particulier pour les zones 1, 5, 6 et 9 (celles où on ne constate pas de convergence nette), ce qui explique pourquoi la maquette ne parvient pas à converger pour ces zones. Plusieurs hypothèses peuvent expliquer que la maquette ait plus de mal à converger :

- Premièrement, la dispersion des ensembles d'ébauche n'est pas réduite d'un cycle à l'autre. En effet, on initialise l'AD avec une erreur sur la valeur d'ébauche égale à 0,25 (voir Section 3.2.3). Après le premier cycle d'assimilation, un ensemble analyse avec une nouvelle dispersion (normalement réduite par rapport à l'ébauche) est généré. L'EnKF classique voudrait que cette nouvelle dispersion réduite soit conservée au cycle d'assimilation suivant pour générer le nouvel ensemble des variables de contrôle d'ébauche. Cependant, au fur et à mesure des cycles d'assimilation, cette dispersion peut devenir très faible. Quand cela arrive, la maquette considère l'ébauche plus sûre que les observations et n'apporte plus de correction. Pour éviter cela, on a choisi, pour l'instant, de simplement ignorer la dispersion de l'ensemble d'analyse et de conserver pour tous les cycles la dispersion de 0,25 pour générer les ensembles d'ébauche. La configuration de l'expérience EP1F est très idéalisée et l'analyse au premier cycle est déjà très proche de la vérité. Pour cette expérience, il n'est sûrement pas nécessaire de conserver une grande erreur sur les variables de contrôle sur la totalité des cycles.
- Une deuxième hypothèse est que la présence d'inondations lors des simulations d'ensemble peut fausser le calcul des covariances dans les matrices d'erreurs d'ébauche. En effet, lors de l'étude des sensibilités des profondeurs d'eau aux paramètres TRIP (voir Chapitre 2, Section 2.3.4.1), on a noté que lorsqu'il n'y a pas d'inondations (expérience AMA8), la profondeur d'eau est

une fonction quasiment linéaire du coefficient de Manning. Mais, lorsqu'il y a des inondations (expérience AMA8_He), ce n'est plus le cas (effets d'interactions non-négligeable). Bien que l'AEnKF (comme l'EnKF) soit plus robuste que le KF (et l'EKF) pour gérer les non-linéarités du modèle, il n'en reste pas moins basé sur les mêmes hypothèses de linéarité et de comportement gaussien des erreurs. La présence d'inondations induit une certaine non-linéarité du comportement du modèle de propagation et pourrait perturber l'étape d'analyse.

- Une dernière hypothèse à vérifier est l'impact du traitement fait sur la matrice $\mathbf{P}^b \mathbf{H}^T_{e,k}$ avant l'étape d'analyse.

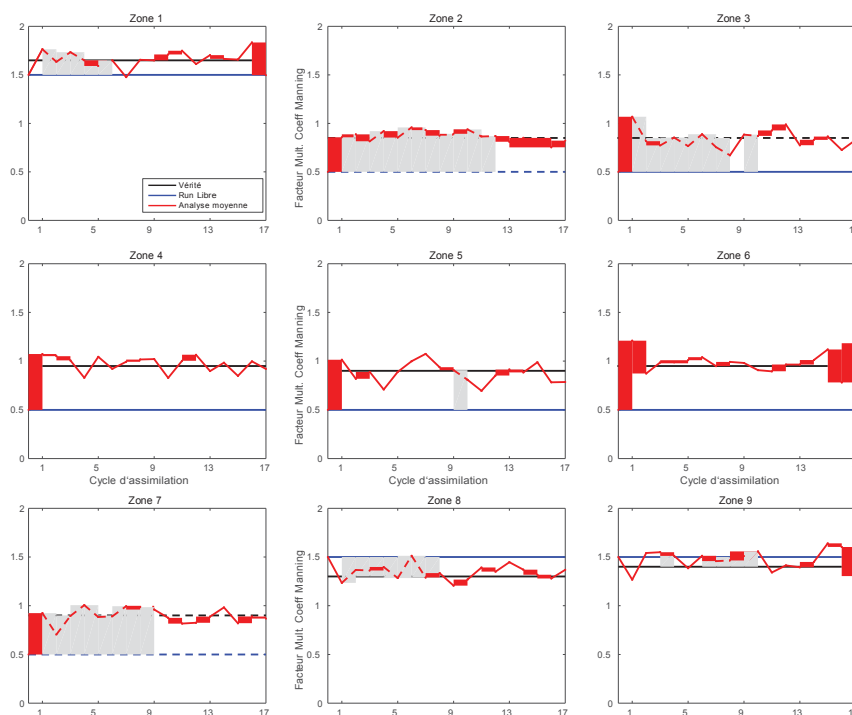


Figure 3.11 – Expérience EP1F : Moyenne de l'ensemble de contrôle analyse au fil des cycles d'assimilation en ligne pleine/tirets rouge, comparée à la valeur vraie en ligne pleine/tirets noire et la valeur a priori au premier cycle en ligne pleine/tirets bleue. Les lignes en tirets matérialisent les cycles où des inondations sont constatées. De même, les zones grisées matérialisent les cycles, pour lesquels les inondations sont constatées dans les runs d'ébauche.

Pour l'expérience EP1F, on ne va détailler que cette dernière hypothèse. Les autres expériences jumelles faites par la suite nous permettront de vérifier les autres hypothèses. Pour cela, on recalcule, pour les trois premiers cycles, les ensembles d'analyse tels qu'ils auraient été calculés si on n'avait pas fait de traitement préalable de la matrice $\mathbf{P}^b \mathbf{H}^T_{e,k}$ (voir Section 3.2.3). Les autres cycles ne sont pas présentés, mais ils ont un comportement similaire aux trois premiers cycles. Les moyennes des ensembles analyse obtenues pour ces trois premiers cycles sont données en Tableau 3.5 où elles sont comparées aux valeurs vraies et aux valeurs d'analyse de l'expérience EP1F. Il apparaît que la maquette fonctionne très bien sans ce traitement. Mis à part pour quelques rares cas où l'analyse initiale (celle de EP1F obtenue avec traitement de $\mathbf{P}^b \mathbf{H}^T_{e,k}$) est équivalente (par exemple, zone 2 cycle 2 ou zone 7 cycle 3), voir meilleure (zone 2 cycle 1), l'analyse obtenue à partir de la matrice

$\mathbf{P}^b \mathbf{H}^T_{e,k}$ complète est bien plus proche du vrai. C'est particulièrement le cas pour l'ensemble des zones « sources » (c'est-à-dire toutes sauf les zones 2 et 3). Malgré tout, il faut garder à l'esprit que les résultats de EP1F sont quand même bons. Ainsi, l'hypothèse faite sur $\mathbf{P}^b \mathbf{H}^T_{e,k}$ ne s'avère pas nécessaire pour la configuration idéalisée de EP1F. On conserve tout de même ce traitement de la matrice de covariance $\mathbf{P}^b \mathbf{H}^T_{e,k}$ pour les autres expériences afin de continuer à explorer son rôle dans la production de l'analyse dans d'autres configurations.

$\mathbf{x}^t$	Cycle 1		Cycle 2		Cycle 3	
	$\overline{\mathbf{x}}^a$	$\overline{\mathbf{x}}^a$	$\overline{\mathbf{x}}^a$	$\overline{\mathbf{x}}^a$	$\overline{\mathbf{x}}^a$	$\overline{\mathbf{x}}^a$
1,65	1,764	1,663	1,634	1,653	1,733	1,630
0,85	0,856	0,835	0,886	0,834	0,816	0,851
0,85	1,069	0,800	0,817	0,840	0,773	0,855
0,95	1,073	0,922	1,051	0,943	1,008	0,967
0,9	1,014	0,852	0,819	0,912	0,889	0,908
0,95	1,210	0,925	0,873	0,948	0,987	0,965
0,9	0,924	0,892	0,705	0,906	0,895	0,911
1,3	1,234	1,283	1,368	1,305	1,361	1,304
1,4	1,268	1,399	1,541	1,464	1,550	1,526

Table 3.5 – Analyse moyenne sur l'ensemble des variables de contrôle sur les trois premiers cycles d'assimilation pour l'expérience EP1F avec traitement de la matrice de covariance des erreurs $\mathbf{P}^b \mathbf{H}^T_{e,k}$ (colonnes 2, 4, 6 ($\overline{\mathbf{x}}^a$)) et sans traitement (colonnes 3, 5, 7 ($\overline{\mathbf{x}}^a$)). Les éléments en verts sont ceux qui sont les plus proches de la vérité pour une zone et un cycle donné. Pour les cycles et les zones sans élément en vert, cela signifie que les deux éléments sont aussi proches l'un que l'autre de la vérité.

La Figure 3.12, quant à elle, détaille pour l'expérience EP1NF l'évolution au fil des cycles des valeurs moyennes d'analyse par rapport à la valeur vraie et à la valeur moyenne d'ébauche au premier cycle, sur le même schéma que la Figure 3.11. Premièrement, conformément à la configuration de l'expérience, on ne constate aucune inondation dans aucune des zones du bassin et ce tout au long de l'année. Malgré cela, le comportement de l'analyse moyenne au fil des cycles est très similaire à celui observé lors de l'expérience EP1F. Ces résultats montrent que, lors de l'assimilation de profondeurs d'eau sans erreurs de bathymétrie, la présence ou non d'inondations n'a pas d'impact sur l'analyse et sur les performances de la maquette d'assimilation. Si on revient aux hypothèses émises lors de l'interprétation des résultats de l'expérience EP1F pour expliquer les oscillations de l'analyse moyenne, les résultats de l'expérience EP1NF permettent de réfuter l'hypothèse faite sur l'impact des inondations.

L'objectif, en corrigeant les coefficients de Manning, est d'améliorer les estimations de profondeurs d'eau et de débit. Le Tableau 3.6 donne les pourcentages d'erreurs par rapport à la simulation vraie pour les facteurs multiplicatifs de la distribution des coefficients de Manning, les profondeurs d'eau et les débits simulés avant et après assimilation. En termes d'erreurs de profondeurs, l'expérience EP1F a permis de passer d'une erreur moyenne (sur l'ensemble du bassin et sur la période complète d'assimilation) de 26 cm sur la profondeur du run libre à une erreur de seulement 4,5 cm après assimilation. L'expérience EP1NF a des statistiques similaires en réduisant l'erreur sur les profondeurs à 4,9 cm. Pour ces expériences, on assimile directement les profondeurs d'eau et on n'a pas d'erreur sur la bathymétrie. Par conséquent, la correction de la distribution des coefficients de Manning

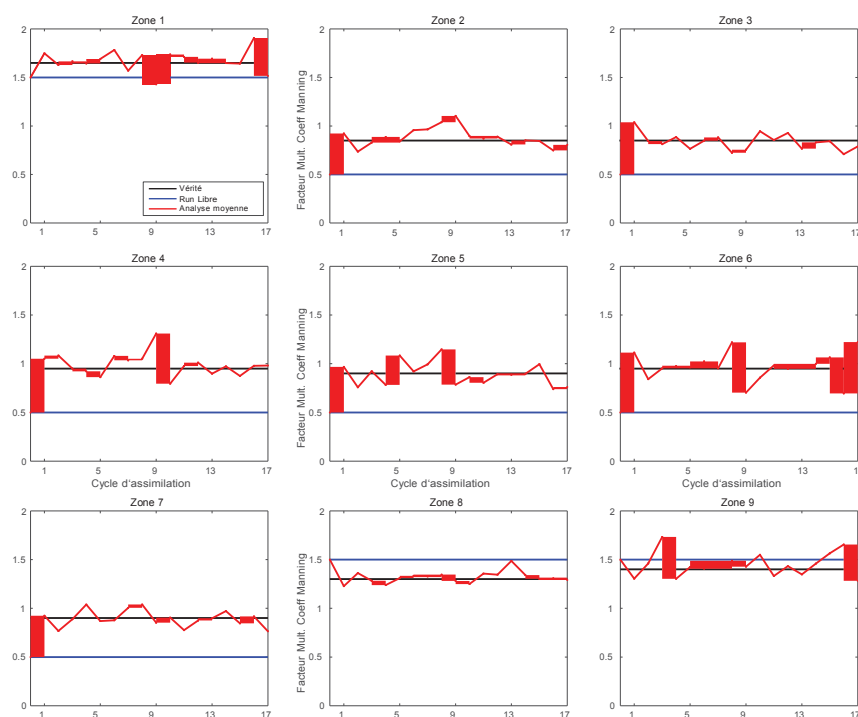


Figure 3.12 – Expérience EP1NF : Moyenne de l’ensemble de contrôle analyse au fil des cycles d’assimilation en ligne pleine/tirets rouge, comparée à la valeur vraie en ligne pleine/tirets noire et la valeur a priori au premier cycle en ligne pleine/tirets bleue.

permet automatiquement de retrouver les profondeurs d’eau et les débits vrais : les Figures 3.13-3.14 comparent, au fil des cycles d’assimilations, les profondeurs d’eau vraies, observées et simulées sur une sélection de 9 cellules TRIP du bassin de l’Amazonie. Chaque cellule est située dans une zone hydrogéomorphologique distincte (leurs noms correspondent à des noms de stations in situ proches et servent à les distinguer).

Ces figures permettent aussi de comprendre concrètement comment la maquette fonctionne. Au cours d’un cycle d’assimilation donné, les runs d’ensemble d’ébauche servent à estimer l’innovation, c’est-à-dire l’écart des observations modèle aux vraies observations du système. Le signe de l’innovation détermine le sens de correction des variables de contrôle, ici les facteurs multiplicatifs de la distribution des coefficients de Manning. Si l’innovation est positive, cela signifie que les profondeurs d’eau modèle sont inférieures aux profondeurs observées. La maquette doit ajuster les variables de contrôle pour faire augmenter les profondeurs d’eau modèle. Dans cette configuration, il apparaît que la maquette d’assimilation augmente le facteur multiplicatif donc augmente les coefficients de Manning. En effet, en augmentant le coefficient de Manning, la vitesse d’écoulement simulée par le modèle va diminuer. En conséquence, les masses d’eau vont s’accumuler dans le réservoir de surface et vont permettre à la profondeur d’eau d’augmenter pour rattraper les profondeurs observées. Dans le cas où l’innovation est négative, la maquette va alors réduire le coefficient de Manning pour accélérer l’écoulement et baisser la profondeur d’eau.

Ces premières expériences ont permis de montrer la possibilité d’assimiler les mesures de profon-

	Err. relative par rapport à la vérité		
	Sans assimilation	Après assimilation EP1F	EP1NF
variables de contrôle	33,07 %	12,19 %	12,04 %
profondeurs d'eau	22,39 %	4,08 %	4,41 %
	25,75 cm	4,5 cm	4,9 cm
débits	4,53 %	0,83 %	0,84 %

Table 3.6 – EP1F et EP1NF : Erreur relative (moyenne dans le temps) des facteurs multiplicatifs (ligne 2), des anomalies (ligne 3), des hauteurs d'eau (ligne 4) et des débits (ligne 5) par rapport à la simulation vraie. Pour les profondeurs d'eau, on indique l'équivalent de cette erreur en cm. Les valeurs analysées en vert montrent qu'il y a eu amélioration par rapport à la valeur a priori.

deurs d'eau pour corriger la distribution des coefficients de Manning. De plus, dans la configuration idéalisée de ces expériences où il n'y pas d'incertitudes sur le référentiel des profondeurs simulées et observées, le fait de retrouver la bonne distribution des coefficients de Manning permet d'automatiquement améliorer les estimations de profondeurs d'eau et de débits. À noter malgré tout que l'on s'est aussi rendu compte que, pour ces expériences, le traitement de la matrice de covariance des erreurs $\mathbf{P}^b \mathbf{H}^T_{e,k}$ n'était pas nécessaire. L'expérience qui suit va nous permettre de tester les limites de la maquette en y insérant des erreurs sur la distribution de la profondeur critique dans le modèle TRIP.

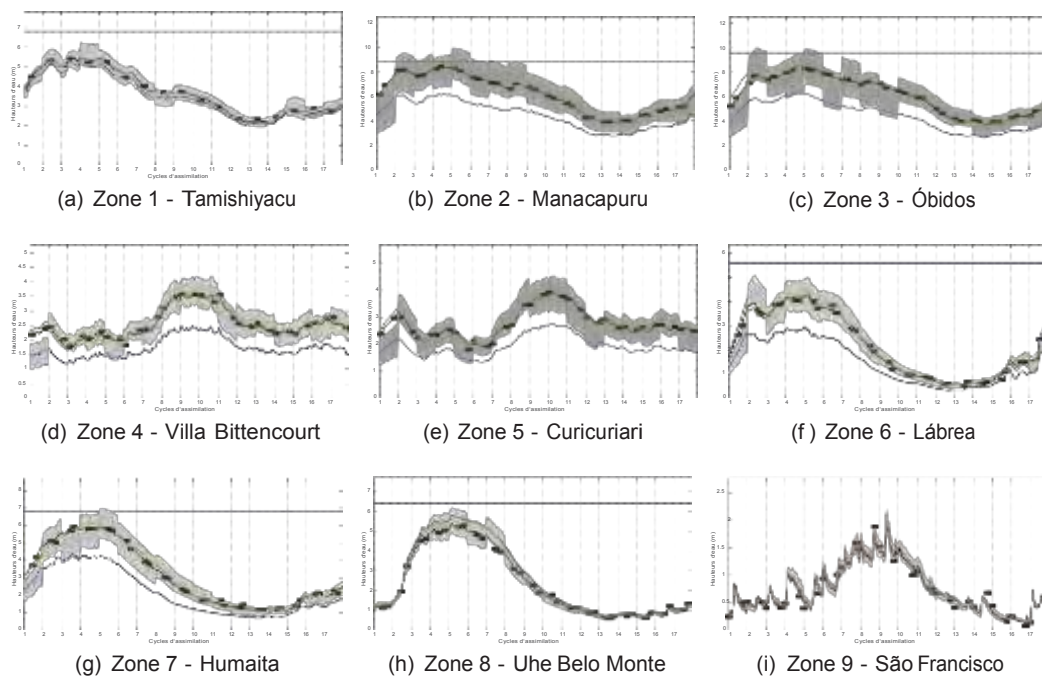


Figure 3.13 – Expérience EP1F : Évolution dans le temps des profondeurs d'eau vraies (ligne noire), des profondeurs d'eau du run libre (ligne bleue), des profondeurs d'eau observées avec leur erreur (marqueurs carrés verts), de l'ensemble des profondeurs d'eau d'ébauche (zone grisée), des profondeurs d'eau analyses moyennes (ligne orange) avec la dispersion de l'ensemble analyse (hachures rouges).

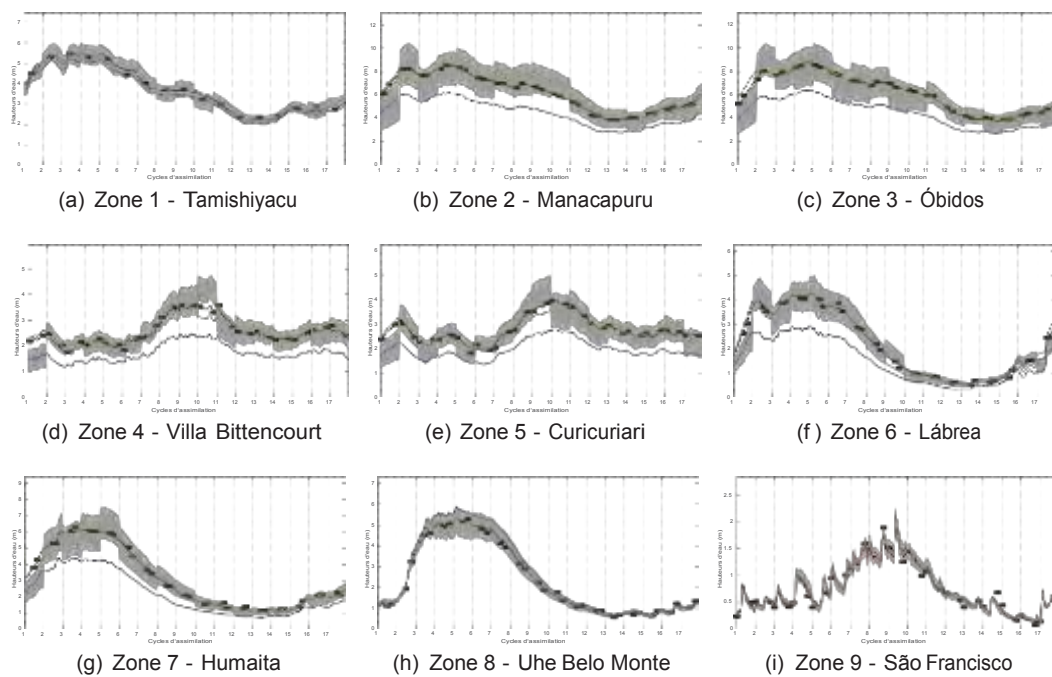


Figure 3.14 – Expérience EP1NF : Évolution dans le temps des profondeurs d’eau vraies (ligne noire), des profondeurs d’eau du run libre (ligne bleue), des profondeurs d’eau observées avec leur erreur (marqueurs carrés verts), de l’ensemble des profondeurs d’eau d’ébauche (zone grisée), des profondeurs d’eau analyses moyennes (ligne orange) avec la dispersion de l’ensemble analyse (hachures rouges).

3.3.2 Résultats de l'expérience EP2F

Pour l'expérience EP2F, on reprend la même configuration que pour l'expérience EP1F mais on utilise des profondeurs critiques différentes entre l'ensemble d'ébauche et la simulation utilisée pour générer le run vrai et les observations. Pour rappel, la distribution de profondeurs critiques des simulations d'ébauche H_c^b est obtenue en réduisant de 25% la distribution profondeur critique vraie H_c^t (voir Section 3.2.3). Lors de la présentation du modèle ISBA-TRIP (voir Chapitre 2), on a insisté sur l'importance de ce paramètre qui permet l'activation du schéma d'inondation dans TRIP, si la profondeur d'eau dans le réservoir de surface excède cette valeur critique. L'objectif de cette expérience est d'étudier le comportement de la maquette face à ces erreurs et comment elle adapte la « correction » de la distribution des coefficients de Manning en réponse à une erreur sur les hauteurs critiques, alors qu'aucune erreur sur ce paramètre est considéré dans la génération de l'ensemble. L'estimation du Manning va donc être affectée, puisque la maquette tentera (en période d'inondation) de corriger le frottement pour corriger ces différences de hauteurs induites par une hauteur critique différente.

Pour analyser les résultats de l'expérience EP2F, la Figure 3.15 donne d'abord les valeurs moyennes de l'ensemble des variables de contrôle analyse au dernier cycle d'assimilation (cercles rouges) en comparaison aux valeurs moyennes d'ébauche au premier cycle d'assimilation (carrés bleus) et aux valeurs vraies (losanges noirs). À première vue, la valeur moyenne analyse est rapprochée de la valeur vraie par rapport à l'ébauche pour toutes les zones, exceptées pour les zones 1 et 9. Les valeurs exactes sont détaillées dans le Tableau 3.7 et confirment ces observations.

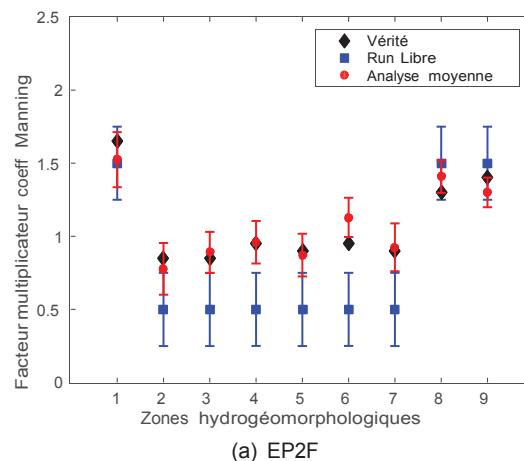


Figure 3.15 – Expériences EP2F (a) : Comparaison des valeurs de facteurs multiplicatifs de la distribution de coefficients de Manning par zones : valeur vraie (losanges noirs), valeur a priori au 1er cycle d'assimilation avec erreur (carrés bleus) et valeur d'analyse moyenne au dernier cycle d'assimilation avec dispersion de l'ensemble (cercles rouges).

Les résultats sont, à première vue, similaires à ceux obtenus avec l'expérience EP1F. La Figure 3.16 détaille alors l'évolution, au fil des cycles d'assimilation, de la valeur analyse moyenne des variables de contrôle par rapport à la valeur vraie et à la valeur d'ébauche. Comme pour les Figures 3.11-3.12, les différents graphiques affichent la présence ou non d'inondations sur le réseau hydrologique principal pour le run vrai, le run libre et les runs d'ensemble d'ébauche et d'analyse. On observe un

Zones	Vrai	A priori (Sans assim)	Analyse (Après assim) EP2F
1	1,65	1,5 (0,25)	1,524 (0,188)
2	0,85	0,5 (0,25)	0,778 (0,177)
3	0,85	0,5 (0,25)	0,890 (0,140)
4	0,95	0,5 (0,25)	0,960 (0,145)
5	0,9	0,5 (0,25)	0,872 (0,146)
6	0,95	0,5 (0,25)	1,129 (0,134)
7	0,9	0,5 (0,25)	0,923 (0,164)
8	1,3	1,5 (0,25)	1,409 (0,111)
9	1,4	1,5 (0,25)	1,300 (0,101)

Table 3.7 – Comparaison, pour chaque zone, des valeurs vraies (colonne 2), des valeurs a priori (colonne 3) avec dispersion (entre parenthèses) avec les valeurs analyses moyennes finales de l'expérience EP2F avec dispersion (colonne 4). Les valeurs analyses en vert montrent qu'il y a eu amélioration par rapport à la valeur a priori. Celles en noir n'ont été ni améliorées, ni dégradées.

comportement bien différent de celui de l'expérience EP1F avec de très fortes oscillations de l'analyse moyenne tout au long de l'expérience et pour toutes les zones hydrogéomorphologiques. Certaines zones semblent malgré tout converger vers la valeur vraie sur les derniers cycles, c'est-à-dire les zones 2, 3, 4, 5 et 7. Les autres zones (les zones 1, 6, 8 et 9) présentent des fortes oscillations même sur les derniers cycles. La majorité des plus grosses variations de l'analyse moyenne se trouvent sur des cycles où à la fois le run vrai, le run libre et les runs d'ensemble d'ébauche sont inondés, par exemple les zones 2 et 3 ainsi que les premiers cycles des zones 1, 7, 8 et 9.

Ces résultats chaotiques s'expliquent facilement par la présence d'un écart entre le référentiel des profondeurs simulées et celui des profondeurs observées. Plus particulièrement, cet écart fait que les simulations d'ensemble inondent plus vite que la simulation vraie (utilisée pour générer les observations). Physiquement, lorsque l'eau est confinée au réservoir de surface (quand il n'y a pas d'inondations), l'étude de sensibilité a montré que la profondeur d'eau augmente quasi-linéairement avec le coefficient de Manning. Une fois que la profondeur a atteint la valeur critique et que le modèle commence à inonder, la masse d'eau excédentaire se déverse dans les plaines d'inondations et le niveau d'eau augmente plus lentement dans le réservoir de surface que lorsque celui-ci n'est pas inondé. De ce fait, des modifications plus importantes du coefficient de Manning sont nécessaires pour compenser ces différences de variation de profondeurs d'eau entre l'ébauche et les observations. Dans le cas de l'expérience EP2F, les observations et le modèle n'inondent pas à partir de la même profondeur critique. Les modifications faites par la maquette sont adaptées à la dynamique des simulations d'ébauche et non à celle des observations, ce qui entraîne des corrections souvent inadaptées. D'autre part, l'erreur ajoutée par l'écart entre les profondeurs critiques n'est pas prise en compte lors de la génération de l'ensemble d'ébauche $\mathbf{X}_e^b$. En effet, comme pour EP1F et EP1NF, l'ensemble des variables de contrôle d'ébauche $\mathbf{X}_e^b$ est généré en ne perturbant que les facteurs multiplicatifs de la distribution des coefficients de Manning. Par conséquent, du point de vue de la maquette, la bathymétrie du modèle est considérée comme parfaite. Néanmoins, il nous faut vérifier l'impact de ces coefficients de Manning analyses sur les profondeurs d'eau simulées pour pouvoir vraiment juger des performances de l'expérience. En effet, les valeurs excessives des variables de contrôle pourrait permettre de malgré tout retrouver les profondeurs d'eau vraies. Figure 3.17 illustre alors le comportement des profondeurs d'eau simulées analyse au fil des cycles et des corrections de l'AEnKF.

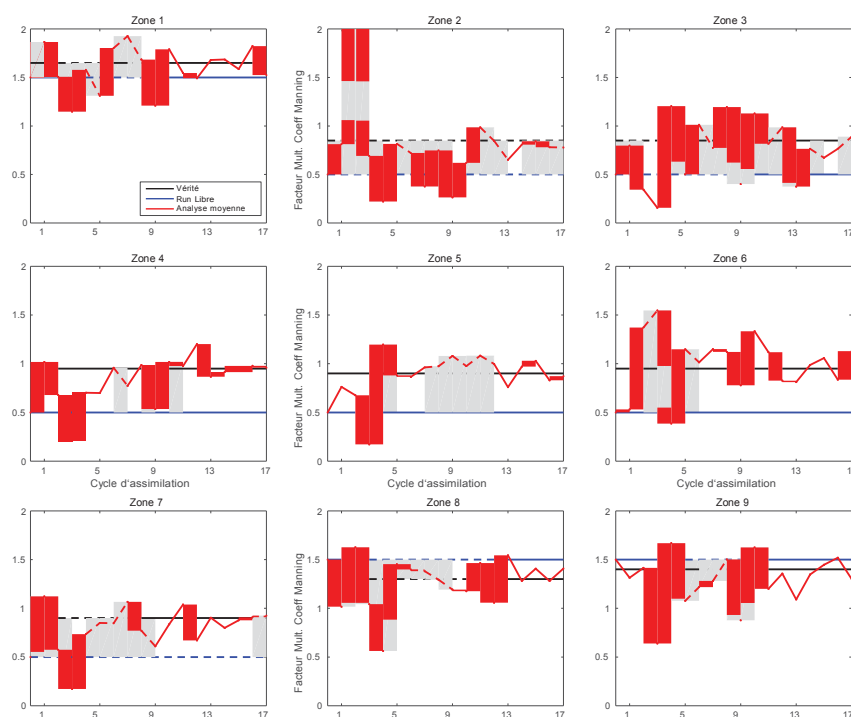


Figure 3.16 – Expérience EP2F : Moyenne de l'ensemble de contrôle analyse au fil des cycles d'assimilation en ligne pleine/tirets rouge, comparée à la valeur vraie en ligne pleine/tirets noire et la valeur a priori au premier cycle en ligne pleine/tirets bleue. Les lignes en tirets matérialisent les cycles où des inondations sont constatées. De même, les zones grisées matérialisent les cycles, pour lesquels les inondations sont constatées dans les runs d'ébauche.

Sur les différents graphiques, on représente en ligne horizontale bleue la profondeur critique du run libre et des runs d'ensemble, distincte de la profondeur critique du run vrai et des observations, représentée par une ligne horizontale noire. On rappelle que cette profondeur critique contrôle l'activation ou non du schéma d'inondations TRIP et donc régule la présence ou non d'inondations dans les différentes simulations. On identifie facilement les cycles lors desquels l'AEnKF va avoir des difficultés à construire une analyse correcte, c'est-à-dire ceux où les observations (symboles verts) et/ou les simulations d'ébauche (zone grise) inondent. En revanche, sur les derniers cycles d'assimilation (c'est-à-dire à partir des cycles 12/13), toutes les zones ne sont plus du tout inondées à la fois dans le run vrai/les observations et le run libre/simulations d'ensemble. À partir de là, l'erreur sur la profondeur critique n'a plus aucun impact et on se retrouve dans la même configuration que les expériences EP1F et EP1NF. C'est d'ailleurs à partir de ces cycles qu'on constate une amélioration et une convergence des analyses (Figure 3.16).

Par ailleurs, il faut tenir compte du fait que les profondeurs d'eau représentées en Figure 3.17 ne sont les profondeurs qu'à une seule cellule parmi toutes les cellules de la zone alors que le facteur multiplicatif du coefficient de Manning dépend des cellules de toute la zone (et aussi des corrélations avec les autres zones). Par conséquent, certaines profondeurs d'eau analyse, comme celles des zones 4 et 5 au cycle 3 (voir Figure 3.17d-e), peuvent se retrouver très éloignées des observations alors qu'il n'y a pas d'inondations dans la cellule et sont peut-être liées à des inondations ayant lieu ailleurs dans

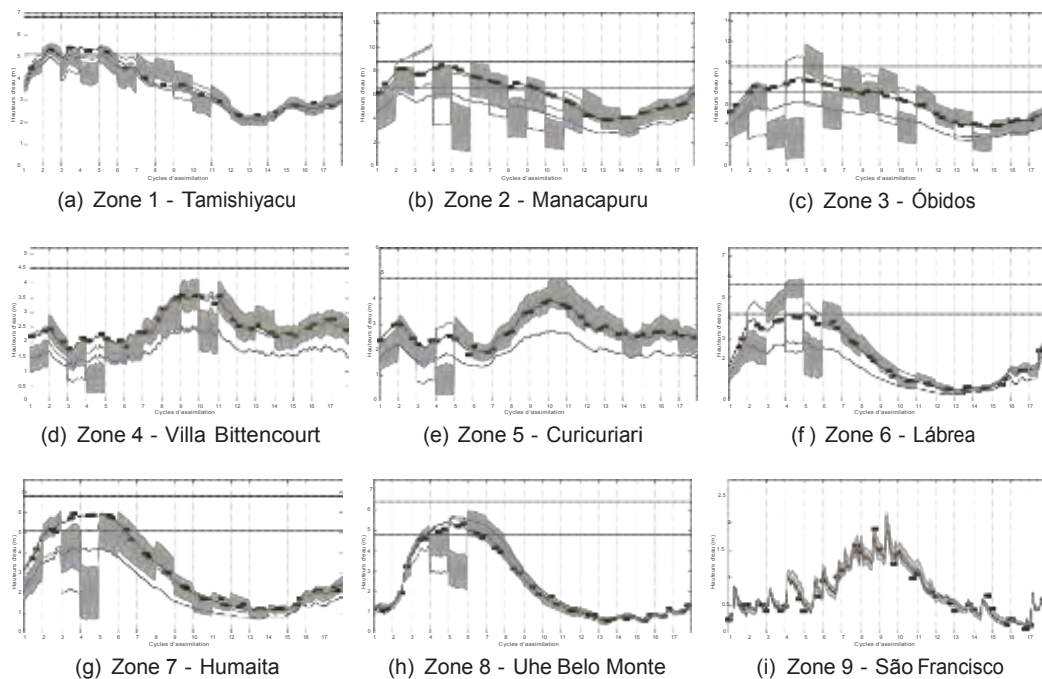


Figure 3.17 – Expérience EP2F : Évolution dans le temps des profondeurs d’eau vraies (ligne noire), des profondeurs d’eau du run libre (ligne bleue), des profondeurs d’eau observées avec leur erreur (marqueurs carrés verts), de l’ensemble des profondeurs d’eau d’ébauche (zone grisée), des profondeurs d’eau analyses moyennes (ligne orange) avec la dispersion de l’ensemble analyse (hachures rouges).

la zone.

Enfin, les autres mauvais résultats peuvent être expliqués par l’écart entre les référentiels de profondeurs d’eau. Prenons l’exemple de la zone 2 au cycle d’assimilation 3 (voir Figure 3.17b). L’ensemble des profondeurs d’ébauche (zone grisée) suit directement les profondeurs d’analyse très hautes du cycle précédent et tous les membres de l’ensemble sont inondés. Par contre, les observations sont toutes au-dessous des profondeurs d’ébauche et de la hauteur critique vraie (et ne sont donc pas du tout inondées). La maquette, qui cherche à se rapprocher des observations, va alors nécessairement fortement diminuer le coefficient de Manning. On a vu qu’en période d’inondation, comme les hauteurs critiques entre l’ébauche et les observations sont différentes, l’algorithme d’assimilation ne peut que jouer sur les coefficients multiplicatifs du coefficient de Manning pour rapprocher les hauteurs analyses des observations, ce qui va donc produire des variations plus importantes du coefficient de Manning en période d’inondation (et donc un écart plus important par rapport à la distribution vraie du coefficient de Manning). Puisque la maquette « voit » que les membres de l’ensemble d’ébauche sont tous inondés, il produit une importante correction sur les facteurs multiplicatifs du Manning. La dynamique des observations et du modèle étant différentes (les observations ne sont pas inondées), la correction produite est beaucoup trop forte.

Finalement, l’expérience parvient malgré tout à un peu diminuer l’erreur sur les profondeurs et les débits, comme le montre le Tableau 3.8 qui présente les erreurs pour le run libre et l’analyse.

	Err. relative par rapport à la vérité	
	Sans assimilation	Après assimilation EP2F
variables de contrôle	33,07 %	21,84 %
profondeurs d'eau	22,37 %	11,25%
	25,76 cm	15,82 cm
débits	4,62 %	2,27 %

Table 3.8 – EP2F : Erreur relative (moyenne dans le temps) des facteurs multiplicatifs (ligne 2), des anomalies (ligne 3), des hauteurs d'eau (ligne 4) et des débits (ligne 5) par rapport à la simulation vraie. Pour les profondeurs d'eau, on indique l'équivalent de cette erreur en cm. Les valeurs analysées en vert montrent qu'il y a eu amélioration par rapport à la valeur a priori.

L'expérience EP2F a permis de mettre en évidence certaines limitations de l'assimilation des profondeurs d'eau directes, en particulier quand on ne considère pas toutes les erreurs présentes dans le système lors de la génération de l'ensemble des variables de contrôle d'ébauche. Parallèlement, on n'a pas mené d'expérience « EP2NF », qui correspondrait à une configuration similaire à l'expérience EP2F mais à partir du run vrai et d'ébauche sans inondations mais avec des hauteurs critiques différentes. Mais cela aurait donné les mêmes résultats que EP1NF, puisque l'expérience EP2NF n'aurait pas eu d'inondations au cours de l'année.

Outre les erreurs sur la hauteur critique, on peut aussi anticiper des erreurs entre les cotes d'eau TRIP et celles de SWOT, dues à un biais de la cote du lit du fleuve TRIP. Dans ce cas, l'assimilation d'anomalies de cote d'eau devrait permettre de supprimer cette source d'erreur. C'est pourquoi dans la section suivante, on s'intéresse à l'assimilation non plus de profondeurs d'eau, mais d'anomalies de cote d'eau (ou anomalies de profondeurs/hauteurs d'eau, ce qui est similaire).

3.4 Résultats pour l'assimilation des anomalies de hauteurs d'eau

3.4.1 Résultats des expériences EP3-1 et EP3-2

Les expériences EP3-1 et EP3-2 permettent de tester le comportement de la maquette d'AD lors de l'assimilation d'anomalies de profondeurs d'eau. On commence par la configuration où il n'y a pas d'erreurs sur la bathymétrie entre la simulation vraie et les simulations d'ensemble. Les deux expériences ont la même configuration, à ceci près qu'elles débutent à une date différente : l'expérience EP3-1 débute au 4 Mars 2008 (saison de hautes eaux) et l'expérience EP3-2 débute au 9 Septembre 2008 (saison de basses eaux). Les deux expériences sont menées sur un nombre réduit de 3 fenêtres d'assimilation de 21 jours. Les dates de début et la durée des simulations ont été choisies à partir des résultats de l'Analyse de Sensibilité sur les anomalies de profondeurs d'eau (voir Chapitre 2, Section 2.3.4.2), afin que les expériences soient effectuées sur des périodes pendant lesquelles les anomalies sont sensibles aux paramètres et en particulier aux coefficients de Manning.

La Figure 3.18 détaillent, pour les deux expériences, les valeurs moyennes sur l'ensemble d'analyse au dernier cycle d'assimilation des variables de contrôle en fonction des zones hydrogéomorpho-

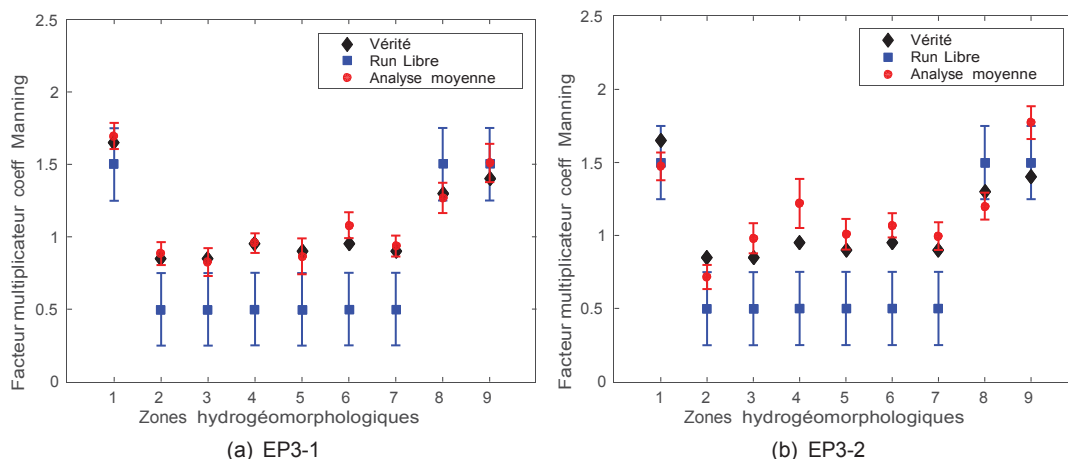


Figure 3.18 – Expériences EP3-1 (a) et EP3-2 (b) : Comparaison des valeurs de facteurs multiplicatifs de la distribution de coefficients de Manning par zones : valeur vraie (losanges noirs), valeur a priori au 1er cycle d’assimilation avec erreur (carrés bleus) et valeur d’analyse moyenne au dernier cycle d’assimilation avec dispersion de l’ensemble (cercles rouges).

logiques. Sur ces figures sont aussi indiquées la valeur vraie et la valeur a priori au première cycle d’assimilation. Ensuite, le Tableau 3.9 donne explicitement ces valeurs avec la dispersion de l’ensemble associée. Sur la Figure 3.18, on constate, pour les deux expériences, que la valeur finale analyse des variables de contrôle s’est rapprochée de la valeur vraie et que la dispersion de l’ensemble d’analyse est réduite par rapport à celle de l’ébauche. Ces observations sont alors confirmées par le Tableau 3.9. Par conséquent, la maquette d’assimilation des anomalies fonctionnent correctement et est capable de retrouver les valeurs vraies des variables de contrôle à partir des anomalies de profondeurs d’eau. Parallèlement, le Tableau 3.10 compare l’erreur relative moyenne des variables d’ébauche (sans assimilation) et d’analyse (après assimilation) pour les variables de contrôle (les facteurs multiplicatifs), les variables observées (les anomalies), ainsi que les profondeurs d’eau et les débits. Ce tableau confirme le bon fonctionnement de la maquette, car les erreurs relatives d’analyse sont toujours inférieures aux erreurs relatives d’ébauche. De surcroît, on remarque que les statistiques de l’expérience EP3-1 sont similaires à celles des expériences EP1F/EP1NF et meilleures que les statistiques de l’expérience EP3-2.

Pour expliquer ces différences, les Figures 3.19-3.20 illustrent, pour chaque zone, l’évolution au fil des cycles d’assimilation des variables de contrôle analyse (couleur rouge) par rapport à la valeur vraie (couleur noire) et la valeur d’ébauche au premier cycle d’assimilation (couleur bleue). Si les valeurs analyses de l’expérience EP3-1 (voir Fig. 3.19) convergent bien vers la valeur vraie (excepté pour la zone 9), cette convergence est bien moins nette pour l’expérience EP3-2 (voir Fig. 3.20) dans laquelle les variables de contrôle analyse ne font que se rapprocher des valeurs vraies. De plus, les différents graphiques sur les Figures 3.19 et 3.20 matérialisent la présence ou non d’inondations au cours des cycles d’assimilation comme pour les Figures 3.11, 3.12 et 3.16. On peut noter la présence d’inondations sur 5 zones de l’expérience EP3-1 (zones du cours principal et affluents droits) et aucune pour l’expérience EP3-2. Cette différence est évidemment liée aux périodes différentes sur lesquelles sont faites les expériences : l’expérience EP3-1 est menée sur les mois de Mars-Avril qui correspondent à la saison des hautes eaux et l’expérience EP3-2 est menée sur les mois de Septembre-Octobre qui

Zones	Vrai	A priori (Sans assim)	Analyse (Après assim)	
			EP3-1	EP3-2
1	1,65	1,5 (0,25)	1,696 (0,091)	1,474 (0,095)
2	0,85	0,5 (0,25)	0,884 (0,079)	0,716 (0,082)
3	0,85	0,5 (0,25)	0,826 (0,096)	0,982 (0,103)
4	0,95	0,5 (0,25)	0,956 (0,068)	1,219 (0,168)
5	0,9	0,5 (0,25)	0,865 (0,124)	1,001 (0,105)
6	0,95	0,5 (0,25)	1,080 (0,089)	1,069 (0,083)
7	0,9	0,5 (0,25)	0,936 (0,072)	0,995 (0,095)
8	1,3	1,5 (0,25)	1,268 (0,104)	1,202 (0,092)
9	1,4	1,5 (0,25)	1,509 (0,132)	1,773 (0,112)

Table 3.9 – Comparaison, pour chaque zone, des valeurs vraies (colonne 2), des valeurs a priori (colonne 3) avec dispersion (entre parenthèses) avec les valeurs analyse moyenne finale de l'expérience EP3-1 avec dispersion (colonne 4) et les valeurs analyse moyenne finale de l'expérience EP3-2 avec dispersion (colonne 5). Les valeurs analyses en vert montrent qu'il y a eu amélioration par rapport à la valeur a priori. Celles en noir n'ont été ni améliorées, ni dégradées et les valeurs analyses en rouge montrent qu'il y a eu un dégradation par rapport à la valeur a priori.

	Err. relative par rapport à la vérité		
	Sans assimilation	Après assimilation EP3-1	EP3-2
variables de contrôle	33,07 %	11,27 %	18,04 %
anomalies	EP3-1 : 48,61 %	6,21 %	34,05 %
	15,66 cm		
	EP3-2 : 59,34 %	3,53 cm	5,28 cm
hauteurs d'eau	11,99 cm		
	EP3-1 : 22,41 %	4,53 %	8,45 %
	36,28 cm		
débits	EP3-2 : 21,63 %	7,31 cm	7,24cm
	15,06 cm		
	EP3-1 : 2,95 %	0,61 %	2,39 %
	EP3-2 : 6,45 %		

Table 3.10 – EP3-1 et EP3-2 : Erreur relative (moyenne dans le temps) des facteurs multiplicatifs (ligne 2), des anomalies (ligne 3), des hauteurs d'eau (ligne 4) et des débits (ligne 5) par rapport à la simulation vraie. Les valeurs analyses en vert montrent qu'il y a eu amélioration par rapport à la valeur a priori. Celles en noir n'ont été ni améliorées, ni dégradées et les valeurs analyses en rouge montrent qu'il y a eu un dégradation par rapport à la valeur a priori.

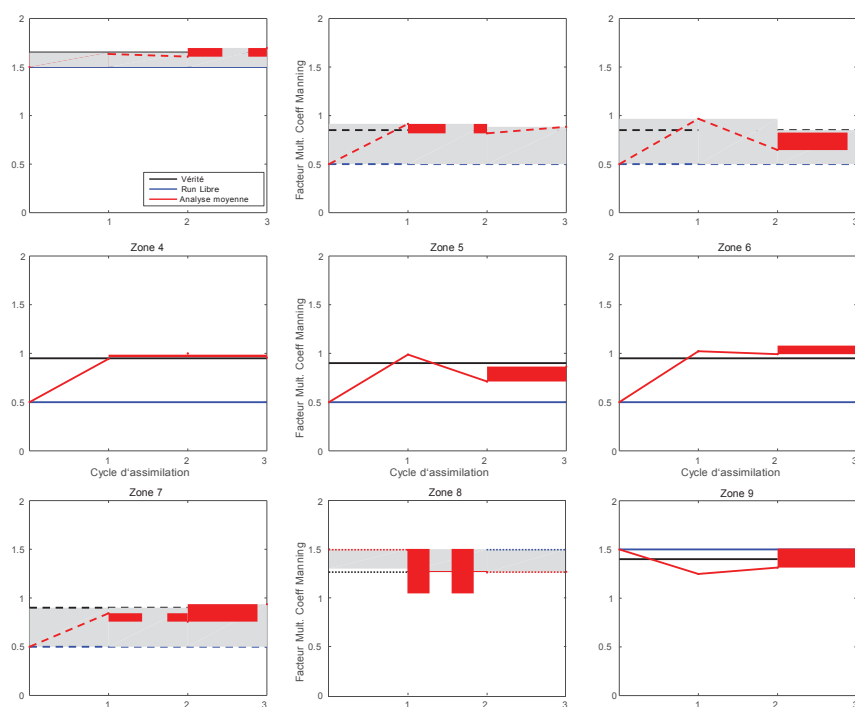


Figure 3.19 - Expérience EP3-1 : Moyenne de l'ensemble de contrôle analyse au fil des cycles d'assimilation en ligne pleine/tirets rouge, comparée à la valeur vraie en ligne pleine/tirets noire et la valeur a priori au premier cycle en ligne pleine/tirets bleue. Les lignes en tirets matérialisent les cycle où des inondations sont constatées. De même, les zones grisées matérialisent les cycles, pour lesquels les inondations sont constatées dans les runs d'ébauche.

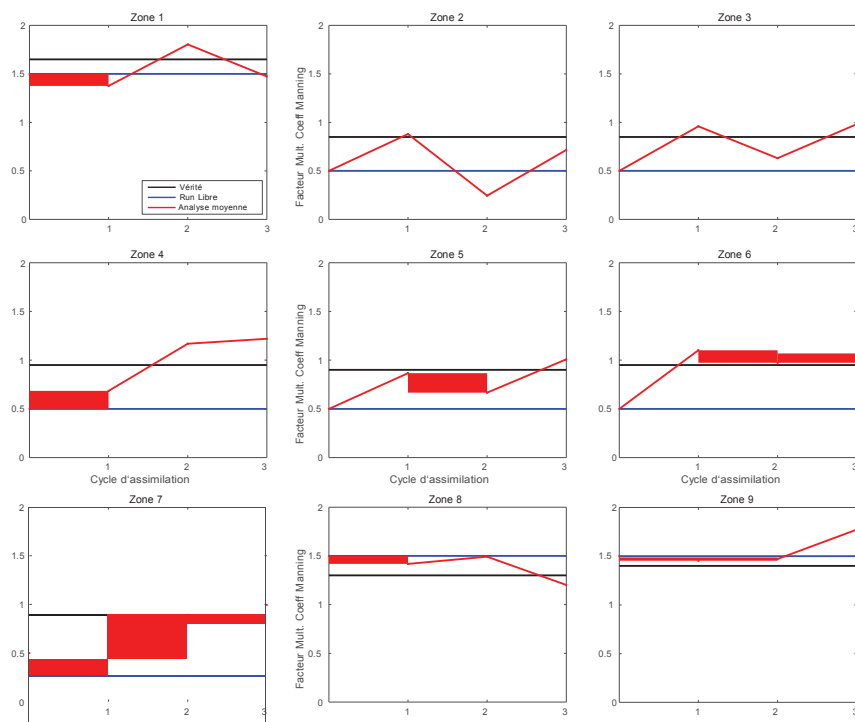


Figure 3.20 - Expérience EP3-2 : Moyenne de l'ensemble de contrôle analyse au fil des cycles d'assimilation en ligne pleine/tirets rouge, comparée à la valeur vraie en ligne pleine/tirets noire et la valeur a priori au premier cycle en ligne pleine/tirets bleu.

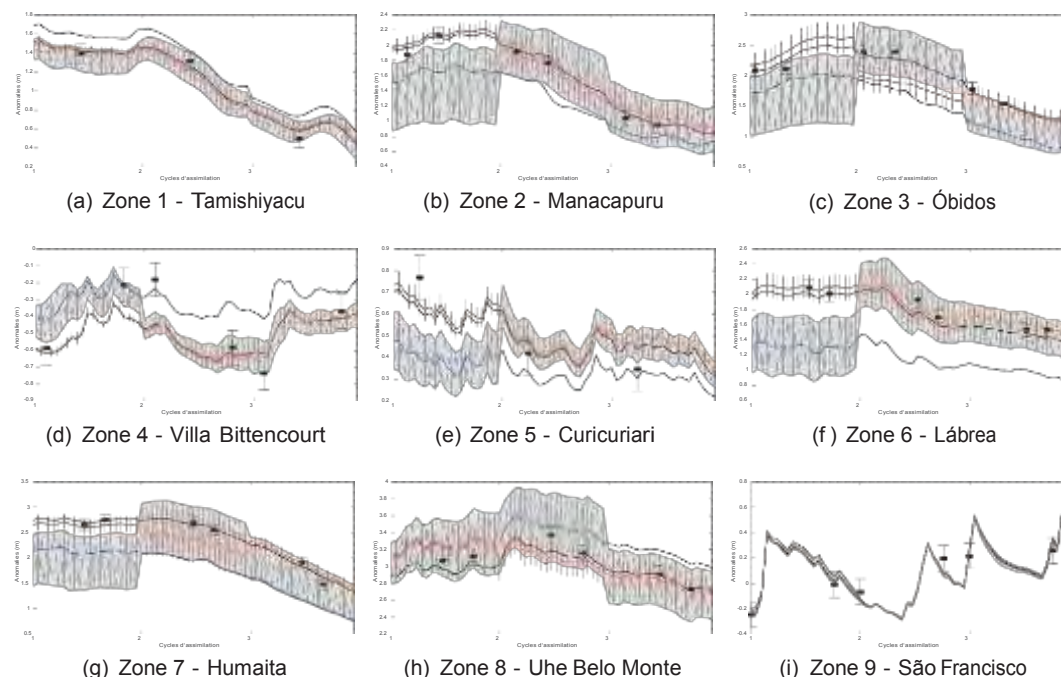


Figure 3.21 – Expérience EP3-1 : Évolution dans le temps des anomalies vraies (ligne noire), des anomalies du run libre (ligne bleue), des anomalies observées avec leur erreur (marqueurs carrés verts), de l'ensemble des anomalies d'ébauche (zone grisée), de la moyenne des anomalies analyse (ligne orange) avec la dispersion de l'ensemble analyse (hachures rouges).

correspondent à la saison des basses eaux. Il semble alors que la maquette soit plus efficace en saison de hautes eaux. Pour confirmer et expliquer cette hypothèse, les Figures 3.21-3.22 illustrent, pour chaque zone, le comportement des anomalies observées et simulées à une station in situ test de la zone.

Pour l'expérience EP3-1 (voir Fig. 3.21), à partir du cycle 3, la moyenne des anomalies après analyse de toutes les zones sont proches des anomalies vraies et présentent une dispersion d'ensemble réduite par rapport à celle des ensembles d'ébauche. Il est important de noter que la dispersion de l'ensemble d'ébauche, elle, est particulièrement importante sur toutes les zones (excepté la zone 9). Dans l'EnKF (et par extension dans l'AEnKF), la dispersion de l'ensemble d'ébauche modélise l'erreur des estimations modèle et doit être comparée à la dispersion de l'ensemble des observations. Quand la dispersion de l'ébauche est plus grande que la dispersion des observations, l'algorithme d'assimilation considère que les observations sont plus sûres que l'ébauche et va alors essayer de s'en rapprocher. Le fait que l'expérience soit réalisée pendant la saison de hautes eaux permet alors de générer cette dispersion importante. En effet, les fortes précipitations et les forts débits amont vont générer une dispersion plus grande de l'ensemble, pour une dispersion similaire des coefficients de Manning qu'aux basses eaux. Le cas particulier de la zone 9 peut s'expliquer par le fait que la dispersion de l'ensemble d'ébauche est petite par rapport à la dispersion des observations et que la valeur d'ébauche du coefficient multiplicatif du Manning soit déjà très proche de la vérité. De ce fait, la maquette ne parvient pas à correctement corriger le facteur multiplicatif dans cette zone. En ce

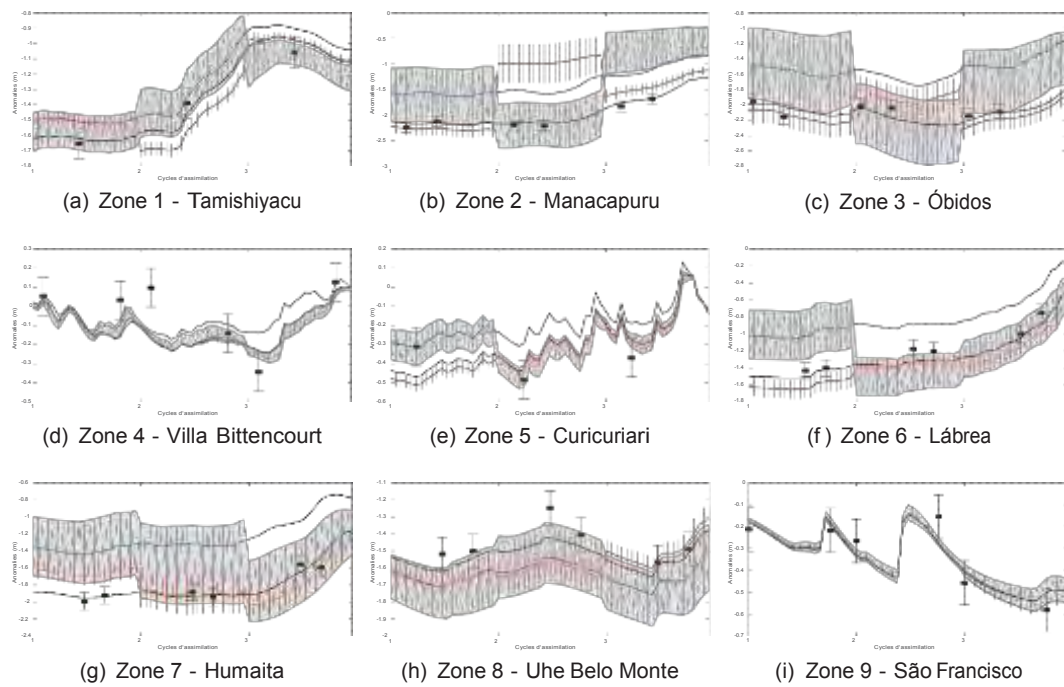


Figure 3.22 – Expérience EP3-2 : Évolution dans le temps des anomalies vraies (ligne noire), des anomalies du run libre (ligne bleue), des anomalies observées avec leur erreur (marqueurs carrés verts), de l'ensemble des anomalies d'ébauche (zone grisée), de la moyenne des anomalies analyse (ligne orange) avec la dispersion de l'ensemble analyse (hachures rouges).

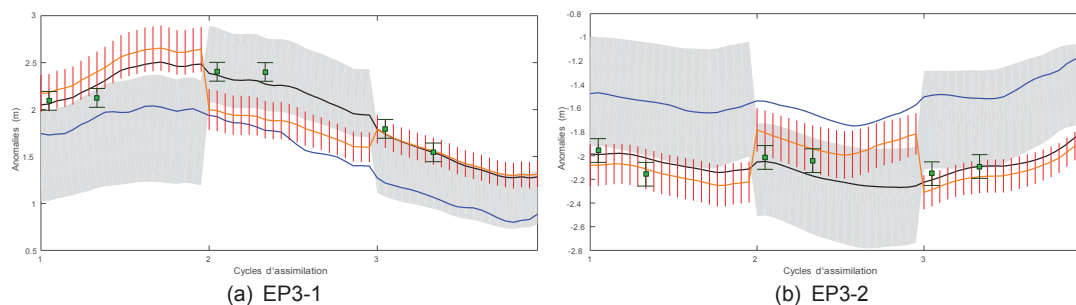


Figure 3.23 – Expériences EP3-1 et EP3-2 à la station d’Óbidos : Évolution dans le temps des profondeurs d’eau vraies (ligne noire), des profondeurs d’eau du run libre (ligne bleue), des profondeurs d’eau équivalentes aux observations avec leur erreur (marqueurs carrés verts), de l’ensemble des profondeurs d’eau d’ébauche (zone grisée), des profondeurs d’eau analyses moyennes (ligne orange) avec la dispersion de l’ensemble analyse (hachures rouges).

qui concerne l’expérience EP3-2 (voir Fig. 3.22), on remarque que les anomalies analyses moyennes sont finalement assez proches des anomalies vraies pour l’ensemble des zones bien qu’un léger écart subsiste. Cependant, la dispersion des ensembles d’ébauche sont aussi un peu moins importantes que celles générées par l’expérience EP3-1, en particulier dans les zones 4, 5 et 9 où la dispersion des observations est clairement plus importante que celle du modèle. Ces différences de dispersion expliquent alors les moins bons résultats de l’expérience.

Ces deux premières simulations sont très encourageantes et montrent le potentiel des anomalies pour corriger la distribution des coefficients de Manning. Elles montrent aussi que certaines périodes de l’année sont plus adaptées à l’assimilation des anomalies (dûes notamment à la variabilité temporelle des sensibilités des anomalies mis en évidence au Chapitre 2, Section 2.3.4.2). Ici, la période de hautes eaux qui permet de générer plus de variabilités dans les ensembles est particulièrement adaptée. Aussi, bien qu’on assimile des anomalies, on cherche avant tout à corriger les estimations de profondeurs d’eau et de débits des fleuves. Dans le cas des expériences EP3-1 et EP3-2, on ne considère aucune erreur de bathymétrie entre la simulation vraie et les simulations d’ensemble. Ainsi, retrouver les anomalies vraies permet automatiquement de retrouver les profondeurs d’eau et débits vrais, comme l’illustre la Figure 3.23, qui montrent l’évolution au fil des cycles des profondeurs d’eau à Óbidos pour les deux expériences.

Les expériences EP4F et EP4NF qui suivent permettent, comme pour EP2F, de tester le comportement de la maquette lorsqu’il y a une erreur de bathymétrie entre les observations et le modèle.

3.4.2 Résultats des expériences EP4F et EP4NF

Au vu des résultats sur les expériences EP3-1 et EP3-2, on choisit de conserver la période de simulation en saison de hautes eaux. Les expériences EP4F et EP4NF sont alors basées sur la même configuration que EP3-1. Cependant, pour ces simulations, on ré-introduit les erreurs de bathymétrie entre la simulation vraie et les simulations d’ensemble comme pour l’expérience EP2F (voir Section 3.2.3 et 3.3.2). Pour distinguer les deux expériences, l’une est basée sur la simulation vraie avec inondations (EP4F) et l’autre sur la simulation vraie sans inondations (EP4NF). Comme seule la hauteur critique est modifiée, le cas EP4NF devrait fonctionner aussi bien, si ce n’est mieux, que

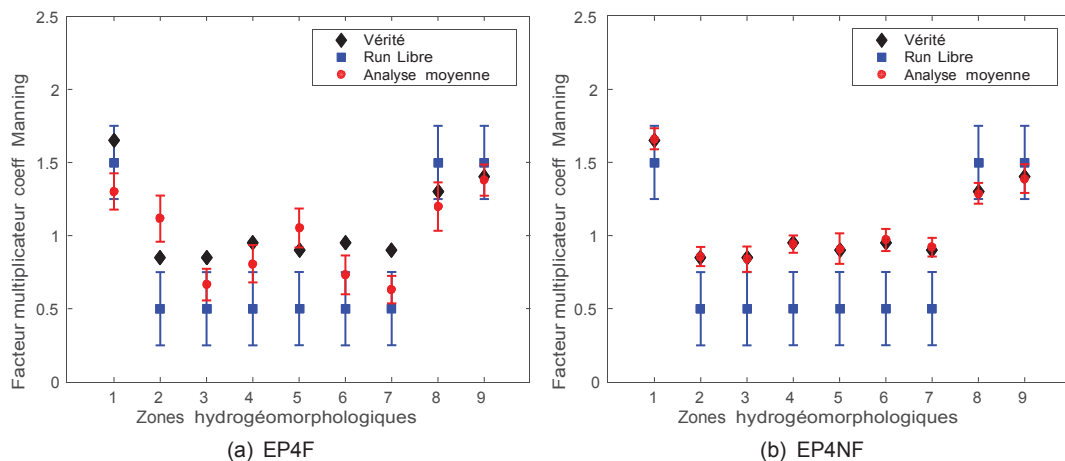


Figure 3.24 – Expériences EP4F (a) et EP4NF (b) : Comparaison des valeurs des facteurs multiplicatifs de la distribution de coefficients de Manning par zones. Les valeurs vraies correspondent aux losanges noirs, les valeurs a priori au 1er cycle d’assimilation avec erreur correspondent aux carrés bleus et les valeurs d’analyse moyenne au dernier cycle d’assimilation avec dispersion de l’ensemble correspondent aux cercles rouges.

l’expérience EP3-1 (car il n’y a pas d’inondations). En effet, dans ce cas, la bathymétrie complète se résume à la largeur et la pente qui sont identiques entre la simulation vraie et les simulations d’ensemble.

La Figure 3.24 présente les valeurs moyennes d’analyse des variables de contrôle au dernier cycle d’assimilation comparées aux valeurs vraies et aux valeurs a priori au premier cycle pour les deux simulations et le Tableau 3.11 résume ces valeurs avec la dispersion des ensembles. Pour l’expérience EP4F, mis à part pour la zone 1, les variables de contrôle de la moyenne d’ensemble analyse de chaque zone sont rapprochées des valeurs vraies, avec une dispersion de l’ensemble d’analyse réduite par rapport au run libre. Parallèlement, l’expérience EP4NF donne les meilleurs résultats (toutes expériences confondues) avec seulement 9,8 % d’erreurs sur les valeurs analyse finales des facteurs multiplicateurs de la distribution de Manning sur tout le bassin (voir Tableau 3.12). Les erreurs relatives moyennes de l’expérience EP4F sont proches de celles de EP3-2 alors que celles de l’expérience EP4NF sont équivalentes (si ce n’est meilleures) à celles de EP3-1. Pour étudier plus en détails le comportement de la maquette pour ces expériences, les Figures 3.25-3.26 montrent l’évolution au fil des cycles de la valeur moyenne analyse des variables de contrôle pour chaque zone et chaque expérience.

	Err. relative par rapport à la vérité		
	Sans assimilation	Après assimilation	
		EP4F	EP4NF
variables de contrôle	33,07 %	20,88 %	9,84 %
anomalies	EP4F : 48,78 % 15,78 cm	17,72 %	6,24 %
	EP4NF : 54,92 % 15,83 cm	8,41 cm	3,35 cm
hauteurs d'eau	EP4F : 22,42 % 36,33 cm	11,43 %	3,87 %
	EP4NF : 22,43 % 36,53 cm	18,04 cm	6,74 cm
débits	EP4F : 3,02 % EP3-2 : 2,83 %	1,62 %	0,42 %

Table 3.12 – EP4F et EP4NF : Erreur relative (moyenne dans le temps) des facteurs multiplicatifs (ligne 2), des anomalies (ligne 3), des hauteurs d'eau (ligne 4) et des débits (ligne 5) par rapport à la simulation vraie.

Zones	Vrai	A priori (Sans assim)	Analyse (Après assim)	
			EP4F	EP4NF
1	1,65	1,5 (0,25)	1,302 (0,124)	1,662 (0,072)
2	0,85	0,5 (0,25)	1,115 (0,158)	0,857 (0,066)
3	0,85	0,5 (0,25)	0,665 (0,108)	0,839 (0,087)
4	0,95	0,5 (0,25)	0,808 (0,128)	0,942 (0,059)
5	0,9	0,5 (0,25)	1,052 (0,133)	0,911 (0,104)
6	0,95	0,5 (0,25)	0,731 (0,132)	0,970 (0,076)
7	0,9	0,5 (0,25)	0,630 (0,094)	0,921 (0,064)
8	1,3	1,5 (0,25)	1,199 (0,166)	1,289 (0,071)
9	1,4	1,5 (0,25)	1,379 (0,107)	1,390 (0,098)

Table 3.11 – Comparaison, pour chaque zone, des valeurs vraies (colonne 2), des valeurs a priori (colonne 3) avec dispersion (entre parenthèses), des valeurs analyses moyennes finales de l'expérience EP4F avec dispersion (colonne 4) et des valeurs analyses moyennes finales de l'expérience EP4NF avec dispersion (colonne 5).

Pour l'expérience EP4F (voir Fig. 3.25), les valeurs analyses se rapprochent de la valeur vraie sans vraiment l'atteindre. On constate aussi d'importantes inondations sur l'ensemble des simulations (vraie, libre, ensemble d'ébauche et d'analyse) et sur la majorité des zones. Finalement, les résultats de cette expérience sont proches de ceux de l'expérience EP2F. D'autre part, si on regarde pour l'expérience EP4NF (voir Fig. 3.26), les variables de contrôle convergent toutes vers la valeur vraie et pour la plupart des zones, dès le premier cycle d'assimilation. Cette expérience combine en effet toutes les configurations favorables pour converger rapidement : (1) l'expérience a lieu en période de hautes eaux, engendrant une grande dispersion de l'ensemble d'ébauche comme vu pour l'expérience EP3-1 et (2) la bathymétrie particulière de cette expérience fait qu'on ne constate pas d'inondations sur le bassin (et ce malgré que l'on soit en saison de hautes eaux).

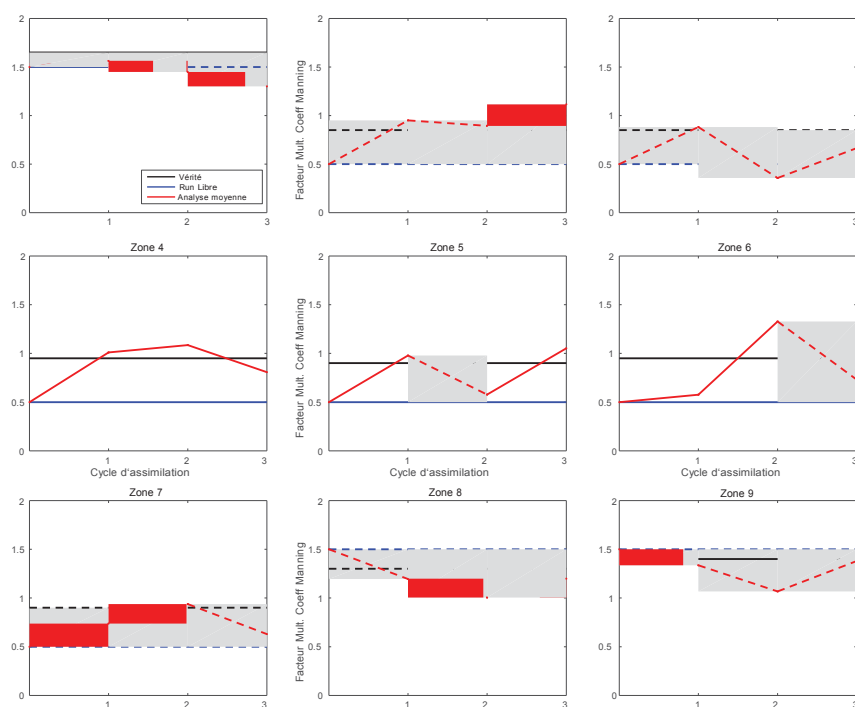


Figure 3.25 - Expérience EP4F : Moyenne de l'ensemble de contrôle analyse au fil des cycles d'assimilation en ligne pleine/tirets rouge, comparée à la valeur vraie en ligne pleine/tirets noire et la valeur a priori au premier cycle en ligne pleine/tirets bleue. Les lignes en tirets matérialisent les cycles où des inondations sont constatées. De même, les zones grisées matérialisent les cycles, par zone, où des inondations sont constatées.

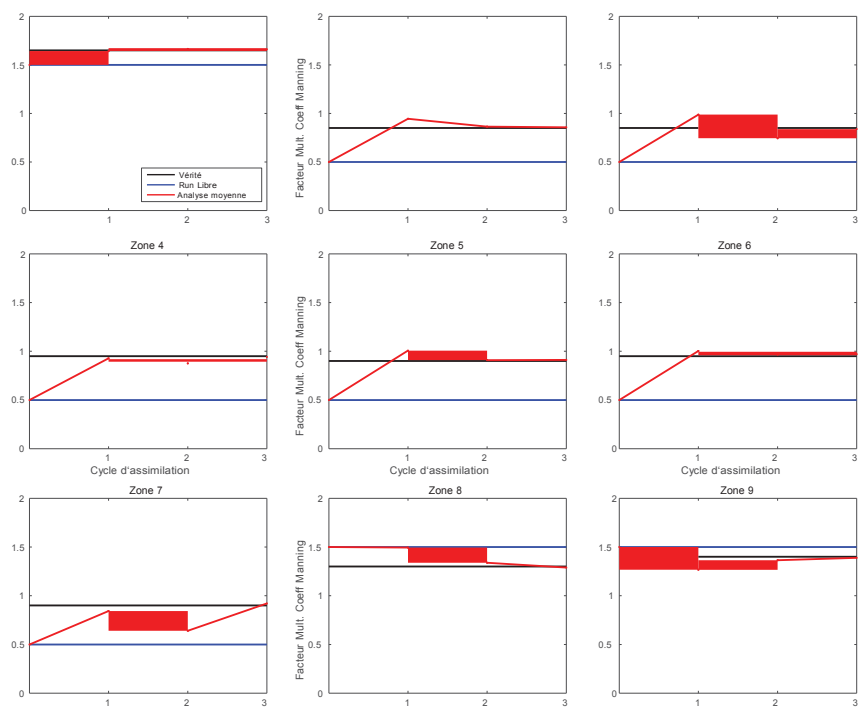


Figure 3.26 - Expérience EP4NF : Moyenne de l'ensemble de contrôle analyse au fil des cycles d'assimilation en ligne pleine/tirets rouge, comparée à la valeur vraie en ligne pleine/tirets noire et la valeur a priori au premier cycle en ligne pleine/tirets bleue.

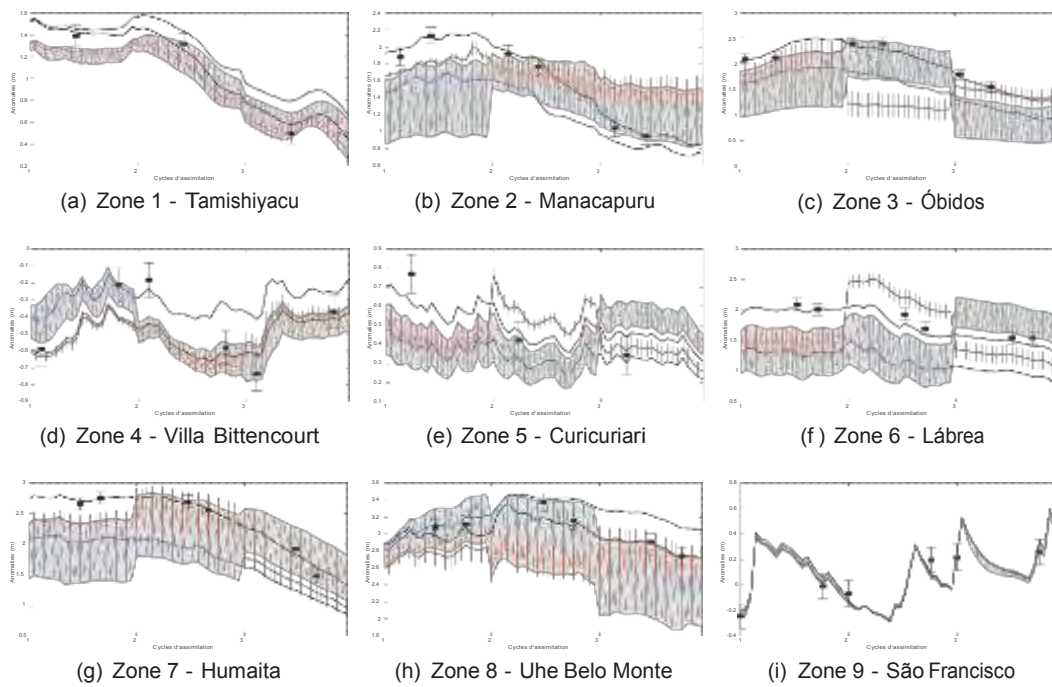


Figure 3.27 – Expérience EP4F : Évolution dans le temps des anomalies vraies (ligne noire), des anomalies du run libre (ligne bleue), des anomalies observées avec leur erreur (marqueurs carrés verts), de l'ensemble des anomalies d'ébauche (zone grisée), de la moyenne des anomalies analyse (ligne orange) avec la dispersion de l'ensemble analyse (hachures rouges).

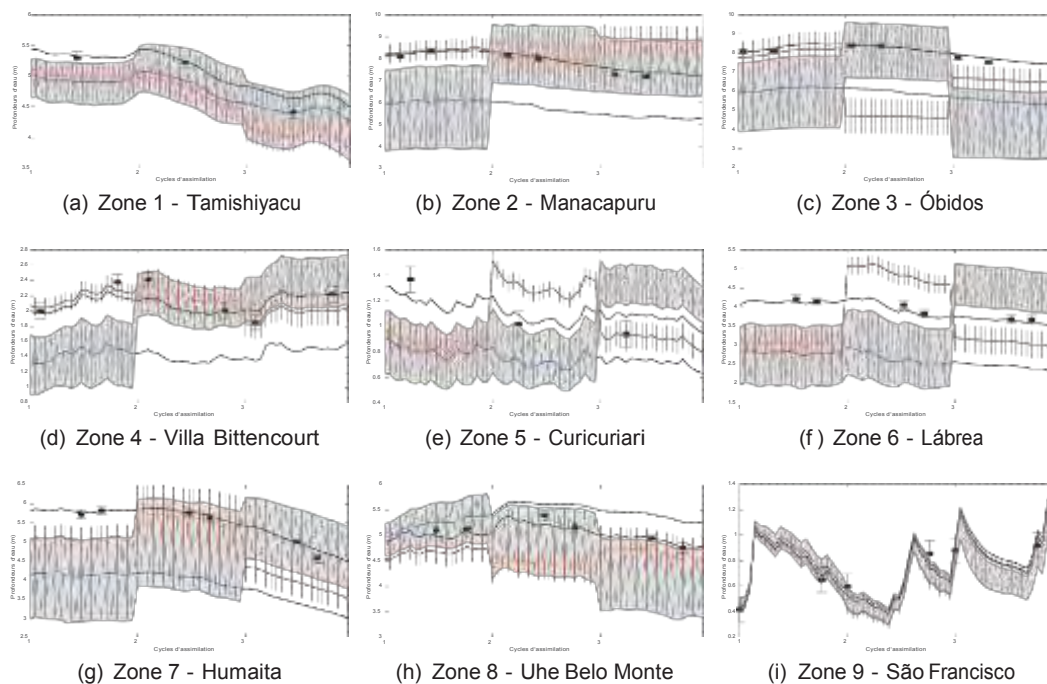


Figure 3.28 – Expérience EP4F : Évolution dans le temps des profondeurs d’eau vraies (ligne noire), des profondeurs d’eau du run libre (ligne bleue), des profondeurs d’eau observées avec leur erreur (marqueurs carrés verts), de l’ensemble des profondeurs d’eau d’ébauche (zone grisée), des profondeurs d’eau analyses moyennes (ligne orange) avec la dispersion de l’ensemble analyse (hachures rouges).

Maintenant, la Figure 3.27 illustrent, pour chaque zone, les anomalies simulées et observées de l’expérience EP4F. Mis à part pour la zone 2 (voir Fig. 3.27b), la maquette parvient à se rapprocher des anomalies vraies. De ce point de vue, la maquette fonctionne correctement, c’est-à-dire que la simulation adapte le Manning pour que les anomalies modèle se rapprochent des anomalies observées. Cependant, si l’on regarde après coup les profondeurs d’eau en Figure 3.28, on constate un écart, parfois important, entre les profondeurs d’eau vraies et les profondeurs d’eau analyse. Cet écart est directement dû aux erreurs de bathymétrie introduites pour cette simulation.

Au contraire, pour l’expérience EP4NF, non seulement les anomalies analyse se rapprochent voir se confondent avec les anomalies vraies, voir Figure 3.29 mais il en est de même pour les profondeurs d’eau, voir Figure 3.30. Ceci s’explique simplement par le fait que ni la simulation vraie (et donc les observations) et ni les simulations d’ensemble ne sont inondées. En effet, les profondeurs d’eau, et par conséquent les anomalies, ne dépendent en fait que de la largeur W du réservoir ainsi que de sa longueur L . La bathymétrie (via la profondeur de plein bord H_c) n’intervient qu’en présence d’inondations. Bien qu’il y ait une erreur sur la valeur de H_c , cette erreur n’a aucun impact sur les profondeurs et anomalies d’eau simulées s’il n’y a pas d’inondations. Ainsi, pour l’expérience EP4NF, l’erreur de bathymétrie n’a aucune incidence et la maquette retrouve à la fois la distribution de coefficients de Manning vraie et les profondeurs/débits vrais.

Pour conclure sur ces 4 dernières expériences, l’assimilation des anomalies offrent de très inté-

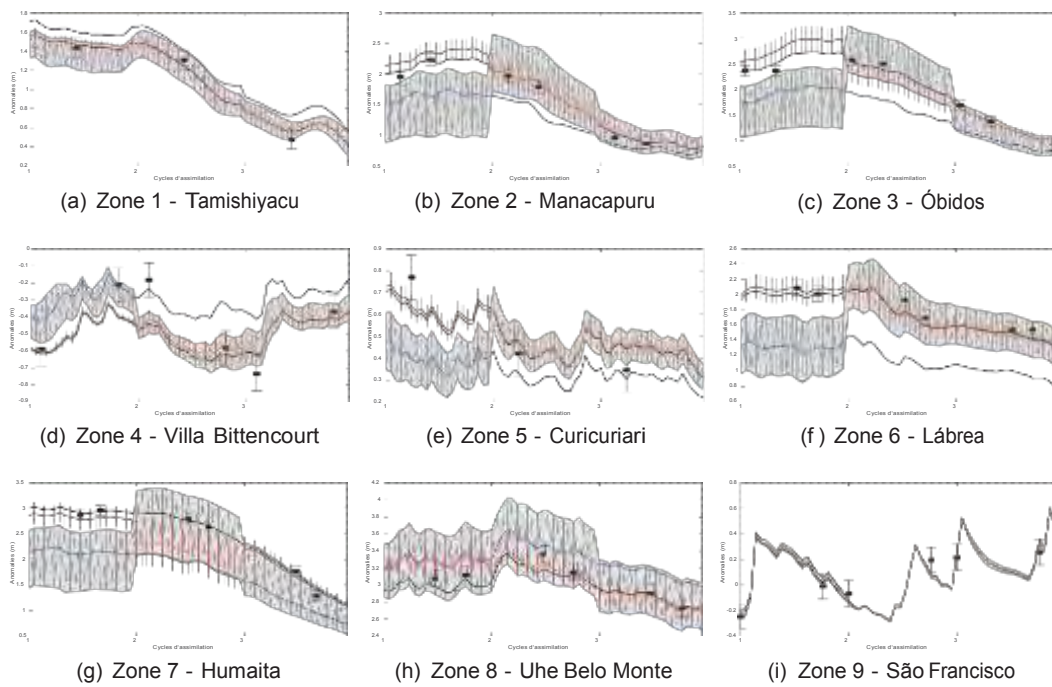


Figure 3.29 – Expérience EP4NF : Évolution dans le temps des anomalies vraies (ligne noire), des anomalies du run libre (ligne bleue), des anomalies observées avec leur erreur (marqueurs carrés verts), de l'ensemble des anomalies d'ébauche (zone grisée), de la moyenne des anomalies analyse (ligne orange) avec la dispersion de l'ensemble analyse (hachures rouges).

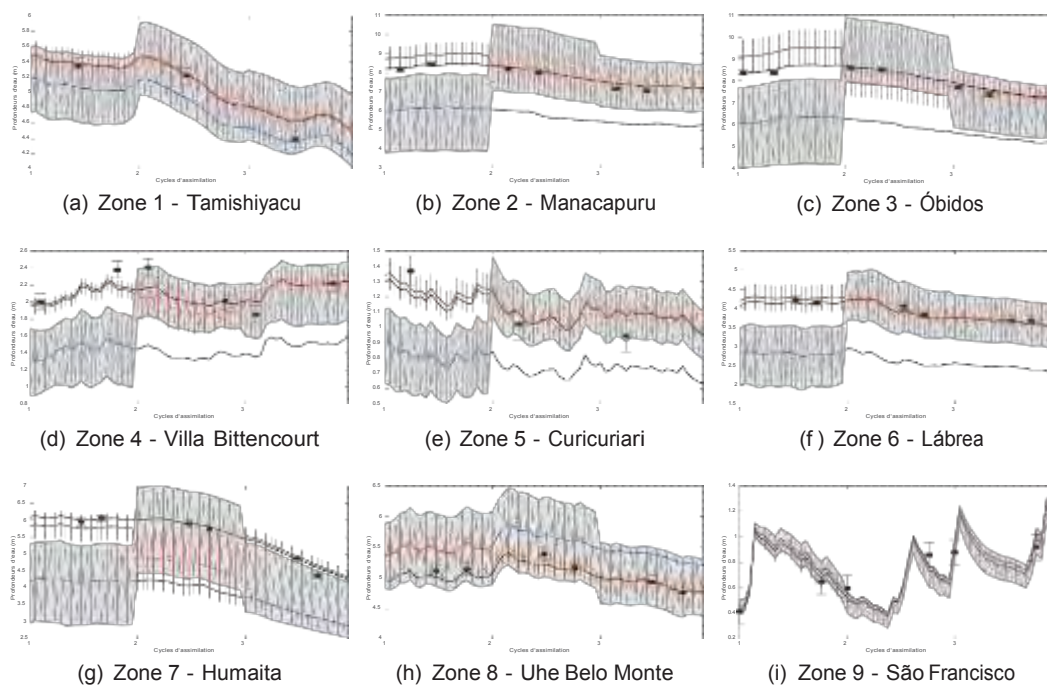


Figure 3.30 – Expérience EP4NF : Évolution dans le temps des profondeurs d'eau vraies (ligne noire), des profondeurs d'eau du run libre (ligne bleue), des profondeurs d'eau observées avec leur erreur (marqueurs carrés verts), de l'ensemble des profondeurs d'eau d'ébauche (zone grisée), des profondeurs d'eau analyses moyennes (ligne orange) avec la dispersion de l'ensemble analyse (hachures rouges).

ressantes perspectives pour corriger les paramètres. La maquette est capable, même dans la pire des configurations - c'est-à-dire la simulation EP4F - d'améliorer les estimations de profondeurs d'eau et de débit. L'assimilation des anomalies doit cependant être menée sur des périodes coïncidant plutôt avec les saisons de hautes eaux. Il est cependant difficile voir impossible, sur un bassin comme celui de l'Amazonie, de retrouver les conditions idéalisées de l'expérience EP4NF. On retrouvera plus une configuration comme celle de l'expérience EP4F qui cumule inondations et erreurs sur la bathymétrie. La maquette a démontré sa capacité à retrouver les anomalies observées, mais, si aucune information extérieure sur la bathymétrie n'est disponible, il sera compliqué de retrouver les bonnes profondeurs d'eau sans y insérer un biais.

Conclusions et perspectives

Dans ce chapitre, on s'est intéressé à l'apport d'un produit de profondeurs ou d'anomalies d'eau présentant des caractéristiques proches de celles du produit de cotes d'eau que fournira SWOT, pour corriger la distribution des coefficients de Manning utilisée par la composante RRM du modèle ISBA-TRIP. Le satellite SWOT n'étant pas lancé avant 2021, on s'est nécessairement placé dans un cadre d'expériences jumelles qui permettent de contrôler l'intégralité des incertitudes présentes dans le système. De là, on a utilisé un AEnKF, qui fonctionne comme un EnKF classique mais qui permet d'assimiler des observations ayant lieu entre l'instant de début du cycle et l'instant d'analyse. L'intérêt de cet algorithme est d'assimiler toutes les observations de type SWOT disponibles sur une fenêtre de 21 jours (égale à la répétitivité du satellite). Plusieurs expériences ont alors été menées : en assimilant d'abord des profondeurs d'eau sans erreurs sur la profondeur critique mais avec des inondations (EP1F) puis sans inondations (EP1NF), en ajoutant des erreurs sur la profondeur critique tout en conservant les inondations (EP2F), en assimilant ensuite des anomalies de profondeurs d'eau sur des périodes différentes de l'année avec des inondations et sans erreurs sur la profondeur critique (EP3-1 et EP3-2) et enfin en ré-insérant des erreurs de profondeur critique avec inondations (EP4F) et sans inondations (EP4NF).

Les Figures 3.31-3.32 présentent une comparaison des performances de chacune des 7 expériences jumelles. Pour chaque cellule dans le réseau hydrologique principal et pour chacune des 9 zones hydrogéomorphologiques, la série temporelle de la profondeur d'eau analyse moyenne est comparée à la série temporelle de la profondeur vraie (calcul de la valeur absolue de la différence). À partir de l'ensemble des séries temporelles des différences, calculées pour chaque cellule du réseau hydrologique principal, on en calcule les statistiques pour les représenter sous forme de boîtes à moustache. Les mêmes calculs sont réalisés pour les observations et les runs libres pour comparaison. La Figure 3.31 représente les performances de l'ensemble des expériences avec inondations en séparant les expériences avec et sans erreurs de bathymétrie c'est-à-dire, de gauche à droite en omettant les boîtes à moustaches vertes et bleues (observations et runs libres) : EP1F (rouge), EP3-1 (beige), EP3-2 (rose), EP2F (violet) et EP4F (jaune). La Figure 3.32 représente alors les performances des expériences sans inondations c'est-à-dire, toujours de gauche à droite en ometant les boîtes à moustaches vertes et bleues (observations et runs libres) : EP1NF (rouge foncé) et EP4NF (orange).

Les Figures 3.31-3.32 confirment que, pour les configurations des expériences (même les plus pessimistes), la maquette d'assimilation est capable d'améliorer les estimations de profondeurs d'eau pour les rapprocher des profondeurs vraies. Comme le montrent les ordres de grandeurs des statistiques pour les expériences représentées en Figure 3.31, la maquette a de meilleures performances lorsqu'il n'y a pas d'inondations. En effet, les inondations perturbent le régime d'écoulement de l'eau dans le réservoir de surface et cela se traduit par l'ajout de non-linéarités dans la dynamique du modèle. Or,

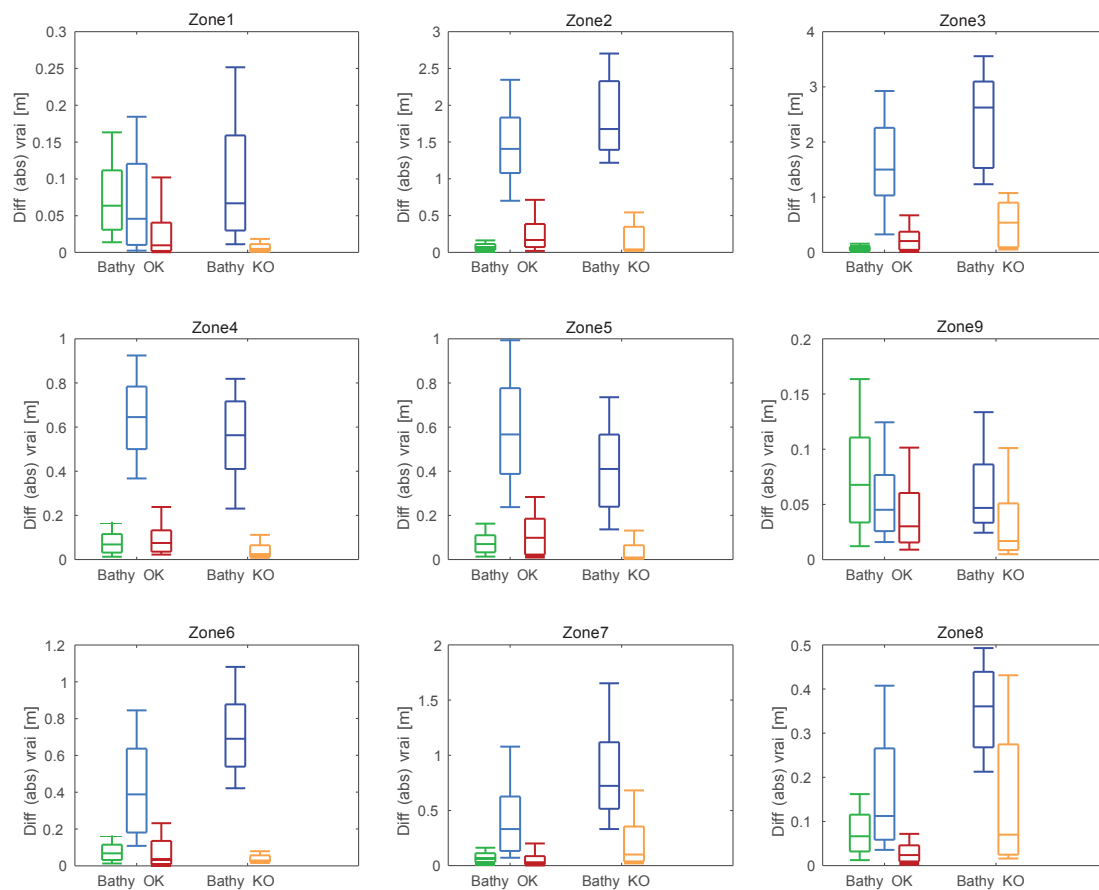


Figure 3.31 – Statistiques, pour les expériences sans inondation, par zone (sous forme de boîtes à moustaches) de l'ensemble des écart des profondeurs simulées et observées au profondeurs vraies, sur l'ensemble de la période d'assimilation et sur les cellules du réseau hydrologique principal : observations (vert), run libre sans erreurs de bathymétrie (bleu clair), EP1NF (rouge foncé), run libre avec erreurs de bathymétrie (bleu foncé) et EP4NF (orange).

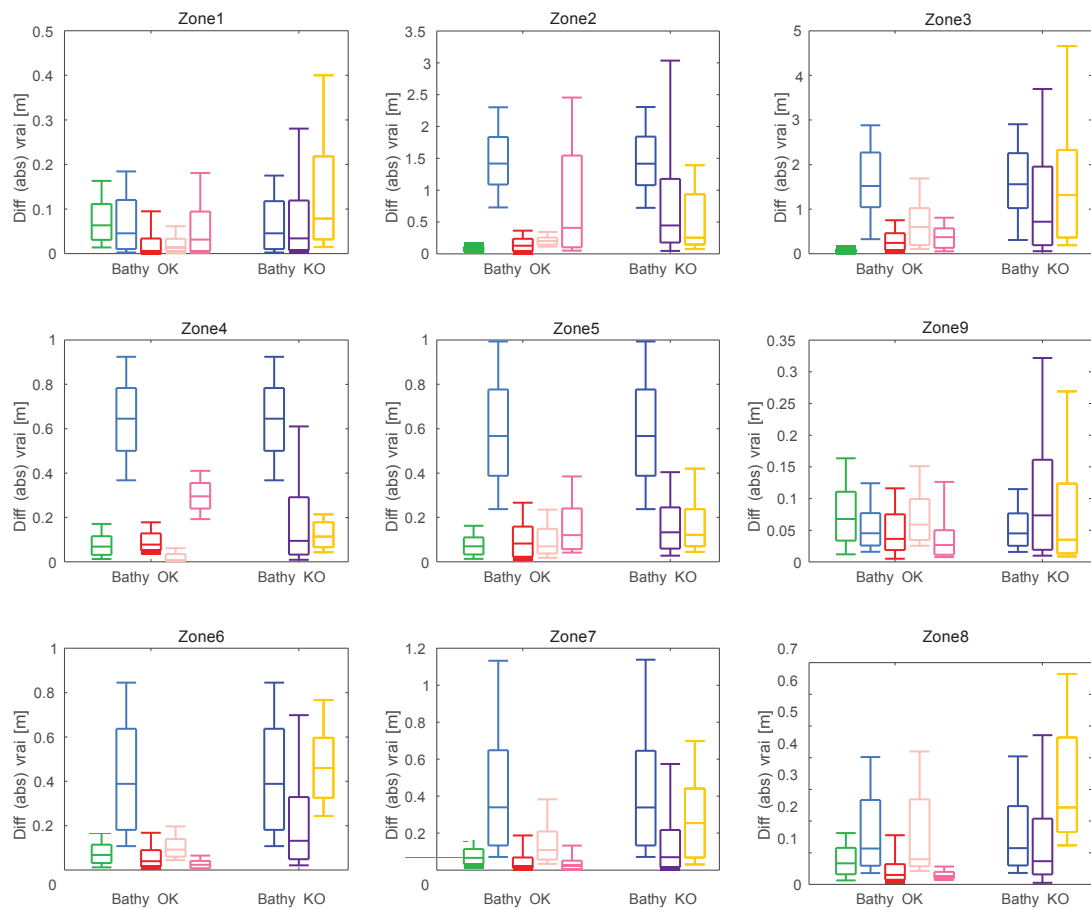


Figure 3.32 – Statistiques, pour les expériences avec inondations, par zone (sous forme de boîtes à moustaches) de l'ensemble des écart des profondeurs simulées et observées au profondeurs vraies, sur l'ensemble de la période d'assimilation et sur les cellules du réseau hydrologique principal : observations (vert), run libre sans erreurs de bathymétrie (bleu clair), EP1F (rouge), EP3-1 (beige), EP3-2 (rose), run libre avec erreurs de bathymétrie (bleu foncé), EP2F (violet) et EP4F (jaune).

même si l'AEnKF utilisé ici est plus robuste que le Filtre de Kalman classique (ou encore le Filtre de Kalman Étendu) pour gérer les non-linéarités du modèle, il n'en reste pas moins aussi basé sur une hypothèse de linéarité du modèle. Ces observations nous incitent à envisager de nouveaux algorithmes d'assimilation de données plus à même de gérer les non-linéarités induites par les inondations comme le Filtre à Particules. La Figure 3.32, quant à elle, montre les performances des expériences ayant une configuration avec inondations (et donc plus réalistes étant donné le bassin d'étude de l'Amazonie). Pour les zones 1, 8 et 9, l'apport de l'assimilation n'est pas évident car il s'avère que l'erreur est déjà très basse dans le run libre (sans assimilation). Ceci explique pourquoi, dans toutes les expériences, ces zones étaient celles dont la distribution des coefficients de Manning était la moins corrigée. Sur les autres zones, les expériences sans erreurs de bathymétrie donnent évidemment des meilleures statistiques que celles avec erreurs de bathymétrie mais il est difficile d'extraire une configuration qui donne pour toutes les zones des meilleurs résultats. L'ensemble de ces expériences montrent la possibilité d'utiliser des observations de profondeurs/d'anomalies d'eau de type SWOT pour corriger les paramètres géomorphologiques du modèle ISBA-TRIP.

Il est important de noter que les expériences assimilant des anomalies de profondeur d'eau (EP3-1, EP3-2, EP4F et EP4NF) fournissent des résultats relativement similaires aux expériences assimilant les profondeurs d'eau (EP1F, EP1NF et EP2F). Ce qui signifie qu'il sera possible de s'affranchir d'éventuel biais sur les cotes d'eau SWOT par rapport aux cotes d'eau TRIP en assimilant des anomalies de cote d'eau SWOT pour corriger la distribution des coefficients de Manning dans TRIP.

Ces premières expériences posent de nouvelles questions et de nouvelles perspectives à explorer. Tout d'abord, en conservant l'objectif de corriger la distribution des coefficients de Manning, plusieurs pistes sont à explorer :

- Une première étape consisterait à insérer de nouvelles sources d'incertitudes dans la maquette. Des erreurs sur les précipitations sont inévitables et peuvent facilement être ajoutées en utilisant deux produits de précipitations différents : un pour le run vrai et les observations, et un pour le run libre et les runs d'ensemble. Seulement, pour ces nouvelles expériences (et contrairement aux expériences EP2F, EP4F et EP4NF), il faudra tenir compte de ces incertitudes lors de la génération des différents ensembles d'ébauche. Des méthodes comme celles utilisées par Clark et al. (2008) ou de Paiva et al. (2013) permettent de perturber les champs de précipitations pour perturber les ensembles d'observations modèle. Par ailleurs, disposer de champs d'erreurs avec les jeux de précipitations directement fournis par les producteurs de ces données serait un apport conséquent pour ce type d'étude. Dans ce cadre, on pourrait, par exemple, utiliser le nouveau produit de précipitations multi-satellites TAPEER (Tropical Amount of Precipitation with an Estimation of ERrors ; Chambon et al., 2012) qui fournit non seulement des précipitations, mais aussi une estimation de l'erreur associée. De plus, d'autres paramètres de TRIP comme la profondeur critique H_c devraient aussi être perturbés pour générer l'ensemble d'ébauche.
- Les expériences jumelles, bien qu'utiles pour générer des pseudo-observations SWOT avant le lancement du satellite, souffrent d'un inconvénient majeur : le même modèle est utilisé pour générer les observations et les simulations d'ensemble. Le modèle ISBA-TRIP, et plus particulièrement TRIP, est basé sur un modèle d'onde cinématique seule. Il serait intéressant de regarder le comportement de la maquette quand on inclut certains processus non modélisés par TRIP comme les effets de remous ou les ondes diffusives. Pour cela, il faut utiliser un second modèle hydrologique, différent de TRIP, pour générer les observations. L'intérêt sera alors de regarder comment la maquette adapte les paramètres de contrôle pour se rapprocher des observations, alors que le modèle ne contient pas toute la physique présente dans les observations.

Dans un second temps, l'analyse de sensibilité (voir Chapitre 2) a mis en évidence une sensibilité des profondeurs d'eau non seulement au coefficient de Manning, mais aussi à d'autres paramètres de TRIP. Les expériences réalisées au cours de ce chapitre peuvent être étendues à ces paramètres, notamment la largeur ou la pente du lit de la rivière. On peut conserver la même configuration en ne perturbant et en ne corrigeant qu'un seul paramètre à la fois, tout en continuant à assimiler les profondeurs d'eau et/ou les anomalies. Contrairement au coefficient de Manning, les autres paramètres évoqués sont plus facilement observables et mesurables. La stratégie d'assimilation pourrait alors évoluer vers l'assimilation d'autres observables SWOT. Il se posera alors le problème de l'échelle des produits SWOT et de l'échelle des processus représentés dans TRIP. En effet, certains paramètres actuels de TRIP ont des valeurs différentes de celles préconisées dans la littérature. Par exemple, dans la plupart des modèles hydrologiques, le coefficient de Manning prend des valeurs entre 0,02 et 0,04 (bien que la gamme de variation de ce coefficient puisse être bien plus importante, en fonction de la nature du lit du fleuve). Or, dans TRIP, il varie entre 0,04 et 0,06. Il était nécessaire de prendre ses valeurs plus importantes dans TRIP afin de pallier aux approximations dues à la basse résolution, qui ne représente pas tous les processus physiques, et ainsi pouvoir malgré tout modéliser correctement le comportement du bassin. Par conséquent, assimiler directement des observations de certains paramètres comme la largeur ou la pente ne garantirait pas forcément la bonne reproduction des profondeurs et des débits observés. En effet, TRIP simule un seul tronçon de rivière par cellule, équivalente à l'ensemble du réseau fluvial contenu dans cette cellule (qui a une taille de 50km x 50 km). Il se posera donc le problème de la représentativité de la largeur de ce tronçon TRIP par rapport aux largeurs observées par SWOT. En dernier lieu, on peut peut-être envisager de faire de l'estimation multi-paramètres, tout en faisant attention aux problèmes d'équifinalité que cela risque d'engendrer.

Enfin, on rappelle que l'objectif de toutes ces expériences est d'étudier l'apport des observables SWOT pour la modélisation hydrologique à grande échelle. Les produits SWOT seront à une résolution bien plus fine que le modèle ISBA-TRIP et cet aspect a été ignoré jusqu'à présent. Il sera alors intéressant de considérer des observations à plus petite résolution et de développer le traitement nécessaire pour pouvoir les assimiler malgré tout dans ISBA-TRIP.

D'autre part, il sera utile de considérer des erreurs d'observation SWOT plus réalistes qu'un simple bruit blanc, comme notamment des erreurs corrélées le long de la trace ou des erreurs qui augmentent avec la distance au nadir. Des travaux comme ceux de Lion (2012) ont notamment travaillé sur la définition de ces erreurs SWOT. De plus, depuis juin 2016, le CNES a financé un simulateur de données SWOT grand échelle adapté aux modèles hydrologiques. Ce simulateur n'a évidemment pas pu être utilisé dans le cadre de cette thèse, mais il serait pertinent de l'utiliser pour les futures études SWOT à grande échelle.

Ce chapitre s'est ainsi concentré sur l'utilisation du produit de cotes d'eau de SWOT pour corriger les paramètres de TRIP. Or, comme présenté au Chapitre 1, Section 1.3.5.5, un produit de débit à partir des observables SWOT est actuellement en développement. En anticipation de ce produit et pour tester le comportement de la maquette en expériences réelles, le chapitre suivant va se concentrer sur l'assimilation d'un produit de débit altimétrique pour corriger l'état du modèle ISBA-TRIP.

Chapitre 4

Préparation à l'assimilation du produit de débits SWOT : estimation d'état avec des débits altimétriques

Sommaire

Introduction	182
4.1 Comparaison du modèle ISBA-TRIP aux débits observés	184
4.1.1 Débits in situ	184
4.1.2 Débits altimétriques	184
4.1.3 Performances d'ISBA-TRIP sur l'Amazonie	188
4.2 Méthode d'assimilation des débits ENVISAT	200
4.2.1 Définition des variables d'observations	200
4.2.2 Stratégies d'assimilation de données et choix du vecteur de contrôle	201
4.2.3 Génération de l'ensemble	203
4.2.4 Présentation de la stratégie d'assimilation et des expériences d'assimilation	211
4.3 Résultats de l'expérience EE1-direct	212
4.3.1 Résultats généraux	212
4.3.2 Comportement pour certaines stations tests	215
4.3.3 Le cas de la confluence Negro-Branco	219
4.3.4 Bilan sur l'expérience EE1-direct	224
4.4 Performances de l'EnKF avec localisation	225
4.4.1 Expérience préliminaire : EE1-diag	225
4.4.2 Expérience-test du masque de localisation : EE1-local	231
4.4.3 Conclusions sur la localisation	236
4.5 Sensibilité à l'opérateur d'observations	238
4.5.1 Évaluation des incertitudes de l'opérateur Z	238
4.5.2 Comparaison des expériences EE1-local, EE2-local et EE3-local	239
Intercomparaison des simulations, conclusion et perspectives	244

Introduction

En Assimilation de Données (AD), l'estimation d'état consiste à assimiler des observations pour directement corriger les sorties du modèle, en d'autres termes l'état des variables pronostiques (voir Chapitre 1) obtenues en intégrant le modèle ISBA-TRIP. On parle aussi de correction de la condition initiale. En effet, après assimilation, le modèle est propagé à partir de l'état corrigé. En estimation d'état, on part ainsi du constat que le modèle est nécessairement imparfait. Les différentes incertitudes (dans la paramétrisation, les conditions aux limites, la modélisation) se répercutant dans les sorties, on choisit de corriger directement l'état ou ces sorties du modèle (qui regroupent indirectement ces incertitudes).

L'estimation d'état est très couramment employée en océanographie et météorologie (Evensen and Leeuwen, 1996; Tanajura et al., 2015; Houtekamer and Mitchell, 2001; Gustaffson et al., 2012). En hydrologie, par contre, l'assimilation de données télédéteçtées pour corriger l'état d'un modèle hydrologique est beaucoup plus récente et bien moins fréquente. Elle s'est surtout développée dans les LSMs (*Land Surface Models* ou modèles de surfaces continentales). À l'échelle globale, on peut citer le NASA GEOS-5 LDAS (*version 5 of the Goddard Earth Observing System - Land Data Assimilation System*) qui assimile des produits d'humidité du sol (notamment de la mission SMOS), de fraction de couvert neigeux (de l'instrument MODIS) et de variations intégrées verticalement de stock d'eau terrestre (de la mission GRACE) pour corriger l'état de plusieurs LSMs, en utilisant un EnKF (Reichle et al., 2014). Pour plus de détails sur ces données spatiales, voir le Chapitre 1. La validation d'un tel système est évidemment très difficile. Il semblerait que l'estimation du couvert neigeux pour de faibles épaisseurs de neige bénéficie grandement de l'assimilation de données optiques MODIS (mais pas pour les couches de neige profondes). L'humidité du sol superficielle semble améliorée en assimilant les données SMOS. Mais Reichle et al. (2014) note, entre autre, que les données de télédétection n'observant pas directement la grandeur physique à corriger et ayant des échelles spatiales assez différentes, il est nécessaire de faire des progrès dans la définition de l'opérateur d'observation. Il existe aussi d'autres études assimilant ce type de données spatiales (et éventuellement des mesures in situ) dans des modèles hydrologiques à l'échelle d'un bassin versant, voire à l'échelle locale (par exemple Trudel et al., 2014).

En ce qui concerne les eaux de surface et, plus particulièrement, les systèmes de routage des modèles hydrologiques, a priori il n'existe pas (encore) de RRM associé à un LSM qui assimile des données de télédétection ou in situ à l'échelle globale. Toutes les études d'assimilation se font pour le moment à des échelles régionales, voire locales. Elles utilisent de plus des systèmes de routage plus fine échelle que les RRM globaux. En ce qui concerne les données de télédétection utilisées, comme le souligne Schumann and Domeneghetti (2016), les données d'altimétrie satellitaires sont encore peu utilisées. On peut toutefois citer Michailovsky et al. (2013), qui assimilent des mesures de cotes d'eau Envisat dans un modèle hydrologique, via un EKF, pour corriger les volumes d'eau sur le Brahmapoutre en Asie du Sud. Paiva et al. (2013) ont assimilé (avec un EnKF) à la fois des anomalies (la moyenne temporelle de la série est soustraite à la série complète) de cote d'eau Envisat et des débits in situ pour corriger les cotes d'eau et les débits d'un modèle hydrologique, sur le bassin de l'Amazonie. Ces deux travaux montrent une certaine amélioration des cotes d'eau après assimilation des données altimétriques, mais une amélioration modérée sur le débit. Mais ce type d'étude reste encore relativement rare (peu de modèles hydrologiques peuvent simuler et donc assimiler directement des cotes d'eau). D'autres études assimilent seulement des données in situ dans des modèles hydrologiques. Le plus souvent, ces études sont réalisées sur des petits bassins, comme par exemple Clark et al. (2008). On peut toutefois citer le travail réalisé par Pan and Wood (2013), qui ont utilisé des débits provenant d'un réseau de stations in situ pour inverser ces débits en un

champ 2D de ruissellement de surface (variable directement simulée par un modèle hydrologique) sur le bassin de l'Ohio. Pour ce faire, ils ont utilisé un a priori du ruissellement de surface provenant d'un modèle hydrologique et un algorithme équivalent à une variante du lisseur de Kalman.

En ce qui concerne les modèles hydrauliques, il existe un peu plus d'études assimilant des données de télédétection, notamment des masques de zones en eau Schumann and Domeneghetti (2016). On peut par exemple citer Neal et al. (2009), qui combinent des images SAR d'Envisat de surface d'eau avec un MNT LiDAR pour en retirer des mesures d'élévation des eaux sur un tronçon d'une rivière au Luxembourg (l'Alzette). Ces mesures de hauteurs d'eau sont ensuite assimilées dans le modèle hydrodynamique HEC-RAS (*Hydrologic Engineering Centers - River Analysis System*) via un EnKF et permettent la réduction des incertitudes sur le débit. Il existe d'autres études d'estimation d'état basées sur des expériences jumelles, dans le cadre de SWOT (donc d'altimétrie à fauchée) et qui n'utilisent donc pas de données réelles. Par exemple, Andreadis et al. (2007) et Biancamaria et al. (2011) utilisent un EnKF pour corriger les profondeurs d'eau dans le modèle Lisflood-FP (Bates and Roo, 2000). Enfin, il existe un plus grand nombre d'études assimilant des données in situ à des échelles plus petites, notamment pour améliorer les prévisions de crues (comme par exemple Barthélémy et al., 2013). Mais comme cette thèse s'intéresse à l'échelle des grands bassins et à l'amélioration de RRM couplé à un LSM, ce type d'étude ne sera pas abordé dans cette thèse.

Dans une formulation très simple, le bilan hydrologique terrestre de n'importe quel bassin peut s'écrire

$$\Delta S = P - ET - Q,$$

où ΔS est la variation de stock/volume d'eau douce dans le bassin, P sont les précipitations (entrées du système), ET est le flux d'évapotranspiration et Q regroupe le débit de surface et les écoulements souterrains (sorties du bassin). La connaissance de Q (via la connaissance des débits) et, par conséquence, une meilleure connaissance de ΔS est très importante pour notre compréhension de la dynamique des eaux continentales à l'échelle des grands bassins mondiaux et à l'échelle globale.

La future mission globale SWOT (voir Chapitre 1, Section 1.3.5) a justement pour objectif scientifique de fournir un inventaire complet de la composante de surface de ΔS (donc une estimation partielle de ΔS) et Q (Rodríguez, 2016; Fu et al., 2012a). Plus particulièrement, SWOT fournira un produit global de débits dérivés des autres observations SWOT - cotes d'eau, largeur, pente avec une estimation de l'incertitude associée. SWOT ne mesurera pas directement le débit. C'est pourquoi, contrairement aux produits de pentes et de cotes, il n'y a pas de contraintes sur la précision du produit de débit SWOT dans le *SWOT Science Requirements Document* (Rodríguez, 2016).

L'objectif de ce chapitre est alors d'anticiper comment combiner le produit de débit SWOT avec les LSMs grande échelle comme ISBA-TRIP. La maquette d'assimilation de données introduite au Chapitre 3, basée sur le filtre de Kalman d'Ensemble (EnKF), est reprise dans une configuration d'estimation d'état. L'objectif de la maquette est d'assimiler des observations de débits (issus d'un traitement de données altimétriques de cotes d'eau) pour corriger l'état du modèle ISBA-TRIP. De plus, pour sortir du cadre idéalisé des expériences jumelles, un produit de débits issus de la télédétection sera assimilé. Ce produit, décrit en Section 4.1, n'est pour l'instant disponible que sur le bassin de l'Amazonie. Les expériences d'assimilation menées dans le cadre de cette étude sont présentées en Section 4.2 avec la configuration de la maquette. Enfin, les Sections 4.3 à 4.5 présentent les résultats et les conclusions.

L'objectif, in fine, des expériences d'AD est de fournir de meilleures estimations du débit des

fleuves dans les grands bassins mondiaux, et plus particulièrement, du débit à l'exutoire. La section suivante présente les différentes sources de mesures de débits utilisées dans le cadre des expériences d'AD en estimation d'état. Cependant, certaines des données n'étant disponibles que sur le bassin de l'Amazonie, on concentrera les études sur ce bassin en particulier.

4.1 Comparaison du modèle ISBA-TRIP aux débits observés

4.1.1 Débits in situ

Les données de débits in situ utilisées dans le cadre de cette étude sont les mesures de débits réalisées et distribuées par l'ANA (Agencia Nacional das Aguas, *hidroweb.ana.gov.br*), l'agence nationale des eaux brésilienne. L'agence effectue et rassemble des relevés de cotes d'eau via des méthodes limnimétriques (voir Chapitre 1, Section 1.3.3.1) et transforme les cotes d'eau obtenues en débit via l'utilisation de courbes de tarage. Au total, les débits de 145 stations in situ sont disponibles sur une période commençant au début des années 60 jusqu'à aujourd'hui (voir la Figure 4.1). Les séries temporelles de débits in situ sont distribuées avec un pas de temps journalier mais peuvent présenter des trous de données sur des périodes plus ou moins longues.

Dans le cadre de l'AD, les débits in situ ne seront pas assimilés. Comme l'AD réelles nous permet de sortir du contexte des expériences jumelles, les débits in situ seront considérés comme la "vérité" et utilisés pour valider les résultats et quantifier les performances de la maquette.

4.1.2 Débits altimétriques

Les mesures de débits utilisées pour les expériences d'AD sont des données réelles issues de la télédétection. Il est important de rappeler qu'on ne sait pas mesurer directement le débit via l'observation satellite. La section suivante présente comment les cotes d'eau issues de l'altimètre nadir du satellite Envisat (mesurées au-dessus du bassin de l'Amazonie) ont été transformées en débits équivalents à partir de courbes de tarage construites à l'aide d'un modèle hydrologique (différent d'ISBA-TRIP).

L'ensemble des données altimétriques utilisées dans cette étude proviennent de l'altimètre RA-2 (Advanced Radar Altimeter) embarqué sur le satellite Envisat. Cette mission a été opérationnelle de Septembre 2002 à Octobre 2010 sur son orbite nominale, ayant une répétitivité de 35 jours et caractérisée par une distance entre les traces à l'équateur d'environ 80 km (voir Tableau 1.2 au Chapitre 1 pour plus d'informations sur le satellite Envisat). Les mesures de cotes d'eau dérivées des relevés Envisat sur le bassin de l'Amazonie ont été initialement générées par da Silva et al. (2010). L'algorithme de retracking Ice-1 (Bamber, 1994) a été appliqué aux formes d'ondes Envisat pour générer les cotes d'eau correspondantes ; Frappart et al. (2006) ayant préalablement démontré que cet algorithme était le plus efficace et donnait les meilleurs résultats sur les surfaces continentales. Le produit final de cotes d'eau est référencé par rapport au géoïde GGM02C (Tapley et al., 2004) et la précision verticale des mesures va de 12 cm à 30-40 cm (les stations présentant les plus mauvaises précisions verticales ont des erreurs de plusieurs mètres).

Pour transformer les cotes d'eau en débit, l'utilisation de courbes de tarage est requise, voir Chapitre 1, Section 1.3.3.1. Celles utilisées dans le cadre de cette thèse ont été construites et validées par Paris et al. (2016) à partir des cotes d'eau Envisat et Jason-2 (da Silva et al., 2010, 2012, 2014) et



Figure 4.1 – Carte du bassin de l’Amazon avec le réseau fluvial principal (chevelu bleu) et la localisation des 145 stations in situ disponibles (cercles jaunes).

des débits simulés par le modèle hydrologique-hydrodynamique MGB-IPH (*Model de Grandes Bacias-Instituto de Pesquisas Hidráulicas* ou Modèle de Grands Bassins-Institut de Recherches Hydrauliques ; Collischon et al., 2007). La version originale du modèle, développée sur le bassin de l’Amazon, consiste en un modèle hydrologique distribué grande échelle, couplé à un module hydrodynamique et un schéma simple de plaines d’inondations (de Paiva et al., 2011). Le bassin est découpé en un ensemble de plus 5763 mini-bassins élémentaires, ayant une superficie variant de 100 à 5000 km², voir la Figure 4.2a. Un schéma de surface - proche de celui d’ISBA - est appliqué dans chaque mini-bassin pour estimer les principaux écoulements et un réseau de routage est utilisé pour diriger les flux d’un mini-bassin à l’autre jusqu’à l’exutoire. Le routage des cours d’eau qui estime entre autres les débits est réalisé via la méthode de Muskingum-Cunge (MC) ou la résolution des équations de Saint-Venant (HD pour hydrodynamique). La méthode MC est utilisée dans les têtes de bassin et les tous petits affluents alors que la méthode HD est appliquée sur les cours d’eau principaux. Les données nécessaires à la représentation de l’hydrodynamique de la rivière proviennent du MNT SRTM (Farr et al., 2007) et les paramètres tels que la largeur et la profondeur sont calculés en fonction de l’aire drainée selon des relations géomorphologiques calibrées pour les bassins (de Paiva et al., 2013). De plus, la version de MGB-IPH utilisée pour la génération des courbes de tarage est celle développée par Paiva et al. (2013) où des débits in situ sont assimilés dans le modèle via un EnKF, sur la période de 1998 à 2010, pour corriger les débits simulés aussi bien sur les bassins jaugés que non-jaugés. En plus de proposer des meilleures estimations des débits, Paiva et al. (2013) fournit une estimation de l’incertitude sur le débit de chaque bassin sous la forme d’un bruit blanc.

Dans la suite, on notera $Q_{mgb,j}(t)$ le débit MGB-IPH dans le sous-bassin j à l’instant t , et $\sigma_{mgb,j}$ l’écart-type de l’incertitude sur $Q_{mgb,j}$ (indépendante du temps). Parallèlement, on notera $H_{alti}(t)$ les cotes d’eau Envisat au temps t et σ_{alti} l’écart-type de l’incertitude sur H_{alti} . À noter aussi que, contrairement à da Silva et al. (2010), le géoïde EGM2008 (Palvis et al., 2012) est utilisé comme référence des cotes d’eau. MGB-IPH sert donc de base de données de débits pour des courbes de tarage liant les débits aux cotes altimétriques sous une forme :

$$Q(t) = a \times (H(t) - z_0)^b, \quad (4.1)$$

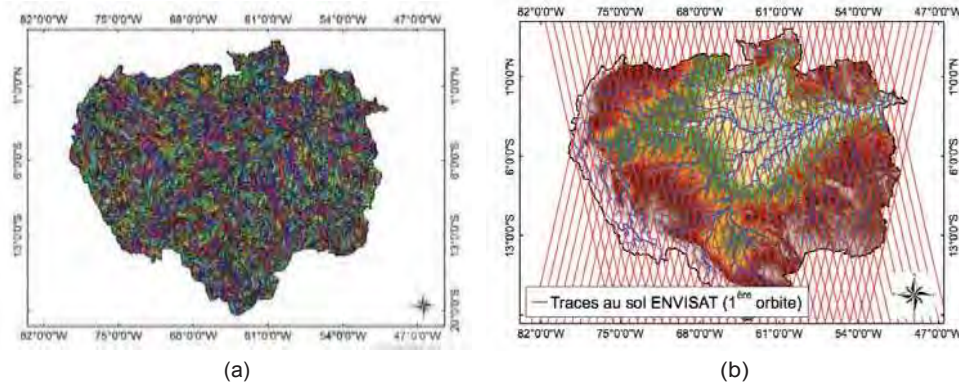


Figure 4.2 – (a) Découpe du bassin de l'Amazonie en 5765 mini-bassins (b) Traces au sol de l'altimètre nadir d'Envisat (lignes rouges) sur le bassin de l'Amazonie.

où Q sont les débits, a , b , z_0 sont les paramètres (constants dans le temps mais variables dans l'espace) de la courbe de tarage à estimer et H est la cote d'eau altimétrique. Cette formule peut être liée à l'équation de Manning (Manning, 1891) par :

$$Q(t) = n^{-1} A R^{2/3} s^{1/2},$$

où n est le coefficient de Manning, A est l'aire de la section, R est le rayon hydraulique dépendant de la hauteur d'eau et s est la pente. En ré-écrivant l'expression sous l'hypothèse que la largeur de la section W est bien plus grande que la hauteur d'eau h , il apparaît que

$$Q(t) = n^{-1} s^{1/2} W h^{5/3}, \text{ avec } h(t) = H(t) - z_0.$$

Finalement le paramètre a est fonction de n , W et s ; b est un indicateur de la géométrie du tronçon et z_0 est l'élévation du lit du tronçon pour convertir la cote d'eau en profondeur d'eau.

Dans le cadre de l'étude de Paris et al. (2016), les débits utilisés sont ceux fournis par le modèle MGB-IPH dans sa version développée par Paiva et al. (2013). Les débits sont journaliers et propres à chaque sous-bassin. Les cotes altimétriques proviennent des 767 stations virtuelles Envisat (une station virtuelle étant la zone d'intersection entre la trace du satellite et le fleuve) disponibles sur le bassin de l'Amazonie (Figure 4.3). Ces séries temporelles ont un pas de temps de 35 jours (période de répétitivité du satellite).



Figure 4.3 – Carte du bassin de l’Amazonie avec le réseau fluvial principal calculé par Stéphane Calmant du LEGOS à partir d’images *Google Earth* (chevelu bleu) et la localisation des 767 stations virtuelles Envisat disponibles (cercles bleus).

Afin d’appliquer la formule 4.1, les cotes d’eau Envisat H_{alti} sont couplées avec le débit provenant du sous-bassin j dans lequel se trouve la station virtuelle, donnant :

$$Q_{mgb,j}(t_{alti}) = a \times (H_{alti}(t_{alti}) - z_0)^b, \quad (4.2)$$

où t_{alti} sont les instants d’observation Envisat. Il est cependant possible que plusieurs stations virtuelles proches soient dans le même sous-bassin. Dans ce cas-là, les cotes d’eau sont associées au même débit MGB-IPH. Les paramètres a , b , z_0 sont uniques pour chaque station virtuelle et dépendent des caractéristiques physiques du tronçon comme évoqué précédemment.

Un algorithme d’optimisation global est utilisé pour établir chaque courbe de tarage. Cet algorithme cherche à minimiser l’erreur quadratique définie par :

$$\sum_{t_{alti}} Q_{mgb,j}(t_{alti}) - a(H_{alti}(t_{alti}) - z_0)^b)^2. \quad (4.3)$$

L’algorithme employé est le Shuffled Complex Evolution Metropolis développé par Vrugt et al. (2003). Pour tenir compte de l’incertitude dans les observations de débits et de cotes, les mesures sont préalablement perturbées en leur ajoutant une erreur gaussienne de moyenne nulle et d’écart-type leur incertitude respective :

$$\begin{aligned} \forall j, \forall t_{alti} \quad \tilde{Q}_{mgb,j}(t_{alti}) &= Q_{mgb,j}(t_{alti}) + E_Q, \quad E_Q \sim N(0, \sigma_{mgb,j}), \\ \forall j, \forall t_{alti} \quad \tilde{H}_{alti}(t_{alti}) &= H_{alti}(t_{alti}) + E_H, \quad E_H \sim N(0, \sigma_{alti}), \quad \sigma_{alti} = 0,35 \text{ cm}. \end{aligned}$$

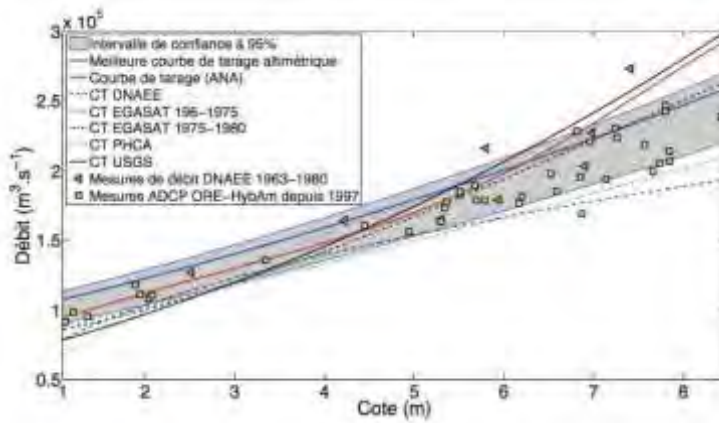


Figure 4.4 – Comparaison des différentes courbes de tarage existantes à la station d’Óbidos sur le cours principal de l’Amazone (Paris, 2015) relevés sur le terrain (marqueurs), courbe de tarage de l’ANA (rouge), courbes de tarage de relevés plus anciens (tracés noirs), meilleure courbe de tarage altimétrique (bleu) et son intervalle de confiance à 95% (zone grisée). Cette image est extraite de Paris (2015).

Les observations perturbées $\tilde{Q}$ et $\tilde{H}$ sont alors utilisées dans la minimisation de 4.3. Pour tenir compte de l’incertitude des paramètres a , b , z_0 lors de l’optimisation, l’algorithme optimise non pas un mais un ensemble de triplets (a, b, z_0) . Lors de l’initialisation, un ensemble de 100 triplets répartis sur 5 groupes aléatoires (appelés complexes dans Vrugt et al., 2003) est généré. Les paramètres suivent alors une loi uniforme sur leur intervalle de définition, c’est-à-dire

$$a \in [0, 1000] \quad b \in [A, 3] \quad z_0 \in [\min(H_{\text{alti}}) - 100, \min(H_{\text{alti}}) + 0,01].$$

L’optimisation se fait ensuite de manière itérative. À chaque itération, selon le principe des chaînes de Markov, un triplet par complexe (donc 5) est sélectionné pour être optimisé. L’algorithme s’arrête lorsque la convergence est considérée atteinte ou quand le nombre d’itération maximal est atteint. Au final, une densité de probabilité a posteriori est générée pour chaque paramètre permettant de construire la courbe de tarage avec une mesure de son incertitude, voir la Figure 4.4. La méthode a ainsi été appliquée aux 767 stations Envisat, disponibles sur le bassin de l’Amazone fournissant une importante source de données de débits "altimétriques" pour l’assimilation, même si certaines séries temporelles peuvent être ponctuellement incomplètes.

Les débits in situ et altimétriques que l’on a introduit dans cette section doivent alors servir de base de comparaison aux débits simulés par le modèle ISBA-TRIP. Or, si la répartition spatiale de ces différentes stations (in situ ou virtuelles) est complètement irrégulière sur le bassin, le modèle est, lui, défini sur une maille régulière à une échelle grossière de $0,5^\circ \times 0,5^\circ$. La section suivante présente la méthode appliquée pour associer une cellule TRIP à une station in situ ou à une station virtuelle et les performances du modèle ISBA-TRIP par rapport aux débits mesurés.

4.1.3 Performances d’ISBA-TRIP sur l’Amazone

Contrairement aux expériences jumelles du Chapitre 3, on va être amené à assimiler des données réelles dans le modèle ISBA-TRIP. Il nous faut alors, au préalable, évaluer les performances du modèle sans assimilation.

Lors des expériences d'AD, plus particulièrement pendant l'étape d'assimilation, on va devoir comparer les débits simulés par ISBA-TRIP aux débits Envisat. Après l'assimilation, les débits corrigés seront aussi comparés aux débits in situ. Cependant, pour pouvoir comparer les débits simulés aux débits observés, il faut pouvoir associer à une maille TRIP une mesure in situ et/ou une mesure Envisat qui soit cohérente hydrologiquement. Ce traitement préalable peut être compliqué étant donné la résolution grossière de TRIP et le fait qu'il ne simule qu'une rivière équivalente. On rappelle que le modèle ISBA-TRIP, utilisé dans le cadre de cette thèse, est défini sur une grille régulière de résolution $0,5^\circ \times 0,5^\circ$ (soit environ $50 \text{ km} \times 50 \text{ km}$). Au contraire, les stations in situ/virtuelles sont situées sur le vrai réseau hydrographique. De ce fait, les stations virtuelles et in situ ne peuvent pas être directement associées à la maille ISBA-TRIP dans laquelle elles se trouvent. La Figure 4.5 illustre la différence des deux réseaux en les superposant.

Ainsi, un traitement particulier a été appliqué aux stations Envisat et in situ pour les associer proprement à la meilleure cellule TRIP en fonction de leur position et du réseau de drainage. L'algorithme de base employé est le suivant :

1. Le réseau TRIP est comparé au réseau fluvial de l'Amazonie (produit par Stéphane Calmant sur GoogleEarth, voir le chevelu sur la Figure 4.5a) afin d'associer les cellules TRIP à un affluent bien déterminé du bassin.
2. On connaît au préalable le nom de l'affluent sur lequel chaque station (virtuelle ou in situ) est située. On associe chaque station à la cellule TRIP la plus proche sur le même affluent (qui peut ne pas être le pixel TRIP contenant la station, mais un pixel voisin en fonction du réseau fluvial réel et de celui de TRIP).

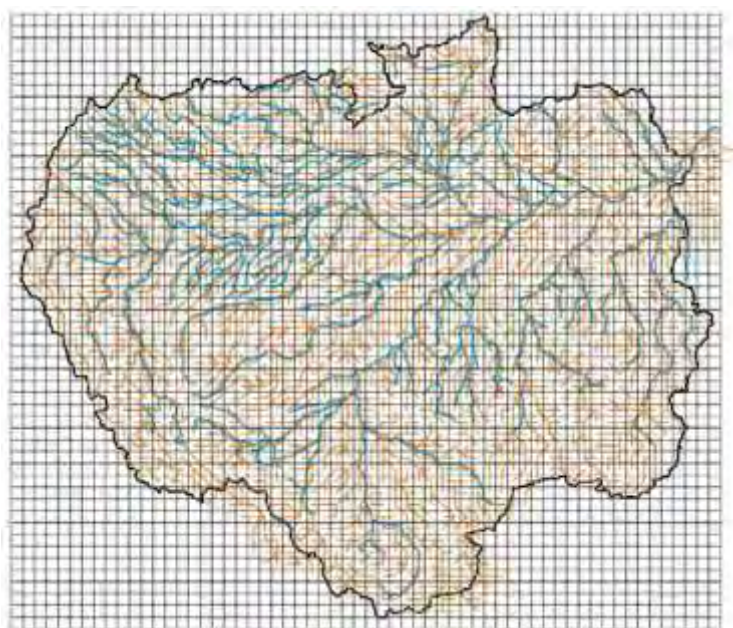
La méthode permet d'associer la majorité des stations à une unique cellule TRIP. Les cas particuliers suivants sont alors apparus :

- Certaines stations étaient associées à des affluents trop petits pour être vus par le réseau hydrologique de TRIP, voir par exemple le cas du Rio Novo (un petit affluent du Madeira) en Figure 4.6. Ces stations n'ont alors pas été conservées dans l'ensemble final des stations Envisat définies dans la maille TRIP.
- Certaines stations très proches ont été associées à la même cellule TRIP lors du traitement initial, voir par exemple sur le Rio Negro en Figure 4.7. Dans ce cas, la station virtuelle avec la plus grande aire drainée a été conservée aux dépens des autres.

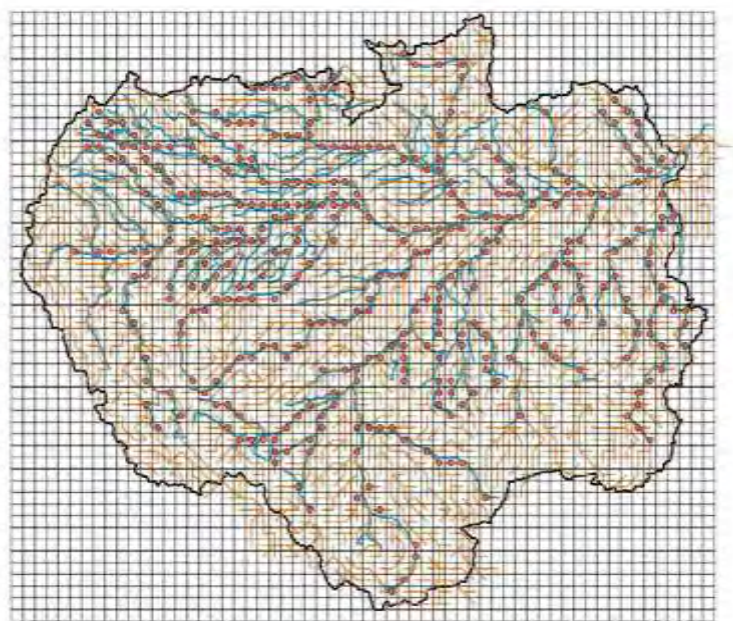
Au final, sur les 767 stations virtuelles Envisat initiales, 368 stations sont conservées et associées à une cellule TRIP, voir Figure 4.5b. Le même traitement a été appliqué aux stations in situ disponibles, 120 stations sur les 145 initiales ont été conservées. La Figure 4.8 représente ces stations in situ à leur position géolocalisée (coordonnées longitude/latitude) avec leur équivalent au centre d'une cellule TRIP.

Maintenant que les stations virtuelles/in situ ont été associées à une cellule TRIP, on est en mesure de comparer ces différentes observations de débits avec les débits simulés par ISBA-TRIP sans aucune étape d'assimilation (simulation que l'on appellera le *run libre*). Pour commencer, un échantillon de 15 stations in situ, réparties sur l'ensemble du bassin (principaux affluents et le cours principal) a été sélectionné. Les Figures (4.9-4.13) compare alors les débits du run libre ISBA-TRIP avec les débits Envisat et in situ aux 15 stations pendant un an de simulation à partir du 25 Septembre 2002 :

- Mises à part à Tamashiyacu (station a, en Figure 4.9), Itaituba (station n, en Figure 4.13) et Uhe Belo Monte (station o, en Figure 4.13), les débits Envisat sont en accord avec les débits in situ. Ces résultats sont le résultat de l'assimilation des débits in situ directement dans MGB-IPH pour corriger les débits (Paiva et al., 2013) et l'utilisation de ce même modèle pour estimer les courbes de tarage utilisées pour les cotes altimétriques (Paiva et al., 2013). Les



(a)



(b)

Figure 4.5 – (a) Superposition du réseau fluvial calculé par Stéphane Calmant du LEGOS (chevelu bleu) et du réseau hydrologique TRIP (lignes oranges). (b) Représentation des 368 stations virtuelles Envisat à la résolution du réseau ISBA-TRIP (cercles rouges).

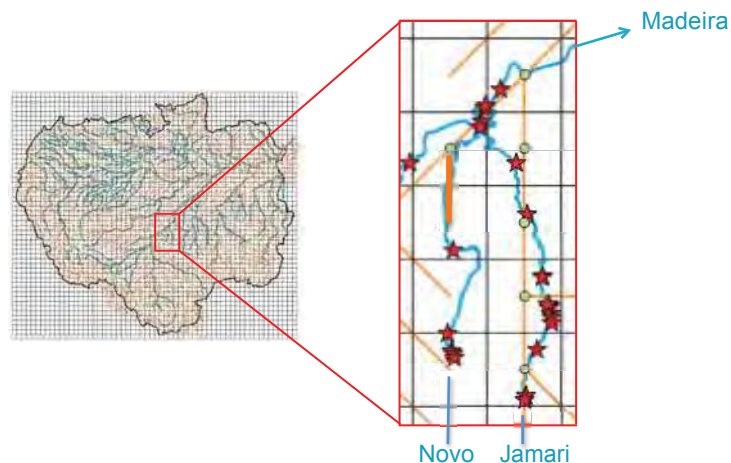


Figure 4.6 – Association des stations virtuelles Envisat avec les cellules du réseau TRIP (traits oranges) : illustration du cas, ici sur le Rio Novo un petit affluent du Madeira, où les petits cours d’eau ne sont pas visibles à la résolution de TRIP. Les 5 stations virtuelles (étoiles rouges) sur cet affluent sont ignorées et donc absentes des stations virtuelles vues par TRIP (cercles verts).

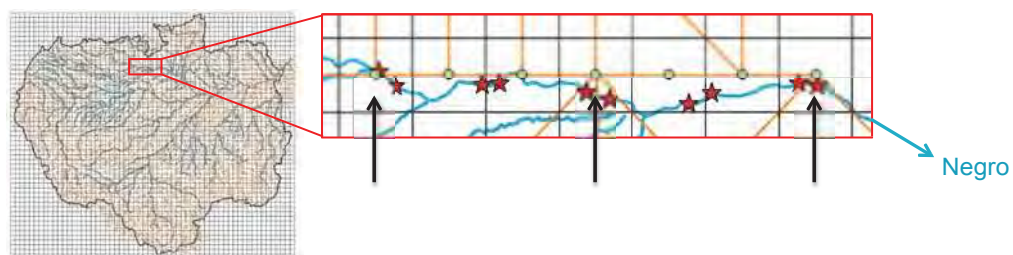


Figure 4.7 – Association des stations virtuelles Envisat avec les cellules du réseau TRIP (traits oranges) : Illustration du cas, ici sur le Rio Negro, où plusieurs stations virtuelles sont associées à la même cellule TRIP : cellules signalées par une flèche noire.

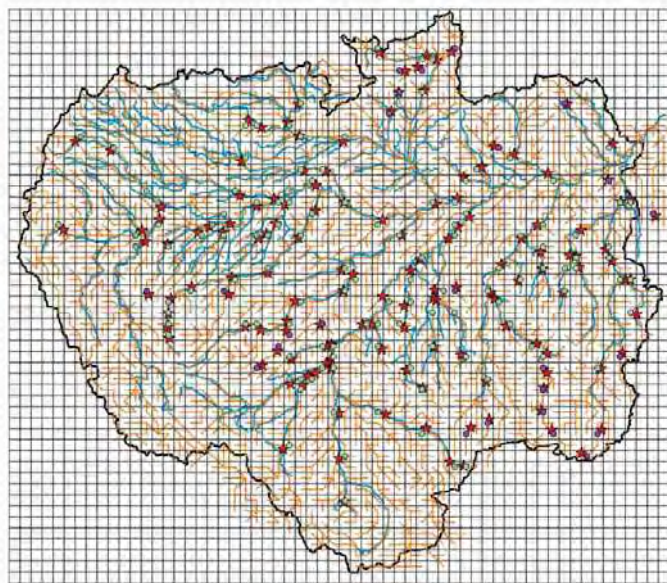


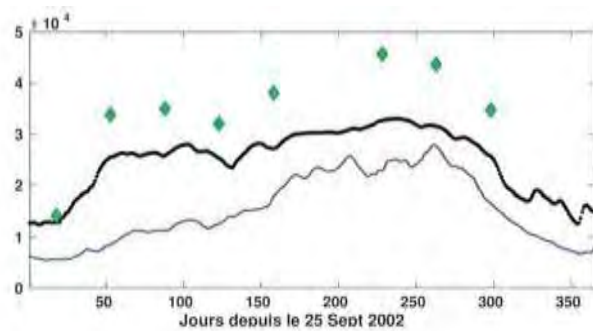
Figure 4.8 – Représentation des stations in situ sur le bassin de l’Amazone : les 145 stations in situ initiales à leur localisation réelle (étoiles rouges), stations in situ à la résolution TRIP (cercles verts), stations in situ dont on ne connaissait pas le nom de l’affluent associé et donc couplées à la plus proche cellule TRIP (cercles violets).

fortes disparités entre ces deux sources de débits signifient probablement que l’optimisation des paramètres de la courbe de tarage n’a pas fonctionné correctement ou que les données initiales (débits MGB-IPH ou cotes Envisat) ont de fortes erreurs. De plus, il faut noter que seuls les débits mensuels sont disponibles à Uhe Belo Monte (Figure 4.13o).

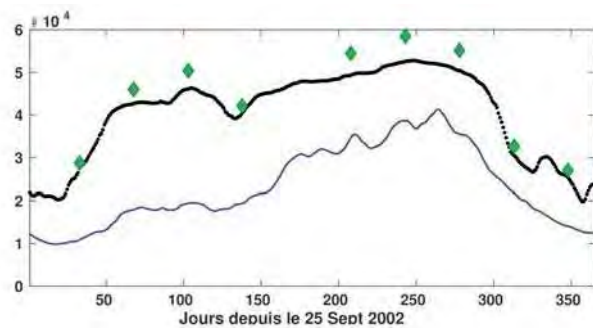
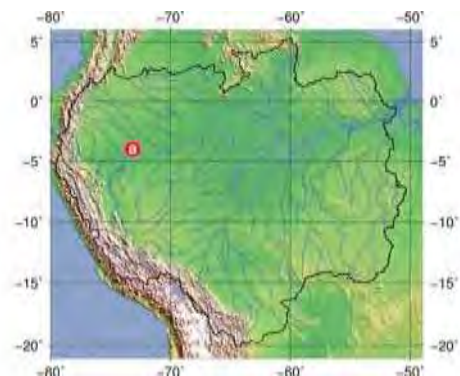
- Lors de la présentation des débits in situ, il a été énoncé que les débits in situ ne seront pas assimilés mais seront utilisés pour évaluer les performances de l’assimilation de données, ce qui équivaut, en quelques sortes, à considérer les débits in situ la meilleure estimation disponible. Il est alors important d’être conscient que ces débits, tout comme les débits Envisat et ISBA-TRIP, ne sont pas parfaits et qu’il peut y avoir aussi des erreurs lors de l’estimation des courbes de tarage in situ. Ils sont aussi sujets à des incertitudes et peuvent avoir des valeurs aberrantes, comme par exemple à Itaituba (Figure 4.13n) où la courbe de débit est discontinue et les débits s’annulent. Cet aspect devra être pris en compte lors de l’analyse des résultats.
- Enfin, les débits simulés par ISBA-TRIP présentent de fortes différences avec les débits observés. Aux stations en aval du cours principal, de Itapéua à Óbidos (Figure 4.9c-4.10f), le modèle n’est pas capable de reproduire les inondations observées au niveau des stations in situ et virtuelles. Dans le run libre, le pic de débit arrive plus tôt dans l’année et surtout décroît très vite, alors que le comportement des débits observés est beaucoup plus lent. Le pic de débit est plus tardif et surtout, les phénomènes de montée et descente de débit sont beaucoup plus longs.

Ainsi, l’assimilation des débits Envisat apparaît comme une opportunité intéressante pour corriger les débits simulés par le modèle en run libre.

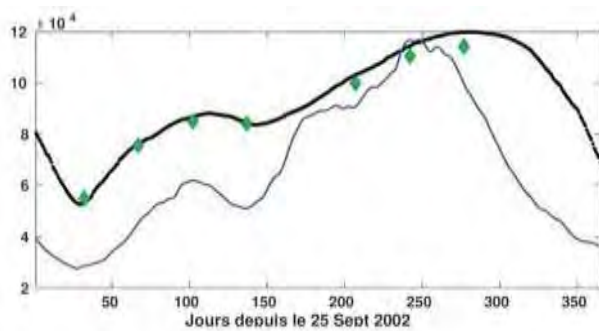
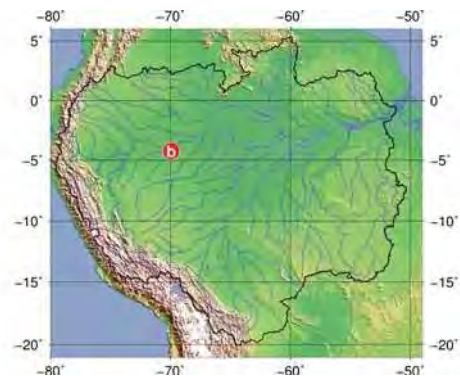
Dans les expériences d’assimilation qui vont suivre, il sera nécessaire de comparer les différents



(a) Tamishiyacu



(b) Tabatinga



(c) Itapéua

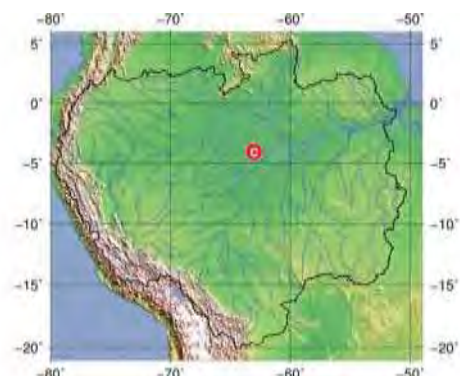
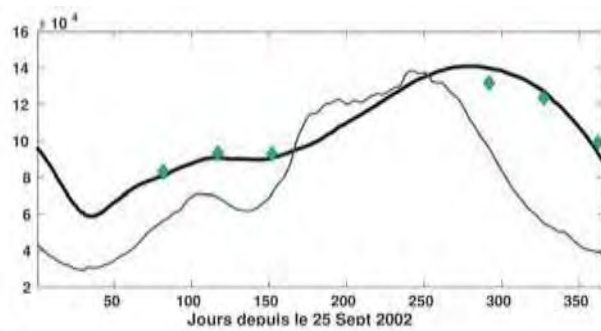
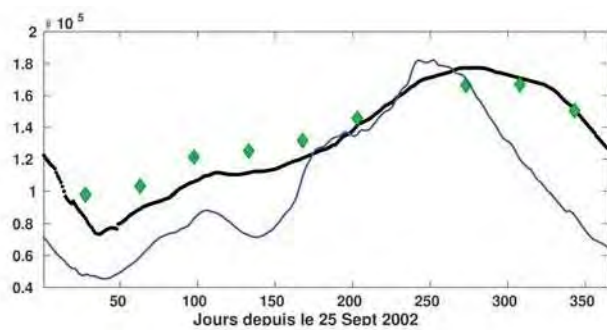
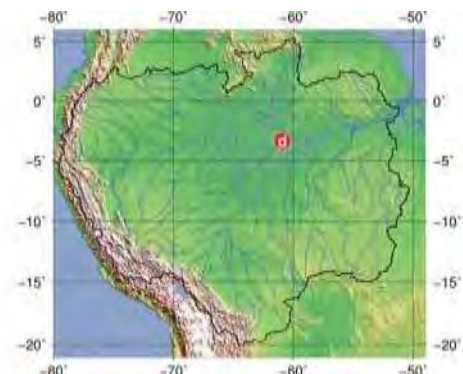


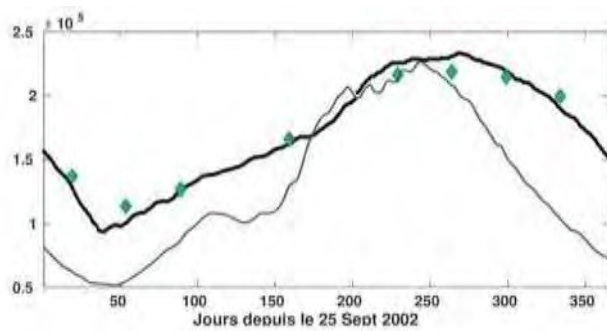
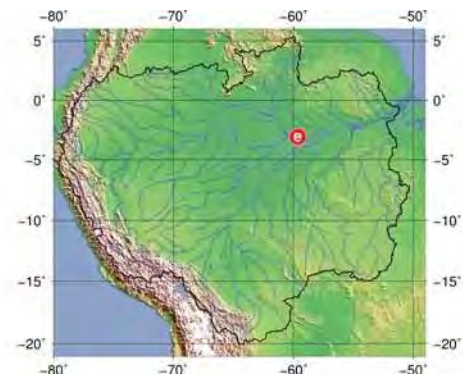
Figure 4.9 – Comparaison des débits simulés par ISBA-TRIP en run libre (ligne bleue) avec les débits observés d'Envisat (losanges verts) et les débits in situ (points noirs) sur l'année 2002. Pour chaque panneau, l'axe des abscisses correspond au temps (avec un marqueur en début de chaque mois) et l'axe des ordonnées correspond au débit (en $\text{m}^3 \cdot \text{s}^{-1}$).



(d) Manacapuru



(e) Jatuarana



(f) Óbidos

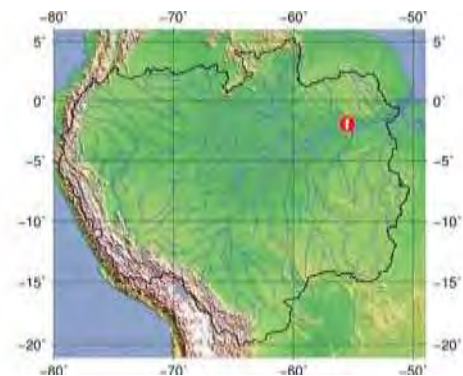
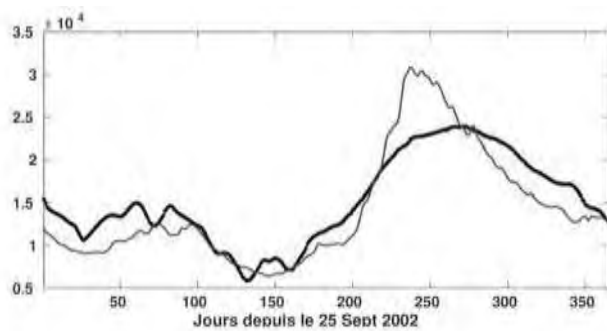
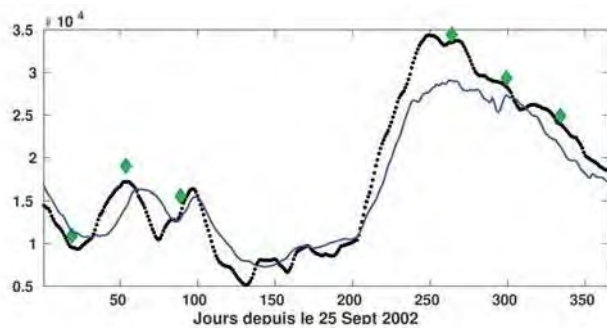
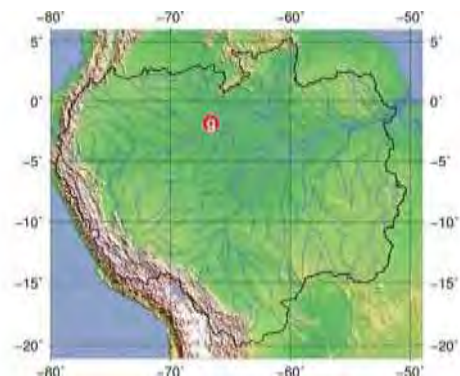


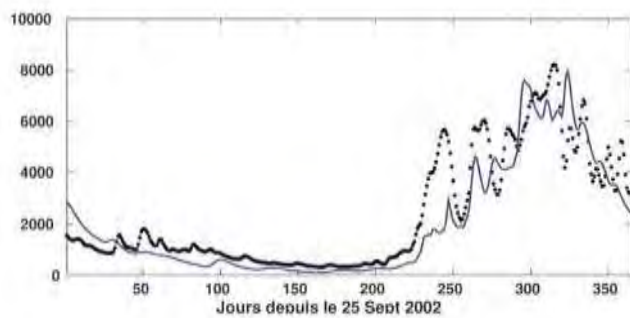
Figure 4.10 – Comparaison des débits simulés par ISBA-TRIP en run libre (ligne bleue) avec les débits observés d'Envisat (losanges verts) et les débits in situ (points noirs) sur l'année 2002. Pour chaque panneau, l'axe des abscisses correspond au temps (avec un marqueur en début de chaque mois) et l'axe des ordonnées correspond au débit (en $\text{m}^3.\text{s}^{-1}$) (Suite).



(g) Acanai



(h) Serrinha



(i) Caracarai

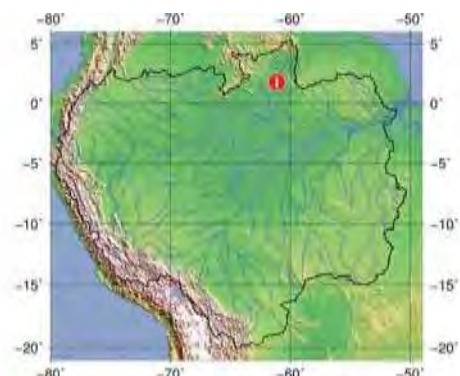
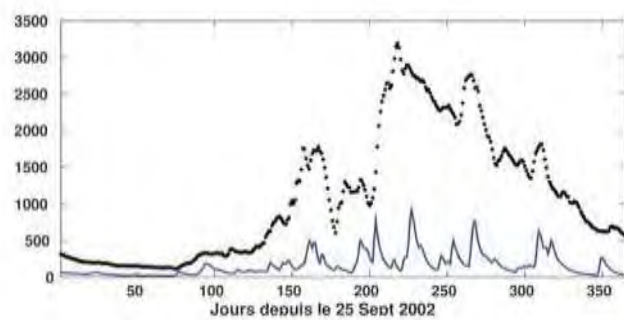
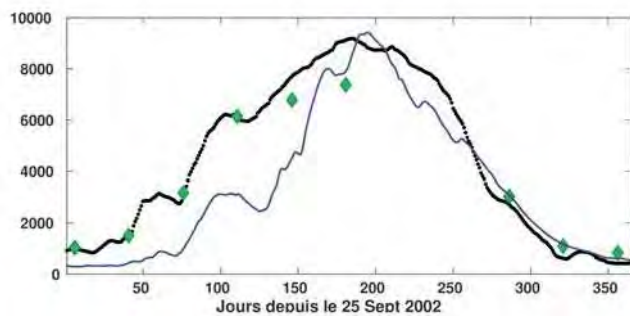
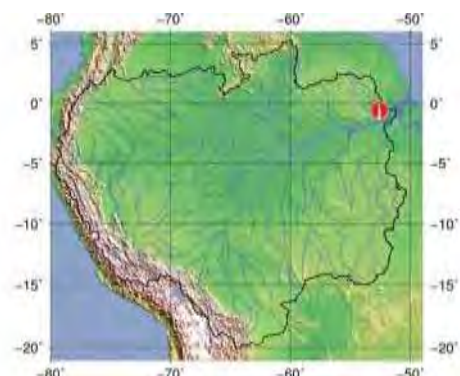


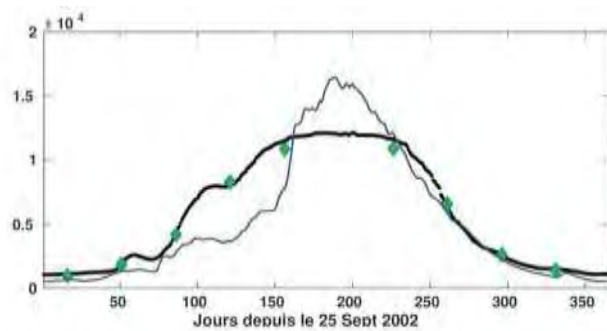
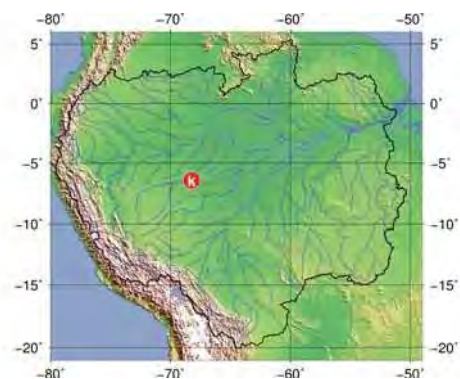
Figure 4.11 – Comparaison des débits simulés par ISBA-TRIP en run libre (ligne bleue) avec les débits observés d'Envisat (losanges verts) et les débits in situ (points noirs) sur l'année 2002. Pour chaque panneau, l'axe des abscisses correspond au temps (avec un marqueur en début de chaque mois) et l'axe des ordonnées correspond au débit (en $\text{m}^3.\text{s}^{-1}$) (Suite).



(j) São Francisco



(k) Santos Dumont



(l) Lábrea

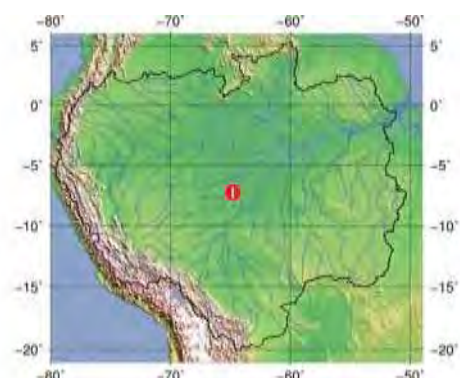
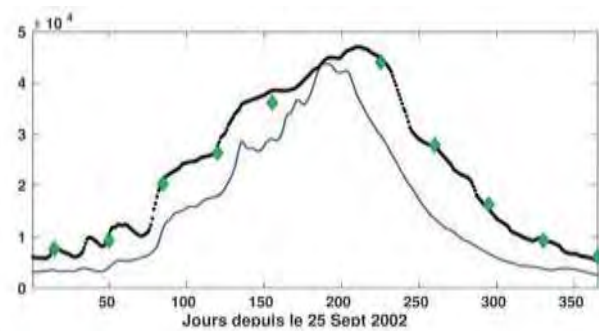
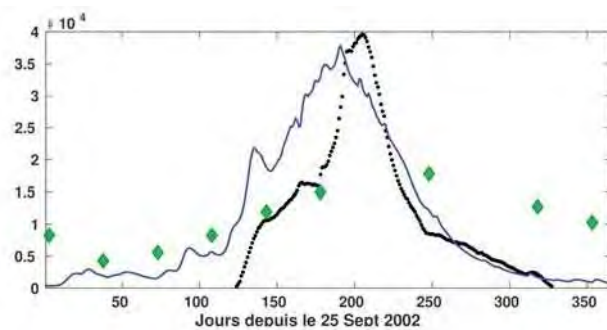
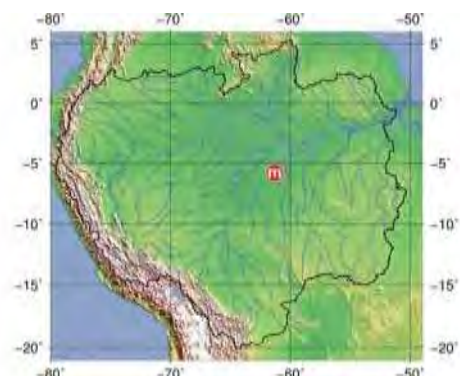


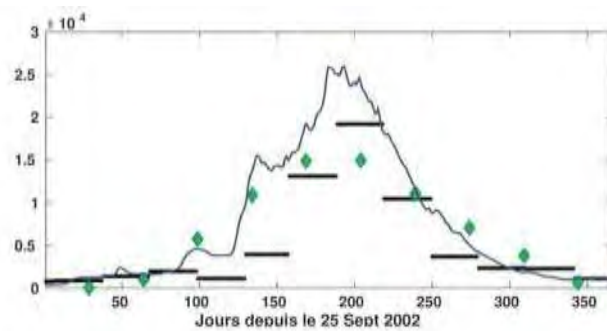
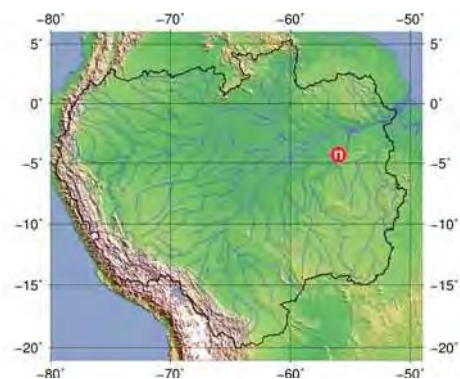
Figure 4.12 – Comparaison des débits simulés par ISBA-TRIP en run libre (ligne bleue) avec les débits observés d'Envisat (losanges verts) et les débits in situ (points noirs) sur l'année 2002. Pour chaque panneau, l'axe des abscisses correspond au temps (avec un marqueur en début de chaque mois) et l'axe des ordonnées correspond au débit (en $\text{m}^3.\text{s}^{-1}$) (Suite).



(m) Manicoré



(n) Itaituba



(o) Uhe Belo Monte

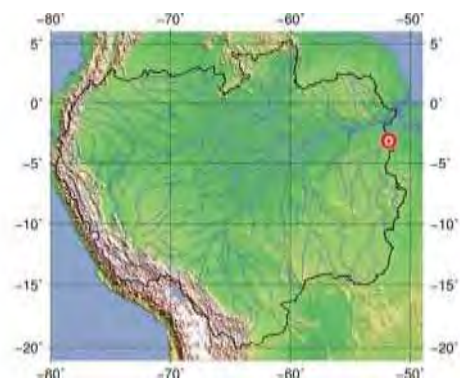


Figure 4.13 – Comparaison des débits simulés par ISBA-TRIP en run libre (ligne bleue) avec les débits observés d'Envisat (losanges verts) et les débits in situ (points noirs) sur l'année 2002. Pour chaque panneau, l'axe des abscisses correspond au temps (avec un marqueur en début de chaque mois) et l'axe des ordonnées correspond au débit (en $\text{m}^3.\text{s}^{-1}$) (Fin).

débits simulés aux débits observés aux stations virtuelles Envisat pour estimer les performances de l'expérience d'assimilation. La qualité de l'assimilation est évaluée cellule par cellule en comparant un débit simulé Q_i - l'indice i est l'indice de cellule - avec un débit observé à la même cellule - débit in situ Q_i^{situ} ou débit Envisat Q_i^o (l'exposant o pour *observé* est utilisé pour désigner les débits Envisat qui sont effectivement les données observées dans le cadre de l'AD). L'exposant k correspondra à la variable de temps. La comparaison est réalisée par le biais de coefficients de performances qui sont :

- L'erreur quadratique moyenne RMSE (*Root Mean Square Error* en anglais) définie par :

$$RMSE_i^* = \frac{\sqrt{\sum_{k=1}^{K^*} (Q_{i,k} - Q_{i,k}^*)^2}}{K^*}, \quad (4.4)$$

- L'erreur quadratique moyenne normalisée RMSEn définie par :

$$RMSEn_i^* = 100 \times \frac{RMSE_i^*}{\overline{Q_{i,\bullet}^*}}, \quad (4.5)$$

- Le coefficient d'efficacité de Nash-Sutcliffe E défini par :

$$E_i^* = 1 - \frac{\sum_{k=1}^{K^*} (Q_{i,k} - \hat{Q}_{i,k})^2}{\sum_{k=1}^{K^*} (Q_{i,k} - \overline{Q_{i,\bullet}^*})^2}. \quad (4.6)$$

Dans les équations (4.4-4.6), l'exposant $*$ peut représenter l'exposant "situ" ou l'exposant "o", pour indiquer que les statistiques sont calculées avec, comme référence, les débits in situ ou les débits Envisat, respectivement. De plus, on rappelle que les débits in situ sont normalement disponibles à un pas de temps journalier mais les séries temporelles peuvent être discontinues sur certaines périodes et les débits Envisat sont disponibles à une fréquence de 35 jours. La constante K^* est ainsi le nombre total de pas de temps auxquels le débit Q^* est disponible sur la période totale d'étude. Le débit simulé utilisé peut être soit le débit simulé lors du run libre Q_i^f (" f " pour *forecast*) ou le débit analyse moyenné sur les membres de l'ensemble obtenu après assimilation $\overline{Q_{i,k}^{a,[.]}}$ (" a " pour *analyse*). Enfin, $\overline{Q_{i,\bullet}^*}$ correspond à la moyenne de $Q_{i,k}^*$ sur tous les pas de temps k de la période d'étude. Les statistiques pour le run libre sont illustrées en Figure 4.14-4.15.

Les cartes présentées en Figure 4.14 correspondent aux RMSEn pour le run libre. Elles montrent des erreurs importantes en tête de bassin de la majorité des affluents, ainsi que des erreurs élevées au niveau des cellules de confluences. Les principaux affluents du cours principal, en particulier le Negro, sont généralement bien modélisés hors zones de confluences et zones très amonts. Les cartes présentées en Figure 4.15 pour le coefficient de Nash-Sutcliffe (E), concordent avec ces observations. Les Nash du cours principal et du sous-bassin des rivières Japurá et Putumayo apparaissent néanmoins tous de mauvaise qualité. La statistique globale sur toutes les stations in situ donne un RMSE moyen de $2,7898 \times 10^3 \text{ m}^3 \cdot \text{s}^{-1}$, soit un RMSEn de 73,61% et un coefficient de Nash moyen de -11,57. La statistique globale sur les stations Envisat donne un RMSE moyen de $5,7075 \times 10^3 \text{ m}^3 \cdot \text{s}^{-1}$, soit un RMSEn de 75,29% et un coefficient de Nash moyen de -134,32. Ces erreurs proviennent probablement des forçages (les précipitations notamment), de l'absence d'inondations (mauvaise calibration du paramètre H_c , voir Chapitre 2, Section 2.1.4) et des autres paramètres des modèles ISBA et TRIP.

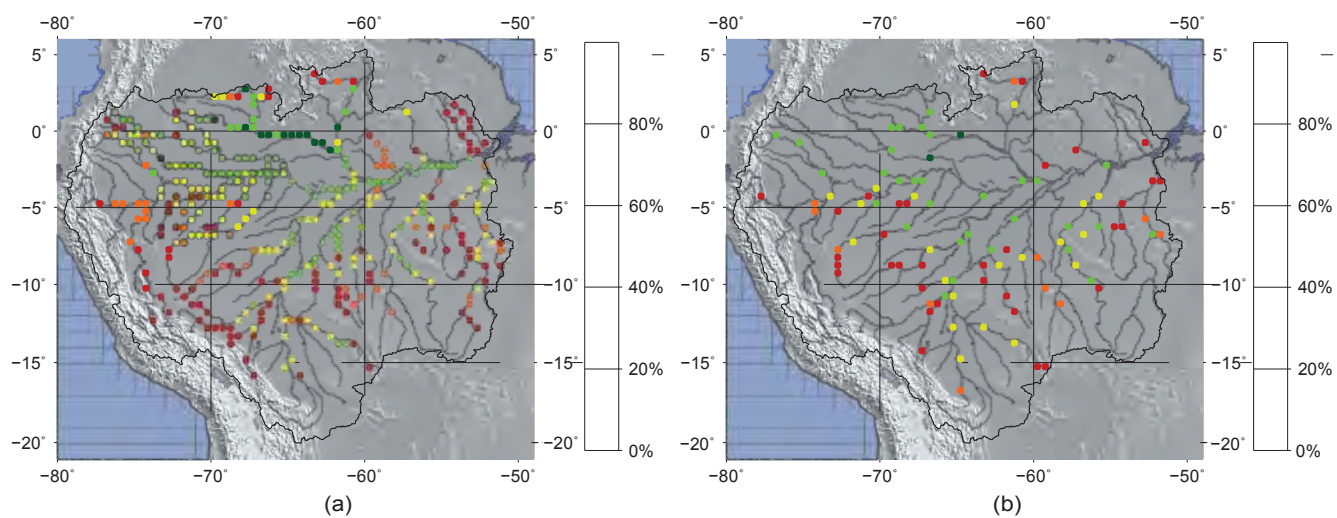


Figure 4.14 – Coefficients d’erreur quadratique moyenne normalisée RMSEn de la simulation run libre par rapport aux débits Envisat (a) et par rapport aux débits in situ (b).

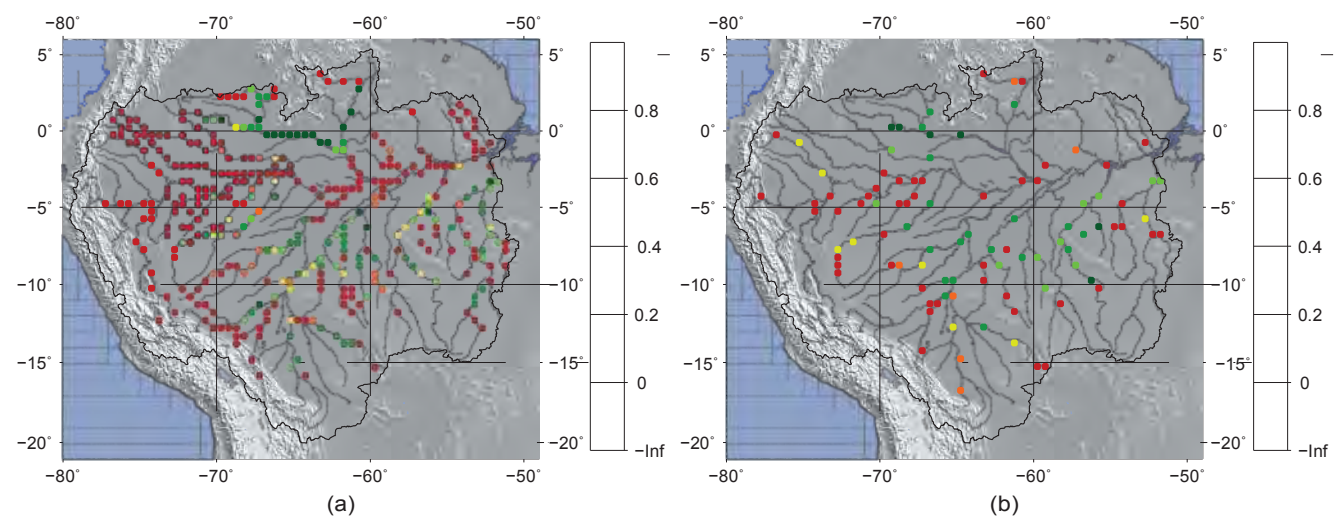


Figure 4.15 – Coefficients de Nash-Sutcliffe E de la simulation run libre par rapport aux débits Envisat (a) et par rapport aux débits in situ (b).

4.2 Méthode d'assimilation des débits ENVISAT

4.2.1 Définition des variables d'observations

Les données d'observations utilisées dans le cadre de l'estimation d'état sont les mesures de débits des stations virtuelles dérivées des cotes d'eau ENVISAT et préalablement associées à une cellule de la maille TRIP. La répétitivité d'ENVISAT est de 35 jours. Ainsi, une même station ne fournira une observation que tous les 35 jours. Malgré tout, l'exploitation simultanée de toutes les stations virtuelles fait que des observations différentes sont disponibles tous les jours quelque part sur le bassin et l'assimilation des observations jour par jour est possible. Par ailleurs, il s'agit ici des toutes premières expériences d'AD d'estimation d'état construites autour du modèle ISBA-TRIP. Pour mener ces premières expériences, on a choisi de se baser sur un Filtre de Kalman d'Ensemble séquentiel classique avec des étapes d'assimilation journalières (donc des cycles d'assimilation d'1 jour). Le vecteur d'observation $\mathbf{y}_k^o$ au cycle k (donc au jour de simulation k) sera composé des $n_{y,k}$ mesures de débits disponibles au jour k :

$$\mathbf{y}_k^o = y_{k,1}^o, y_{k,2}^o, \dots, y_{k,n_{y,k}}^o, \quad (4.7)$$

où $y_{k,j}^o, j = 1 \dots n_{y,k}$ est la j -ème observation parmi les $n_{y,k}$ au cycle k .

Les erreurs d'observations $\mathbf{E}_k^m$ proviennent des erreurs des cotes d'eau ENVISAT, des erreurs sur les débits MGB et des incertitudes dans l'estimation des paramètres des courbes de tarage qui sont utilisées pour transformer les côtes d'eau en débits. Sorooshian and Dracup (1980); Clark et al. (2008) ont fait remarquer que la concavité de la relation hauteur-débit fait que plus les eaux sont hautes, plus les erreurs sur les débits sont importantes. Par ailleurs, l'erreur de représentativité $\mathbf{E}_k^r$ induite par l'association des stations virtuelles à une maille du réseau hydrologique TRIP est négligée dans le reste de l'étude. De ce fait, l'écart-type de l'erreur $\sigma_{k,j}^o$, associée à la j -ème observation au cycle k , est définie en fonction de la mesure de débit $y_{k,j}^o$, i.e :

$$\sigma_{k,j}^o = \eta^o \times y_{k,j}^o, j = 1 \dots n_{y,k} \quad (4.8)$$

où η^o est une constante telle que $\eta^o \in]0, 1[$. L'écart-type de l'erreur d'observation est alors un certain pourcentage de la valeur de débit mesurée. Selon les études et les données utilisées, $\sigma_{k,j}^o$ varie de 0, 1 (Clark et al., 2008) à 0, 3 (Durand et al., 2016). On utilisera alors la valeur médiane de 0, 2 pour nos expériences, voir Tableau 4.1.

De plus, au cours d'un même cycle d'assimilation et entre les différents cycles, les observations sont considérées indépendantes et non-corrélées. La matrice de covariance des erreurs d'observations $\mathbf{R}_k$ est alors une matrice carrée diagonale définie positive telle que :

$$\mathbf{R}_k = \begin{bmatrix} (\sigma_{k,1}^o)^2 & 0 & \dots & 0 \\ 0 & (\sigma_{k,2}^o)^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & (\sigma_{k,n_{y,k}}^o)^2 \end{bmatrix} \quad (4.9)$$

Maintenant que les observations sont définies, il faut choisir une variable de contrôle à corriger via l'AD et définir comment projeter le vecteur de contrôle dans l'espace des observations, c'est-à-dire comment construire l'opérateur d'observation (défini au Chapitre 1, Section 1.4.2.3).

4.2.2 Stratégies d'assimilation de données et choix du vecteur de contrôle

Contrairement aux modèles hydrodynamiques, qui résolvent directement les équations de Saint-Venant et où le débit est une variable pronostique du modèle (voir la définition au Chapitre 1, Section 1.4.2.1, Eq. 1.1), les modèles hydrologiques comme ISBA-TRIP résolvent des équations d'évolution de stocks d'eau dans les différents réservoirs modélisés : pour TRIP, on trouve le réservoir de surface, le réservoir des plaines d'inondations et celui des eaux souterraines. Pour TRIP, les débits et les hauteurs d'eau sont des variables diagnostiques du modèle (voir la définition au Chapitre 1, Section 1.4.2.1, Eq. 1.2) dépendant de l'évolution du stock d'eau dans le réservoir de surface S . Pour rappel, l'équation différentielle d'évolution du stock d'eau de surface, qui est résolue pour chaque cellule, s'écrit :

$$\frac{dS(t)}{dt} = Q_{in,ISBA}^S + Q_{in,TRIP}^S + Q_{out}^G + Q_{out}^F - Q_{in}^F - Q_{out}^S, \quad (4.10)$$

avec $Q_{in,ISBA}^S$ le ruissellement de surface ISBA dans la cellule, $Q_{in,TRIP}^S$ le débit entrant dans la cellule via le réseau hydrologique TRIP, Q_{out}^S le débit sortant de la cellule, Q_{out}^G le débit entrant dans la cellule via le réservoir d'eau souterraine G de la cellule, Q_{in}^F/Q_{out}^F le débit entrant/sortant dans la cellule via le réservoir de plaines d'inondations F de la cellule. Le débit associé au débit ENVISAT est alors le débit de sortie du réservoir Q_{out}^S défini par :

$$Q_{out}^S = L^{-1} v S \text{ [kg.s}^{-1}\text{]}, \quad (4.11)$$

avec L [m] la longueur du tronçon, v la vitesse d'écoulement - estimée par l'équation de Manning - et S le stock d'eau.

Pour lancer une intégration d'ISBA-TRIP - c'est-à-dire la résolution numérique des équations différentielles définissant l'évolution temporelle du système sur un intervalle de temps fini, seuls sont nécessaires les forçages atmosphériques sur le temps d'intégration et surtout les stocks initiaux d'eau de chacun des réservoirs dont celui du réservoir de surface S_{ini} . Une fois l'intégration du modèle terminée, les variables pronostiques disponibles sont les stocks d'eau finaux de chaque réservoir, dont celui du réservoir de surface S_{end} , ainsi que le diagnostique du débit dans chaque cellule Q .

Ainsi, trois types de variables possibles peuvent être considérés comme variables de contrôle du problème d'AD, chacun ayant leurs avantages et leurs inconvénients : le débit Q , le stock final S_{end} ou le stock initial S_{ini} . La définition et la complexité de l'opérateur d'observation H_k dépend alors de la nature des variables de contrôle.

4.2.2.1 Option 1 : Assimilation des débits ENVISAT pour corriger... les débits ISBA-TRIP

Pour cette première option, les variables de contrôle, rassemblées dans le vecteur $\mathbf{x}_k$ sont les $n_x = 2028$ (nombre de cellules TRIP formant le bassin de l'Amazonie) débits estimés à chaque cellule au cycle d'assimilation k .

L'opérateur d'observation H_k se résume à un simple opérateur de sélection S_k qui sélectionne les cellules TRIP observées au cycle d'assimilation courant :

$$H_k = S_k. \quad (4.12)$$

Cet opérateur d'observation est direct et surtout complètement linéaire.

La difficulté associée à ce choix de variables de contrôle est, qu'une fois l'étape d'analyse réalisée, il faut nécessairement transformer le débit analyse Q_k^a en stock final analyse $S_{k,end}^a$ équivalents.

En effet, le stock final $S_{k,\text{end}}^a$ sert de condition initiale pour l'intégration d'ISBA-TRIP au cycle d'assimilation suivant ($S_{k,\text{end}}^a = S_{k+1,\text{ini}}^b$). En reprenant les notations du Chapitre 1, il faut alors trouver une expression explicite de l'opérateur Z^{-1} .

Une solution consiste à inverser l'Eq. 4.11. On suppose alors que le débit de sortie Q_k^a d'ISBA-TRIP est le débit instantané à l'instant final de la fenêtre d'intégration du modèle :

$$Q_{k,[\text{kg.s}^{-1}]}^a = L^{-1} v S_{k,\text{end},[\text{kg}]} \iff Q_{k,[\text{m}^3.\text{s}^{-1}]}^a = \rho^{-1} L^{-1} v S_{k,\text{end},[\text{kg}]}.$$

Pour faciliter les écritures, on omettra d'indiquer les unités dans les équations suivantes sachant que les débits sont en $\text{m}^3.\text{s}^{-1}$ et les stocks en kg.

$$Q_k^a = L^{-1} \rho^{-1} s^{1/2} n^{-1} \frac{Wh}{W+2h}^{2/3} S_k,$$

On suppose : $W \gg h$

$$Q_k^a \approx L^{-1} \rho^{-1} s^{1/2} n^{-1} h^{2/3} S_{k,\text{end},[\text{kg}]},$$

Or $S = \rho L W h$, soit $h = \frac{S}{\rho L W}$ et en remplaçant dans l'équation ci-dessus, on obtient :

$$Q_k^a \approx L^{-5/3} \rho^{-5/3} W^{2/3} s^{1/2} n^{-1} S_k^{5/3},$$

$$S_k^a \approx \rho L W^{2/5} s^{-3/10} n^{3/5} (Q_k^a)^{3/5} \quad (4.13)$$

avec ρ la masse volumique de l'eau [$\text{m}^3.\text{kg}^{-1}$], L la longueur du tronçon dans la cellule [m], W la largeur de la rivière équivalente dans la cellule [m], s la pente de fond de rivière [m.m^{-1}] et n le coefficient de Manning du lit de la rivière [-]. Les incertitudes de l'Eq. 4.13 résident dans deux hypothèses :

- D'abord, le débit ISBA-TRIP se résume au débit instantané au dernier pas de temps de la fenêtre d'intégration du modèle.
- Ensuite, on fait aussi l'hypothèse que W , la largeur de la rivière, est bien plus grande que h la hauteur d'eau dans la rivière.

Ces incertitudes peuvent être regroupées en *incertitude de formule*.

4.2.2.2 Option 2 : Assimilation des débits ENVISAT pour corriger... les stocks finaux ISBA-TRIP

Pour cette deuxième option, les variables de contrôle sont les $n_x = 2028$ (nombre de cellules TRIP formant le bassin de l'Amazonie) stocks d'eau finaux dans le réservoir de surface estimés à chaque cellule au cycle d'assimilation k .

Le coût en calcul numérique est le même que pour l'option 1. Dans ce cas de figure, l'opérateur d'observation s'écrit comme

$$H_k = S_k \circ Z_k. \quad (4.14)$$

Bien que H_k soit maintenant non-linéaire, cette option a l'avantage de directement corriger le stock final des réservoirs $S_{k,\text{end}}^a$ pour le cycle d'assimilation suivant. Ainsi, il n'y a pas d'incertitude supplémentaire introduite par un traitement post-analyse nécessaire pour poursuivre l'assimilation. En revanche, le débit analyse Q_k^a correspondant au stock $S_{k,\text{end}}^a$ est maintenant inconnu.

Une expression explicite de H_k , et plus particulièrement Z_k , est nécessaire pour calculer Q_k^a et être capable de faire des statistiques comparatives avec les débits ENVISAT pour évaluer les performances

de l'AD. En effet, on rappelle que dans le cadre d'expériences réelles (en opposition aux expériences jumelles), le débit et le stock vrai ne sont pas connus. Pour cela, une solution possible réside dans l'utilisation des expressions de Z_k déduites de l'Eq. 4.13. Cependant, les incertitudes de formule et de flux seront introduites dans la valeur des débits analysés et seront à prendre en compte dans le calcul des performances de la maquette d'assimilation.

4.2.2.3 Option 3 : Assimilation des débits ENVISAT pour corriger... les stock initiaux ISBA-TRIP

Pour cette dernière option, les variables de contrôle sont les $n_x = 2028$ (nombre de cellules TRIP formant le bassin de l'Amazonie) stocks d'eau initiaux dans le réservoir de surface estimés à chaque cellule au cycle d'assimilation k .

On utilise les observations de débits pour corriger le stock d'eau dans le réservoir de surface au temps précédent celui de l'observation ou, en d'autres termes, au temps initial de l'intervalle sur lequel le modèle est intégré (instant k_1 pour l'opérateur $M_{[k-1,k]}$). On doit alors utiliser le modèle pour propager le stock initial au temps $(k-1)$ pour le convertir en débit au temps k . L'opérateur d'observation s'écrit comme la composition du modèle $M_{[k-1,k]}$ avec l'opérateur défini en Eq. 4.15 :

$$H_k = S_k \circ D_k \circ M_{[k-1,k]}. \quad (4.15)$$

L'opérateur d'observation est le plus complexe (et aussi "le plus non-linéaire") des trois options présentées et le lien entre les variables de contrôle et les variables observées est aussi le plus indirect. Cependant, c'est la seule option où la relation entre le stock final analyse $S_{k,\text{end}}^a$ et le débit analyse Q_k^a est certaine. En effet, ces deux quantités sont directement obtenues en ré-intégrant le modèle, une fois la condition initiale (stock initial) analyse $S_{k,\text{ini}}^a$ calculée via l'analyse de Kalman.

Il peut être remarqué que l'option 3 se rapproche d'une expérience d'estimation de paramètres, aux vues de la définition donnée au Chapitre 3. Par ailleurs, dans chacune des 3 options, le vecteur de contrôle se composera, respectivement, des débits/stocks finaux/stocks initiaux en chacune des $n_x = 2028$ cellules constituant le bassin de l'Amazonie dans ISBA-TRIP. Le problème est largement sous-dimensionné car le nombre d'observations disponibles par jour n'excède jamais 15. De ce fait, la qualité de l'assimilation est liée à la capacité de la maquette à propager l'information apportée par les observations aux cellules non-observées. Les matrices de covariance des erreurs $[PH^T]_{e,k}$ et $[HPH^T]_{e,k}$ - voir leur définition et leur utilisation au Chapitre 1, Section 1.4.3.1 - traduisent, sous la forme de covariances, les liens entre les différentes variables de contrôle et leurs incertitudes.

4.2.3 Génération de l'ensemble

Les matrices de covariance des erreurs d'ébauche $[PH^T]_{e,k}$ et $[HPH^T]_{e,k}$ sont estimées à partir de la définition proposée par Evensen (2004); Moradkhani et al. (2005); Durand et al. (2008) selon laquelle :

$$[PH^T]_{e,k} = (n_e - 1)^{-1} \left(\mathbf{X}_{e,k}^b - \overline{\mathbf{X}_{e,k}^b} \right) \cdot \mathbf{1}_{n_e}^T \hat{\mathbf{C}} \left(\mathbf{H}(\mathbf{X}_{e,k}^b) - \overline{\mathbf{H}(\mathbf{X}_{e,k}^b)} \right) \cdot \mathbf{1}_{n_e}^T \hat{\mathbf{I}}_T, \quad (4.16)$$

et

$$[HPH^T]_{e,k} = (N_e - 1)^{-1} \left(\mathbf{H}(\mathbf{X}_{e,k}^b) - \overline{\mathbf{H}(\mathbf{X}_{e,k}^b)} \right) \cdot \mathbf{1}_{n_e}^T \hat{\mathbf{C}} \left(\mathbf{H}(\mathbf{X}_{e,k}^b) - \overline{\mathbf{H}(\mathbf{X}_{e,k}^b)} \right) \cdot \mathbf{1}_{n_e}^T \hat{\mathbf{I}}_T, \quad (4.17)$$

avec $\mathbf{X}_{e,k}^b$ la matrice de contrôle contenant les n_e vecteurs de contrôle $\mathbf{x}_k^{b,[l]}$, $l = 1 \dots n_e$ de l'ensemble d'ébauche telle que

$$\mathbf{X}_{e,k}^b = \begin{bmatrix} \mathbf{x}_k^{b,[1]} & \dots & \mathbf{x}_k^{b,[n_e]} \end{bmatrix}$$

et $H(\mathbf{X}_{e,k}^b)$ la matrice de contrôle exprimée dans l'espace des observations telle que

$$H(\mathbf{X}_{e,k}^b) = \begin{bmatrix} H(\mathbf{x}_k^{b,[1]}) & \dots & H(\mathbf{x}_k^{b,[n_e]}) \end{bmatrix}.$$

$\overline{\mathbf{X}_{e,k}^b}$ et $\overline{H(\mathbf{X}_{e,k}^b)}$ sont les moyennes sur l'ensemble, respectivement de la matrice de contrôle $\mathbf{X}_{e,k}^b$ et de son image dans l'espace des observations $H(\mathbf{X}_{e,k}^b)$, telles que

$$\overline{\mathbf{X}_{e,k}^b} = \frac{1}{n_e} \sum_{l=1}^{n_e} \mathbf{x}_k^{b,[l]} \quad \overline{H(\mathbf{X}_{e,k}^b)} = \frac{1}{n_e} \sum_{l=1}^{n_e} H(\mathbf{x}_k^{b,[l]}).$$

Ce sont des vecteurs de taille respective n_x et $n_{y,k}$ éléments - d'où l'utilisation du vecteur $\mathbf{1}_{n_e}$ contenant n_e éléments égaux à 1 pour respecter les dimensions en calcul matriciel.

Il apparaît qu'un terme à la ligne m_1 ($m_1 = 1 \dots n_x$) et la colonne m_2 ($m_2 = 1 \dots n_{y,k}$) de la matrice $[\mathbf{PH}^T]_{e,k}$ correspond à la covariance entre l'élément m_1 des vecteurs de contrôle - $\mathbf{x}_{k,m_1}^{b,[l]}$ $l = 1 \dots n_e$ - et l'élément m_2 des vecteurs de contrôle exprimé dans l'espace des observations - $H(\mathbf{x}_{k,m_2}^{b,[l]})$ $l = 1 \dots n_e$. En effet :

$$([\mathbf{PH}^T]_{e,k})_{m_1, m_2} = \frac{1}{n_e - 1} \sum_{l=1}^{n_e} \left(\mathbf{x}_{k,m_1}^{b,[l]} - \overline{\mathbf{x}_{k,m_1}^{b,[\bullet]}} \right) \hat{\mathbf{I}} \left(H(\mathbf{x}_{k,m_2}^{b,[l]}) - \overline{H(\mathbf{x}_{k,m_2}^{b,[\bullet]})} \right). \quad (4.18)$$

Identiquement, l'élément $([\mathbf{HPH}^T]_{e,k})_{m_1, m_2}$ est la covariance entre les m_1 -ème ($m_1 = 1 \dots n_{y,k}$) et m_2 -ème ($m_2 = 1 \dots n_{y,k}$) éléments du vecteur de contrôle exprimé dans l'espace des observations :

$$([\mathbf{HPH}^T]_{e,k})_{m_1, m_2} = \frac{1}{n_e - 1} \sum_{l=1}^{n_e} \left(H(\mathbf{x}_{k,m_1}^{b,[l]}) - \overline{H(\mathbf{x}_{k,m_1}^{b,[\bullet]})} \right) \hat{\mathbf{I}} \left(H(\mathbf{x}_{k,m_2}^{b,[l]}) - \overline{H(\mathbf{x}_{k,m_2}^{b,[\bullet]})} \right). \quad (4.19)$$

On constate que les éléments des matrices de covariance d'erreurs - explicités en Eq. 4.18 et 4.19 - dépendent entièrement de la définition de l'ensemble d'ébauche stocké dans la matrice de contrôle $\mathbf{X}_{e,k}^b$ et des incertitudes de paramètres d'entrée pris en compte pour générer ensuite $H(\mathbf{X}_{e,k}^b)$. La définition de ces incertitudes à l'aide de variables aléatoires permet de perturber aléatoirement les paramètres/variables incertains et de générer les ensembles de contrôle $\mathbf{X}_{e,k}^b$ et $H(\mathbf{X}_{e,k}^b)$ pour l'assimilation. Ainsi, dans le cadre des expériences d'estimation d'état, on a fait le choix de considérer les incertitudes dans la condition initiale du modèle ainsi que les incertitudes dans les forçages atmosphériques et plus particulièrement les précipitations, en accord avec des études précédentes (Nijssen and Lettenmaier, 2004; Andreadis and Lettenmaier, 2006; Clark et al., 2008; Paiva et al., 2013).

4.2.3.1 Perturbations de la condition initiale d'ISBA-TRIP

La condition initiale au cycle d'assimilation k , $\mathbf{c}_k$, est un vecteur de taille $n_x = 2028$ éléments, chacun représentant la masse d'eau contenue dans le réservoir de surface de TRIP avant intégration du modèle pour chacune des 2028 cellules TRIP du bassin. Pour faciliter les écritures, on omettra autant

que possible l'indice de cycle d'assimilation k , sachant que pour toutes les variables et constantes perturbées stochastiquement, un nouvel ensemble est généré à chaque cycle.

On réintroduit la division du bassin de l'Amazonie en $n_s = 9$ zones hydrogéomorphologiques (voir Figure 4.16) vue au Chapitre 2. La condition initiale est alors simplement perturbée en lui appliquant un coefficient multiplicatif $\eta_s^{c,[l]}$ tel que

$$\forall s = 1 \dots n_s, \forall l = 1 \dots n_e, c_{s,k}^{[l]} = \eta_i^{c,[l]} \cdot c_{s,k}, \quad (4.20)$$

où $c_{s,k}$ est la réduction de la condition initiale c_k à la zone s . Pour que la perturbation varie d'un membre à l'autre de l'ensemble, la valeur $\eta_s^{c,[l]}$ est une réalisation d'une loi normale, différente pour chaque membre $[l] = 1 \dots n_e$ et pour chaque zone hydrogéomorphologique s . Les lois normales utilisées ont une moyenne égale à 1 et un écart-type $\sigma_{\eta_{s,k}^{c,[l]}}$ dont les valeurs sont détaillées dans le Tableau 4.1.

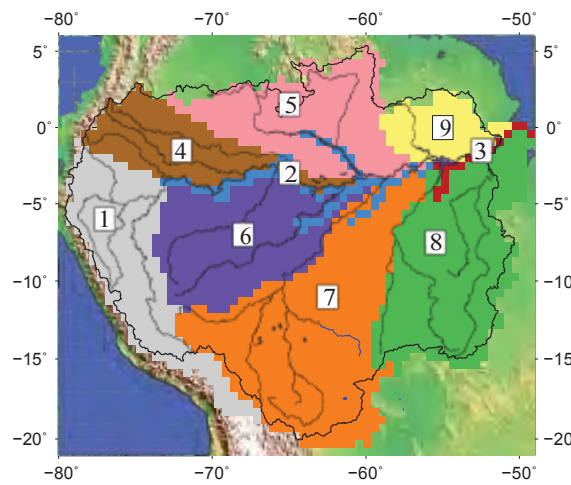


Figure 4.16 – Découpe du bassin de l'Amazonie en zones hydro-géomorphologiques.

Les valeurs de $\eta_i^{c,[l]}$ dépendent donc du cycle d'assimilation courant k ainsi que de la zone hydro-géomorphologique dans laquelle se trouve la cellule i :

- Premièrement, une perturbation plus importante est apportée pour les cellules du cours principal (zones 2 et 3) car on estime que les incertitudes sont plus importantes dans ces zones. En effet, les débits dans ces zones sont les plus importants de tout le bassin et de nombreux points de confluences - qui impliquent des effets de remous - sont présents. Comme ISBA-TRIP ne modélise pas les effets de remous, l'incertitude du stock d'eau dans ces cellules est accentuée.
- Deuxièmement, lors du tout premier cycle d'assimilation, la condition initiale avant perturbation - c_k - est identique pour tous les membres. La variance de l'ensemble des conditions initiales n'est générée que par la seule application de la perturbation définie en Eq. 4.20. Au contraire, sur les autres cycles d'assimilation, les étapes d'analyse successives ont introduit une variabilité de la condition initiale entre les membres, et ce avant les étapes de perturbations. Ainsi, la variance de la condition initiale est plus importante à partir du deuxième cycle d'assimilation et pour les suivants. De ce fait, pour générer une assez grande variabilité lors du premier cycle d'assimilation, l'écart-type $\sigma_{\eta_{s,k}^{c,[l]}}$ de la variable $\eta_i^{c,[l]}$ est plus important lors du premier cycle par rapport au suivant.

4.2.3.2 Perturbations des précipitations

Une autre source d'incertitude pour la génération de l'ensemble $H_k(\mathbf{X}^b)$ réside dans les champs de précipitations. Les forçages utilisés sont ceux du projet GSWP3 introduits et utilisés au Chapitre 2. Pour perturber les forçages de précipitations, on se base sur la technique utilisée par Clark et al. (2008). L'ensemble des champs de précipitations perturbés $\hat{\mathbf{P}}_e$ (perturbés en intensité, temps et espace) est défini tel que :

$$\hat{\mathbf{P}}_e = [\hat{\mathbf{P}}^{[1]}, \hat{\mathbf{P}}^{[2]}, \dots, \hat{\mathbf{P}}^{[n_e]}] = [\phi_p^{[1]} \cdot \mathbf{F}, \phi_p^{[2]} \cdot \mathbf{F}, \dots, \phi_p^{[n_e]} \cdot \mathbf{F}], \quad (4.21)$$

où :

- $\mathbf{F}$ est le champ bi-dimensionnel des forçages de précipitations sans perturbation (qui a un pas de temps de 3h), dont un exemple est illustré en Figure 4.17a,
- $\hat{\mathbf{P}}^{[I]}$ pour $I = 1 \dots n_e$ est la I -ème champ perturbé de précipitations, dont un exemple est illustré en Figure 4.17b,
- $\phi_p^{[I]}$ pour $I = 1 \dots n_e$ est le I -ème champ multiplicateur de $\mathbf{F}$ pour générer $\hat{\mathbf{P}}^{[I]}$.

Le champ de précipitation $\mathbf{F}$ est donc perturbé en lui appliquant un champ multiplicateur aléatoire défini tel que :

$$\phi_p^{[I]} = (1 - \eta^{F,[I]}) + 2\mathbf{U}_F^{[I]} \cdot \eta^{F,[I]}, \quad (4.22)$$

avec :

- $\mathbf{U}_F^{[I]}$ un champ aléatoire suivant une loi uniforme entre 0 et 1,
- $\eta^{F,[I]}$ un scalaire représentant l'erreur relative des précipitations.

De ce fait, $\phi_p^{[I]}$ est un champ aléatoire suivant une loi uniforme de paramètres $(1 - \eta^{F,[I]})$ et $(1 + \eta^{F,[I]})$. Un exemple de champs aléatoire uniforme $\phi_p^{[I]}$ obtenu pour une réalisation de $\eta^{F,[I]} = 0,2522$ est donné en Figure 4.18.

L'erreur relative de précipitations $\eta^{F,[I]}$ quantifie les incertitudes dans l'intensité des précipitations. La valeur de $\eta^{F,[I]}$ est différente pour chaque membre et suit une loi normale de moyenne $\overline{\eta^F} = 30\%$ et d'écart-type $\sigma_{\eta^F} = 0,1\%$. Ces valeurs ont été choisies à partir des études de Clark et al. (2008) et Paiva et al. (2013), voir Tableau 4.1.

Les champs $\mathbf{U}_F^{[I]}$ permettent d'introduire une corrélation en temps et en espace et sont générés avec l'algorithme proposé par Evensen (2003), basé sur la transformée de Fourier. En ce qui concerne la corrélation spatiale, cet algorithme permet de générer des champs bi-dimensionnels aléatoires gaussiens $\mathbf{S}^{[I]}$ de moyenne nulle, de variance égale à 1 et de longueur de corrélation spatiale égale à e^{-1} . Ces champs aléatoires gaussiens sont transformés en champs aléatoires uniformes en leur appliquant la fonction d'erreur complémentaire $\text{erfc}()$:

$$\mathbf{U}_{F,k}^{[I]} = \frac{1}{2} \text{erfc} \left(\frac{\mathbf{S}_k^{[I]}}{\sqrt{\frac{k}{2}}} \right), \quad (4.23)$$

où k^i est le pas de temps propre aux forçages atmosphériques, égal à 3h dans ISBA-TRIP et inférieur au pas de temps des sorties d'ISBA-TRIP qui sont fournies toutes les 24 heures. La PDF en espace de $\mathbf{U}_F^{[I]}$ décroît de e^{-1} quand la distance est égale à la distance de corrélation spatiale τ_x (ici, le x qualifie exceptionnellement la dimension spatiale). Pour les simulations, τ_x est fixé à une valeur de $1,0^\circ$ (Clark et al., 2008; Paiva et al., 2013), soit environ 100 km et est invariant d'un membre à l'autre, d'un cycle à l'autre.

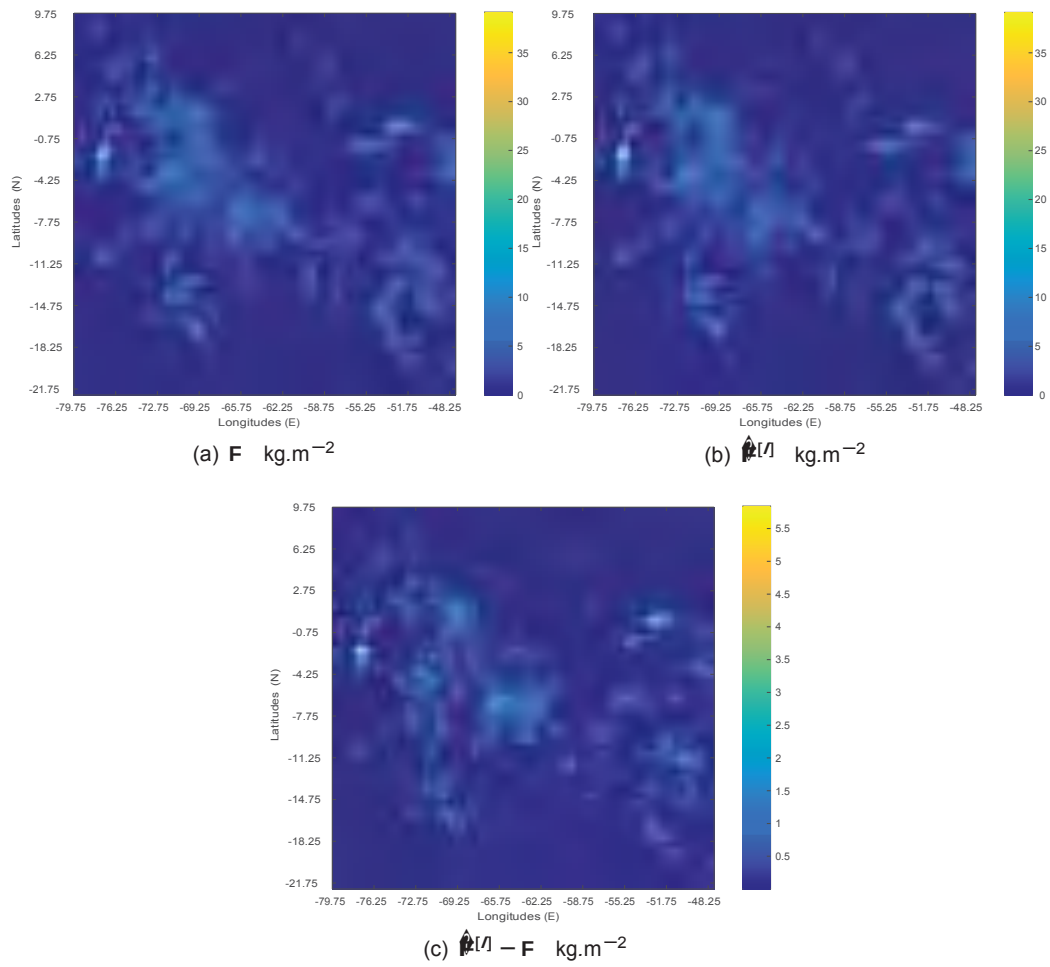


Figure 4.17 – Exemple d'un champ de précipitations $\mathbf{F}$ (a) et du même champ de précipitations après une perturbation $\mathbf{P}$ (b). La différence des deux champs est donnée en figure (c).

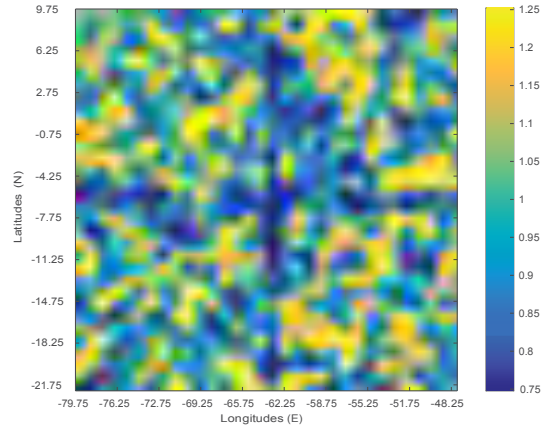


Figure 4.18 – Exemple d'un champ aléatoire uniforme $\phi_p^{[I]}$ obtenu pour une réalisation de $\eta^{F,[I]} = 0,2522$.

En ce qui concerne la corrélation temporelle, le paramètre $\mathcal{J}^{[I]}$

$$\mathcal{J}^{[I]} = 1 - \frac{\Delta k^i}{\tau_k^{[I]}} \quad (4.24)$$

détermine la longueur de corrélation temporelle entre les différents champs $\mathbf{S}_k^{[I]}$. Elle est concrètement générée via l'utilisation du champ aléatoire à l'instant précédent $\mathbf{S}_{k-1}^{[I]}$ et d'un autre champ auxiliaire $\mathbf{W}_k^{[I]}$ ayant les mêmes propriétés tel que :

$$\mathbf{S}_k^{[I]} = \mathcal{J}^{[I]} \mathbf{S}_{k-1}^{[I]} + \sqrt{1 - (\mathcal{J}^{[I]})^2} \mathbf{W}_k^{[I]}, \quad (4.25)$$

avec $\Delta k^i = 3$ heures le pas de temps des forçages atmosphériques et $\tau_k^{[I]}$ la constante de temps caractéristique de $\mathcal{J}^{[I]}$. $\mathcal{J}^{[I]} = 0$ génère un bruit blanc (c'est-à-dire une perturbation décorrélée des autres dans le temps) alors que $\mathcal{J}^{[I]} = 1$ rend la perturbation stochastique constante dans le temps. Elle prend une valeur différente pour chaque membre de l'ensemble et suit une loi normale de moyenne $\tau_k = 12$ heures (43200 secondes dans le Tableau 4.1) et d'écart-type $\sigma_{\tau_k} = 3$ heures (10800 secondes dans le Tableau 4.1). Ces valeurs ont été choisies de telle sorte que la corrélation temporelle apparaisse sur des cycles d'assimilation de seulement 1 jour.

Variable	Description	Valeur	Eq.
n_e	Nbre de membres	101	-
n_x	Nbre de var. de contrôle	2028	-
n_y	Nbre de var. d'obs	$\in [0, 15]$	-
η^o	Erreur d'obs	0, 2	4.8
$\sigma_{\eta_{s,k}^x}$	Erreur sur la condition initiale	$s = 2, 3, k = 1 : 0, 25$ $s = 2, 3, k > 1 : 0, 05$ $s = 1, 4 : 9, k = 1 : 0, 10$ $s = 1, 4 : 9, k > 1 : 0, 02$	4.20
$\tau_x [^\circ]$	Corrélation spatiale précip.	1.0	
η^F	Erreur relative précip. moyenne	0, 3	4.22
$\sigma_{\eta^F} \tau_k$	Erreur relative précip. dispersion	0, 1	4.22
$[s] \sigma_{\tau_k}$	Corrélation temporelle précip. moyenne	43200	4.24
$[s]$	Corrélation temporelle précip. dispersion	10800	4.24

Table 4.1 – Constantes utilisées pour la génération des ensembles d'ébauche $\mathbf{X}_{e,k}^b$, d'observations $\mathbf{Y}_{e,k}^o$ et de modèle $H(\mathbf{X}_{e,k}^b)$, k étant l'indice du cycle d'assimilation et s l'indice de zone hydrogéomorphologique (voir Figure 4.16). Pour chaque ligne, l'unité (quand elle est définie) est donnée entre crochets dans la colonne « Variable ».

4.2.3.3 Localisation des matrices de covariance d'erreurs

Dans le contexte de l'estimation d'état, les erreurs d'échantillonnage (c'est-à-dire liées à la taille limitée de l'ensemble) peuvent introduire des corrélations artificielles dans les matrices de covariance des erreurs d'ébauche/d'analyse, et engendrer par exemple des corrélations irréalistes entre deux points de maille très éloignés sur la grille (Anderson, 2007). La taille de l'ensemble n_e est limitée par les contraintes matérielles et de coûts de calcul. Ainsi, un traitement numérique des matrices, en amont de l'étape d'analyse de l'EnKF, est nécessaire pour supprimer ces corrélations artificielles qui peuvent sensiblement dégrader l'analyse. Ce sont les étapes de *localisation* qui sont justement conçues pour pallier ces problèmes de corrélations irréalistes.

Un premier processus classique de localisation consiste à modifier explicitement la matrice de covariance des erreurs d'ébauche $\mathbf{P}_{e,k}^b$ en annulant directement les corrélations entre deux variables de contrôle éloignées dans l'espace (Houtemakker and Mitchell, 1998; Anderson, 2001). Une autre technique, indirecte, consiste à multiplier la matrice $\mathbf{P}_{e,k}^b$ par une matrice de corrélations générée à partir d'une fonction radiale, c'est-à-dire une fonction des deux/trois dimensions de l'espace qui décroît de façon monotone quand les points s'éloignent des variables de contrôle (Hamill et al., 2001; Houtemakker and Mitchell, 2001, 2005). On obtient alors une matrice $\tilde{\mathbf{P}}_{e,k}^b$ creuse avec les éléments non nuls plutôt centrés autour de la diagonale et cette matrice modifiée $\tilde{\mathbf{P}}_{e,k}^b$ remplace $\mathbf{P}_{e,k}^b$ dans la détermination du gain de Kalman $\mathbf{K}_{e,k}$. Les techniques de localisation permettent d'éviter les problèmes liés à la génération des ensembles lors d'expériences d'AD. Parmi d'autres techniques de localisation, le LETKF (*Local Ensemble Transform Kalman Filter* ; Hunt et al., 2007) combine un EnKF avec une localisation qui consiste à effectuer l'étape d'analyse dans des sous-espaces caractéristiques de l'espace du problème.

Seulement, les techniques de localisation décrites ci-dessus ont toutes été développées pour l'AD en modélisation atmosphérique où les problèmes sont en deux voire trois dimensions. L'utilisation de la localisation en hydrologie est beaucoup plus limitée. Plusieurs études existent pour améliorer la

modélisation des écoulements de sub-surface (Devegowda et al., 2010; Delijani et al., 2014), mais on reste proche de la modélisation atmosphérique dans le sens où on travaille dans un système continu en deux/trois dimensions. D'autres études utilisant la localisation permettent aussi de mieux estimer les paramètres du modèle (Sun et al., 2009; Rasmussen et al., 2015), eux aussi définis en continu sur plusieurs dimensions.

La méthode de localisation utilisée pour un modèle de routage comme TRIP ne peut pas être basée sur une seule fonction radiale en deux dimensions décroissante avec la distance. En effet, l'écoulement des flux ne se fait pas en deux dimensions comme pour les modèles atmosphériques mais selon plusieurs directions d'écoulement en une dimension matérialisée par le réseau de routage. Le processus de localisation doit alors tenir compte du routage pour justement décorréler deux cellules voisines se trouvant alors sur deux sous-bassins versants distincts du bassin global. Malgré tout, le long d'une même direction d'écoulement, la corrélation entre deux cellules distinctes est aussi liée à la distance entre les deux cellules. Ainsi, la localisation consiste à construire, à chaque cycle d'assimilation, un masque de localisation définissant une zone d'influence propre à chaque observation. Ces zones d'influence regroupent alors un certain nombre de cellules amonts et aval à l'observation - selon le réseau de routage.

Pour définir le nombre de cellules définissant la taille de la zone d'influence, on se base sur la découpe du bassin selon les neuf zones hydro-géomorphologiques (voir Figure 4.16) et sur la vitesse d'écoulement moyenne dans ces zones. La zone d'influence, autour d'une cellule observée donnée, est alors définie selon les critères ci-dessous. Pour une zone d'influence dont la taille est fixée à un nombre p de cellules, la zone d'influence se compose de :

- la cellule observée elle-même,
- des p cellules situées en aval de la cellule observée selon le réseau de routage,
- de toutes les cellules situées en amont de la cellule observée en remontant p fois en amont toujours selon le réseau hydrologique. Cependant, la remontée du réseau hydrologique s'arrête si on change de zone hydro-géomorphologique.

Le nombre de cellules dans la zone d'influence dépend de la vitesse de propagation de l'écoulement dans la zone. Les vitesses d'écoulement sont calculées à partir de la simulation run libre. La résolution du modèle TRIP est de $0,5^\circ \times 0,5^\circ$ soit environ $50\text{km} \times 50\text{km}$. Étant donné les méandres des tronçons de rivières, la distance minimale que l'écoulement peut parcourir à travers une cellule est de 50 km. De plus, en comparant le run libre aux débits in situ et débits ENVISAT, il apparaît que le run libre sous-estime le débit (donc la vitesse d'écoulement). Par conséquent, pour choisir le nombre de cellules de la zone d'influence de chaque région hydro-géomorphologique :

- on arrondit au multiple de 50 supérieur la distance maximale (sur la période de simulation) parcourue sur un jour dans la zone,
- le nombre de cellules de la zone d'influence correspond aux nombres de cellules équivalent à la distance arrondie, sachant que $50\text{ km} = 1\text{ cellule}$.

Le nombre de cellules de la zone d'influence est présenté pour chaque zone dans le Tableau 4.2. Le masque de localisation final se présente sous la forme d'une matrice de dimension $n_x \times n_{y,k}$ (avec n_x le nombre de variables de contrôle et $n_{y,k}$ le nombre d'observations au cycle courant k) contenant uniquement des 0 et des 1. Les éléments non nuls se trouvent alors dans la zone d'influence d'une des observations. Le localisation est appliquée aux matrices de covariance d'erreur $[\mathbf{PH}^T]_{e,k}$ et $[\mathbf{HPH}^T]_{e,k}$ avant la détermination du gain de Kalman et l'étape d'analyse. La masque de localisation a exactement la même dimension que $[\mathbf{PH}^T]_{e,k}$, ainsi la localisation est simplement appliquée en multipliant termes à termes le masque de localisation et $[\mathbf{PH}^T]_{e,k}$. Pour étendre la localisation à

Zone	Dist. max réelle [km]	Dist. max arrondie [km]	Zone d'influence [nbre de cellules]
1	84.16	100	2
2	174.24	200	4
3	233.80	250	5
4	93.29	100	2
5	82.03	100	2
6	80.03	100	2
7	69.57	100	2
8	87.86	100	2
9	67.02	100	2

Table 4.2 – Détails des tailles des "zones d'influence", en nombre de cellules TRIP.

$[\mathbf{HPH}^T]_{e,k}$, on extrait du masque de localisation les lignes correspondant aux cellules observées et on multiplie identiquement terme à terme. L'étape de localisation décrite ci-dessus est alors appliquée lors de chaque cycle d'assimilation en fonction des stations ENVISAT activées au cycle courant.

4.2.4 Présentation de la stratégie d'assimilation et des expériences d'assimilation

Les expériences d'estimation d'état permettront de tester les différentes variables de contrôle possibles décrites à la Section 4.2.2. Elles permettront aussi de tester, valider et critiquer le masque de localisation présenté précédemment, en le comparant à deux autres simulations tests : une sans localisation et une seconde en supprimant dans les matrices d'erreurs toutes les corrélations entre les mailles du bassin. Dans la suite, le nom des expériences utilisant la localisation aura le suffixe « -local », celui des expériences sans localisation aura le suffixe « -direct » et celui des expériences sans corrélation entre les mailles dans la matrice d'erreur d'ébauche aura le suffixe « -diag ».

L'objectif de ce chapitre est ainsi de déterminer la meilleure stratégie d'assimilation des débits ENVISAT dans le modèle ISBA-TRIP via un EnKF utilisé pour estimer l'état. Les différentes simulations réalisées sont listées dans le Tableau 4.3 ci-dessous.

Nom de l'exp	Variable de contrôle	Localisation
EE1-direct	stock initial	Non - $[\mathbf{PH}^T]_{e,k}$ et $[\mathbf{HPH}^T]_{e,k}$ définies aux Eq. 4.16-4.17
EE1-diag	stock initial	Non - $[\mathbf{PH}^T]_{e,k}$ et $[\mathbf{HPH}^T]_{e,k}$ diagonales
EE1-local	stock initial	Oui - masque décrit en Section 4.2.2.3
EE2-local	stock final	Oui - masque décrit en Section 4.2.2.2
EE3-local	débit	Oui - masque décrit en Section 4.2.2.1

Table 4.3 – Présentation des différentes expériences d'estimation d'état. La dénomination "EE" correspond à "Estimation d'État", la numération "1/2/3" différencie les variables de contrôle utilisées et l'appellation "direct/diag/local" traduit la méthode de localisation utilisée ou non.

Pour toutes les expériences d'AD, les erreurs d'observations et de modèle sont modélisées telles qu'elles ont été décrites en Section 4.2.1 pour les observations et en Section 4.2.3 pour les erreurs modèles et les valeurs des différents paramètres impliqués dans l'assimilation sont données en Tableau 4.1. Aussi, pour toutes les expériences, on réalise 365 cycles d'assimilation d'un jour (soit un an

de simulation) à partir du 25 Septembre 2002 (date à laquelle les premières données ENVISAT sont disponibles sur le bassin de l'Amazonie) et donc jusqu'au 24 Septembre 2003. Les nombreuses simulations élémentaires d'ISBA-TRIP nécessaires pour générer les ensembles modèles ont été menées sur la plateforme de calcul hautes performances CALMIP (<https://www.calmip.univ-toulouse.fr/spip/>) et sur son super ordinateur EOS.

Dans l'expérience EE1-direct, on assimile les débits ENVISAT pour corriger les stocks d'eau initiaux de TRIP et intrinsèquement les débits simulés par TRIP. Pour cette première simulation, un EnKF classique, sans aucune localisation des matrices de covariance d'erreur $[\mathbf{PH}^T]_{e,k}$ et $[\mathbf{HPH}^T]_{e,k}$, est réalisé. Cette première simulation servira alors de base de comparaison pour les deux simulations suivantes EE1-diag et EE1-local qui permettront de mettre en évidence les apports et éventuelles limitations de la méthode de localisation choisie. Enfin, les deux dernières simulations, EE2-local et EE3-local, permettront de comparer le choix de différentes variables de contrôle sur le résultat d'analyse (voir la deuxième colonne du Tableau 4.3). En particulier, EE2-local utilise l'option 2 pour le vecteur de contrôle (voir Section 4.2.2.2) : assimiler les débits ENVISAT pour corriger les stocks finaux du réservoir de surface de TRIP. EE3-local, quant à lui, utilise l'option 1 (voir Section 4.2.2.1) pour le vecteur de contrôle, à savoir utiliser les débits ENVISAT pour corriger les débits TRIP.

4.3 Résultats de l'expérience EE1-direct

Pour cette première expérience d'estimation d'état, on choisit pour variables de contrôle le stock d'eau initial dans le réservoir de surface, c'est-à-dire $S_{k,ini}$. L'expérience est réalisée sur un an de simulation, soit 365 cycles d'assimilation d'un jour, du 25 Septembre 2002 au 24 Septembre 2003. Les matrices de covariance d'erreur modèle $[\mathbf{PH}^T]_{e,k}$ et $[\mathbf{HPH}^T]_{e,k}$ sont déterminées en appliquant les Eq. 4.16 et 4.17, c'est-à-dire en appliquant les matrices de covariance d'erreur complètes, sans localisation.

4.3.1 Résultats généraux

Pour étudier le comportement de la maquette, on se base sur les statistiques de la moyenne de l'ensemble du run analyse $\overline{Q_{e,k}^a}$, que l'on appellera dans la suite le *run analyse moyen*. L'ensemble des débits analyse au cycle k , $Q_{e,k}^a$ est obtenu en propageant le modèle ISBA-TRIP sur la fenêtre d'assimilation $[k - 1, k]$ en prenant comme condition initiale le vecteur de contrôle analyse $\mathbf{x}_{e,k}^a$. $\overline{Q_{e,k}^a}$ est alors la moyenne du débit sur l'ensemble des membres. Les résultats sur l'ensemble du bassin de l'expérience EE1-direct sont présentés en Figure 4.19-4.22.

Les Figures 4.19-4.20 présentent les statistiques (voir Eq. 4.4, 4.5 et 4.6), calculées sur la période d'étude d'1 an, du run analyse moyen $\overline{Q_{e,k}^a}$ selon la même configuration que la Figure 4.14-4.15. Elles sont calculées par rapport aux débits ENVISAT Q_k^o d'une part (voir Figure 4.19a, 4.20a) et par rapport aux débits in situ Q_k^{situ} qui n'ont pas été utilisés dans l'étape d'assimilation) d'autre part (voir Figures 4.19b, 4.20b). Ensuite, pour mesurer concrètement l'apport de l'AD, la Figure 4.21 affiche le pourcentage d'augmentation ou de réduction des RMSEn entre les statistiques du run libre Q_k^f et celles du run analyse moyen $\overline{Q_{e,k}^a}$ la Figure 4.22 affiche la différence de Nash entre celui du run analyse moyen $\overline{Q_{e,k}^a}$ et celui du run libre Q_k^f (une différence positive montre une augmentation du Nash, alors qu'une différence négative montre une détérioration de celui-ci).

De manière générale, les RMSEn sur la Figure 4.19 sont plus bas que ceux de la Figure 4.14.

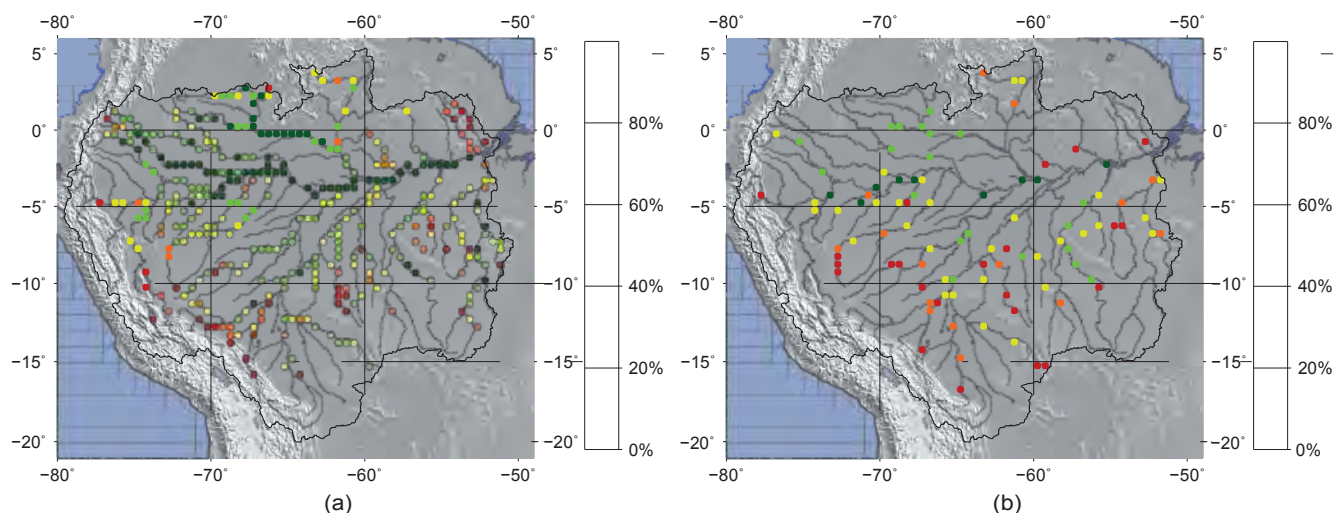


Figure 4.19 – Coefficients d’erreurs moyennes quadratiques normalisées RMSEn de l’expérience d’AD EE1-direct par rapport aux débits ENVISAT (a) et par rapport aux débits in situ (b).

Statistiques globales	Run libre	Run analyse EE1-direct	Unité
$RMSE^{situ}$	$2,79 \times 10^3$	$1,86 \times 10^3$ (–33%)	$m^3.s^{-1}$
$RMSEn^{situ}$	73,61	56,01 (–24%)	%
E^{situ}	–11,57	–3,90	-
$RMSE^o$	$5,71 \times 10^3$	$2,48 \times 10^3$ (–56%)	$m^3.s^{-1}$
$RMSEn^o$	75,29	41,87 (–44%)	%
E^o	–134,32	–32,13	-

Table 4.4 – Comparaison des statistiques globales du run libre et de l’expérience EE1-direct

On peut observer des RMSEn très bas (moins de 20%) sur le cours principal et sur les principaux affluents gauches, notamment le Japurá et le Negro. Les RMSE sont plus élevées (entre 20 et 40%) sur les principaux affluents droits, c’est-à-dire le Jutai, le Juruá, le Purús, le Madeira, le Tapajós et le Xingu, et leur confluence avec le cours principal affichent des RMSE entre 40 et 60%. Enfin, les têtes de bassins ainsi que des petits affluents, tels que le Paru et le Jari, conservent des RMSE supérieures à 60%. De manière similaire, les Nash sont généralement augmentés (Figure 4.20) par rapport au run libre (Figure 4.15). Les confluences apparaissent avec des coefficients de Nash négatifs, ainsi que le haut bassin du Japurá. Les Figures 4.21 et 4.22 confirment une amélioration générale des statistiques après assimilation, avec des RMSE globalement réduits et des Nash globalement augmentés. Pour mieux les quantifier, le Tableau 4.4 compare les statistiques globales sur l’ensemble des cellules entre le run libre d’une part et la simulation EE1-direct d’autre part, par rapport aux débits ENVISAT (ce qui correspond à 368 stations) et par rapport aux débits in situ (120 stations).

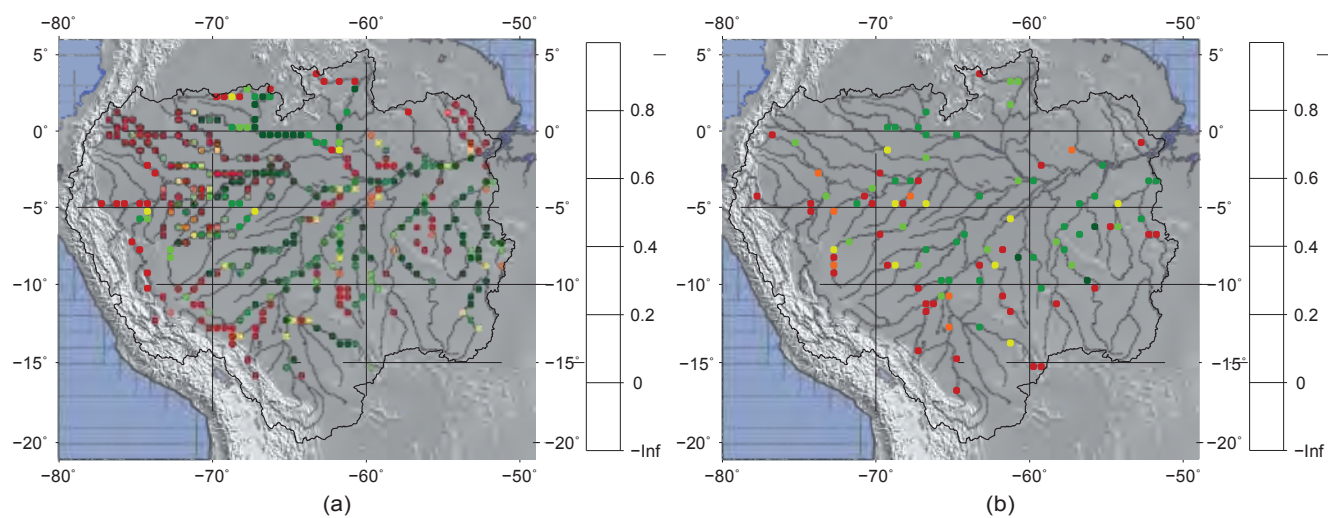


Figure 4.20 - Coefficients de Nash-Sutcliffe E de l'expérience d'AD EE1-direct par rapport aux débits ENVISAT (a) et par rapport aux débits in situ (b).

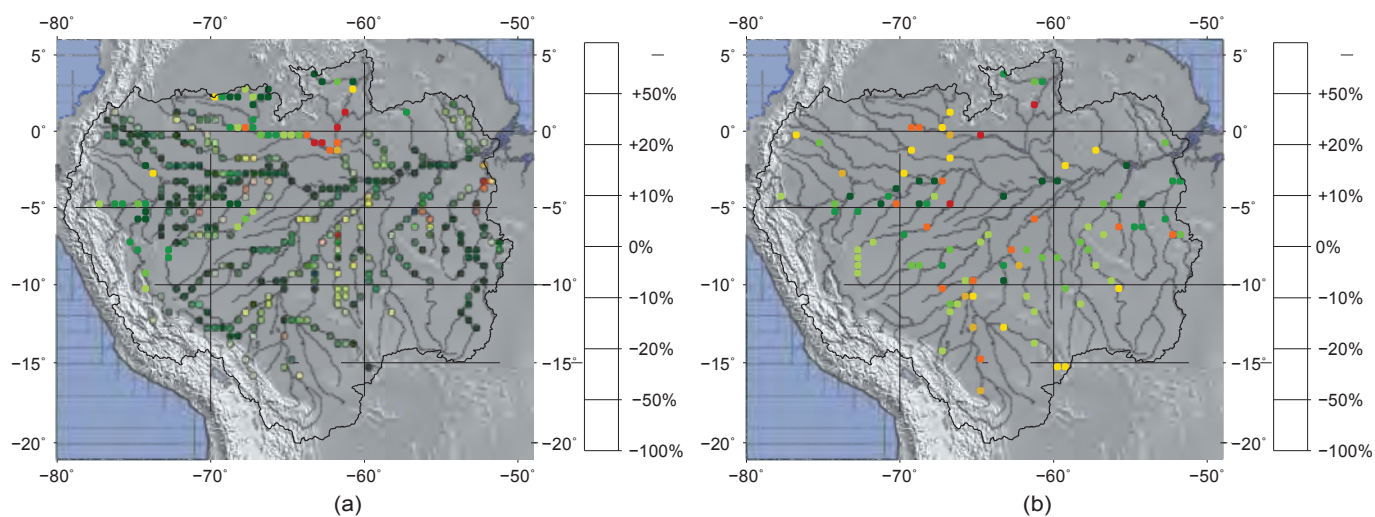


Figure 4.21 - Pourcentage d'augmentation/de réduction du RMSEn pour l'expérience d'AD EE1-direct par rapport au run libre pour les débits ENVISAT (a) et pour les débits in situ (b).

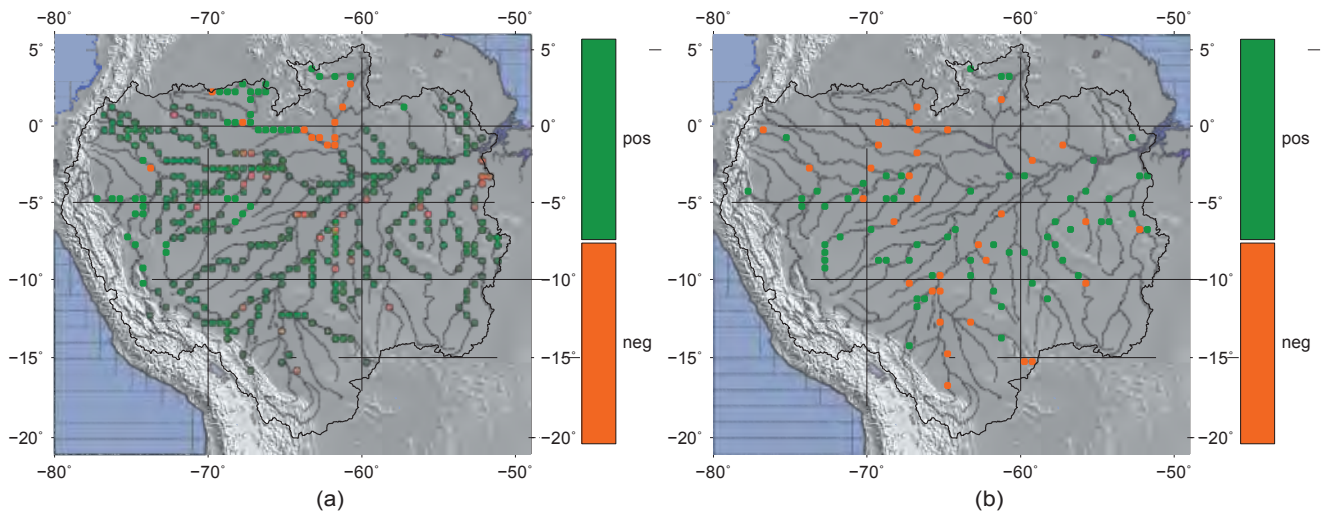


Figure 4.22 – Augmentation (vert)/réduction (rouge) du Nash E pour l'expérience d'AD EE1-direct par rapport au run libre pour les débits ENVISAT (a) et pour les débits in situ (b).

4.3.2 Comportement pour certaines stations tests

La comparaison des différents débits est présentée en Figure 4.23 pour une quinzaine de cellules TRIP auxquelles sont associées une station de débits in situ et, pour la plupart, une station virtuelle ENVISAT. Ces 15 stations sont réparties sur les principaux affluents de l'Amazone ainsi que le long du cours principal, voir Figure ?? . Aussi, les résultats à ces stations sont représentatifs de l'ensemble des résultats obtenus. Le Tableau 4.5 détaille, pour les différentes stations-test, les statistiques du débit analyse moyen comparées à celles du run libre.

Le débit analyse moyen $\overline{Q_{\bullet,k}^a}$ présente, pour la majorité des stations (voir Figure 4.23), un comportement chaotique avec de nombreux pics et de nombreux creux parfois très importants - voir par exemple à Serrinha (Figure 4.23h) ou sur les affluents droits (Figures 4.23k-o). Malgré tout, la tendance saisonnière des débits analyses, que l'on distingue aux travers des pics/creux, est rapprochée de la tendance des débits observés. Les creux de débits observés à une station donnée ne sont pas synchronisés avec les dates d'observations de la station. Par exemple à Serrinha, les différentes données de débits sont présentées pour les 35 premiers cycles d'assimilation en Figure 4.24. On constate une chute importante du débit analyse moyen au cycle d'assimilation 6 alors que la station virtuelle associée à Serrinha est active (c'est-à-dire qu'il y a une observation ENVISAT sur cette cellule) au cycle 19 seulement. Ce sont alors les covariances non-nulles entre les cellules observées et les cellules non-observées, stockées dans la matrice de covariance d'erreur $[\mathbf{PH}^T]_{e,k}$, qui permettent de contraindre les variables de contrôle non-observées. En conséquence, la baisse de débit au cycle 6 est liée à une observation hors de la cellule considérée. Il apparaît alors que les "corrections" apportées à chaque cycle ne sont pas systématiquement judicieuses du fait des corrélations non-nulles et peuvent générer ces pics/creux qui perturbent fortement le cycle hydrologique.

En effet, décomposons l'équation d'analyse de l'EnKF au cycle 6 pour extraire la "correction"

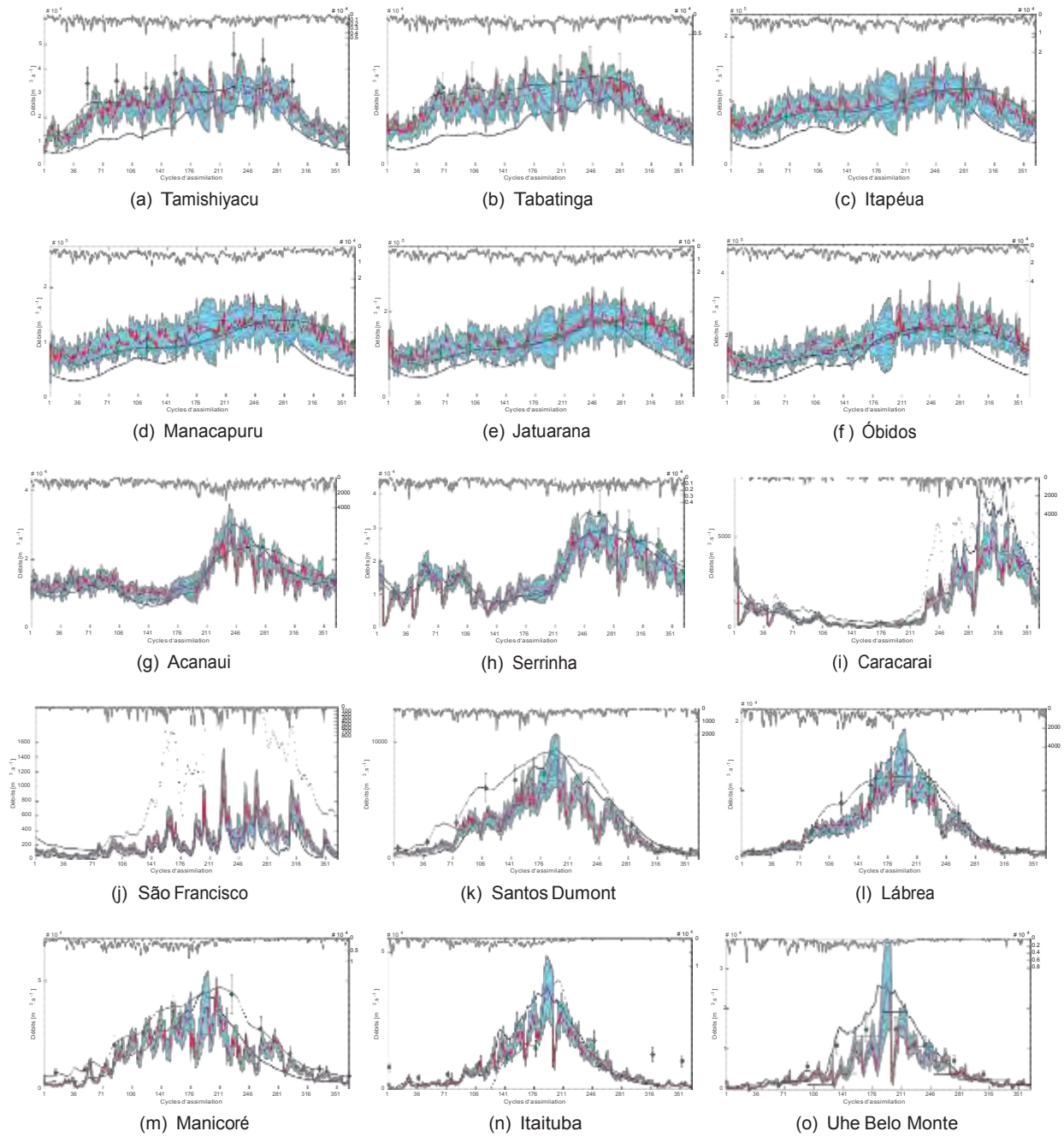


Figure 4.23 – Expérience EE1-direct : Comparaison des débits d'intérêt (en $\text{m}^3.\text{s}^{-1}$) aux stations in situ : ENVISAT Q_k^o et leur incertitude (losanges verts), in situ Q_k^{situ} (points noirs), run libre Q_k^f (ligne bleu), dispersion de l'ensemble d'ébauche $Q_{e,k}^b = H(\mathbf{x}_{e,k}^b)$ (zone bleue du 1^{er} au 9^{ème} décile) et run analyse moyen $\overline{Q_{\bullet,k}^a}$ (rouge). L'ensemble des précipitations $\hat{\mathbf{P}}_e$ est en zone grise sur l'axe des ordonnées inversé. L'axe des abscisses correspond au temps (un marqueur tous les 35 jours, numéroté de 1 à 351).

Indice de la station	RMSEn Q_k^f	RMSEn $Q_{\bullet,k}^a$	E Q_k^f	E $Q_{\bullet,k}^a$
Tamashiyacu	0,401	0,162	-1,783	0,563
Tabatinga	0,449	0,162	-2,221	0,580
Itapêua	0,332	0,126	-1,641	0,620
Manacapuru	0,364	0,164	-1,123	0,523
Jatuarana	0,270	0,129	-0,210	0,724
Obidos	0,291	0,133	-0,315	0,724
Acanaui	0,195	0,213	0,661	0,596
Serrinha	0,156	0,283	0,902	0,677
Caracarai	0,431	0,651	0,788	0,517
São Francisco	1,106	0,993	0,945	-0,569
Santos Dumont	0,380	0,498	0,684	0,456
Lábrea	0,371	0,368	0,740	0,744
Manicoré	0,360	0,468	0,635	0,384
Itaituba	0,558	0,470	0,526	0,666
Uhe Belo Monte	0,852	0,600	0,402	0,704

Table 4.5 – Performances de l'EnKF lors de l'expérience EE1-direct : comparaison des erreurs quadratiques moyennes normalisées (RMSEn) et des coefficients d'efficacité de Nash-Sutcliffe (E) du débit analyse moyen $\overline{Q_{\bullet,k}^a}$ et du débit du run libre Q_k^f , calculés à partir des débits observés in situ.

faite à Serrinha (m_1 de l'Eq. 4.18 égal à 984 pour Serrinha) :

$$\mathbf{X}_{e,6}^a(984,-)_{\{1,n_e\}} = \mathbf{X}_{e,6}^b(984,-)_{\{1,n_e\}} + [\mathbf{P}\mathbf{H}^T]_{e,6}(984,-)_{\{1,n_{y,6}\}} \cdot \left([\mathbf{H}\mathbf{P}\mathbf{H}^T]_{e,6} + \mathbf{R}_6 \right)^{\hat{\mathbf{I}}_{-1}}_{\{n_{y,6},n_{y,6}\}} \cdot \mathbf{Y}_{e,6}^o(984,-)_{\{n_{y,6},n_e\}} - H(\mathbf{X}_{e,6}^b(984,-)_{\{1,n_e\}})$$

Plus encore, on peut décomposer la "correction" apportée à l'ébauche

$$(\delta\mathbf{X})_{(984,-)} = [\mathbf{P}\mathbf{H}^T]_{e,6}(984,-)_{\{1,n_{y,6}\}} \cdot \left([\mathbf{H}\mathbf{P}\mathbf{H}^T]_{e,6} + \mathbf{R}_6 \right)^{\hat{\mathbf{I}}_{-1}}_{\{n_{y,6},n_{y,6}\}} \cdot \mathbf{Y}_{e,6}^o(984,-)_{\{n_{y,6},n_e\}} - H(\mathbf{X}_{e,6}^b(984,-)_{\{1,n_e\}})$$

en fonction de l'information apportée par chacune des observations $j = 1 \dots n_{y,6}$ - $(\delta\mathbf{X})_{(984,-)} = \sum_{j=1}^{n_{y,6}} (\delta\mathbf{X}_j)_{(984,-)}$:

$$(\delta\mathbf{X}_j)_{(984,-)} = [\mathbf{P}\mathbf{H}^T]_{e,6}(984,j)_{\{1,n_{y,6}\}} \cdot \left([\mathbf{H}\mathbf{P}\mathbf{H}^T]_{e,6} + \mathbf{R}_6 \right)^{\hat{\mathbf{I}}_{-1}}_{\{j,n_{y,6}\}} \cdot \mathbf{Y}_{e,6}^o(j,-)_{\{n_{y,6},n_e\}} - H(\mathbf{X}_{e,6}^b(984,-)_{\{1,n_e\}})$$

ce qui donne concrètement, moyenné sur les n_e membres de l'ensemble :

$$\begin{aligned} (\delta\mathbf{X}_1)_{(984,\bullet)} &= -5,09 \times 10^{10}, \\ (\delta\mathbf{X}_2)_{(984,\bullet)} &= -1,39 \times 10^{11}, \\ (\delta\mathbf{X}_3)_{(984,\bullet)} &= 7,86 \times 10^9, \\ (\delta\mathbf{X}_4)_{(984,\bullet)} &= -3,30 \times 10^7, \\ (\delta\mathbf{X}_5)_{(984,\bullet)} &= 2,62 \times 10^7, \\ (\delta\mathbf{X}_6)_{(984,\bullet)} &= -1,26 \times 10^9, \\ (\delta\mathbf{X}_7)_{(984,\bullet)} &= 2,82 \times 10^{10}. \end{aligned}$$

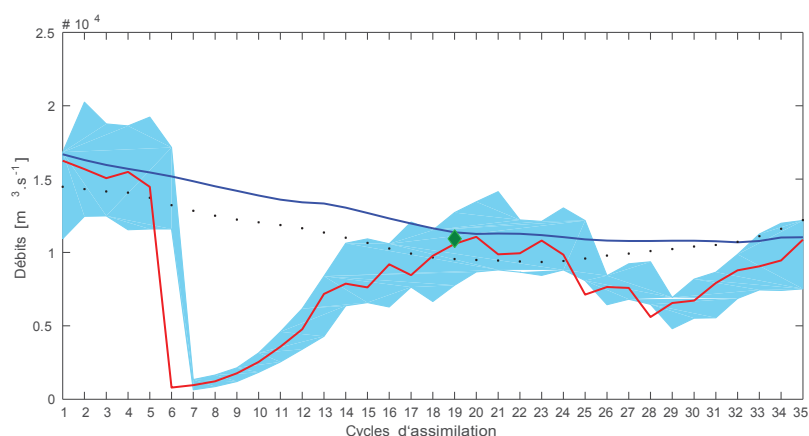


Figure 4.24 – Comparaison des différents débits d'intérêt à Serrinha sur les 35 premiers cycle d'assimilation : débit ENVISAT Q_k^o et leur incertitude (losanges verts), débits in situ Q_k^{situ} (points noirs), débits run libre Q_k^f (ligne bleu), dispersion de l'ensemble d'ébauche $Q_{e,k}^b = H(\mathbf{X}_{e,k}^b)$ (zone bleue du 1^{er} au 9^{ème} décile) et run analyse moyen $\overline{Q_{*,k}^a}$

Les stocks d'eau générés par ISBA-TRIP sont de l'ordre de 10^{12} kg, ce qui implique que les "corrections" apportées par l'EnKF, ici à Serrinha, sont loin d'être négligeables, en particulier les corrections négatives des deux premières observations. Ces "corrections" négatives sont à l'origine de la forte chute de débit au cycle 6. Ainsi, la chute de débit au cycle 6 est bien liée à des observations hors de la cellule considérée et est liée à des problèmes d'échantillonnage de l'ensemble qui induit de fausses corrélations. Une étude plus approfondie du comportement de l'EnKF à Serrinha et sur le rio Negro en général sera faite à la Section 4.3.3.

Malgré tout, à une station virtuelle donnée, les jours où une observation est disponible, on remarque que les corrections de débits réalisées par l'EnKF sont systématiquement adaptées. D'ailleurs, les statistiques analyse par rapport au débit ENVISAT sont nettement améliorées (voir Figures 4.21-4.22 sauf pour quelques exceptions que l'on détaillera dans un prochain paragraphe), ce qui montre le bon fonctionnement de la maquette d'assimilation.

Les stations se trouvant le long du cours principal - Figures 4.23a-f - présentent les meilleures résultats. Les observations ENVISAT sont très proches des débits in situ, par exemple le coefficient de Nash entre les débits ENVISAT et les débits in situ est de 0,966 à Óbidos, et les séries ENVISAT à ces stations sont quasiment complètes (au moins 8 observations sur les 11 possibles au cours de la période d'assimilation). Par ailleurs, le cours principal correspond aux flux de masse les plus importants du bassin. De ce fait, ces stations profitent des corrections en amont faites à des cycles antérieurs et qui se sont propagées jusqu'à ces stations et, le débit analyse est moins impacté par les corrélations croisées avec les débits observés à des stations éloignées et sur des cours d'eau plus petits. Ainsi, la station in situ du cours principal la moins performante - Tamishiyacu, c.f. Figure 4.23a - est aussi la plus en amont, avec les débits les plus faibles et la dispersion de l'ensemble d'ébauche ne contient pas d'observations ENVISAT.

Les statistiques globales sont généralement moins bonnes sur les affluents droits du bassin. Les stations situées sur ces grands affluents (Figures 4.23k-o) - i.e. Santos Dumont sur le Juruá, Lábrea sur

le Purus, Manicoré sur le Madeira, Itaituba sur le Tapajós et Uhe Belo Monte sur le Xingu - présentent un comportement similaire : un débit analyse moyen $\overline{Q_{\bullet,a}^a}$ très bruité et qui se rapproche difficilement des débits observés/in situ. Exceptées les stations de Itaituba et Belo Monte où les débits in situ sont problématiques (courbe de tarage erronée pour Itaituba et débit non-journalier mais mensuel pour Belo Monte), les débits ENVISAT et in situ sont en accord et présentent un hydrographe avec pics de débit modérés mais surtout une croissance/décroissance du débit très lente. Le débit run libre, au contraire, présente un pic de débit plus marqué avec un hydrographe moins étalé dans le temps. L'EnKF parvient plus ou moins à capter le début de croissance des débits observés dans l'analyse mais l'ensemble est trop peu dispersé pour permettre au débit analyse de se rapprocher des observations.

Certaines stations in situ n'ont pas de station virtuelle associée - ici Acanai (Figure 4.23g), Caracarai (Figure 4.23i) et São Francisco (Figure 4.23j) - mais on observe malgré tout un débit analyse moyen $\overline{Q_{\bullet,k}^a}$ différent du run libre. Selon les cas, les statistiques par rapport au run libre peuvent être soit améliorées - voir à Acanai ou Caracarai (Figures 4.23g,i) - soit détériorées - voir à São Francisco (Figure 4.23j). Ici, les différences entre le run libre et le run analyse ne sont pas dues à l'assimilation directe de mesures de débits ENVISAT mais peuvent être néanmoins expliquées par deux phénomènes :

- la propagation d'une correction ayant été apportée en amont (selon le réseau hydrologique) de ces cellules ou,
- une "correction" directe du stock initial par le biais des observations actives (plus ou moins éloignées de la cellule in situ considérée) et surtout des corrélations non-nulles entre les cellules, ce qui sous-entend que la cellule étudiée et les cellules observées sont liées (via les matrices $[\mathbf{PH}^T]_{e,k}$ et $[\mathbf{HPH}^T]_{e,k}$).

Même si ces cellules - c'est-à-dire celles correspondant à Acanai, Caracarai et São Francisco - ne sont pas directement observées lors de l'assimilation, des cellules proches, voire adjacentes, le sont et peuvent apporter de l'information. Cependant, comme pour les cellules sur le cours principal, certaines "corrections" apportées par l'EnKF ne sont pas adaptées et modifient drastiquement le stock d'eau initial (et donc le débit), détériorant le run libre plutôt que de l'améliorer.

L'impact inadapté des corrélations croisées entre les cellules observées et les cellules de contrôle non-observées a ainsi été mis en évidence systématiquement lors de l'analyse des résultats de l'expérience EE1-direct. La section suivante est consacrée à l'étude plus en détails de ce phénomène pour un cas particulier : la confluence Negro-Branco.

4.3.3 Le cas de la confluence Negro-Branco

La confluence entre le Negro et le Branco au sein du bassin de l'Amazonie est présentée sur la Figure 4.25. Dans un premier lieu, il faut rappeler que le run libre modélise très bien le Negro et le Branco sans assimilation (Figure 4.14-4.15) : les RMSEn par rapport aux débits in situ sont de 15, 6% à Serrinha sur le Negro et de 43, 1% à Caracarai sur le Branco ou encore, le coefficient de Nash-Sutcliffe, toujours par rapport aux débits insitu, est de 0, 9 à Serrinha et de 0, 79 à Caracarai. Malgré la détérioration du débit analyse moyen, les statistiques restent très correctes, c.f. Figures 4.19-4.20 : à Serrinha, les RMSEn par rapport aux débits in situ sont de 28, 3% et le Nash est de 0, 68.

Les débits in situ, ENVISAT, du run libre et d'analyse sont alors comparés en Figure 4.26 aux 9 cellules TRIP formant la confluence Negro-Branco. Ces cellules sont localisées par des cercles rouges



Figure 4.25 – Localisation du Negro et du Branco dans le bassin de l'Amazone et leur confluence.

en Figure 4.28. Sur l'ensemble de ces cellules, on note une forte diminution de la valeur moyenne du débit moyen d'analyse $\overline{Q_{e,k}^a}$ aux cycles d'assimilation 6 - c.f. Figure 4.24 qui représente les 35 premiers cycles d'assimilation à Serrinha ou encore la Figure 4.27 qui représente les 35 premiers cycles d'assimilation de la cellule confluence entre le Negro et le Branco. Le même phénomène est observé sur les cycles suivants, i.e. cycles 41, 76 (non représentés ici). Ces fortes baisses de débits perturbent le cycle hydrologique dans ces cellules mais on remarque qu'ISBA-TRIP retourne vers le régime du run libre directement après la modification en faisant réaugmenter les débits. Enfin, lors des observations directes de ces cellules, l'analyse générée par l'EnKF se rapproche plus ou moins près des observations. Lorsque l'ensemble de débits d'ébauche reste trop éloigné du débit observé et n'est pas assez dispersé, l'EnKF ne parvient pas à apporter une correction suffisante pour ramener le débit vers les observations, générant une analyse détériorée par rapport au run libre. Ce phénomène est à l'origine de la détérioration des statistiques d'analyse pour ces cellules.

Par contre, le phénomène observé au cycle 6 et après - voir Figure 4.27 - est dû à une modification des stocks initiaux réalisés par l'EnKF. Les stations virtuelles actives ce jour-là (le 30 Septembre 2002) sont représentées par des cercles bleus sur la Figure 4.28. Du haut vers le bas, elles se situent sur les cours du Casiquiare et du Siapa (deux affluents du Negro), puis sur les cours du Solimões, Jutai, Juruá, Purús et Madre de Dios en rive droite. En détaillant la contribution de chacune des observations lors de l'étape d'analyse de l'EnKF, il apparaît que la forte réduction de stocks dans ces cellules est essentiellement due à la contribution des deux premières observations, situées en tête du bassin du Negro. La covariance de ces observations avec les cellules de contrôle formant la confluence est très élevée indiquant que ces cellules sont fortement liées. De plus, l'innovation calculée sur ces observations est négative et de très grande amplitude ($-1,6439 \times 10^3 \text{ m}^3 \cdot \text{s}^{-1}$ en moyenne sur l'ensemble pour l'observation 1), ce qui indique alors que le modèle sur-estime fortement le débit vrai a priori. L'accumulation de ces comportements fait que l'action de l'EnKF est de réduire fortement les stocks du modèle pour réduire les débits alors qu'ils devraient être augmentés. Le phénomène décrit ci-dessus est en fait observé sur l'ensemble des cellules du Negro. En effet, la Figure 4.29 représente les débits in situ, ENVISAT, du run libre et analyse moyen au niveau de trois stations in situ le long du cours du Negro (ces stations sont représentées par des carrés noirs sur la Figure 4.28. On remarque toujours la baisse des ensembles de débits au cycle 6 modulo 35. Cependant, le débit moyen de ces stations étant moins élevé qu'à la confluence, ISBA-TRIP retourne plus vite vers le régime du run libre après le cycle 6.

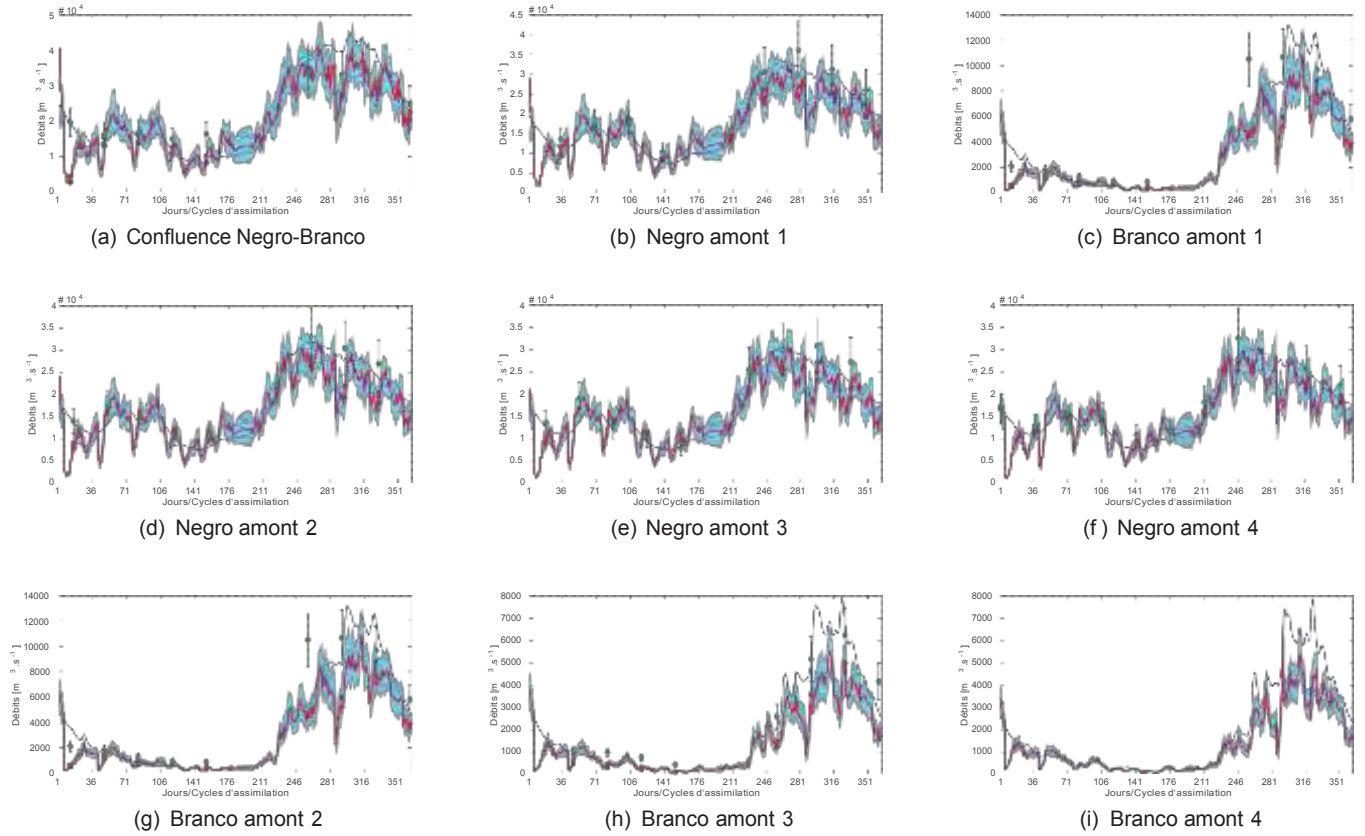


Figure 4.26 – Comparaison des débits de EE1-direct au niveau des cellules confluence entre le Negro et le Branco : Q_k^o (marqueurs losanges verts) et son incertitude, le run libre Q_k^f (ligne fine bleu foncé), $Q_{e,k}^a$ (ligne bleue) et l'ensemble d'ébauche $H(\mathbf{x}_e^b)$ associé (zone bleue délimitant le 1^{er} et le 9^{ème} décile de l'ensemble) et $Q_{e,k}^a$ (ligne rouge). Les marqueurs carrés bleus (respectivement cercles rouges) marque les valeurs des débits moyens d'ensemble d'ébauche (respectivement d'analyse) aux temps d'observations. La zone bleue, délimitant les 1^{er} et 9^{ème} déciles de l'ensemble d'ébauche au cycle k , correspond à l'ensemble juste avant assimilation. Par contre cet ensemble est le résultat des assimilations effectuées aux cycles précédents. C'est pourquoi elle est différente du run libre (ligne fine bleu foncé).

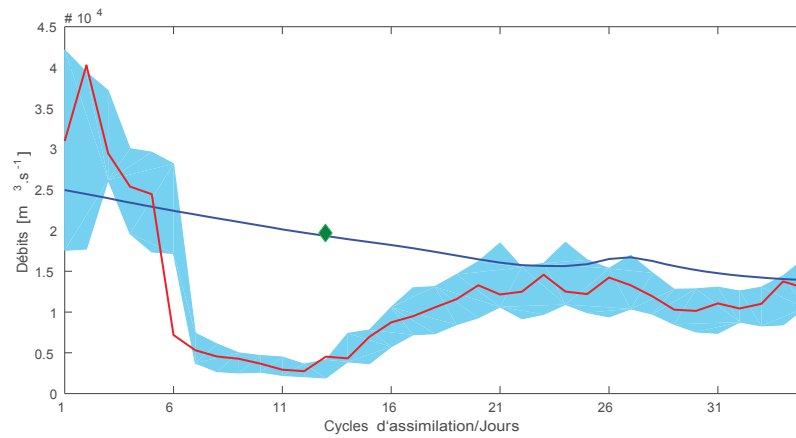


Figure 4.27 – Comparaison des différents débits d'intérêt à la confluence Negro-Branco sur les 35 premiers cycle d'assimilation : débit ENVISAT Q_k^o et leur incertitude (losanges verts), débits run libre Q_k^f (ligne bleue), dispersion de l'ensemble d'ébauche $Q_{e,k}^b = H(\mathbf{x}_{e,k}^b)$ (zone bleue du 1^{er} au 9^{ème} décile) et run analyse moyen $\overline{Q_{e,k}^a}$.

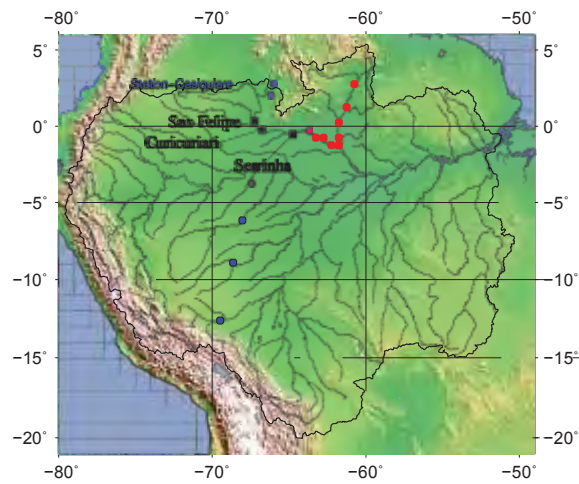


Figure 4.28 – Localisation des stations et cellules impliquées lors du cycle d'assimilation 6 de l'expérience EE1-direct : les stations virtuelles ENVISAT actives au cycle 6 (cercles bleus), les cellules TRIP de la confluence Negro-Branco (cercles rouges) et des stations in situ test (carrés noirs).

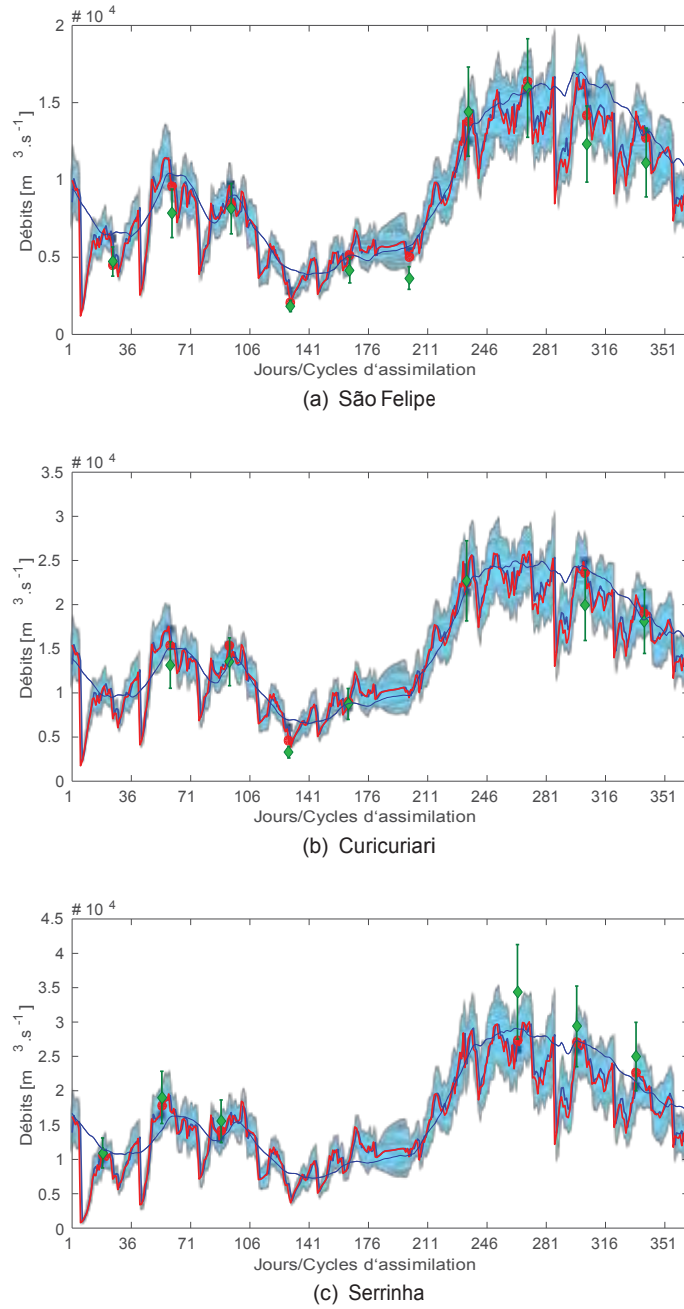


Figure 4.29 – Comparaison des débits de EE1-direct au niveau de stations in situ le long du cours du Negro : Q_k^o (marqueurs losanges verts) et son incertitude, le run libre Q_k^f (ligne fine bleu foncé), $Q_{e,k}^b$ (ligne blue) et l'ensemble d'ébauche $H(\mathbf{x}_e^b)$ associé (zone blue délimitant le 1^{er} et le 9^{ème} décile de l'ensemble) et $Q_{e,k}^a$ (ligne rouge). Les carrés bleus (respectivement cercles rouges) marquent les valeurs des débits moyens d'ensemble d'ébauche (respectivement d'analyse) aux temps d'observations.

Cette courte étude illustre les problèmes que peuvent engendrer l'utilisation de matrices de covariance d'erreurs purement stochastiques : la limitation de la taille n_e de l'ensemble (pour rappel, $n_e = 101$) ou les incertitudes non prises en compte contribuent à l'apparition de corrélations erronées dans les matrices $[\mathbf{PH}^T]_{e,k}$ et $[\mathbf{HPH}^T]_{e,k}$. Aussi, l'estimation de ces matrices à partir d'un ensemble engendre de très hautes corrélations entre les différentes cellules du réseau, dues à la forte ressemblance des cycles hydrologiques saisonniers de certaines cellules. Par exemple, les grands affluents droits que sont le Purus, le Madeira, le Tapajós et le Xingu, ont un cycle saisonnier très similaire. L'utilisation des matrices pleines fait que chaque cellule du bassin est contrainte à chaque cycle d'assimilation, même si les observations sont très éloignées et/ou n'ont aucun rapport du point de vue du réseau hydrologique. D'une part, une observation qui est sur un sous-bassin différent d'une cellule de contrôle donnée aura une influence sur celle-ci. D'autre part, les étapes d'analyse de l'EnKF à chaque cycle contraignent et réduisent systématiquement la dispersion des ensembles de contrôle (voire la faible dispersion des ensembles d'ébauche sur les Figures 4.26 et 4.29, ce qui peut être problématique pour les cycles suivants.

4.3.4 Bilan sur l'expérience EE1-direct

Cette première simulation montre la capacité de l'EnKF à retrouver les stocks d'eau initiaux au sein des cellules du bassin afin de contraindre le modèle ISBA-TRIP à simuler les débits réels, en utilisant des débits dérivés de cotes d'eau ENVISAT et de courbes de tarage. Dans cette étude, on a fait l'hypothèse que les débits des stations in situ peuvent être considérés comme égaux aux débits réels. De manière plus détaillée, l'EnKF est particulièrement efficace sur les cellules du cours principal mais son action est plus limitée sur les autres cellules du bassin. Les observations sur une même station ne sont disponibles que tous les 35 jours (répétitivité du satellite ENVISAT), mais le débit est quotidiennement contraint par le biais des corrélations entre les cellules induites par la construction stochastique des matrices de covariance des erreurs $[\mathbf{PH}^T]_{e,k}$ et $[\mathbf{HPH}^T]_{e,k}$. Les moins bonnes performances de la maquette sont expliquées par les raisons ci-dessous :

- Les performances sont estimées à l'aide de statistiques qui comparent le débit analyse aux débits observés (in situ et ENVISAT). Les statistiques par rapport aux débits ENVISAT sont généralement légèrement meilleures que celles par rapport aux débits in situ. Ceci s'explique facilement car on assimile les débits ENVISAT et non in situ. Si les séries temporelles de débits observés ne sont pas cohérentes - voir les stations d'Itaituba et Belo Monte sur les Figures 4.23n-o pour les stations in situ et Paris et al. (2016) pour l'élaboration des débits ENVISAT - les statistiques analyses peuvent apparaître mauvaises.
- L'ensemble des stations in situ et ENVISAT sont initialement définies à une échelle bien plus fine qu'ISBA-TRIP. Un traitement leur a alors été appliqué pour les associer à une cellule de la grille du modèle ISBA-TRIP en fonction du réseau hydrologique et des aires drainées au niveau des stations. Cependant, certaines stations ont pu être mal associées et la comparaison de débits trop différents entre contrôle et observations entraînent des modifications aberrantes des stocks initiaux et débits.
- Les ensembles d'ébauche apparaissent des fois pas ou très peu dispersés. Quand les observations sont très éloignées de l'ensemble d'ébauche, les corrections apportées par l'EnKF sont alors insuffisantes.
- Enfin, la limitation qui apparaît la plus importante dans la simulation EE1-direct est l'utilisation des matrices de covariance des erreurs purement stochastiques. Ces matrices génèrent des débits analyses très chaotiques avec de nombreux pics et creux autour de la tendance saisonnière. Ces déformations brutales, qui apparaissent entre les temps d'observations, sont le fait de "corrections" apportées sur le stock de la cellule et pondérées par les corrélations non-nulles entre les cellules observées et les autres non-observées. Comme pour le cas de la

confluence entre le Negro et le Branco, on remarque des observations situées sur des petits affluents éloignés, voire sur des cours d'eau dans des sous-bassins complètement indépendants qui influencent, à tort, les débits des stations non-observées. Cette influence illogique perturbe alors complètement le cycle hydrologique à ces cellules.

L'idée est alors de développer un algorithme de traitement de ces matrices de covariance d'erreurs pour limiter la zone d'influence des observations.

4.4 Performances de l'EnKF avec localisation

Dans un premier temps, on étudie comment se comporte la maquette d'assimilation quand, au cours d'un cycle d'assimilation, seules les cellules d'ISBA-TRIP qui sont observées sont corrigées par l'EnKF. Cette hypothèse constitue l'expérience EE1-diag. On étudiera ensuite l'influence du procédé de localisation dans l'expérience EE1-local.

4.4.1 Expérience préliminaire : EE1-diag

4.4.1.1 Configuration de l'expérience

Cette expérience possède la même configuration que EE1-direct, voir Section 4.2.4 et Tableau 4.1. La différence réside dans la construction des matrices de covariance d'erreurs $[\mathbf{PH}^T]_{e,k}$ et $[\mathbf{HPH}^T]_{e,k}$. De ce fait, les matrices sont forcées "diagonales". La matrice $[\mathbf{HPH}^T]_{e,k}$ est carrée, il est alors facile de récupérer sa diagonale. $[\mathbf{PH}^T]_{e,k}$, au contraire, est une matrice non-carrée de dimension n_x lignes $\times n_{y,k}$ colonnes, avec $n_x > n_{y,k}$. L'algorithme pour construire ces matrices se déroule selon les étapes suivantes :

- $[\mathbf{PH}^T]_{e,k}$ et $[\mathbf{HPH}^T]_{e,k}$ sont d'abord calculées stochastiquement suivant Eq. 4.16-4.17.
- $[\mathbf{HPH}^T]_{e,k}$ représente les covariances des variables de contrôle entre elles, exprimées dans l'espace des observations. La matrice étant carrée, on ne conserve que la diagonale de la matrice.
- $[\mathbf{PH}^T]_{e,k}$ rassemble les covariances entre les n_x variables de contrôle dans l'espace de contrôle et ces mêmes variables de contrôle, exprimées dans l'espace des observations de taille $n_y < n_x$. Cette matrice n'est donc pas carrée. Le traitement de la matrice consiste à forcer la majorité de ses éléments à 0 sauf 1 élément par colonne. L'élément non-nul d'une colonne donnée j se trouve à la ligne de la cellule de contrôle correspondant à l'observation j .

À chaque cycle d'assimilation, les matrices $[\mathbf{PH}^T]_{e,k}$ et $[\mathbf{HPH}^T]_{e,k}$ sont alors d'abord estimées stochastiquement (comme pour EE1-direct) puis traitées pour être forcées sous forme "diagonale" - d'où le nom de la simulation. Les matrices "diagonales" sont alors utilisées au sein de l'analyse de l'EnKF pour estimer la matrice de gain de Kalman $\mathbf{K}_e$ (voir Chapitre 1, Section 1.4.5.3, Eq. 1.45).

4.4.1.2 Résultats généraux

Le Tableau 4.6 compare les pourcentages de variations des RMSEn entre le run libre et l'expérience EE1-diag en global (c'est-à-dire en moyennant spatialement les comparaisons pour chaque station in situ ou pour chaque station virtuelle ENVISAT). Ensuite, les Figures 4.30-4.33 illustrent, selon le même schéma que la Figure 4.19-4.22, les statistiques générales de l'expérience EE1-diag. Ces figures montrent une amélioration des statistiques des débits analyse par rapport à celles du run libre sur

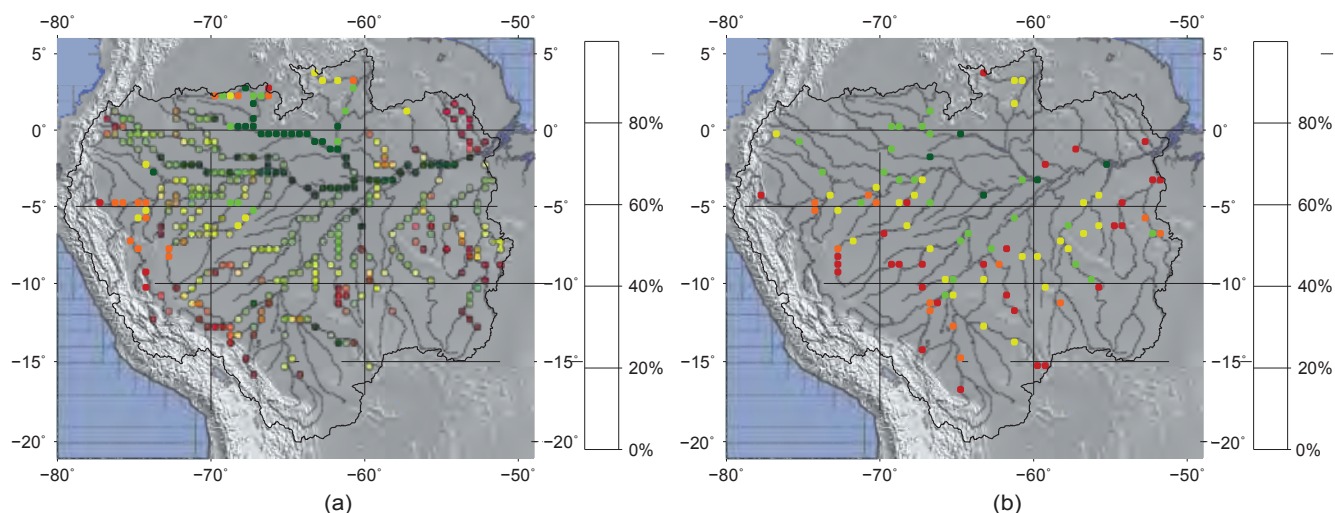


Figure 4.30 – Coefficients d’erreurs moyennes quadratiques RMSEn de l’expérience d’AD EE1-diag par rapport aux débits ENVISAT (a) et par rapport aux débits in situ (b).

certaines des zones critiques de EE1-direct : on n’observe plus de dégradations des statistiques sur les confluences Negro-Branco et Xingu-Amazonas. Les débits analyse du cours principal apparaissent toujours bien corrigés (moins de 20% de RMSEn) de l’embouchure jusqu’à la confluence avec le Juruá en remontant le cours du fleuve, ainsi que le long du cours du Negro. Par rapport à l’expérience EE1-direct, le cours principal avec le cours du Putulayo, au Nord-Ouest du bassin, sont moins bien corrigés, alors que les cours des principaux affluents droits, c’est-à-dire le Juruá, le Purus, le Madeira et le Xingu, sont mieux gérés par l’expérience EE1-diag.

Statistiques globales	Run libre	Run analyse EE1-direct	Run analyse EE1-diag	Unité
RMSE ^{situ}	$2,79 \times 10^3$	$1,86 \times 10^3$ (–33%)	$2,26 \times 10^3$ (–19%)	$\text{m}^3.\text{s}^{-1}$
RMSEn ^{situ}	73,61	56,01 (–24%)	59,17 (–20%)	%
E ^{situ}	–11,57	–3,90	–4,06	–
RMSE ^o	$5,71 \times 10^3$	$2,48 \times 10^3$ (–56%)	$3,21 \times 10^3$ (–44%)	$\text{m}^3.\text{s}^{-1}$
RMSEn ^o	75,29	41,87 (–44%)	48,00 (–36%)	%
E ^o	–134,32	–32,13	–49,930	–

Table 4.6 – Comparaison des statistiques globales du run libre et les expériences EE1-direct et EE1-diag.

Cette configuration est moins efficace que EE1-direct pour les cellules du cours principal mais semble être plus adaptée pour les cellules des affluents (exceptés Caracará, Itaituba et Uhe Belo Monte). Le paragraphe suivant se concentre sur les tendances des débits analyse aux stations in situ.

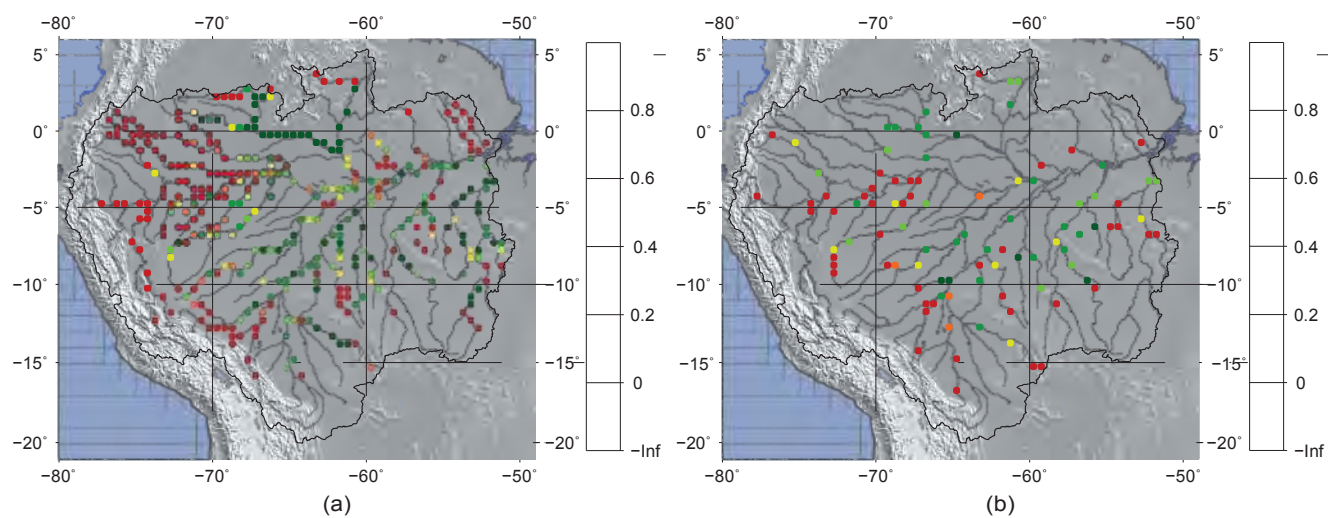


Figure 4.31 – Coefficients de Nash-Sutcliffe E de l'expérience d'AD EE1-diag par rapport aux débits ENVISAT (a) et par rapport aux débits in situ (b).

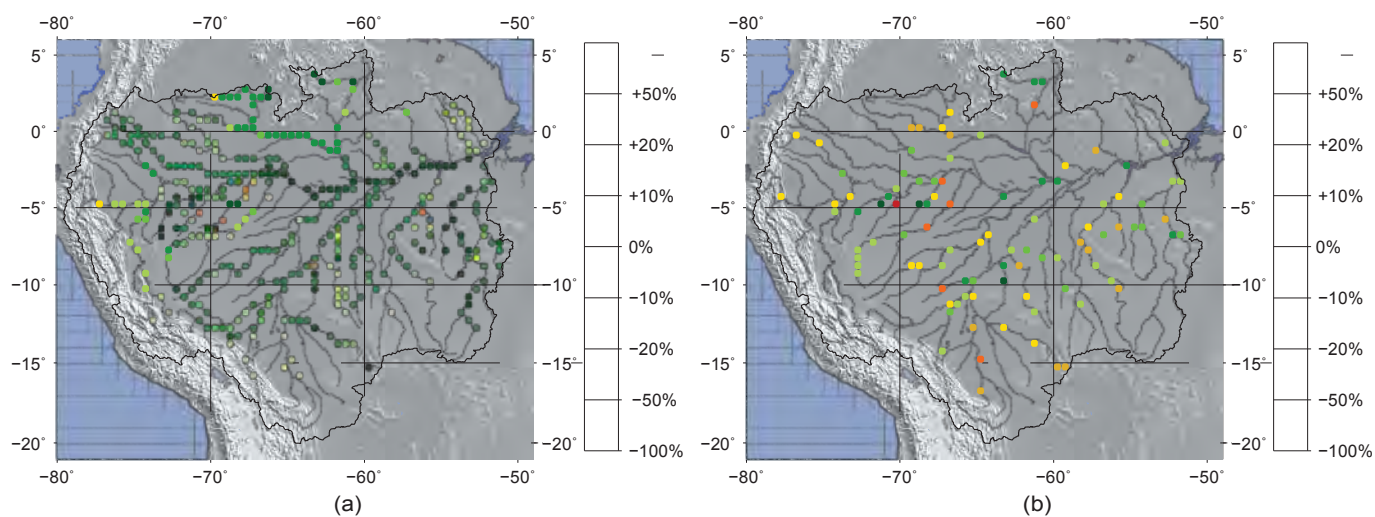


Figure 4.32 – Pourcentage d'augmentation/de réduction du RMSEn pour l'expérience d'AD EE1-diag par rapport aux débits ENVISAT (a) et par rapport aux débits in situ (b).

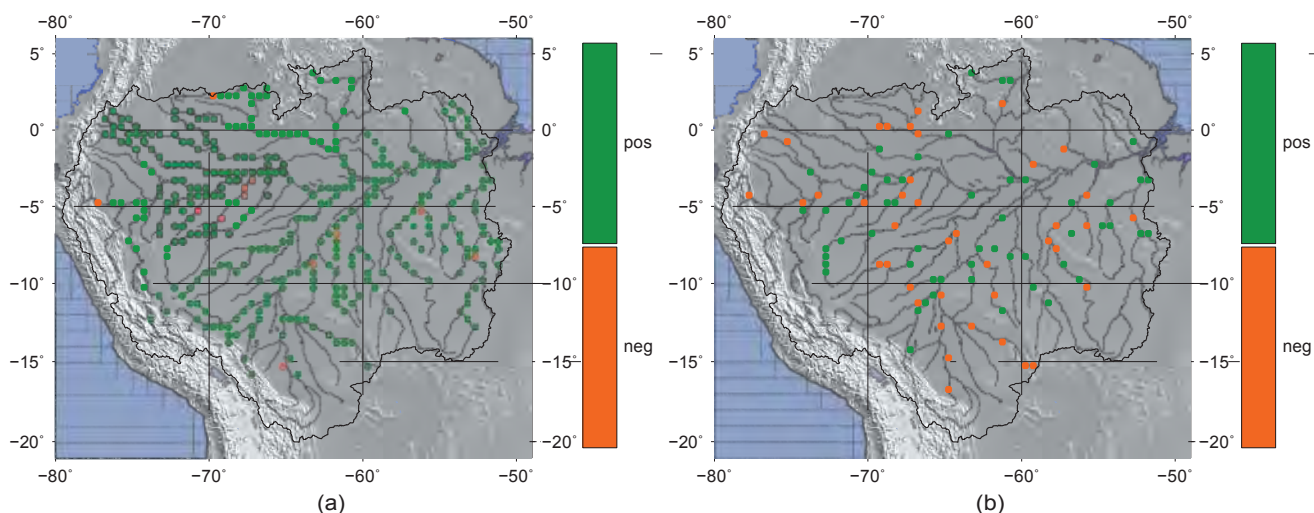


Figure 4.33 – Augmentation (vert)/réduction (rouge) du Nash E pour l'expérience d'AD EE1-diag par rapport aux débits ENVISAT (a) et par rapport aux débits in situ (b).

4.4.1.3 Comportement aux stations tests

La Figure 4.34 montre les séries temporelles des débits d'intérêt pour la simulation EE1-diag aux stations in situ test utilisées pour EE1-direct. Le Tableau 4.7 compare les variations de RMSEn et de Nash aux différentes stations entre le run libre et l'analyse pour la simulation EE1-diag.

Premièrement, on remarque que la dispersion de l'ensemble d'ébauche $H(\mathbf{x}_e^b)$ (zone bleue sur les différents graphiques de la Figure 4.34) est généralement bien plus importante que celle de EE1-direct. Contrairement à EE1-direct, qui contraignait l'ensemble des variables de contrôle à chaque cycle et restreignait la dispersion des ensembles de contrôle, le fait d'utiliser des matrices diagonales dans EE1-diag ne réduit plus autant la dispersion des ensembles de contrôle à chaque cycle. Ceci est probablement dû au fait que la correction est moins forte et contraint donc moins l'ensemble.

Ensuite, les débits moyens analysés présentent des tendances différentes de celles de EE1-direct. Les hydrographes de débits sont bien plus réguliers et suivent exactement les variations du run libre avec un biais plus ou moins important. Pour toutes les stations observées, et en particulier à Tamashiyacu (Figure 4.34a), Tabatinga (Figure 4.34b) ou encore Santos Dumont (Figure 4.34k) et Manicoré (Figure 4.34m), le débit analyse présente une variation brutale (pics ou creux) les jours d'observations. Le débit se rapproche très brièvement de la valeur du débit observé. Ce comportement est le résultat direct d'une correction de stock générée par l'EnKF. La localisation des corrections est caractéristique de l'utilisation des matrices diagonales.

Cependant d'autres stations ne sont pas ou très peu affectées par les observations comme par exemple à Serrinha (Figure 4.34h), Labreá (Figure 4.34l) ou Itaituba (Figure 4.34m), et le débit moyen analyse $\overline{Q_{\bullet,k}^a}$ reste très proche du run libre. A Serrinha (Figure 4.34h), le run libre est déjà très proche des observations (0, 16% de RMSEn et un Nash égal à 0, 90). La norme du vecteur d'innovation est alors très faible et très peu de corrections sont apportées aux temps d'observation. Pour les stations sur les affluents droits - Labreá (Figure 4.34l) et Itaituba (Figure 4.34m) notamment - les observations

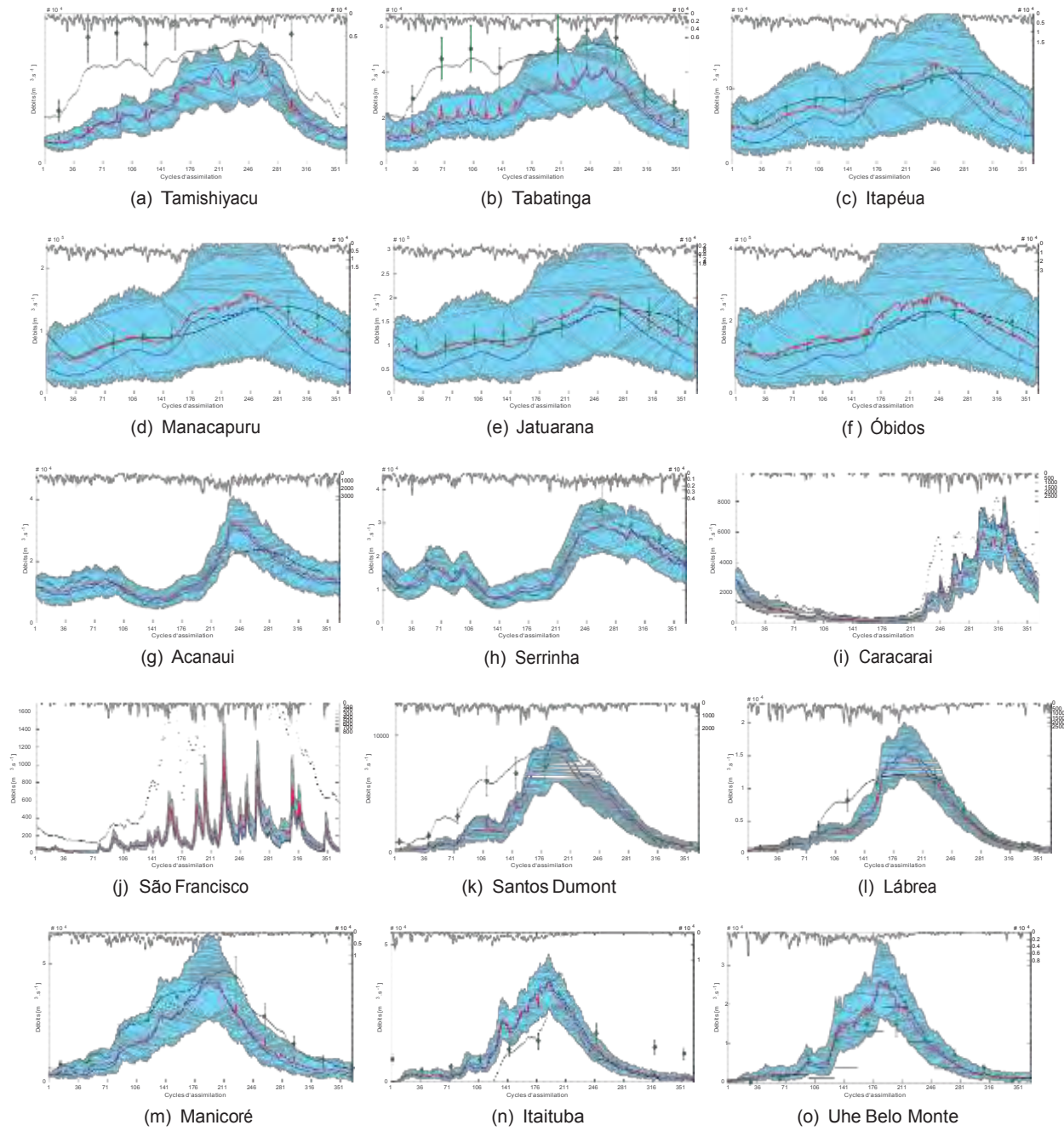


Figure 4.34 – Comparaison des différents débits d'intérêt aux stations in situ pour l'expérience EE1-diag : débit ENVISAT Q_k^o et leur incertitude (losanges verts), débits in situ Q_k^{situ} (points noirs), débits run libre Q_k^f (ligne bleu), dispersion de l'ensemble d'ébauche $Q_{e,k}^b = H(\mathbf{x}_{e,k}^b)$ (zone bleue du 1^{er} au 9^{ème} décile) et run analyse moyen $\overline{Q_{e,k}^a}$. L'ensemble des précipitations $\mathbf{P}_e$ est tracé en zone grise sur l'axe des ordonnées inversé.

Indice de la station	RMSEn Q_k^f	RMSEn $Q_{\bullet,k}^a$	E Q_k^f	E $Q_{\bullet,k}^a$
Tamashiyacu	0,401	0,441	-1,783	-2,230
Tabatinga	0,449	0,404	-2,221	-1,612
Itapêua	0,332	0,192	-1,641	0,115
Manacapuru	0,364	0,212	-1,123	0,207
Jatuarana	0,270	0,154	-0,210	0,608
Obidos	0,291	0,157	-0,315	0,617
Acanauí	0,195	0,192	0,661	0,671
Serrinha	0,156	0,150	0,902	0,909
Caracará	0,431	0,529	0,788	0,681
São Francisco	1,059	0,993	0,945	-0,782
Santos Dumont	0,380	0,472	0,684	0,513
Lábrea	0,371	0,372	0,740	0,738
Manicoré	0,360	0,336	0,635	0,681
Itaituba	0,558	0,561	0,526	0,523
Uhe Belo Monte	0,852	0,834	0,402	0,428

Table 4.7 – Performances de l’EnKF lors de l’expérience EE1-diag : comparaison des erreurs quadratiques moyennes normalisées (RMSEn) et des coefficients d’efficacité de Nash-Sutcliffe (E) du débit analyse moyen $\overline{Q_{\bullet,k}^a}$ et du débit du run libre Q_k^f , calculés en comparaison aux débits observés in situ.

sont trop éloignées de l’ensemble d’ébauche pour permettre à l’EnKF d’apporter une correction. Dans ces cas, il apparaît que l’information apportée par seulement l’observation à la station in situ n’est pas suffisante.

Dans la même idée, les stations in situ qui n’ont pas de stations ENVISAT associées - ici Acanauí (Figure 4.34g), Caracará (Figure 4.34i) et São Francisco (Figure 4.34j) - sont très peu déviées du run libre. Cependant, des différences entre le débit analyse moyen et le débit du run libre apparaissent clairement à ces stations. Ceci est aussi le cas aux autres stations in situ, où le débit analyse dévie du débit du run libre à des temps qui ne sont pas des temps d’observations propres de la station. Ces déviations illustrent la propagation d’une correction apportée par l’EnKF sur une cellule observée vers les cellules avalées. En d’autres termes, il existe des corrélations entre les cellules du réseau TRIP, intrinsèques au modèle, qui proviennent de l’utilisation du schéma de routage.

Cette expérience a permis de mettre en évidence les corrélations intrinsèques à TRIP qui lient une cellule avec les cellules en aval de celle-ci. Cette corrélation fait que, sur un cycle d’assimilation, une correction apportée sur une cellule observée, va se propager et être observée sur les cellules avalées. Une cellule donnée est alors systématiquement corrélée à des cellules en amont et en aval de celle-ci. Ce phénomène est aussi directement lié au fait que l’on cherche à corriger le stock d’eau initial lors de chaque cycle d’assimilation. En effet, après une étape d’analyse, et avant de lancer le cycle d’assimilation suivant, ISBA-TRIP doit être propagé à partir de la condition initiale d’analyse jusqu’au prochain temps d’assimilation, propageant automatiquement les modifications faites par l’EnKF. Si cette modification de l’algorithme introduite dans l’expérience EE1-diag est moins efficace que l’expérience EE1-direct sur les cellules du cours principal, il semblerait qu’une correction localisée soit plus adaptée pour les cellules situées sur les affluents. Malgré tout, utiliser des matrices seulement diagonales n’est pas suffisant, surtout pour les cellules qui ne sont jamais observées lors de l’assimilation - e.g. Acanauí (Figure 4.34g), Caracará (Figure 4.34i) ou encore São

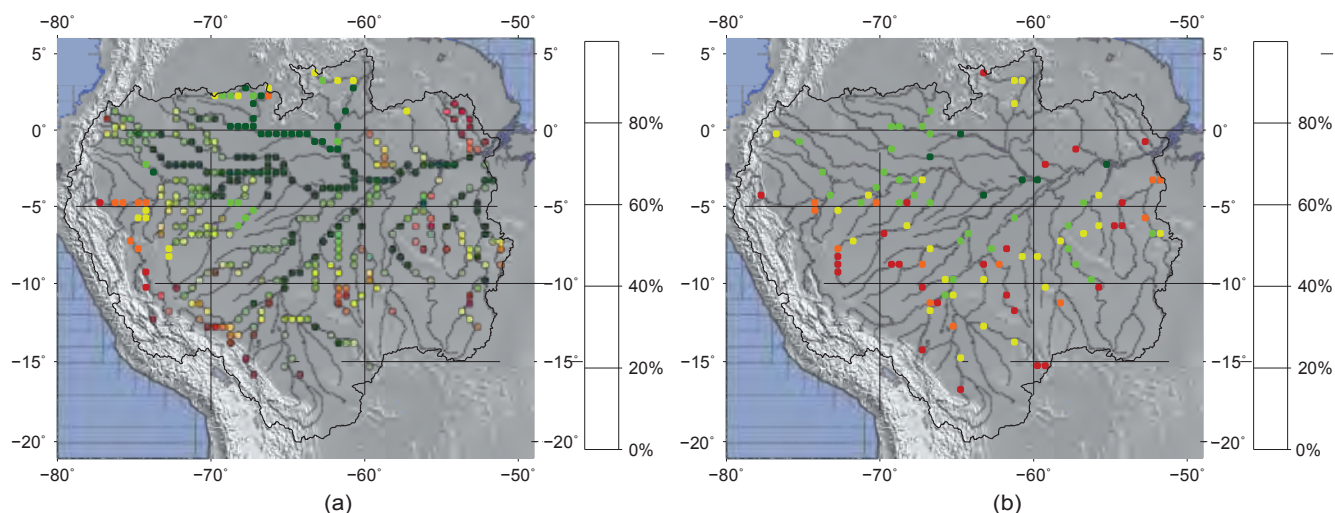


Figure 4.35 – Coefficients d’erreurs moyennes quadratiques RMSEn de l’expérience d’AD EE1-local par rapport aux débits ENVISAT (a) et par rapport aux débits in situ (b).

Francisco (Figure 4.34j). L’idée est alors de contruire des matrices de covariance des erreurs localisées dans lesquelles on définit pour chaque observation une zone d’influence limitée autour d’elle et suivant le réseau hydrologique.

4.4.2 Expérience-test du masque de localisation : EE1-local

4.4.2.1 Configuration de l’expérience

Cette expérience possède la même configuration que EE1-direct et EE1-diag, voir Section (4.2.4) et Tableau (4.1). Dans cette simulation, on applique maintenant le masque de localisation introduit en Section (4.2.3.3) pour contruire les matrices de covariance d’erreurs $[\mathbf{PH}^T]_{e,k}$ et $[\mathbf{HPH}^T]_{e,k}$.

4.4.2.2 Résultats généraux

Les Figures 4.35-4.38 illustrent, selon le même schéma que les Figures 4.19-4.22, les statistiques générales de l’expérience EE1-local et le Tableau 4.8 donne quantitativement les performances globales de l’expérience et permet la comparaison avec les résultats des autres expériences.

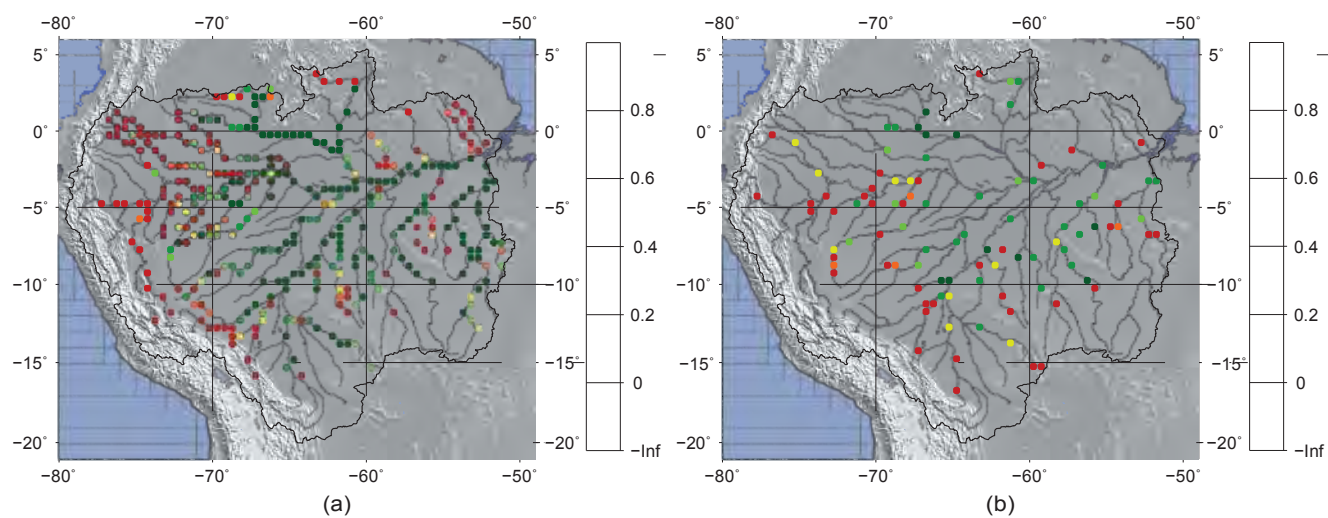


Figure 4.36 - Coefficients de Nash-Sutcliffe E de l'expérience d'AD EE1-local par rapport aux débits ENVISAT (a) et par rapport aux débits in situ (b).

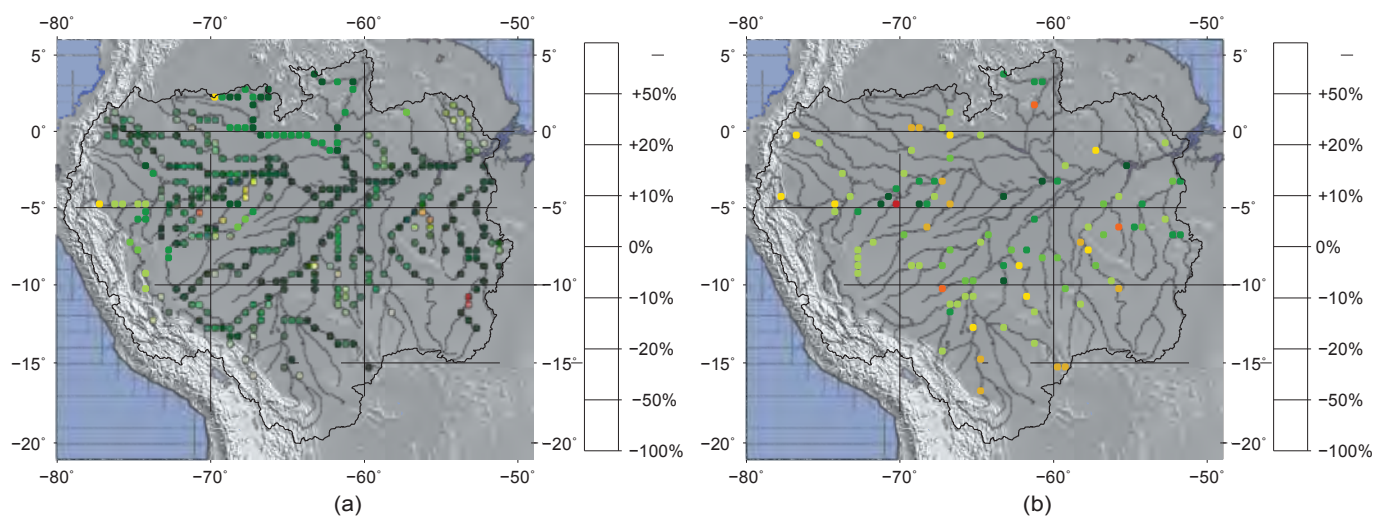


Figure 4.37 - Pourcentage d'augmentation/de réduction du RMSEn pour l'expérience d'AD EE1-local par rapport aux débits ENVISAT (a) et par rapport aux débits in situ (b).

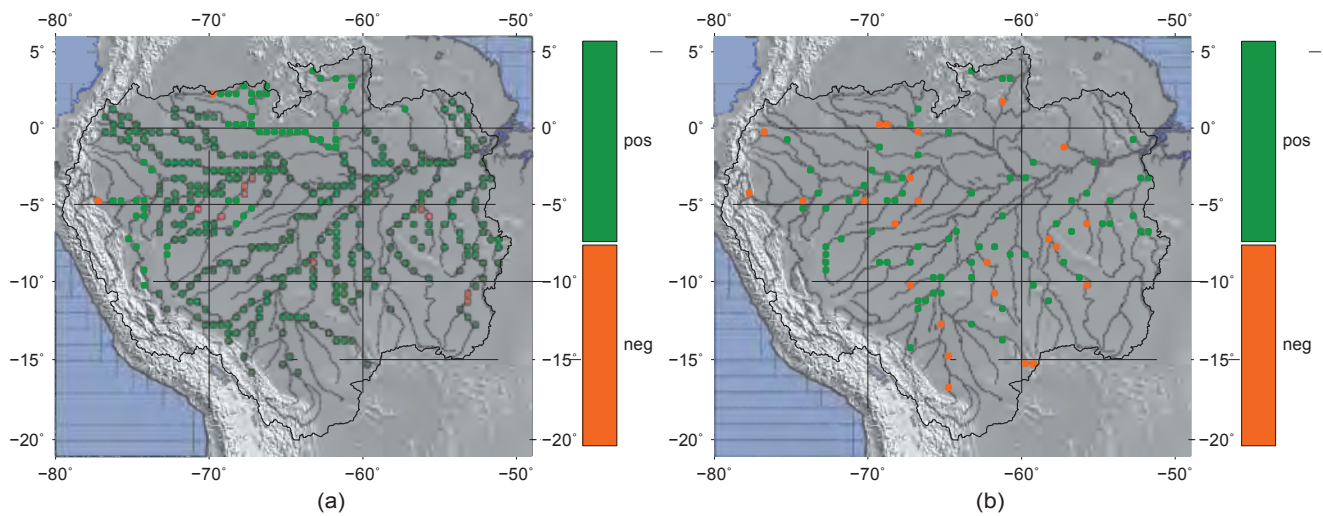


Figure 4.38 – Augmentation (vert)/réduction (rouge) du Nash E pour l’expérience d’AD EE1-local par rapport aux débits ENVISAT (a) et par rapport aux débits in situ (b).

Statistiques globales	Run libre	Run analyse EE1-direct	Run analyse EE1-diag	Run analyse EE1-local	Unité
$RMSE^{situ}$	$2,79 \times 10^3$	$1,86 \times 10^3$ (–33%)	$2,26 \times 10^3$ (–19%)	$1,94 \times 10^3$ (–31%)	$m^3.s^{-1}$
$RMSEn^{situ}$	73,61	56,01 (–24%)	59,17 (–20%)	57,21 (–22%)	%
E^{situ}	–11,57	–3,90	–4,06	–3,90	–
$RMSE^o$	$5,71 \times 10^3$	$2,48 \times 10^3$ (–56%)	$3,21 \times 10^3$ (–44%)	$1,86 \times 10^3$ (–67%)	$m^3.s^{-1}$
$RMSEn^o$	75,29	41,87 (–44%)	48,00 (–36%)	37,19 (–51%)	%
E^o	–134,32	–32,13	–49,93	–27,23	–

Table 4.8 – Comparaison des statistiques globales du run libre et des expériences EE1-direct, EE1-diag et EE1-local.

Les figures montrent une amélioration générale des débits analyse par rapport aux débits du run libre, en particulier le long du cours principal (moins de 20% de $RMSEn$) et des principaux affluents de l’Amazonie (moins de 40% de $RMSEn$) - le Negro, le Purus, le Madeira et aussi le Tapajós. Cependant, l’EnKF ne parvient toujours pas à retrouver les débits observés aux stations situées en têtes de bassins.

Pour évaluer du bon fonctionnement et de la pertinence du masque de localisation choisi pour cette simulation, on va à présent regarder plus en détail l’allure des débits analyse aux différentes stations in situ test.

4.4.2.3 Comportement aux stations tests

Les Figures 4.39 montrent les séries temporelles des débits d’intérêt pour la simulation EE1-local aux stations in situ test utilisées pour EE1-direct/EE1-diag et le Tableau 4.9 compare les variations de $RMSEn$ et de Nash aux différentes stations entre le run libre et l’analyse.

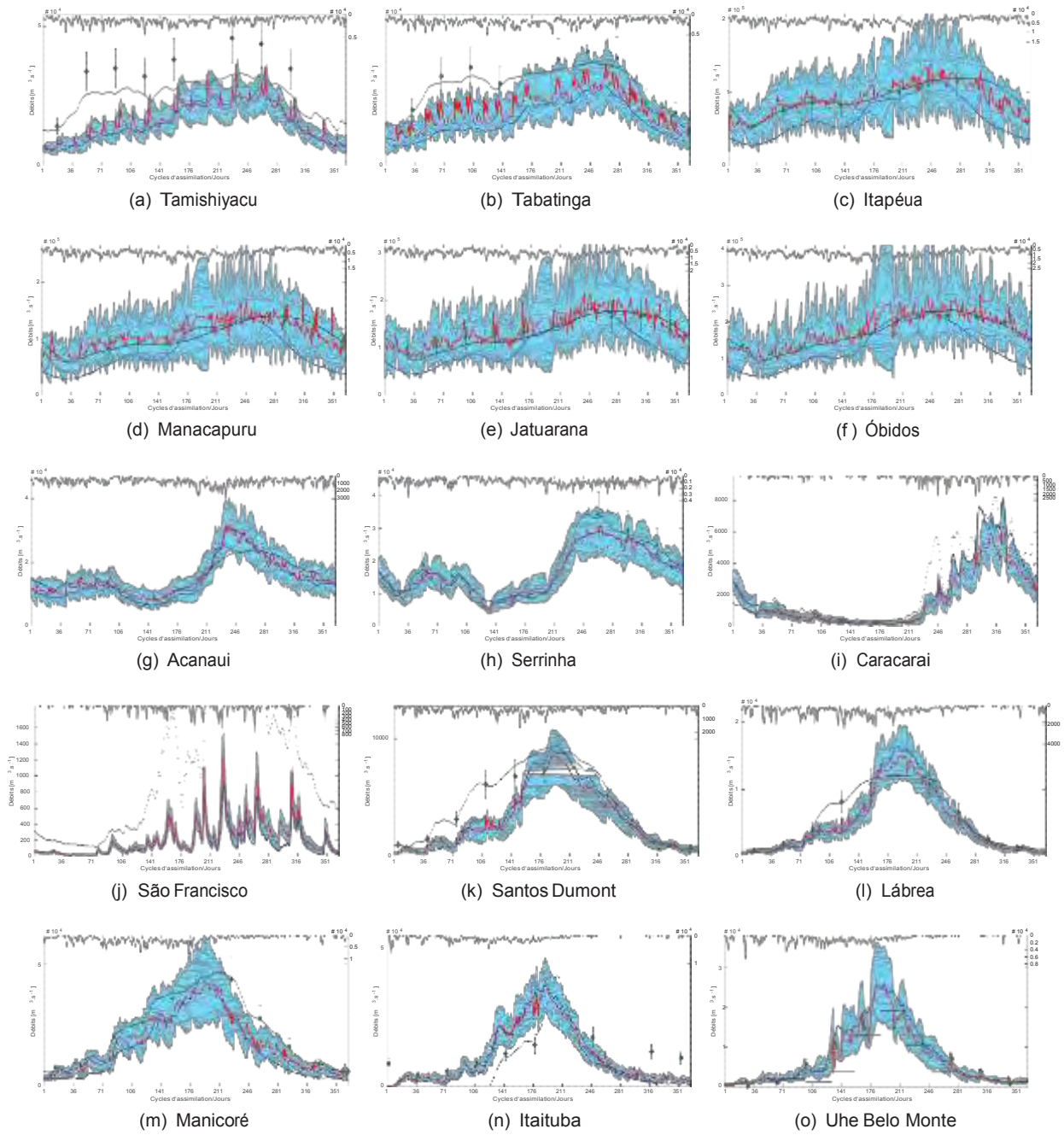


Figure 4.39 – Expérience EE1-local : Comparaison des différents débits d'intérêt aux stations in situ : débit ENVISAT Q_k^o et leur incertitude (losanges verts), débits in situ Q_k^{situ} (points noirs), débits run libre Q_k^f (ligne bleu), dispersion de l'ensemble d'ébauche $Q_{e,k}^b = H(\mathbf{X}_{e,k}^b)$ (zone bleue du 1^{er} au 9^{ème} décile) et run analyse moyen $\overline{Q_{\bullet,k}^a}$ de l'expérience EE1-local. L'ensemble des précipitations $\mathbf{P}_e$ est tracé en zone grise sur l'axe des ordonnées inversé.

Indice de la station	RMSEn Q_k^f	RMSEn $Q_{\bullet,k}^a$	E Q_k^f	E $Q_{\bullet,k}^a$
Tamashiyacu	0,401	0,384	-1,783	-1,450
Tabatinga	0,449	0,314	-2,221	-0,573
Itapéua	0,332	0,120	-1,641	0,654
Manacapuru	0,364	0,169	-1,123	0,496
Jatuarana	0,270	0,139	-0,210	0,681
Obidos	0,291	0,127	-0,315	0,751
Acanauí	0,195	0,172	0,661	0,736
Serrinha	0,156	0,148	0,902	0,913
Caracarai	0,431	0,537	0,788	0,672
São Francisco	1,042	0,993	0,945	-0,726
Santos Dumont	0,380	0,432	0,684	0,591
Lábrea	0,371	0,334	0,740	0,789
Manicoré	0,360	0,280	0,635	0,780
Itaituba	0,558	0,537	0,526	0,563
Uhe Belo Monte	0,852	0,673	0,402	0,628

Table 4.9 – Performances de l’EnKF lors de l’expérience EE1-local : comparaison des erreurs quadratiques moyennes normalisées (RMSEn) et des coefficients d’efficacité de Nash-Sutcliffe (E) du débit analyse moyen $\overline{Q_{\bullet,k}^a}$ et du débit du run libre Q_k^f , calculés à partir des débits observés in situ.

Le comportement du débit analyse moyen aux stations in situ le long du cours principal (voir Figures 4.39a-f) se rapproche de celui observé lors des expériences précédentes :

- Aux stations les plus en amont, à savoir Tamashiyacu et Tabatinga (Figures 4.39a-b), on observe plusieurs pics de débits. Certains pics sont synchronisés avec un temps d’observation (pics alignés avec une observation ENVISAT, représenté par un marqueur vert sur ces figures). Ce comportement se rapproche de ceux observés lors de l’expérience EE1-diag (Figure 4.34). D’autres pics de débits sont observés mais sont désynchronisés des observations ENVISAT. Au même titre que les pics observés lors de l’expérience EE1-direct, ces pics sont dus à des corrélations non-nulles entre les cellules observées et les cellules non-observées. Dans le cadre de la localisation, ces pics indiquent que des cellules de la zone d’influence sont observées.
- Aux autres stations plus en aval, c’est-à-dire de Itapéua à Óbidos (Figure 4.39c-f), le comportement est plutôt chaotique comme celui observé pour l’expérience EE1-direct (Figure 4.23) mais avec des variations moins importantes.

Aux stations in situ situées sur les affluents (voir Figure 4.39g-o), le débit analyse moyen est généralement proche du débit du run libre. Tout comme à Tamashiyacu et Tabatinga, l’allure du débit analyse moyen n’est pas chaotique mais présente des pics - par exemple à Santos Dumont (Figure 4.39k ou à Labréa (Figure 4.39l) - ou creux - par exemple à Serrinha (Figure 4.39h et à Itaituba (Figure 4.39n) - ponctuels qui sont parfois synchronisés avec les observations ENVISAT de la station virtuelle associée à la station in situ, et parfois décalés.

La Figure 4.40 détaillent le comportement du débit analyse moyen à Santos Dumont ainsi qu’en toutes les cellules TRIP se situant dans la zone d’influence de celle de Santos Dumont. Le débit analyse moyen est alors comparé au run libre ainsi qu’aux débits in situ et ENVISAT (si disponibles), sur les trois premières périodes de répétitivité ENVISAT (jour 1 à 105). La station virtuelle à Santos Dumont est active le 6^{ème} jour du cycle de répétitivité ENVISAT (qui est égal à 35 jours) et les autres stations virtuelles ENVISAT se trouvant dans la zone d’influence sont actives le 11^{ème} (cellule 2 en

Figure 4.40a et débits en Figure 4.40c), le 22^{ème} (cellule 3 en Figure 4.40a et débits en Figure 4.40d) et le 14^{ème} (cellule 4 en Figure 4.40a et débits en Figure 4.40e) jour du cycle de répétitivité ENVISAT. Le débit analyse moyen à Santos Dumont, très similaire aux débits aux cellules 2 à 5, présente 4 pics successifs sur chaque cycle de 35 jours. Le premier pic est synchronisé avec l'observation ENVISAT ayant lieu à Santos Dumont et résulte d'une correction apportée par l'EnKF. Les autres pics sont alors associés à des corrections faites dans les cellules voisines qui sont répercutées sur Santos Dumont via les corrélations non annulées par la localisation : le deuxième pic avec l'observation de la cellule 2 (Figure 4.40c) au 11^{ème} jour de la répétitivité, le troisième pic avec l'observation de la cellule 4 (Figure 4.40e) au 14^{ème} jour de la répétitivité et le quatrième et dernier pic avec la cellule 3 (Figure 4.40d) au 22^{ème} jour de la répétitivité.

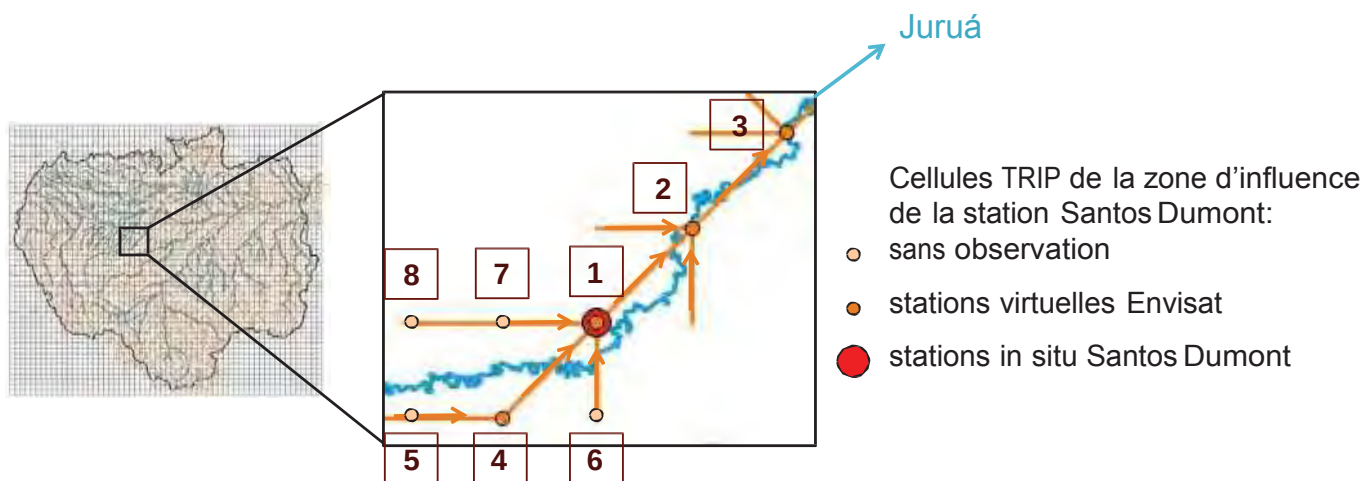
Des comportements similaires sont ainsi observés pour la majorité des stations in situ étudiées. L'expérience EE1-local apparaît alors comme une bonne alternative à l'expérience EE1-direct, qui engendre des débits complètement chaotiques et donc irréalistes. Cependant, entre deux observations consécutives au sein d'une même zone d'influence - c'est-à-dire deux observations situées à deux stations virtuelles différentes mais assez proches pour être dans la zone d'influence de l'une et l'autre - plusieurs jours peuvent se passer sans qu'aucune correction ne soit apportée. Pendant cette période, comme le modèle n'est plus contraint par les observations, le débit retourne vers la tendance du run libre. Ce comportement engendre l'apparition de pics et de creux dans l'hydrogramme.

4.4.3 Conclusions sur la localisation

Le masque de localisation a été construit afin d'éviter l'apparition de corrélations résiduelles entre cellules éloignées et/ou se situant sur des sous-bassins versant différents. Dans ce sens, le masque de localisation, tel qu'il a été construit, répond à cette contrainte. En effet, les seules cellules pouvant impacter une cellule observée quelconque dans le bassin sont celles définies comme étant à l'intérieur d'une certaine zone d'influence comprenant la cellule observée même ainsi qu'un nombre limité de cellules adjacentes selon le réseau hydrologique TRIP. Les résultats de l'expérience EE1-local permettent de valider ce premier masque de localisation. En effet, les statistiques des débits analyses sont globalement améliorées par rapport à celles du run libre.

Cependant, à l'instar des expériences précédentes - EE1-direct et EE1-diag - les performances de l'EnKF sont limitées par la faible couverture temporelle des observations ENVISAT (une observation au plus tous les 35 jours). L'utilisation de la localisation permet de contraindre le modèle plusieurs fois au cours d'un cycle de 35 jours grâce à l'utilisation des covariances non-nulles entre cellules adjacentes. Malgré tout, pendant plusieurs jours sans observations (directes ou indirectes), le modèle retourne vers le run libre.

Par ailleurs, on admet facilement que le masque de localisation actuel n'est pas parfait et peut être amélioré. La difficulté majeure pour construire le masque de localisation, tel qu'il est défini dans cette étude, réside dans le choix de la taille de la zone d'influence. Actuellement, la taille de la zone d'influence est pré-déterminée pour chaque cellule et demeure invariée tout le long de l'année. La taille des zones d'influence a été initialement déterminée en fonction de la vitesse d'écoulement moyenne sur une année dans les différentes zones hydro-géomorphologiques divisant le bassin - pour rappel, voir ces 9 zones en Figure 4.8. Une évolution possible du masque de localisation serait de définir une taille de la zone d'influence propre à chaque cellule et pouvant évoluer dans le temps avec la vitesse d'écoulement et l'année hydrologique. Ainsi, en période de hautes eaux par exemple, où les vitesses d'écoulement sont les plus élevées, les zones d'influence seraient plus grandes et plus de cellules seront influencées par une observation. Cela reviendrait à définir des matrices de covariance d'erreurs



(a) Cellules de la zone d'influence de Santos Dumont

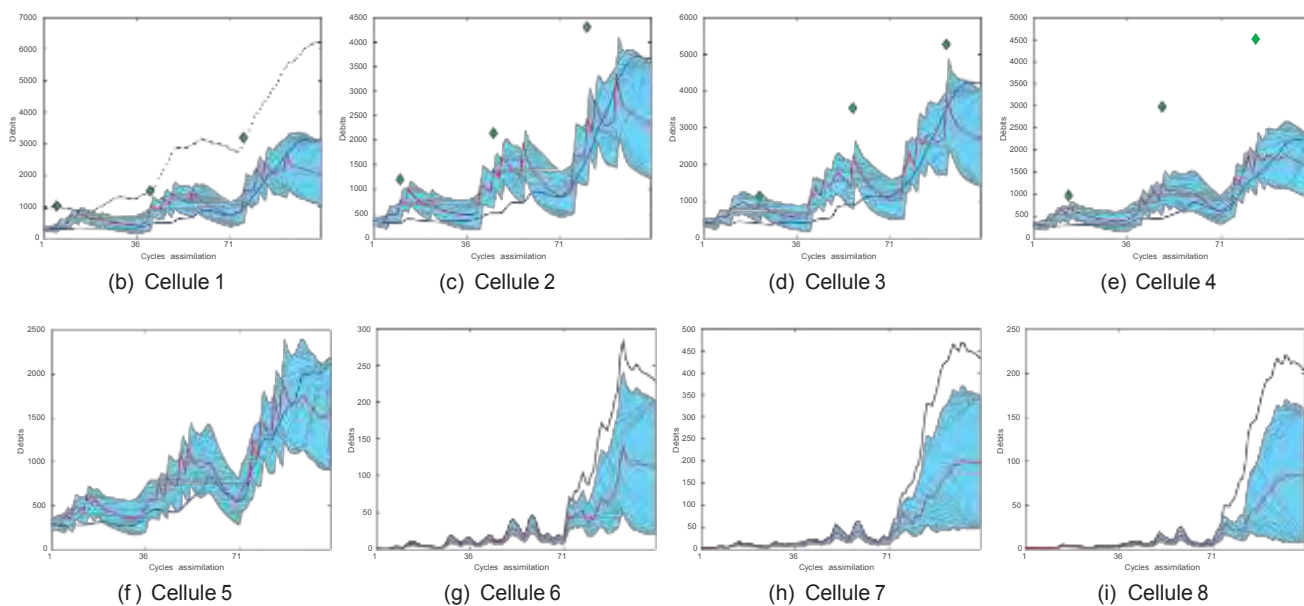


Figure 4.40 - EE1-local à Santos Dumont

dépendantes de la dynamique en espace et en temps, comme le ferait un ensemble correctement échantillonné et représentatif des erreurs. De telles évolutions du masque de localisation ne seront pas développées au cours de cette thèse. Pour la suite de l'étude, on conserve le masque tel qu'il a été défini initialement et utilisé pour l'expérience EE1-local.

4.5 Sensibilité à l'opérateur d'observations

Lors du développement de la maquette d'assimilation pour l'estimation d'état, plusieurs opérateurs d'observation ont été proposés. Les dernières expériences qui suivent vont consister à les comparer à celui utilisé pour les expériences EE1-direct, EE1-diag et EE1-local.

Jusqu'à présent, on assimilait les débits ENVISAT pour corriger les stocks d'eau utilisés comme conditions initiales pour le modèle TRIP. Il est cependant aussi possible d'assimiler les débits ENVISAT pour corriger le stock final d'eau dans le réservoir de surface (option 2 définie à la section 4.2.2.2) ou pour directement corriger le débit TRIP (option 1 définie à la section 4.2.2.1). La difficulté mise en évidence avec ces deux méthodes est qu'il n'y a pas de formule exacte pour passer du stock final S_k - nécessaire pour initialiser le modèle TRIP pour le cycle d'assimilation suivant - au débit Q_k - nécessaire pour évaluer les performances de l'EnKF en le comparant aux débits observés. Une formule approchée pour lier ces deux quantités est basée sur la formule de Manning :

$$S_k \approx \rho L W^{2/5} s^{-3/10} n^{3/5} (Q_k)^{3/5} \iff Q_k \approx \rho^{-5/3} L^{-5/3} W^{-2/3} s^{1/2} n^{-1} (S_k)^{5/3}. \quad (4.26)$$

Cette formule constitue ainsi une approximation de l'opérateur Z_k , (voir Eq 4.13) tel que :

$$Q_k = Z_k(S_k)$$

Dans un premier temps, on va utiliser les résultats de l'expérience EE1-local pour évaluer les incertitudes de l'opérateur Z_k .

4.5.1 Évaluation des incertitudes de l'opérateur Z

Pour évaluer l'incertitude sur la formule de l'opérateur Z_k , évoquée en Section 4.2.2.1, on se base sur les ensembles de stocks finaux analyse $S_{i,k,\text{end}}^{a,[I]}$ et de débits analyse $Q_{i,k}^{a,[I]}$ de l'expérience EE1-local pour chaque cellule $i = 1 \dots n_x$, chaque membre de l'ensemble $I = 1 \dots n_e$ à chaque pas de temps $k = 1 \dots K$:

1. On construit d'abord un nouvel ensemble de débits $\tilde{Q}_{i,k}^{a,[I]}$ à partir de l'ensemble des stocks finaux $S_{i,k,\text{end}}^{a,[I]}$ et de la formule 4.26.
2. Ensuite, on calcule l'erreur relative entre le débit correct $Q_{i,k}^{a,[I]}$ et le débit approché $\tilde{Q}_{i,k}^{a,[I]}$ $\forall i, k, I$:

$$\text{Erreur Relative}_{i,k,I} = 100 \cdot \frac{Q_{i,k}^{a,[I]} - \tilde{Q}_{i,k}^{a,[I]}}{Q_{i,k}^{a,[I]}} \%. \quad (4.27)$$

3. On peut alors construire l'évolution temporelle de l'ensemble des erreurs relatives en chaque cellule du bassin. La Figure B.7 en l'Annexe B illustre les séries temporelles d'ensemble d'erreurs relatives aux 15 stations in situ test.

Statistiques globales	Run libre	Run analyse EE1-local	Run analyse EE2-local	Run analyse EE3-local	Unité
RMSE ^{situ}	$2,79 \times 10^3$	$1,94 \times 10^3$ (-31%)	$2,72 \times 10^3$ (-2%)	$1,98 \times 10^3$ (-29%)	m ³ .s ⁻¹
RMSEn ^{situ}	73,61	57,21 (-22%)	59,94 (-19%)	53,94 (-27%)	%
E ^{situ}	-11,57	-3,90	-4,40	-3,57	-
RMSE ^o	$5,71 \times 10^3$	$1,86 \times 10^3$ (-67%)	$3,65$ (-35%)	$1,07 \times 10^3$ (-81%)	m ³ .s ⁻¹
RMSEn ^o	75,29	37,19 (-51%)	41,78 (-45%)	26,30 (-65%)	%
E ^o	-134,32	-27,23	-20,40	-5,22	-

Table 4.10 – Comparaison des statistiques globales du run libre et des expériences EE1-local, EE2-local et EE3-local

Il apparaît alors que les séries temporelles ne présentent pas de tendance temporelle particulière. Aussi, les erreurs moyennes sont plutôt basses - entre 5% et 20% d'erreur en moyenne. Ainsi, on conserve cette formule pour les dernières expériences d'estimation d'état : EE2-local et EE3-local. On rappelle que EE2-local consiste à assimiler les débits ENVISAT pour corriger le stock final du réservoir de surface de TRIP. La formule 4.26 est utilisée en dehors de l'expérience pour transformer les stocks analyse en débits et évaluer les performances de l'EnKF. Ensuite, l'expérience EE3-local consiste à assimiler les débits ENVISAT pour corriger directement les débits du modèle TRIP. De ce fait, la formule 4.26 est utilisée dans la maquette d'assimilation de données pour convertir les débits analyse en stocks et permettent d'initialiser le cycle d'assimilation suivant. La section qui suit présente les résultats de ces deux simulations.

4.5.2 Comparaison des expériences EE1-local, EE2-local et EE3-local

Les deux expériences suivantes - EE2-local et EE3-local - ont pour objectif de tester la pertinence de la formule 4.26 et d'évaluer les performances de l'EnKF par rapport aux autres expériences. Tableau 4.10 présente les statistiques globales des deux expériences par rapport aux débits in situ et ENVISAT. De plus, les Figures 4.41-4.44 illustrent les variations de RMSEn et de Nash entre le run libre et les analyses des deux simulations, par rapport aux débits ENVISAT ainsi que par rapport aux débits in situ. Les cartes des RMSEn et Nash analyse sont disponibles en Annexe B.

On constate, pour les deux expériences, une amélioration générale des RMSEn entre le run libre et l'analyse. Cependant, l'amélioration réalisée par l'expérience EE2-local est bien plus modérée que l'expérience EE3-local, ce qui était attendu puisque EE3 corrige directement le débit. Pour EE2-local, on observe que les statistiques des cellules situées à l'amont du cours principal - en amont de la station d'Itapéua, voir les Figures 4.41-4.42 - sont sévèrement détériorées (RMSEn et coefficient de Nash-Sutcliffe). Cette détérioration peut être expliquée par deux raisons. Soit l'EnKF ne parvient pas à corriger correctement le stock. Ainsi, une fois converti en débit, le débit analyse est très éloigné des débits observés. Soit la formule 4.26 utilisée pour convertir les stocks analyse en débits analyse n'est pas applicable pour ce tronçon du bassin. Une comparaison des stocks finaux des expériences EE1-local, EE2-local et EE3-local (non présentée ici) pour les stations de Tamishiyacu et Tabatinga, démontre que l'analyse issue de l'EnKF est anormalement basse. De manière générale par la suite, le débit analyse sous-estime systématiquement le débit in situ le long du cours principal. Au contraire, l'expérience EE3-local fournit de très bons résultats avec une réduction globale des RMSEn de 65% en se basant sur les débits ENVISAT et de 27% en se basant sur les débits in situ, c.f. Tableau 4.10.

Pour plus de détails sur ces expériences, la Figure 4.45 compare les débits analyse des deux

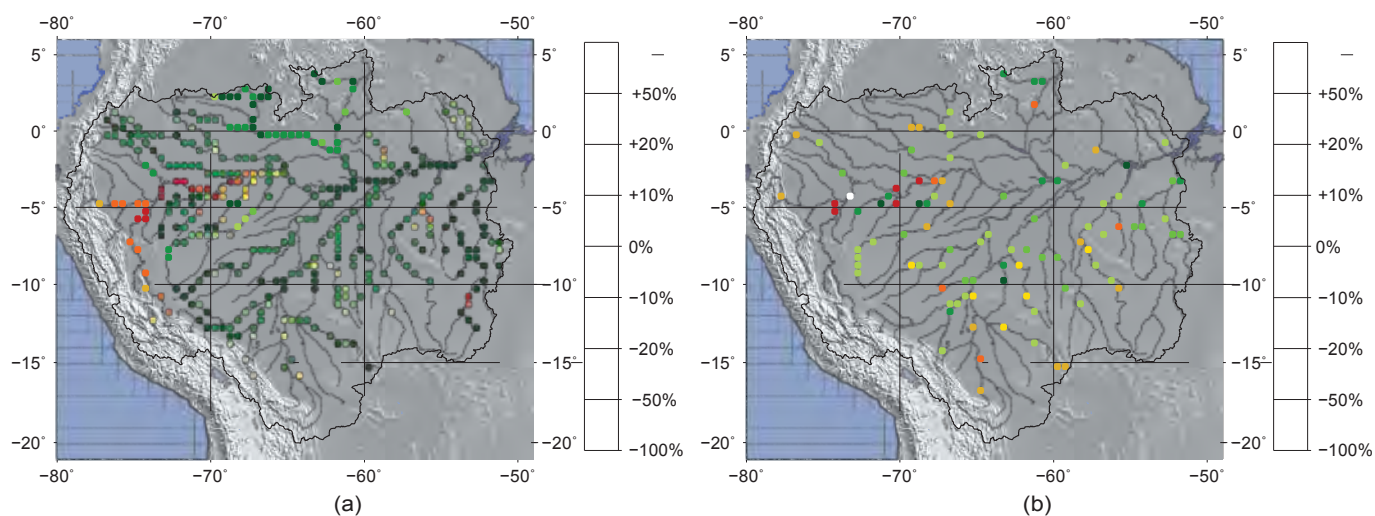


Figure 4.41 - Pourcentage d'augmentation/de réduction du RMSEn pour l'expérience d'AD EE2-local par rapport aux débits ENVISAT (a) et par rapport aux débits in situ (b).

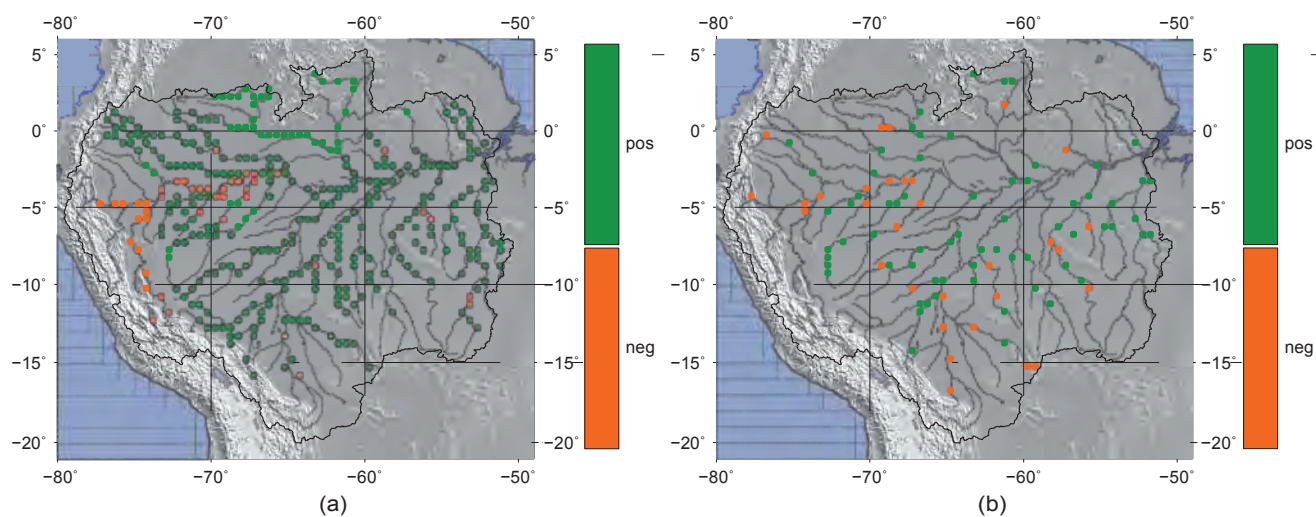


Figure 4.42 - Augmentation (vert)/réduction (rouge) du Nash E pour l'expérience d'AD EE2-local par rapport aux débits ENVISAT (a) et par rapport aux débits in situ (b).

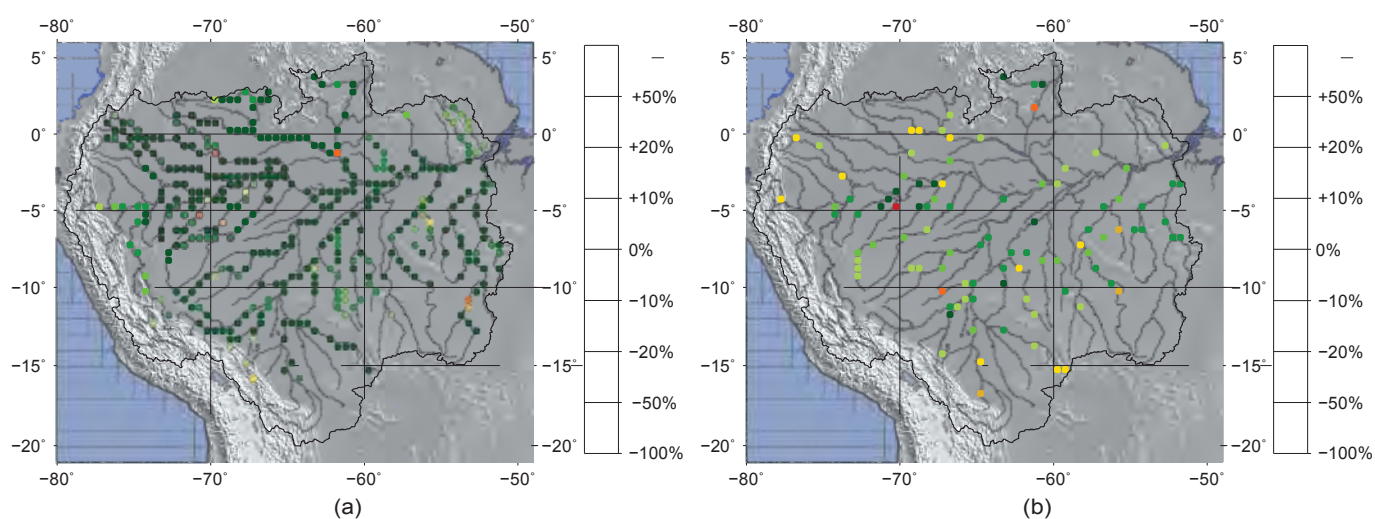


Figure 4.43 - Pourcentage d'augmentation/de réduction du RMSEn pour l'expérience d'AD EE3-local par rapport aux débits ENVISAT (a) et par rapport aux débits in situ (b).

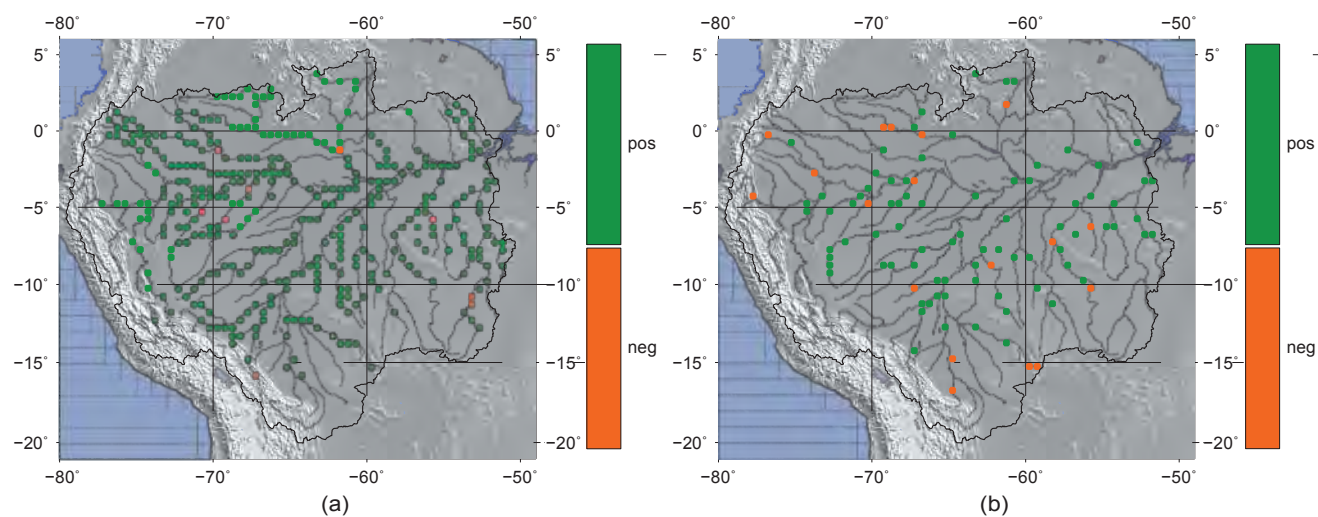
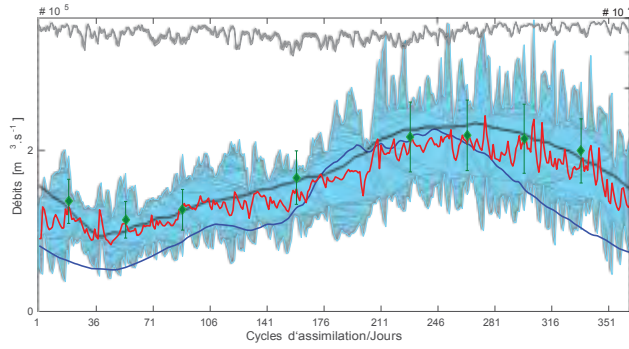


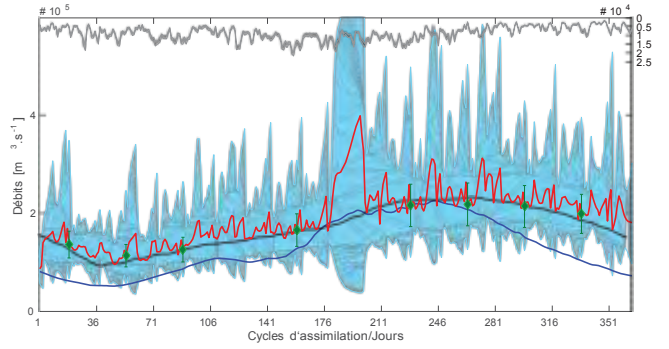
Figure 4.44 - Augmentation (vert)/réduction (rouge) du Nash E pour l'expérience d'AD EE3-local par rapport aux débits ENVISAT (a) et par rapport aux débits in situ (b).

simulations - EE2-local et EE3-local - avec le débit du run libre ainsi que les débits observés ENVISAT et in situ sur un sous-ensemble de trois stations in situ : une sur le cours principal - Óbidos (Figures 4.45a-b) -, sur un affluent gauche - Serrinha sur le Negro (Figures 4.45c-d) -, sur un affluent droit - Manicoré sur le Madeira (Figure 4.45e-f). Pour l'expérience EE2-local, on remarque que le débit analyse moyen sous-estime les débits observés. Les résultats sont très similaires à ceux de l'expérience EE1-local : le débit analyse moyen reste proche du run libre et présente plusieurs pics et creux aux temps d'observations. Pour l'expérience EE3-local, l'assimilation des débits pour corriger directement les débits de TRIP fait que le débit analyse est très proche des débits observés. En particulier pour les stations se situant le long de grands affluents droits, voir l'exemple de Manicoré sur la Figure 4.45f, les expériences précédentes n'avaient pas permis de retrouver le cycle saisonnier observé, contrairement à l'expérience EE3-local.

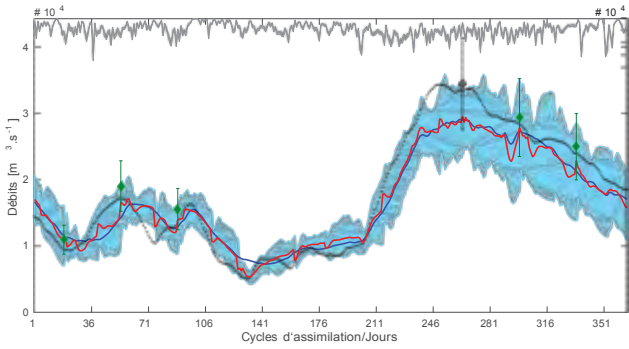
L'expérience EE2-local ne montre pas d'amélioration particulière par rapport aux autres expériences malgré l'utilisation d'un opérateur d'observation plus simple. À l'opposé, l'assimilation des débits pour corriger directement les débits du modèle - l'expérience EE3-local - s'est avérée très efficace et ce malgré l'utilisation d'une formule approchée pour lier les stocks. Pour pouvoir conclure sur l'ensemble de ces simulations, nous allons finir par comparer l'ensemble des expériences les unes par rapport aux autres.



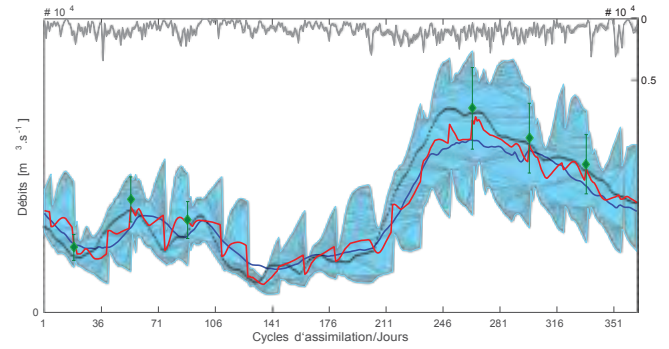
(a) Óbidos - EE2-local



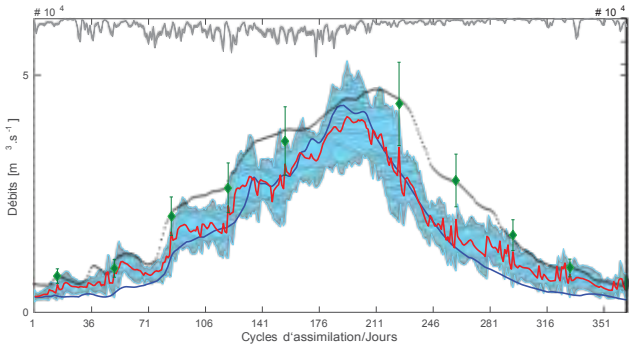
(b) Óbidos - EE3-local



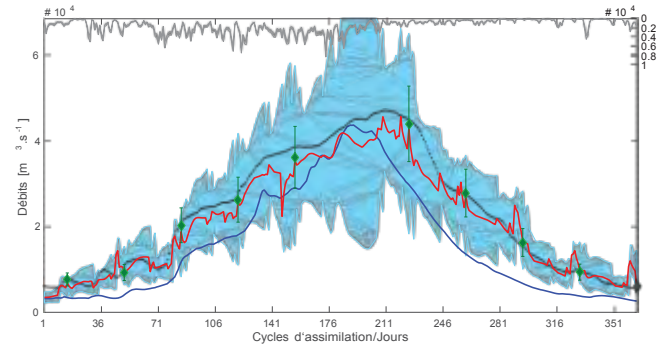
(c) Serrinha - EE2-local



(d) Serrinha - EE3-local



(e) Manicoré - EE2-local



(f) Manicoré - EE3-local

Figure 4.45 – Comparaison des différents débits d'intérêt sur une sélection de stations in situ pour les expériences EE2-local et EE3-local : débit ENVISAT Q_K^o et leur incertitude (losanges verts), débits in situ Q_K^{situ} (points noirs), débits run libre Q_K^f (ligne bleu), dispersion de l'ensemble d'ébauche $Q_{e,k}^b = H(\mathbf{x}_{e,k}^b)$ (zone bleue du 1^{er} au 9^{ème} décile) et run analyse moyen $Q_{\bullet,k}^a$. L'ensemble des précipitations P_e est tracé en zone grise sur l'axe des ordonnées inversé.

Intercomparaison des simulations, conclusion et perspectives

Dans ce chapitre, on s'est intéressé à l'apport d'un produit de débit, estimé à partir de cotes altimétriques ENVISAT et en utilisant des courbes de tarage, sur le bassin de l'Amazonie pour améliorer le débit simulé par ISBA-TRIP. Pour ce faire un EnKF a été utilisé et différents tests ont été réalisés : en utilisant les matrices de covariance d'erreurs générées à partir de tout l'ensemble (expériences EE*-direct), en supprimant toutes les corrélations entre les mailles (expériences EE*-diag) et enfin en utilisant une localisation de l'impact de observations (expériences EE*-local). De plus différentes variables de contrôles ont été utilisées : le stock initial (expériences EE1-*), le stock final (expérience EE2-local) et le débit (expérience EE3-local).

La Figure 4.46 présente une comparaison des performances de chacune de ces cinq expériences d'estimation d'état par rapport au run libre et aux observations ENVISAT. Pour chacune des stations in situ, la série temporelle de la moyenne d'ensemble des débits analyse $\overline{Q_{\bullet,k}^a}$ est comparée à la série temporelle des débits in situ Q_k^{situ} correspondante (qui n'a pas été utilisée pendant l'assimilation), en calculant l'erreur relative à chaque pas de temps k telle que :

$$\text{ErreurRelative}_k = \frac{\overline{Q_{\bullet,k}^a} - Q_k^{situ}}{Q_k^{situ}}.$$

À partir des séries temporelles d'erreurs relatives, on est en mesure de calculer les statistiques de l'ensemble des erreurs relatives et de les représenter sous forme de boîtes à moustache pour chaque simulation et chaque station in situ test, voir la Figure 4.46. On applique alors la même méthode pour calculer les statistiques des débits ENVISAT et des débits simulés lors du run libre. Comme les débits sont comparés aux débits in situ, on ne représente pas ces statistiques pour les stations de Iaituba et Uhe Belo Monte où la série temporelle de débit in situ est mauvaise.

Pour les stations du cours principal (graphique du haut dans la Figure 4.46), toutes les méthodes présentent une amélioration de l'estimation du débit par rapport au run libre, la meilleure simulation étant EE1-direct qui combine une faible erreur relative moyenne et une faible dispersion de cette erreur, suivie par EE3-local qui peut avoir des dispersions plus importantes dues à l'erreur sur la formule de passage entre les débits et les stocks. Les moins bons résultats sont fournis par EE2-local (notamment à Tamishiyacu et Tabatinga). Le débit sur les affluents gauche (graphique du milieu) étudiés ici n'ayant pas de stations ENVISAT, est généralement peu ou pas corrigé. Enfin, pour les affluents droits (graphique du bas), les méthodes utilisant la localisation (EE1-local et EE3-local) fournissent les meilleurs résultats, confirmant la nécessité de réaliser une localisation de l'impact des observations.

Ces expériences offrent de nombreuses perspectives. Premièrement, le masque de localisation proposé dans cette étude n'est qu'une première ébauche au développement d'une méthode de localisation plus adaptée et plus performante. Pour définir la taille des zones d'influence, on s'est basé sur les vitesses maximales d'écoulement de la simulation libre (sans assimilation) et cette taille était définie invariante par zone (indifféremment des têtes de bassin et des cours d'eau principaux) et dans le temps (indifféremment de la période de l'année). Une première amélioration de la localisation serait de définir une taille de zone d'influence dynamique dans le temps et dans l'espace qui dépend de la vitesse d'écoulement au cycle courant pour chacune des observations. La taille de la zone d'influence s'adapterait alors à la période hydrologique de l'observation, à la taille du tronçon de rivière observé et à la zone observée (comportement différent entre le cours principal et les affluents droits par exemple).

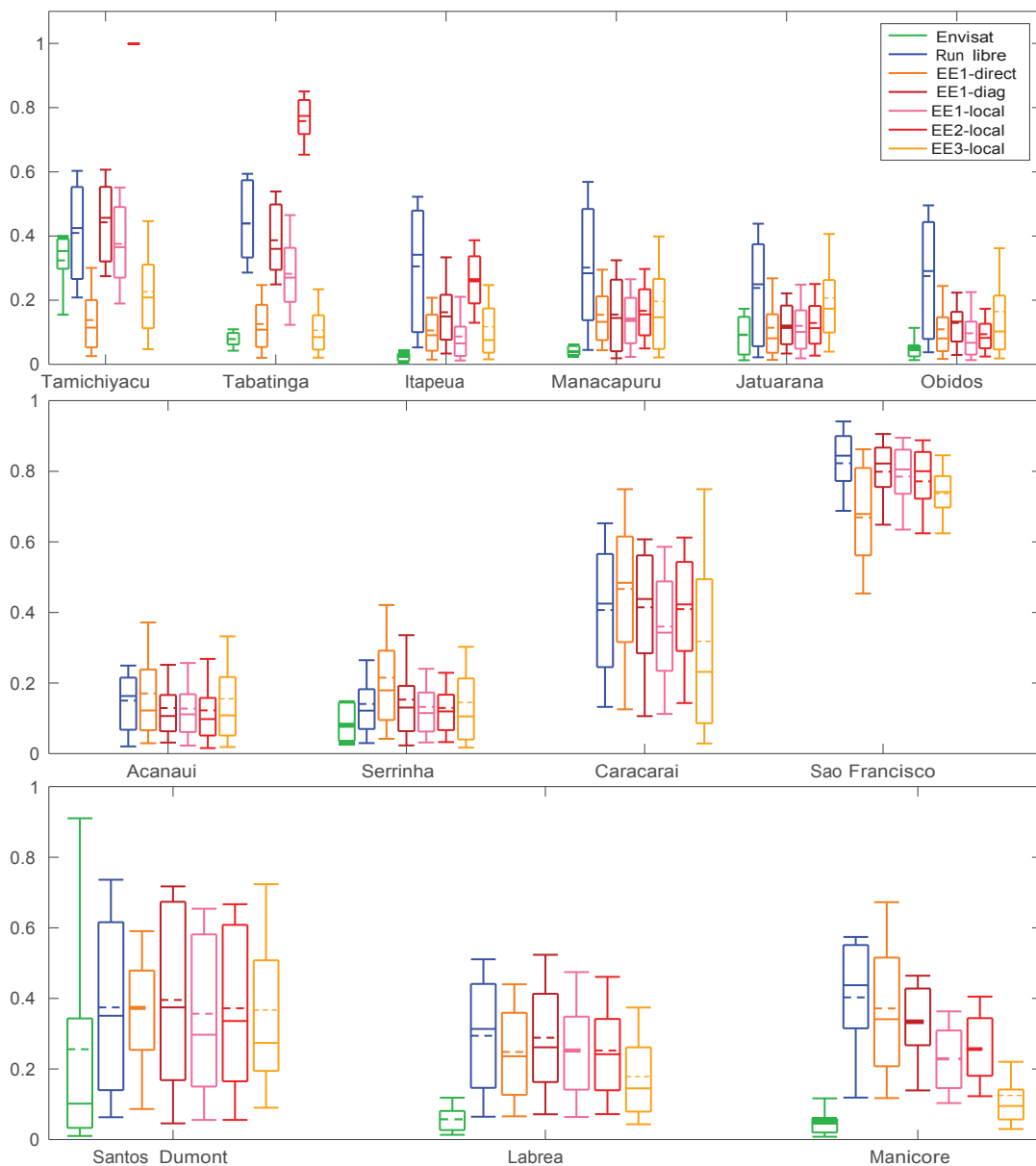


Figure 4.46 - Boîte à moustache, par station insitu test, représentant les statistiques des séries temporelles d'erreurs entre le débit in situ à la station et le débit ENVISAT (vert), le débit du run libre (bleu), le débit analyse moyen de l'expérience EE1-direct (orange), de l'expérience EE1-diag (rouge foncé), de l'expérience EE1-local (rose), de l'expérience EE2-local (rouge) et de l'expérience EE3-local (jaune). Le panneau du haut correspond aux stations sur le cours principal, le panneau du milieu aux stations sur les grands affluents gauches et le panneau du bas aux stations sur les grands affluents droits.

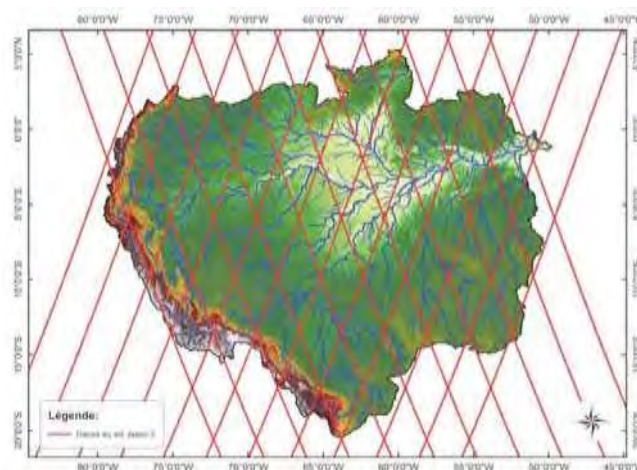


Figure 4.47 – Traces au sol de l’altimètre de Jason-2 sur le bassin de l’Amazonie (lignes rouges). La couverture spatiale est plus faible que celle d’ENVISAT (voir la Figure 4.2b), mais cela lui permet d’avoir une meilleure résolution temporelle (la répétitivité du satellite étant de 10 jours).

La méthode d’assimilation choisie peut aussi évoluer. On a remarqué que certaines approches (entre "directe", "diagonale" et "locale") fonctionnaient mieux sur certaines régions du bassin. Par exemple, l’expérience EE1-direct donne les meilleurs résultats sur le cours principal et la localisation est plus adaptée aux affluents. En effet, la dynamique du cours principal doit être mieux représentée dans l’ensemble, d’où l’intérêt moindre de la localisation sur cette zone. Il serait alors possible de combiner ces deux approches lors d’une même expérience en les appliquant à des cellules différentes du bassin. Par ailleurs, pour les expériences d’estimation de paramètres réalisées au Chapitre 3, on a utilisé un Filtre de Kalman d’Ensemble Asynchrone (AEnKF) qui consiste à propager le modèle sur une fenêtre plus longue que le pas temps des observations et de considérer lors du même cycle d’assimilation toutes les observations depuis le début de l’intégration. Pour les expériences d’estimation de paramètres, on a utilisé une fenêtre de 21 jours correspondant à la répétitivité de SWOT. Ici, on pourrait utiliser l’AEnKF (ou un Lisseur de Kalman, Chapitre 1) sur une période de répétitivité ENVISAT de 35 jours. Cela permettrait d’avoir une meilleure couverture spatiale des observations (car "tout" le bassin serait observé en un seul cycle) et d’atténuer peut-être l’aspect chaotique des débits corrigés.

Dans toutes les expériences menées dans ce chapitre, la principale limitation de la maquette était liée à la faible couverture temporelle des observations ENVISAT (une observation tous les 35 jours, si aucune donnée ne manque). Il serait donc intéressant d’utiliser d’autres observations similaires comme des débits altimétriques issus de l’altimètre de Jason-2 (voir la Figure 4.47 pour la couverture spatiale de Jason-2). Bien que la couverture spatiale soit plus faible que celle d’ENVISAT, Jason-2 permet d’avoir une observation tous les 10 jours. On pourrait donc envisager d’assimiler seulement les données Jason-2 dans un premier temps et de les comparer aux résultats obtenus avec ENVISAT, afin de quantifier l’impact d’une meilleure résolution temporelle mais avec une moins bonne résolution spatiale. Puis, évidemment, on pourrait assimiler les deux jeux de données (ENVISAT et Jason-2) sur leur période d’activité commune afin d’augmenter le nombre d’observations (temporellement et spatialement). Par ailleurs, ces missions altimétriques fournissant des données sur l’ensemble du globe, la méthode développée ici pourrait facilement être exportée sur d’autres grands bassins mondiaux si des courbes de tarage altimétriques peuvent y être définies.

Comme Jason-3 a été lancé en Janvier 2016 sur la même orbite que Jason-2, la continuité des séries de débits pour cette mission est assurée. La mission altimétrique SARAL ayant été lancée en 2013 sur la même orbite qu'ENVISAT, elle pourrait permettre aussi de prolonger ce type d'étude sur une longue période (presque 10 ans). Si des courbes de tarages peuvent être aussi réalisées pour l'altimètre Sentinel-3A (lancé en Février 2016) et la future mission Sentinel-3B (lancement prévu pour 2017), cela permettrait de continuer l'assimilation et fournir une réanalyse du débit de l'Amazone sur une longue période, voire jusqu'au lancement de SWOT. La mission SWOT offrira évidemment une couverture spatio-temporelle encore plus complète que celle de Jason-2 et ENVISAT réunie, ce qui devrait permettre d'obtenir des cartes spatialisées des débits fluviaux et de leur dynamique encore plus précises. L'intérêt de l'assimilation est de disposer d'une estimation de la covariance d'erreur après analyse, ce qui sera essentiel pour ce type de produit de débit multisatellitaire/modèle pour caractériser l'évolution de l'incertitude sur les débits et fonction des observations.

Conclusions et perspectives

L'eau, et en particulier l'eau douce, est une ressource primordiale aux activités humaines (domestiques et économiques) ainsi qu'aux écosystèmes terrestres. Cette eau douce n'est cependant disponible qu'en très faible quantité à la surface des continents et est, de surcroît, très inégalement répartie d'une région à une autre. Malgré tout, ces réserves d'eau sont continuellement renouvelées via le cycle de l'eau. Activé par l'énergie solaire reçue par la surface de la Terre, le cycle hydrologique permet un échange de masses d'eau et d'énergie considérable entre les océans, l'atmosphère et les continents et joue donc un rôle essentiel dans le système climatique terrestre. Les surfaces continentales sont des interfaces particulièrement dynamiques de ce système à des échelles spatio-temporelles très variées. Ces processus sont actuellement suivis et étudiés avec attention, mais de nombreuses incertitudes demeurent notamment aux grandes échelles spatiales. De plus, les changements climatiques actuels, qu'ils soient dus aux activités humaines ou à la variabilité naturelle du climat, entraînent une redistribution des masses d'eau douce disponibles et une redéfinition de leurs flux et de leur variabilité (par exemple, il semblerait qu'il y ait des événements extrêmes plus intenses et plus nombreux). Il est alors primordial d'améliorer notre connaissance de la dynamique spatio-temporelle des eaux de surface étant donné les enjeux scientifiques, sociétaux et économiques qui y sont liés. Cette thèse s'inscrit dans ce contexte et propose de nouvelles solutions pour étudier le cycle de l'eau continental.

On a donc commencé par dresser un aperçu de l'ensemble des techniques déjà disponibles pour étudier le cycle de l'eau continental (Chapitre 1). D'une part, les modèles hydrologiques simulent, avec plus ou moins de détails, l'ensemble ou une partie des processus hydrologiques impliqués dans le cycle de l'eau continental et permettent de couvrir une gamme d'échelles spatio-temporelles assez large. D'autre part, les techniques de mesure in situ et par télédétection observent, avec plus ou moins de précision et de manière plus ou moins directe, ces processus, avec une couverture spatio-temporelle plus limitée (que les modèles). Néanmoins, les mesures satellitaires couvrent des échelles spatiales (et temporelles) complémentaires des réseaux de mesures in situ. C'est pourquoi les données obtenues par télédétection ne remplacent pas les mesures de terrain, mais viennent les compléter. L'assimilation de données permet alors de combiner ces deux informations en tenant compte des incertitudes de chacune. Elle est largement répandue en géosciences, surtout en météorologie et océanographie où les données observées sont très denses. Cependant, les techniques satellitaires actuelles d'observation des eaux à surface libre, bien qu'ayant déjà démontrées leur utilité, ont un certain nombre de limitations. La mission altimétrique SWOT (Surface Water and Ocean Topography), prévue pour un lancement en 2021, sera dédiée à l'observation de ces eaux continentales et en particulier de l'eau bleue (c'est-à-dire l'eau des lacs, fleuves et rivières). Cette mission présente un énorme potentiel car elle permettra l'observation et la mesure d'une majorité des processus inhérents au cycle de l'eau continental. L'objectif de cette thèse était alors d'étudier les apports de la mission SWOT pour l'amélioration de la modélisation hydrologique à grande échelle. Pour cela, on a construit une maquette d'assimilation de données, basée sur un Filtre de Kalman d'Ensemble, afin de combiner le modèle hydrologique grande échelle ISBA-TRIP avec soit, des pseudo-observations SWOT, soit avec d'autres produits de téléde-

tection déjà existants, dans le but d'améliorer les estimations des flux de surface, et en particulier les débits. Cette maquette peut corriger soit certains paramètres du modèle, soit l'état du modèle.

Avant de pouvoir concrètement appliquer l'assimilation de données, il a d'abord été nécessaire d'étudier plus en détails le modèle hydrologique ISBA-TRIP et plus particulièrement sa composante routage de rivière qui modélise les flux d'eau bleue (eau douce) entre la rivière, la nappe souterraine et les plaines d'inondations. En effet, il est important de bien comprendre comment le modèle se comporte et comment les paramètres d'entrée agissent sur les sorties pour construire la maquette d'assimilation le plus judicieusement possible. Pour cela, au cours du Chapitre 2, et après avoir décrit les différentes composantes du modèle ISBA-TRIP, une étude de sensibilité basée sur l'Analyse de Variance (méthode ANOVA) de la composante routage du modèle (c'est-à-dire de TRIP) a été menée. On a alors étudié la sensibilité des profondeurs d'eau, de leurs anomalies et aussi des débits simulés aux principaux paramètres du modèle TRIP. Ces sorties du modèle ISBA-TRIP sont celles qui se rapprochent le plus des produits SWOT et pourront y être comparées dans le cadre d'expériences d'assimilation de données. Les paramètres physiques étudiés sont : soit des paramètres géométriques de la rivière, comme la largeur, la longueur, la profondeur maximale (ou profondeur critique) ou la pente du lit de la rivière, soit des paramètres pilotant les vitesses d'écoulement, comme les coefficients de Manning (frottements) ou le temps de résidence des eaux souterraines. On a ainsi montré que les profondeurs d'eau et leurs anomalies étaient essentiellement sensibles aux paramètres géomorphologiques locaux que sont le coefficient de Manning, la pente ou encore la largeur de la rivière. D'un autre côté, les débits sont beaucoup moins sensibles à ces paramètres mais plutôt aux flux de masses d'eau provenant de l'aire drainée en amont, des eaux souterraines (matérialisées par le temps de résidence) et des plaines d'inondation (matérialisées par la profondeur critique). Au vu de ces résultats, on en a conclu que le produit de cartes bidimensionnelles de cotes d'eau fourni par SWOT pourrait être utilisé pour corriger les paramètres du modèle, en particulier ceux qui sont difficiles à mesurer comme le coefficient de Manning. De plus, le produit de débit, issu des observables SWOT et actuellement en développement, pourrait aussi être utilisé, mais pour corriger directement l'état du modèle.

Ainsi, au cours du Chapitre 3, on a développé et testé une maquette d'assimilation de données permettant l'assimilation d'observations de profondeurs d'eau ou d'anomalies de profondeurs d'eau de type SWOT. Un simulateur d'observations SWOT simplifié a été utilisé pour générer un ensemble de mesures de profondeurs d'eau dans les fauchées du satellites, avec une répétitivité de 21 jours. On s'est alors basé sur l'algorithme d'assimilation de données du Filtre de Kalman d'Ensemble Asynchrone (AEnKF) qui permet d'assimiler des observations ayant lieu avant l'instant effectif de correction. Dans ce chapitre, la maquette a été développée pour faire de l'estimation de paramètres dans le modèle ISBA-TRIP, c'est-à-dire que l'on cherche à corriger les paramètres d'entrée du modèle. Plus particulièrement, on s'est concentré exclusivement sur la correction de la distribution des coefficients de Manning. Par ailleurs, les données SWOT n'existant pas encore, on se place dans le cadre idéalisé des expériences jumelles qui permettent de contrôler l'intégralité des incertitudes présentes dans le système (le même modèle est utilisé pour générer l'ensemble d'ébauche et les observations). Ainsi, on a testé le comportement et les performances de la maquette en présence d'inondations et d'erreurs de bathymétrie (plus précisément d'erreurs sur la hauteur critique). Pour l'ensemble des expériences menées, l'assimilation a permis de corriger la distribution des coefficients de Manning, permettant l'amélioration des estimations de profondeurs d'eau et de débits et démontrant la faisabilité de ces expériences. L'assimilation d'anomalies de cotes d'eau a aussi été testée et a permis de corriger, là aussi, les coefficients de Manning. C'est un résultat très encourageant, puisque l'utilisation des anomalies de cotes d'eau permettra de s'affranchir d'éventuels biais dans les surfaces de références utilisées pour calculer les cotes d'eau entre SWOT et le modèle. La physique simplifiée du modèle limite cependant de meilleures performances surtout en présence d'inondations. Pour un modèle

comme ISBA-TRIP, basé sur un modèle d'onde cinématique, il apparaît plus efficace d'effectuer des étapes d'assimilation hors des périodes d'inondations. Malgré tout, les expériences menées dans ce chapitre ne sont que préliminaires. Elles peuvent facilement être améliorée en prenant en compte d'autres sources d'incertitudes que la seule incertitude des coefficients de Manning et mieux gérer les inondations.

Parallèlement, lors du Chapitre 4, la même maquette d'assimilation, a été utilisée pour corriger les variables pronostiques et/ou diagnostiques du modèle ISBA-TRIP, en d'autres termes, pour faire de l'estimation d'état. Conformément aux conclusions de l'analyse de sensibilité, on choisit d'assimiler des observations de débits, en préfiguration du produit de débits dérivés des observables SWOT. De plus, pour sortir du cadre idéalisé des expériences jumelles, on utilise un produit de débits altimétriques réels issus des cotes d'eau fournies par l'altimètre du satellite ENVISAT. La maquette est alors basée sur un Filtre de Kalman d'Ensemble classique avec des cycles d'assimilation d'un jour. Lors de ce chapitre, deux aspects de l'EnKF ont été étudiés : (1) d'abord le choix de la variable d'état à corriger et (2) la construction particulière des matrices de covariances des erreurs. Pour le premier point, la construction du modèle TRIP nous permettait de corriger différentes variables pronostiques (stocks d'eau initial et final) ou diagnostiques (débit). Les trois types de variables de contrôle ont été testés et la correction des débits et des stocks d'eau initiaux ont fourni des bons résultats. Ensuite, on propose un premier algorithme pour localiser la matrice de covariances des erreurs d'ébauche qui dépend de la position de l'observation dans le bassin afin d'éviter des liens irréalistes entre observations très éloignées. Les résultats ont montré que selon le tronçon de fleuve observé, différentes localisations devraient être utilisées : les observations du cours principal fournissent de meilleurs résultats si aucune localisation ne leur est appliquée, les observations en tête de bassin sur les tous petits affluents fonctionnent mieux si elles sont complètement indépendantes des autres observations et les observations sur des cours d'eau intermédiaire donnent de meilleurs résultats quand la localisation est appliquée. Finalement, ces expériences montrent la faisabilité de l'assimilation d'un produit de débit (produit à partir de données ENVISAT et qui préfigure celui qui sera fourni par SWOT) pour améliorer l'estimation des flux de surface dans le modèle ISBA-TRIP.

Les perspectives dans le cadre de la mission SWOT sont très nombreuses et très variées. Pour commencer, l'étude de sensibilité présentée au Chapitre 2 ne se concentre que sur les paramètres du modèle de routage de rivière. La plateforme actuelle peut être facilement étendue à l'étude de l'impact du modèle de surface (ici le modèle ISBA), des forçages atmosphériques et en particulier des précipitations sur les sorties TRIP. Le formalisme de la méthode ANOVA peut facilement être exporté à d'autres paramètres/d'autres modèles et donnera certainement des résultats différents.

Les pistes à explorer pour l'estimation de paramètres sont aussi très nombreuses. La maquette actuelle peut facilement évoluer vers des configurations plus complexes. Il suffit d'insérer des nouvelles sources d'incertitudes, récurrentes en assimilation de données réelles, comme les erreurs de précipitations. Et surtout, il nous faudra tenir compte de ces nouvelles incertitudes lors de la génération des ensembles d'ébauche (contrairement aux expériences EP2F et EP4F menées lors du Chapitre 3). Il faudrait aussi tester la correction d'autres paramètres que seulement le coefficient de Manning. Les paramètres potentiels sont la largeur et la pente, puisque les profondeurs d'eau d'ISBA-TRIP sont assez sensibles à ces paramètres. On pourrait aussi penser corriger les hauteurs critiques, voire le temps de résidence des eaux souterraines, en utilisant les cotes d'eau SWOT, mais aussi le produit de débit SWOT. Il faudra d'abord tester la correction de ces paramètres un à un. Il faudra aussi penser à la stratégie d'assimilation permettant la correction de plusieurs paramètres en même temps. Par contre, il n'est pas clair si cela est possible, puisqu'on risque d'être confronté au problème d'équifina-

lité (plusieurs jeux de paramètres très différents peuvent produire le même état du système et donc il n'est pas garanti qu'il soit possible de converger vers les paramètres vrais avec l'assimilation). Il pourrait aussi être intéressant de tester d'autres stratégies d'assimilation, comme le filtre à particule, qui est bien adapté pour des cas où le vecteur de contrôle a une taille pas trop importante et qui est plus adapté que l'EnKF aux problèmes non linéaires. Une autre piste d'étude pourrait être la correction de paramètres non plus que de TRIP, mais aussi d'ISBA, voire en utilisant aussi d'autres données spatiales que les seules données SWOT, comme les variations de stock d'eau total fournies par la mission GRACE, l'humidité superficielle des sols de la mission SMOS.

Les perspectives sont tout aussi nombreuses pour l'estimation d'état. La première étape serait de passer à un AEnKF, comme pour l'estimation de paramètres, afin de profiter d'une couverture spatiale plus globale des observations. La suite du travail se trouve dans l'amélioration de l'algorithme de localisation. Il s'agirait de le rendre plus dynamique dans le sens où il s'adapterait à la cellule observée ainsi qu'à la période de l'année. En outre, l'ensemble des expériences ont été réalisées en assimilant les données provenant d'un seul altimètre nadir. La faible couverture temporelle de ces observations n'est pas représentative de celle attendue pour SWOT. Mais en attendant le lancement de la mission en 2021, on peut facilement combiner les produits de débits provenant de plusieurs missions altimétriques (comme Jason-2 par exemple), pour augmenter la couverture spatiale et temporelle des observations et mieux contraindre le modèle. Tout comme pour l'estimation de paramètres, on pourrait aussi investiguer l'assimilation d'autres données spatiales (comme par exemple les données GRACE, SMOS, du capteur MODIS,...), en complément des produits altimétriques, pour corriger d'autres variables pronostiques du modèle, notamment d'ISBA, pour mieux contraindre le bilan hydrologique.

En outre, il faudrait tester d'autres stratégies pour générer des pseudo-observations SWOT plus réalistes. On pourrait notamment utiliser un modèle différent d'ISBA-TRIP pour générer les observations (et ne plus être en expérience jumelle). Parallèlement, il sera nécessaire de complexifier le simulateur de données SWOT utilisé actuellement. En effet, les observations SWOT seront définies à une résolution bien plus fine que la résolution du modèle ISBA-TRIP et un traitement sera nécessaire pour pouvoir les assimiler dans ISBA-TRIP. Il sera intéressant d'anticiper quelle méthode de changement d'échelle appliquer et quelles sont les conséquences sur le bilan des erreurs SWOT. En même temps, on pourra considérer, lors de l'assimilation, des modèles d'erreurs plus réalistes que le seul bruit blanc. Ceci est désormais possible, puisque depuis juin 2016, le CNES a financé un simulateur de données SWOT grand échelle qui ne nécessite en entrée qu'un masque d'eau au format vectoriel. Ce simulateur n'a évidemment pas pu être utilisé dans le cadre de cette thèse, mais ce nouveau simulateur offre des perspectives intéressantes pour les études SWOT à grande échelle.

Enfin, l'ensemble des méthodes développées sur cette thèse n'ont été appliquées que sur un seul grand bassin hydrologique : celui de l'Amazonie. Pour avoir des résultats à l'échelle plus globale, il serait intéressant d'appliquer toutes ces méthodes (analyse de sensibilité, estimation de paramètres et estimation d'état) à d'autres grands bassins hydrologiques ayant un régime climatique et hydrologique différents comme par exemple le Mississippi, le Niger, l'Ob ou encore le Congo, avant de les appliquer globalement.

Annexe A

Description de la maquette d'assimilation de données réalisée avec le coupleur OpenPALM

Dans le cadre de cette thèse, l'exécution couplée du modèle hydrologique ISBA-TRIP avec les différentes instances du filtre de Kalman d'Ensemble (Asynchrone ou non) est menée par le coupleur OpenPALM Fouilloux and Piacentini (1990); Lagarde et al. (2001); Buis et al. (2006) développé au CERFACS (Centre Européen de la Recherche et de Formation Avancée en Calcul Scientifique, www.cerfacs.fr) et à l'ONERA (Office Nationale d'Études et de Recherches Aérospatiales, www.onera.fr) à Toulouse, en France.

A.1 Présentation du coupleur OpenPALM

Le logiciel OpenPALM, accessible en open source sur www.cerfacs.fr/gobc/PALM_WEB/user est un puissant coupleur dynamique et flexible. Il a été originellement développé par le projet MERCATOR dans le contexte de prédictions opérationnelles en océanographie Ferrya et al. (2007). OpenPALM peut être utilisé pour coupler différents modèles numériques entre eux, par exemple un modèle atmosphérique avec un modèle océanographique, ou pour coupler un modèle physique avec un algorithme de traitement comme un algorithme d'optimisation ou de post-traitement. Actuellement, il est largement utilisé par la communauté scientifique notamment pour des projets d'assimilation de données notamment le projet français ADOMOCA (Assimilation de DONnées pour les MOdèles de Chimie Atmosphérique) pour la chimie de l'atmosphère Massart et al. (2009) ou des applications en hydraulique avec le code MASCARET Ricci et al. (2011). OpenPALM est aussi utilisé comme un coupleur de modèle par l'équipe CFD (Computational Fluid Dynamic) du CERFACS avec les logiciels ELSA en mécanique des fluides Poitou et al. (2011). Plus récemment, OpenPALM a été utilisé en modélisation de propagation des feux de forêts Rochoux et al. (2015) et en hydrologie de surfaces continentales Häfliger et al. (2015).

OpenPALM est un coupleur dynamique, c'est-à-dire qu'une composante couplée du code peut être lancée, envoyer/recevoir des données et partager des ressources à n'importe quel moment pendant la simulation. Le "driver" OpenPALM gère les différentes ressources telles que la mémoire nécessaire ainsi que le nombre de processeurs à allouer. Une application OpenPALM peut être décrite comme

un ensemble d'unités de calcul organisées en un algorithme de couplage. Les différentes unités sont positionnées sur des séquences algorithmiques appelées branches et sont contrôlées par des structures conditionnelles ou itératives. Une branche est organisée comme un programme codé dans un langage informatique de haut niveau et les différentes unités de la branche sont exécutées de manière séquentielle comme des sous-routines d'un programme global. Il est alors possible de définir des points de communications entre les unités de différentes branches pour échanger, envoyer et recevoir des données. Les applications OpenPALM sont implémentées via une interface graphique appelée Pre-PALM. Dans cette interface, l'utilisateur construit visuellement son application couplée à partir des unités (boîtes) et de branches (traits), il définit les sections d'exécution séquentielle (unités sur une même branche) ou parallèle (unités sur des branches distinctes) et gère les ressources. Ensuite, les communications sont matérialisées par des points positionnées au sommet des boîtes-unités (donnée entrante) ou à la base (donnée sortante) reliés entre eux par des lignes pointillées. Figure (A.2) illustre la maquette d'AD utilisée dans le cadre de cette thèse.

A.2 La fonctionnalité Parasol

La fonctionnalité Parasol d'OpenPALM permet de lancer automatiquement et en parallèle un certain nombre d'instances du même programme. Elle engendre n instances du même code source avec chacune un jeu de paramètres d'entrée différents (et en conséquence, des sorties différentes). L'utilisateur définit le nombre de processeurs disponibles et Parasol exécute les instances sur ces processeurs en parallèle. Pendant l'exécution, Parasol gère le rassemblement et la distribution des données d'entrée et de sortie dans le bon ordre. Dans le cadre de cette thèse, Parasol est utilisé pour propager le modèle ISBA-TRIP sur la fenêtre d'assimilation. Informatiquement, Parasol génère deux types de sous-routines :

- Une sous-routine "esclave" qui exécute une des n instances de la fonction à exécuter sur l'ensemble,
- Une sous-routine "maître" qui appelle les esclaves et gère la distribution, l'organisation et la récupération des données d'entrée et de sorties.

La Figure (A.1) présente une exécution générale des runs d'ensemble ISBA-TRIP à l'aide de Parasol.

Nous allons maintenant décrire plus en détail la maquette d'AD utilisée lors de cette thèse.

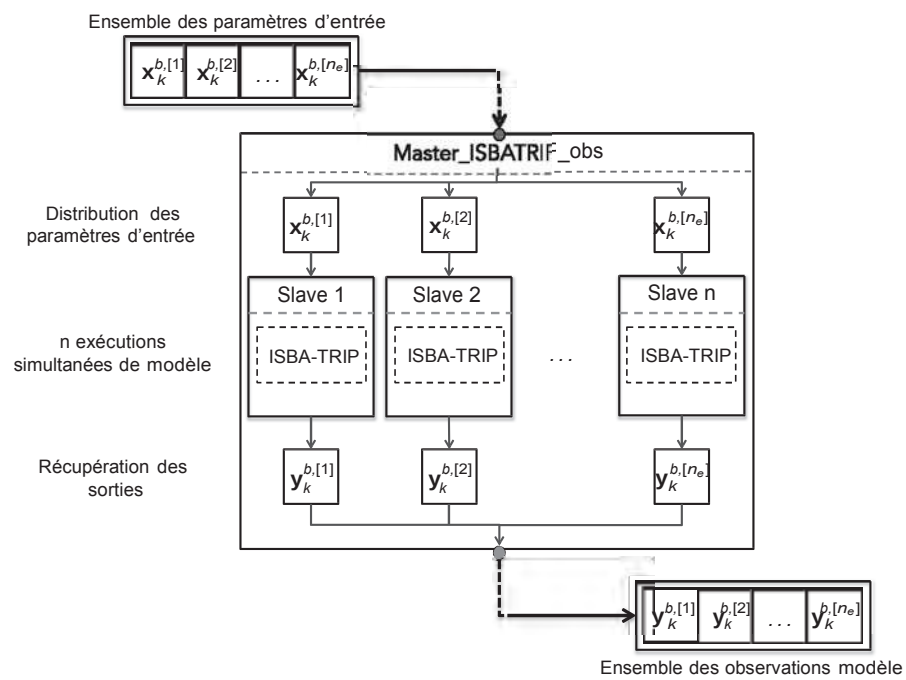


Figure A.1 – Illustration du fonctionnement de Parasol : l'unité maître (master) appelle n esclaves (slave) qui exécute en parallèle n intégration du modèle ISBA-TRIP avant de renvoyer les sorties vers l'unité suivante.

A.3 Description de la maquette d'AD

La maquette illustrée en Figure A.2 est un Filtre de Kalman d'Ensemble construit autour du modèle ISBA-TRIP. Elle peut être utilisée pour de l'estimation de paramètres (voir Chapitre 3) et de l'estimation d'état (voir Chapitre 4), pour assimiler des données synthétiques (voir les expériences jumelles du Chapitre 3) ou des données réelles (voir l'assimilation des données Envisat au Chapitre 4) synchrones (voir Chapitre 4) ou asynchrones (voir Chapitre 3) de hauteurs d'eau, anomalies de hauteurs d'eau ou débits.

La maquette d'assimilation se compose de 5 branches PALM chacune ayant un rôle bien défini :

- La branche MAIN (violet) initialise la simulation d'AD en collectant, à partir de fichiers textes d'entrée, la configuration générale de l'expérience c'est-à-dire entre autre la date de début de simulation, la taille de la fenêtre d'assimilation, le type d'expérience - expérience jumelle ou réelle, en estimation de paramètres ou d'état - la nature des variables de contrôle et des variables observées, la taille de l'ensemble etc ... Elle gère ensuite une boucle sur les cycles d'assimilation qui lance les différentes instances de l'(A)EnKF. À la fin d'un cycle d'assimilation, la branche MAIN est chargée de cycliser l'analyse pour le cycle d'assimilation suivant.
- La branche OBSERVATIONS (vert) génère toutes les variables liées aux observations c'est-à-dire l'ensemble des vecteurs d'observations $\mathbf{Y}_{e,k}^o$, la matrice de covariances des erreurs d'observations $\mathbf{R}_k$ ainsi que d'autres informations relatives aux observations comme la taille des vecteurs d'observations $n_{y,k}$ par exemple.
- La branche BACKGROUND (rouge) crée les différents ensembles d'ébauche nécessaires pour perturber les sorties du modèle ISBA-TRIP. Selon les expériences, on va directement générer les ensembles de paramètres d'entrée perturbés ou générer des ensembles de variables intermédiaires qui vont servir à perturber les champs de forçages et/ou la condition initiale du modèle.
- La branche FORECAST (beige) est chargée de générer les différents ensemble d'ébauche nécessaires à l'EnKF : l'ensemble des variables de contrôle $\mathbf{X}_{e,k}^b$ via l'unité *master_BUILD_XB*, l'ensemble des observations modèle d'ébauche $\mathbf{Y}_{e,k}^b = H_k(\mathbf{X}_{e,k}^b)$ via l'unité *master_RUN_SURFEX_FORECAST* et les différentes matrices de covariances $\mathbf{P}^b \mathbf{H}^T_{e,k}$ et $\mathbf{H}^b \mathbf{P}^T_{e,k}$ ainsi de le gain de Kalman $\mathbf{K}_{e,k}$ via l'unité *EnKF_ALGEBRA*.
- La branche ANALYSIS (bleu) détermine alors l'ensemble des variables de contrôle analyse $\mathbf{X}_{e,k}^a$ et ré-effectue les runs d'ensemble modèle si nécessaire.

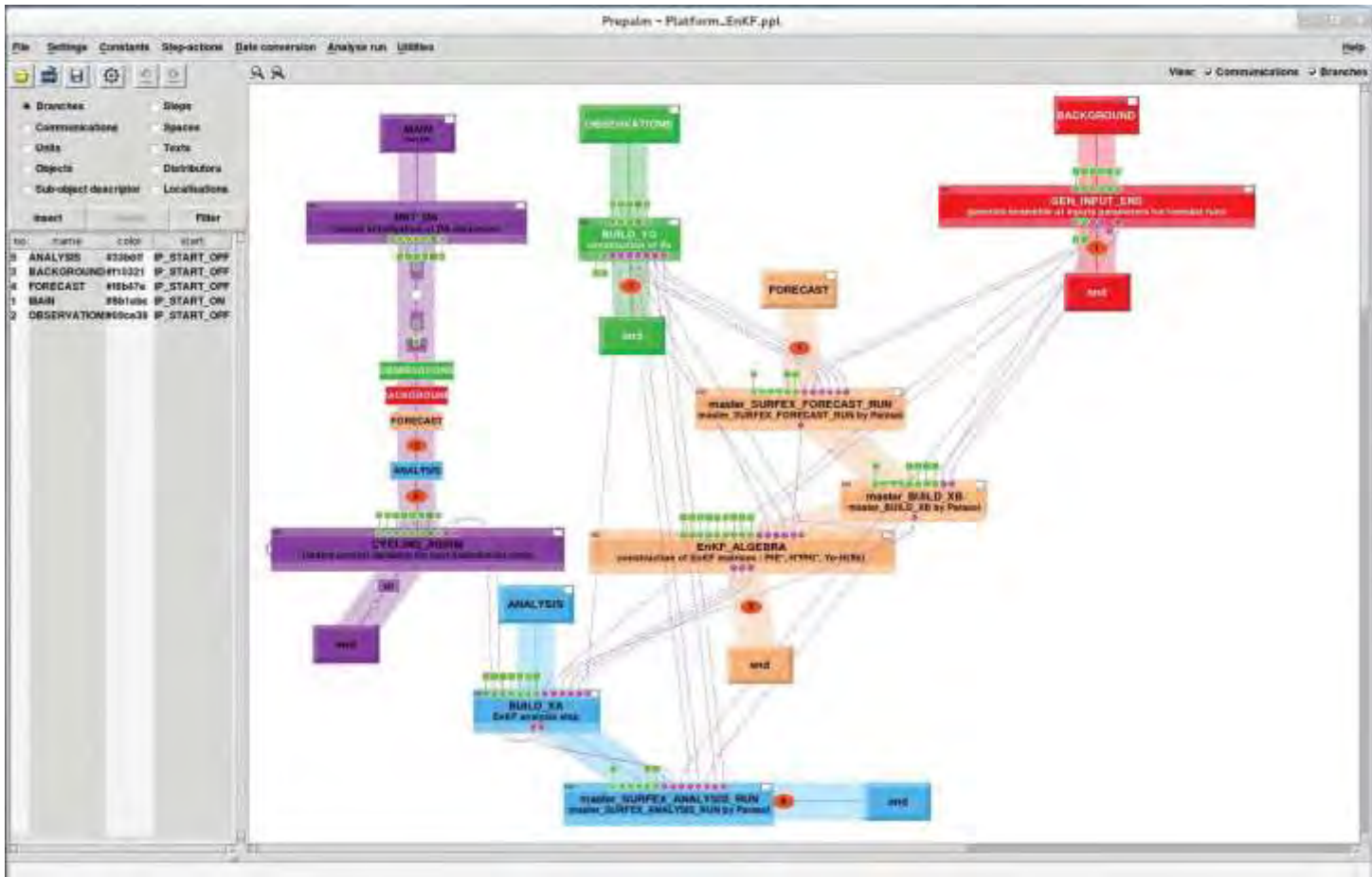


Figure A.2 – Interface graphique PrePALM de la maquette d'assimilation de données combinant le modèle ISBA-TRIP avec un AEnKF (ou EnKF).

Annexe B

Résultats et figures supplémentaires pour les expériences d'estimation d'état

B.1 Statistiques globales

Cette section présente, les statistiques globales des expériences EE2-local et EE3-local sur le même schéma que les Figures 4.19-4.22. Pour chacune des expériences EE2-local et EE3-local, on a quatre figures :

- Cartes des RMSEn par rapport aux débits in situ et Envisat,
- Cartes des E par rapport aux débits in situ et Envisat,
- Cartes des augmentations/réductions de RMSEn entre le run libre et le run analyse moyen,
- Cartes des augmentations/réductions de coefficients de Nash-Sutcliffe entre le run libre et le run analyse moyen.

B.1.1 EE2-local

B.1.2 EE3-local

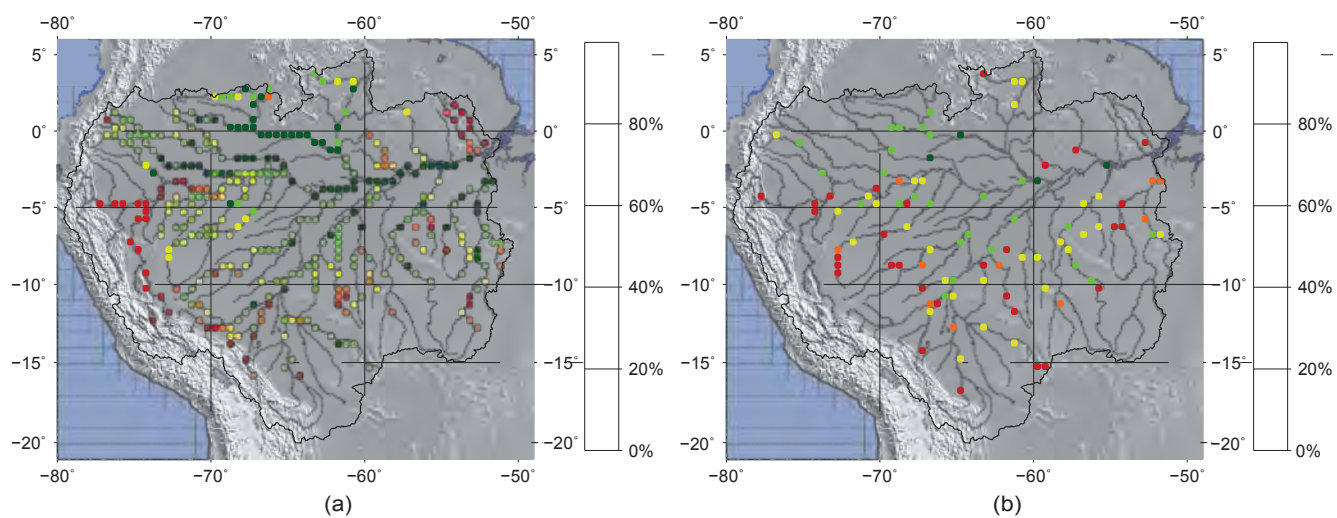


Figure B.1 – Coefficients d'erreurs moyennes quadratiques RMSEn de l'expérience d'AD EE2-local par rapport aux débits Envisat (a) et par rapport aux débits in situ (b).

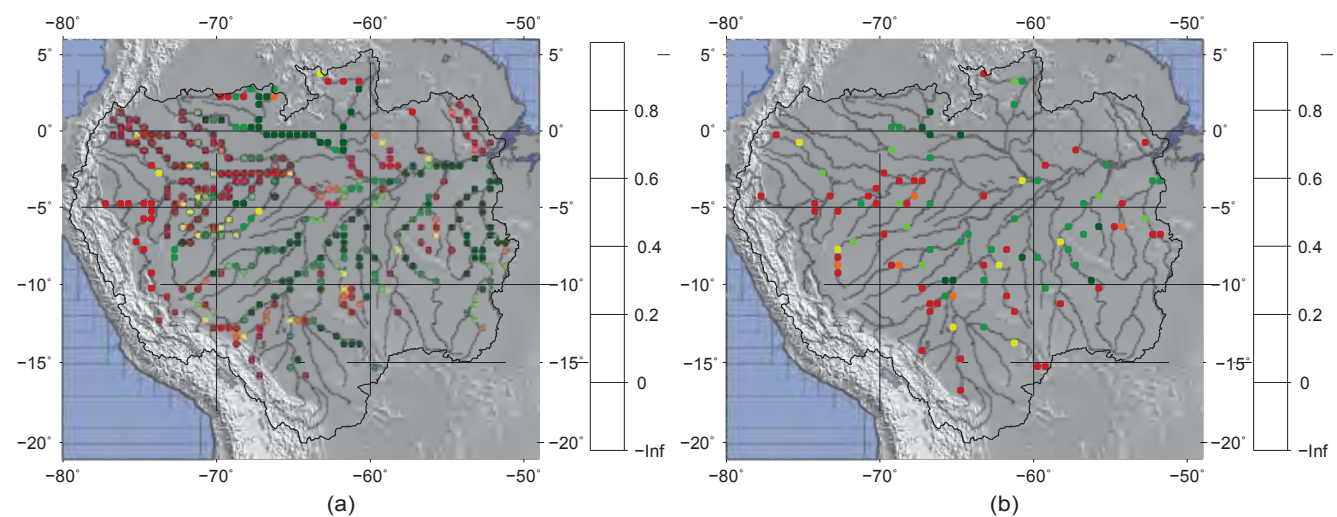


Figure B.2 – Coefficients de Nash-Sutcliffe E de l'expérience d'AD EE2-local par rapport aux débits Envisat (a) et par rapport aux débits in situ (b).

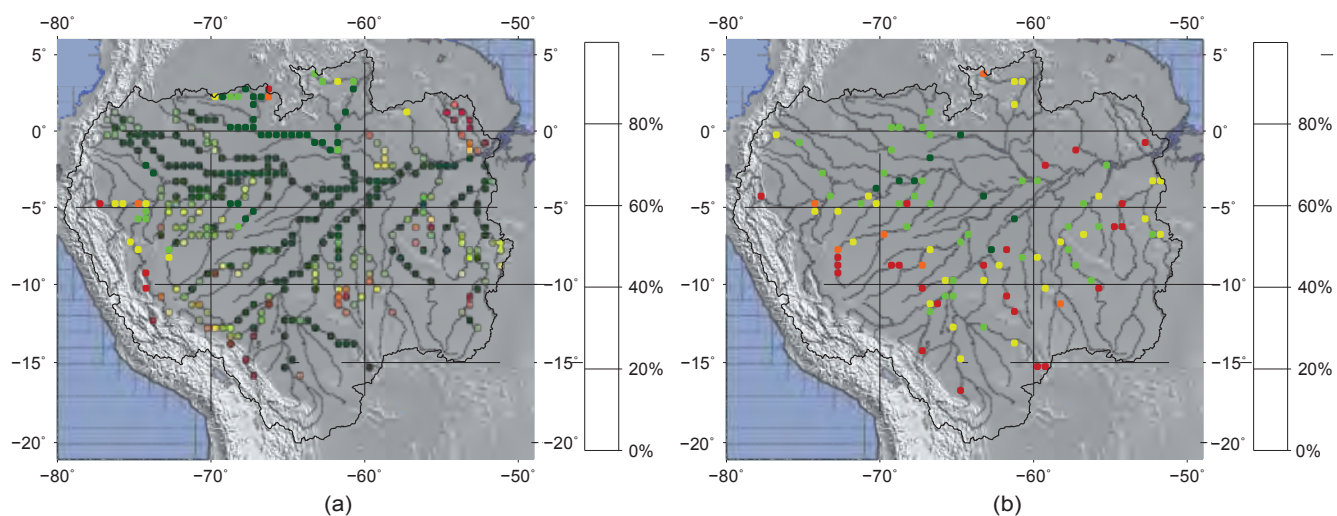


Figure B.3 – Coefficients d'erreurs moyennes quadratiques RMSEn de l'expérience d'AD EE3-local par rapport aux débits Envisat (a) et par rapport aux débits in situ (b).

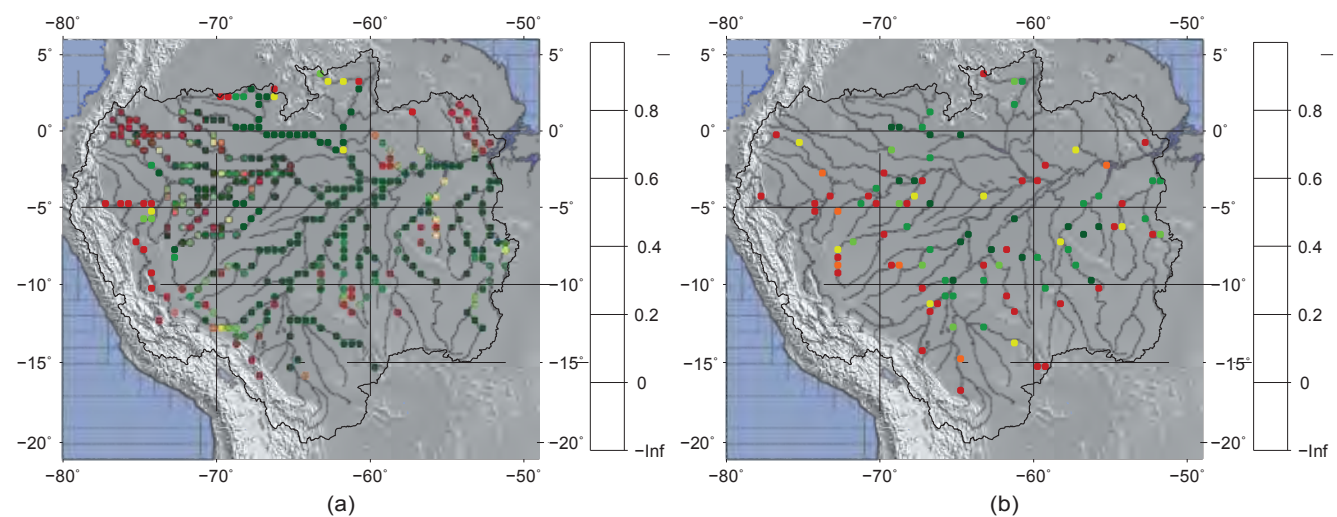


Figure B.4 – Coefficients de Nash-Sutcliffe E de l'expérience d'AD EE3-local par rapport aux débits Envisat (a) et par rapport aux débits in situ (b).

B.2 Débits aux stations in situ tests

Dans cette section, on compare, pour chaque station in situ test et pour les expériences EE2-local et EE3-local, les débits in situ, Envisat avec les débits du run libre et le débit analyse moyen.

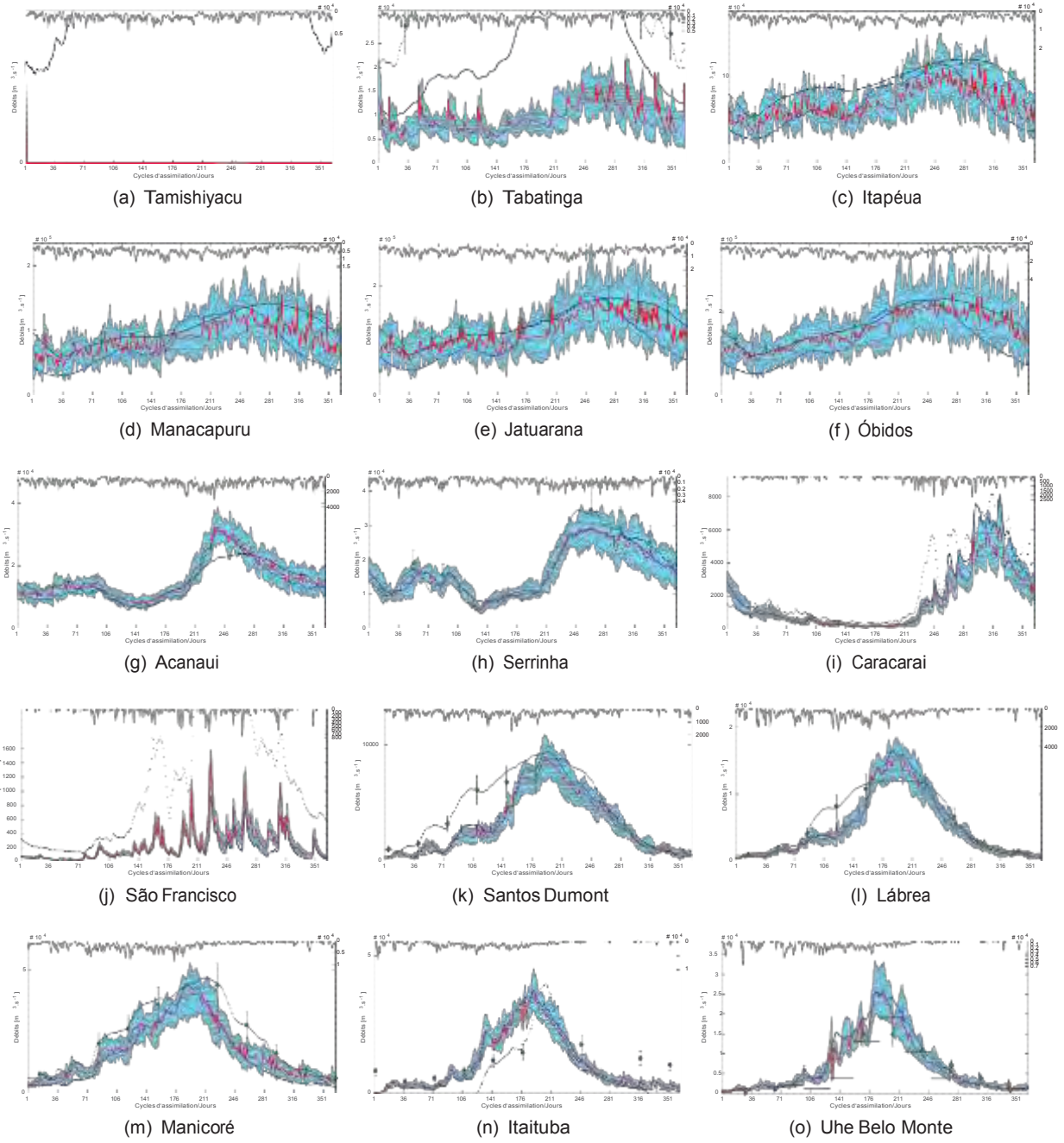


Figure B.5 – Comparaison des différents débits d'intérêt aux stations in situ : débit Envisat Q_k^0 et leur incertitude (losanges verts), débits in situ Q_k^{situ} (points noirs), débits run libre Q_k^f (ligne bleu), dispersion de l'ensemble d'ébauche $Q_{e,k}^b = H(\mathbf{x}_{e,k}^b)$ (zone bleue du 1^{er} au 9^{ème} décile) et run analyse moyen $\overline{Q_{e,k}^a}$ de l'expérience EE2-local. L'ensemble des précipitations $\mathbf{p}_e$ est tracé en zone grise sur l'axe des ordonnées inversé.

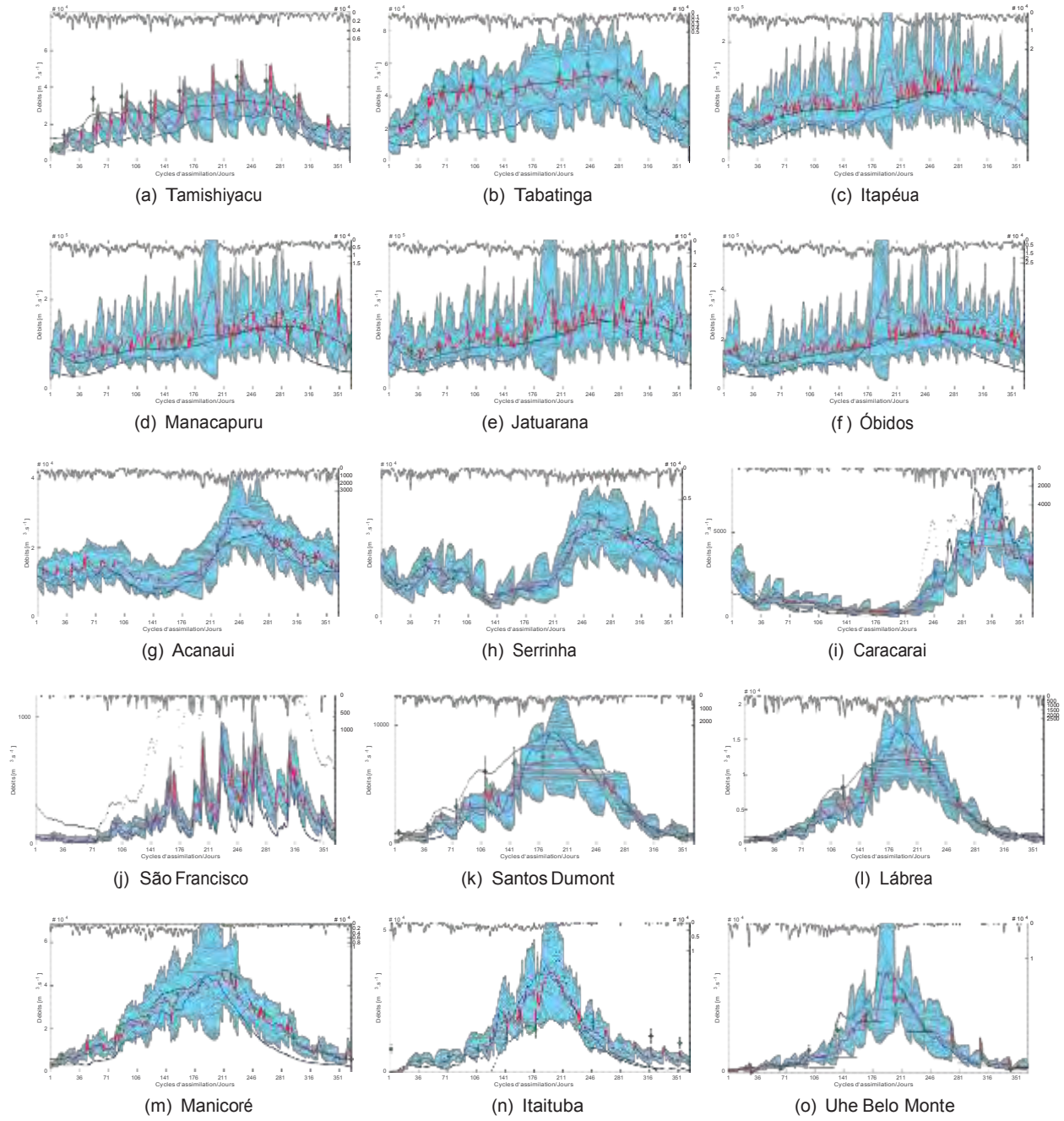


Figure B.6 – Comparaison des différents débits d'intérêt aux stations in situ : débit Envisat Q_k^o et leur incertitude (losanges verts), débits in situ Q_k^{situ} (points noirs), débits run libre Q_k^f (ligne bleu), dispersion de l'ensemble d'ébauche $Q_{e,k}^b = H(\mathbf{x}_{e,k}^b)$ (zone bleue du 1^{er} au 9^{ème} décile) et un analyse moyen $\overline{Q_{e,k}^a}$ de l'expérience EE3-local. L'ensemble des précipitations $\hat{P}_e$ est tracé en zone grise sur l'axe des ordonnées inversé.

Indice de la station	RMSEn	RMSEn	RMSEn	E	E	E
	Q_k^f	$\overline{Q_{\bullet,k}^a}$ EE2-local	$\overline{Q_{\bullet,k}^a}$ EE3-local	Q_k^f	$\overline{Q_{\bullet,k}^a}$ EE2-local	$\overline{Q_{\bullet,k}^a}$ EE3-local
Tamashiyacu	0,401	0,8034	0,128	-1,783	-16,556	-0,129
Tabatinga	0,449	0,282	0,169	-2,221	-9,322	0,740
Itapéua	0,332	1,029	1,009	-1,641	-0,900	0,313
Manacapuru	0,364	0,206	0,279	-1,123	0,251	-0,371
Jatuarana	0,270	0,155	0,254	-0,210	0,602	-0,075
Óbidos	0,291	0,116	0,251	-0,315	0,790	0,022
Acanauí	0,195	0,183	0,158	0,661	0,702	0,777
Serrinha	0,156	0,153	0,148	0,902	0,906	0,912
Caracarái	0,431	0,549	0,564	0,788	0,656	0,637
São Francisco	1,042	1,029	1,009	0,945	0,686	0,619
Santos Dumont	0,380	0,432	0,370	0,684	0,592	0,699
Lábrea	0,371	0,343	0,271	0,740	0,778	0,861
Manicoré	0,360	0,311	0,144	0,635	0,728	0,941
Itaituba	0,558	0,551	0,471	0,526	0,540	0,664
Uhe Belo Monte	0,852	0,679	0,591	0,402	0,621	0,713

Table B.1 – Performances de l'EnKF lors de l'expérience EE2-local et EE3-local : comparaison des erreurs quadratiques moyennes normalisées (RMSEn) et des coefficients d'efficacité de Nash-Sutcliffe (E) du débit analyse moyen $\overline{Q_{\bullet,k}^a}$ et du débit du run libre Q_k^f , calculés à partir des débits observés in situ.

B.3 Erreur de formulation sur l'opérateur Z_k

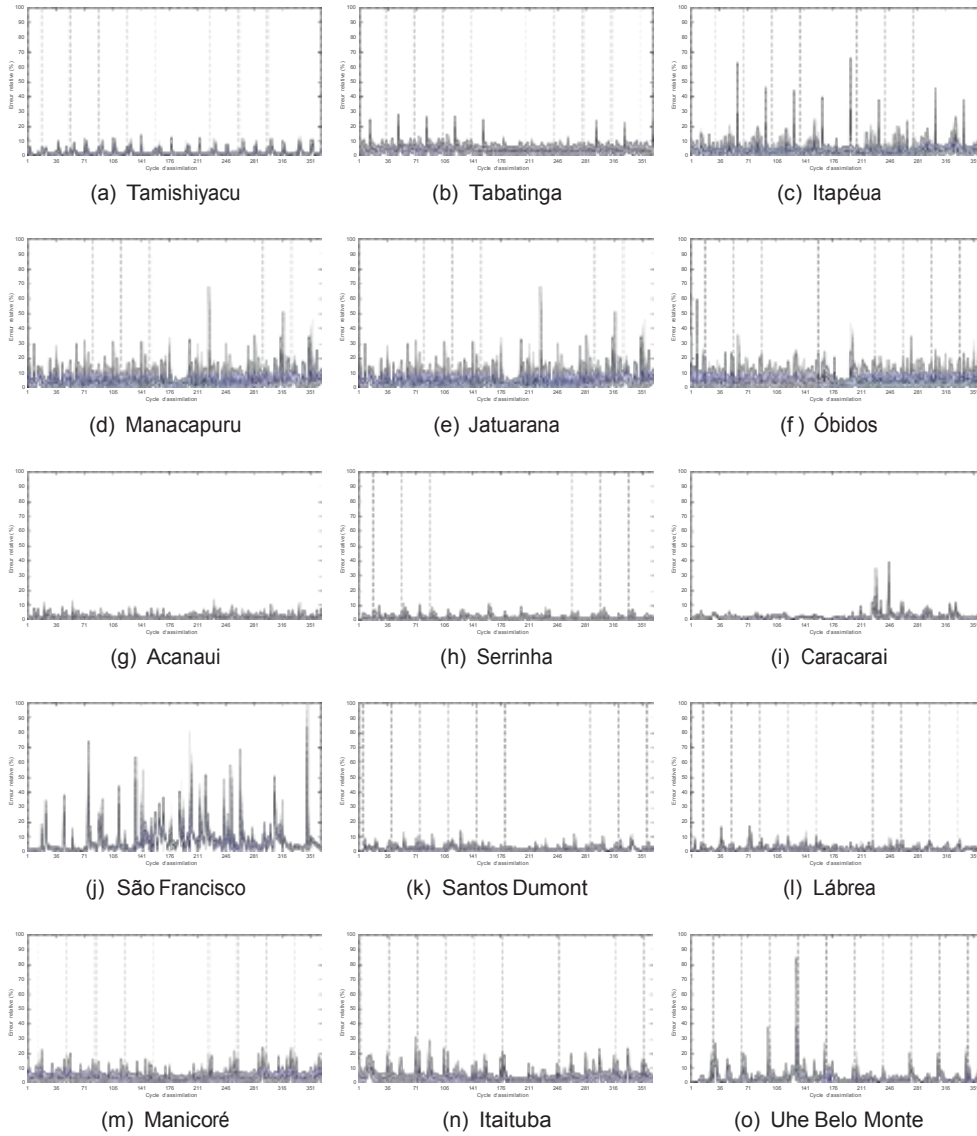


Figure B.7 – Évolution temporelle de l'ensemble des erreurs relatives entre le débit vrai et le débit approché par la formule (4.26) : dispersion entre le 1^{er} et le 9^{ème} décile de l'ensemble (zone grise), moyenne de l'ensemble (ligne bleue). Les lignes verticales en tirets grises représentent les temps d'observation Envisat de la station.

Annexe C

**Article : "Temporal Variance-based
Sensitivity Analysis of the River
Routing Component of the large
scale hydrological model
ISBA-TRIP : Application on the
Amazon Basin**

Temporal Variance-Based Sensitivity Analysis of the River-Routing Component of the Large-Scale Hydrological Model ISBA–TRIP: Application on the Amazon Basin

CHARLOTTE M. EMERY AND SYLVAIN BIANCAMARIA

Laboratoire d'Études en Géophysique et Océanographie Spatiale, Toulouse, France

AARON BOONE

CNRM-GAME, Météo-France, Toulouse, France

PIERRE-ANDRÉ GARAMBOIS

Institut National des Sciences Appliquées Strasbourg, Strasbourg, France

SOPHIE RICCI AND MÉLANIE C. ROCHOUX

Climat, Environnement, Couplages et Incertitudes, CERFACS-CNRS, Toulouse, France

BERTRAND DECHARME

CNRM-GAME, Toulouse, France

(Manuscript received 22 February 2016, in final form 15 June 2016)

ABSTRACT

The continental part of the water cycle is commonly represented with hydrological models. Yet, there are limits in their capacity to accurately estimate water storage and dynamics because of their coarse spatial resolution, simplified physics, and an incomplete knowledge of atmospheric forcing and input parameters. These errors can be diminished using data assimilation techniques. The model's most sensitive parameters should be identified beforehand. The objective of the present study is to highlight key parameters impacting the river-routing scheme Total Runoff Integrating Pathways (TRIP) while simulating river water height and discharge as a function of time focusing on the annual cycle. Thus, a sensitivity analysis based on the decomposition of model output variance (using a method called ANOVA) is utilized and applied over the Amazon basin. Tested parameters are perturbed with correcting factors. First, parameter-correcting coefficients are considered uniform over the entire basin. The results are specific to the TRIP model and show that geomorphological parameters explain around 95% of the water height variance with purely additive contributions, all year long, with a dominating impact of the river Manning coefficient (40%), the riverbed slope (35%), and the river width (20%). The results also show that discharge is essentially sensitive to the groundwater time constant that makes up more than 90% of the variance. To a lesser extent, in rising/falling flow period, the discharge is also sensitive to geomorphological parameters. Next, the Amazon basin is divided into nine subregions and the sensitivity analysis is carried out for regionalized parameter-correcting coefficients. The results show that local-region parameters impact water height, while upstream-region parameters affect discharge.

1. Introduction

The earth's climate is undergoing changes in response to natural variability and also increasing concentrations of greenhouse gases (Stocker et al. 2013). The continental

part of the water cycle is an active component of this system (Alkama et al. 2008). Global climate change will affect the water cycle, likely creating perennial drought in some areas and frequent flooding in others (Trenberth 2011). It is therefore of prime importance to study and understand these changes to continental waters.

Continental water can be studied at the global scale using land surface models (LSMs) coupled with a river-routing model (RRM). LSMs provide lower boundary

Corresponding author address: Charlotte M. Emery, Laboratoire d'Études en Géophysique et Océanographie Spatiale, 16 Av. Édouard Belin, 31400 Toulouse, France.
E-mail: emery@legos.obs-mip.fr

conditions to atmospheric general circulation models, while RRM models compute river discharge, which can be potentially used as boundary conditions to ocean general circulation models.

Several RRM models have been developed at a global scale (Vörösmarty et al. 1989; Coe 1998; Hagemann and Dümenil 1997; Oki and Sud 1998; Arora et al. 1999; Olivera et al. 2000; Ducharme et al. 2003; Ngo-Duc et al. 2007; Lucas-Picher et al. 2003; Yamazaki et al. 2011). In those models, the hydrological network is generally derived from a digital elevation model (DEM) at coarse resolution (from 0.258 3 0.258 to 48 3 48). This network is then used by the RRM to transfer water mass from grid cell to grid cell until the river outlet at the continent–ocean interface. Water mass transfer is based on simple budget equations representing the temporal variations of the water mass stored in each grid cell. The currently available RRM models differ from each other in their method of dealing with river flow velocity, surface parameterization, groundwater, and floodplains.

The description of river flow velocity is addressed in several ways in the literature. For instance, Coe (1998) and Oki and Sud (1998) considered a single surface reservoir with a uniform and constant flow velocity. It is also possible to use a time-constant but spatially distributed flow velocity based on topography and river channel characteristics (Ducharme et al. 2003; Hagemann and Dümenil 1997; Vörösmarty et al. 1989). In contrast, several works rely on a time-varying and spatially distributed flow velocity estimation based on the Manning formula (Manning 1891), for example, Arora et al. (1999), Decharme et al. (2012), Lucas-Picher et al. (2003), and Ngo-Duc et al. (2007).

When a single surface reservoir is considered (Coe 1998; Ngo-Duc et al. 2007; Oki and Sud 1998; Vörösmarty et al. 1989), input water mass to each gridcell reservoir corresponds to surface runoff within the grid cell and river outflow from upstream grid cells. The computed water mass is then transferred to the downstream grid cell along the river network. Several studies (Arora et al. 1999; Decharme et al. 2008; Ducharme et al. 2003; Hagemann and Dümenil 1997; Lucas-Picher et al. 2003) highlighted the necessity of modeling groundwater inflow to the surface reservoir in order to simulate the delayed contribution of groundwater in the river discharge estimation. Inland water bodies such as floodplains are rarely taken in consideration. When it exists, a flooding scheme is triggered when the water stage (Decharme et al. 2008, 2012) or the river discharge (Vörösmarty et al. 1989) exceeds a given threshold and the floodplain within a grid cell exchanges water mass with the associated river reservoir.

Several input parameters are used by RRM models. The parameters can be topographical (such as riverbed slope) and geomorphological (such as river width or roughness)

and can also be based on empirical relationships, in which case other parameters are usually required. These relationships correspond to multiplying and powering coefficients linking parameters between each other or with the model state (Leopold and Maddock 1953; Moody and Troutman 2002; Gleason and Smith 2014). Parameters can be either spatially or uniformly distributed over river catchments according to assumptions and simplifications adopted in each study.

It seems obvious that to better represent observations in hydrological modeling, model complexity can be increased (Grayson and Blöschl 2001). In this context, Sieber and Uhlenbrook (2005) point out that sensitivity analysis (SA) can be a powerful tool to both identify the most sensitive (and therefore important) parameters and to understand the hydrological model structure and its response.

SA can be used for two purposes: either to explore multidimensional parameter spaces and understand the source of model uncertainties (Hornberger and Spear 1981) or to identify key parameters and their role in the model response (Saltelli et al. 2008). The first option has been widely investigated, most notably by using the Generalized Likelihood Uncertainty Estimation (GLUE) method (Beven and Binley 1992) both in catchment modeling (Freer et al. 1996) and rainfall–runoff modeling (Blazkova and Beven 2004). The GLUE method assumes that, for a given catchment numerically represented by a given model, several sets of this model's input parameters can be equally accepted as a simulator of the catchment's real response. By comparing an ensemble of simulated responses with observations of the same nature, a value of likelihood is assigned to each set of parameters of this ensemble. All simulations giving an acceptable—or behavioral—likelihood are resampled to build a distribution function of the parameter sets. The likelihood values are finally used to estimate the uncertainty associated with the model predictions. Other common methods are regional sensitivity analysis (Wagener et al. 2003) for dynamic identifiability analysis and Bayesian total error analysis method (Kavetski et al. 2006) for calibration and uncertainty estimation. Alternatively, to identify the key parameters, SA can be seen as the study of how uncertainty in the model output can be apportioned to different sources of uncertainty in the model inputs (Saltelli et al. 2008). Methods based on analysis of variance (ANOVA) are particularly suited for this purpose (Francos et al. 2003; Hall et al. 2005; van Griensven et al. 2006). Contrary to GLUE, ANOVA does not require system observation. The model output variance is synthetically generated by considering uncertain parameters as random variables. Using an ensemble of parameters sets, ANOVA

determines the contribution of each parameter to the unconditional variance.

Several SAs have been carried out on various RRM (Apel et al. 2004; Aronica et al. 1998; Bates et al. 2004; Pappenberger et al. 2004; Werner et al. 2005). These studies, mainly based on the GLUE method, were usually performed on short-time events (flood) at catchment scale. To the authors' knowledge, only Pappenberger et al. (2010) have performed a global-scale SA of the river-routing Total Runoff Integrating Pathways-2 (TRIP2) model (Ngo-Duc et al. 2007), coupled with the Hydrology Tiled ECMWF Scheme of Surface Exchanges over Land (HTESSEL; Balsamo et al. 2009). The study is based on a 10-yr reanalysis run (1986–95) of ECMWF, which was then compared with an ensemble of in situ discharges from more than 400 worldwide river gauging stations provided by the Global Runoff Data Centre (GRDC; <http://grdc.bafg.de>) by means of a performance measure (Pappenberger and Beven 2004). The ensemble of time-averaged performances was then used to apply both GLUE and ANOVA methods. The sensitivity of crucial parameters for the routing scheme such as slope, river length, groundwater time delay constant, and Manning formula parameters was studied. Results from both GLUE and ANOVA revealed that the groundwater time constant—which models the delayed contribution of drained water into the river—is the most sensitive parameter.

The present study displays an SA using a global method of the TRIP RRM version included in the land surface modeling platform Surface Externalisée (SURFEX; Masson et al. 2013) developed at the Centre National de Recherches Météorologiques (CNRM). For hydrological purposes, TRIP describes river discharge at catchment scale and is coupled to the well-established Interactions between Soil, Biosphere, and Atmosphere (ISBA; Boone et al. 1999) LSM used to simulate thermal and hydrological exchanges between the soil and the atmosphere.

In section 2, we present the ISBA–TRIP version used in this paper along with the Amazon River basin description. Section 3 provides a basic background of variance-based SA methods and details the present ANOVA SA configuration. In sections 4 and 5 we present results for two SA studies differing in terms of number and spatial distribution of perturbed parameters, which are referred to as AMA8 and AMA45. The last section presents our conclusions and perspectives.

2. The hydrological model and study zone

a. The ISBA–TRIP model classical configuration

The ISBA model (Boone et al. 1999) is a relatively standard LSM, including the force–restore method

(Blackadar 1976), which is adopted in the current study. The model is globally defined over a regular mesh grid at a 0.58° × 0.58° resolution. The model's equations are solved in each grid cell independently from the others. All grid cells are only correlated through the spatial patterns of vegetation cover, soil composition, and external forcing, that is, atmospheric (especially precipitation) and radiative inputs. By taking into account heterogeneity in precipitation, topography, and vegetation within each grid cell (Decharme and Douville 2006), ISBA gives a diagnosis of the water and energy budgets in each grid cell (Noilhan and Planton 1989).

The present study relies on the ISBA configuration used in Alkama et al. (2010) and Decharme et al. (2012). The soil is divided into three layers: the superficial layer, the root zone, and the subroot zone (Boone et al. 1999). Precipitation can either fall directly on the soil surface or be intercepted by the canopy. When the canopy is saturated, water drips from the canopy to the soil and then either flows on the surface or infiltrates. Hydraulic conductivity through soil layers is determined following an exponential profile of the saturated hydraulic conductivity (Decharme et al. 2006). The soil water content varies with surface infiltration, soil evaporation, plant evapotranspiration, surface runoff, and deep drainage [for more details, see Alkama et al. (2010) and Decharme et al. (2012)]. ISBA then gives a diagnosis of each water budget component, in particular the surface runoff Q_s and gravitational drainage Q_{sub} .

The TRIP RRM is also defined over a regular mesh grid. The present study is at the same resolution of 0.58° × 0.58° as ISBA. Because ISBA works for each grid cell separately, TRIP is dedicated to the lateral transfer of water from one cell to another up to the continent–ocean interface following a river network. The original version of TRIP was developed by Oki and Sud (1998) and consisted of a single linear surface reservoir. Surface runoff data from any LSM were needed to force TRIP and were then converted into river discharge under the assumption of a 0.5 m s⁻²¹ constant flow velocity. Subsequent developments by Decharme et al. (2010, 2012) first coupled TRIP with ISBA by taking ISBA simulated surface runoff as input. Then, a groundwater and floodplain reservoir was added and the constant flow velocity was changed to a time- and space-dependent flow velocity based on the following Manning formula:

$$v = \frac{p_{\text{eff}}}{n} R^{2/3}, \quad (1)$$

where s is the channel slope, n is the friction coefficient at the bottom of the channel, and R is the hydraulic radius depending on the channel geometry. In the present work, s , n , and the channel geometry parameters are constant in time but spatially distributed. The actual

TRIP model (Fig. 1), consists then in a system of three reservoirs: the surface reservoir S modeling the river, the groundwater reservoir G (Decharme et al. 2010),

and the floodplain reservoir F, solved using Eq. (2) through the estimation of the water mass stored in each reservoir (kg):

$$\begin{aligned} \frac{dS(t)}{dt} &= Q_s(t) - Q_{in,TRIP}^S(t) - Q_{out}^G(t) - Q_{out}^F(t) + 2 \frac{y(t)}{L} S(t), \\ \frac{dG(t)}{dt} &= Q_{sub}(t) - 2 Q_{out}^G(t), \quad \text{and} \\ \frac{dF(t)}{dt} &= Q_{in}^F(t) - 2 Q_{out}^F(t) - [P_F(t) - I_F(t) - E_F(t)]. \end{aligned} \quad (2)$$

Only the surface reservoir sends water from cell to cell based on the TRIP routing network. A cell can receive water from several upstream cells but sends water into a unique downstream cell. For any given cell, TRIP inputs are the TRIP outflow from upstream cells $Q_{in,TRIP}^S(t)$ (kg s^{-21}) and the ISBA surface runoff for that cell Q_s (kg s^{-21}). Moreover, S receives water from groundwater $Q_{out}^G(t)$ (kg s^{-21}) and can exchange water mass with the floodplain $Q_{out}^F(t) - 2 Q_{in}^F(t)$ (kg s^{-21} ; Fig. 1).

With the implementation of G in TRIP, ISBA Q_{sub} (kg s^{-21}) now flows into the groundwater reservoir, whose outflow goes to river reservoir. This simple modeling assumption does not represent the real groundwater dynamics, but only the delay of drainage contribution to the river. The groundwater outflow [i.e., $Q_{out}^G(t)$] (kg s^{-21}) is then estimated by

$$Q_{out}^G(t) = \frac{G(t)}{t_G}, \quad (3)$$

where t_G (s) is the time delay factor. When introduced in Decharme et al. (2010), t_G was taken as a constant over river catchment with values between 30 and 60 days.

The floodplain scheme activates when the water height in the river h_s (m) exceeds a given critical bankfull height H_c (m). The terms P_F , I_F , and E_F are the precipitation intercepted by the floodplain, the infiltration, and the direct free water surface evaporation over the floodplain, respectively. A detailed description of the flood scheme is given in Decharme et al. (2008, 2010, 2012).

b. TRIP specific parameters

Within a cell, S corresponds to an equivalent river that may represent in reality several river branches. The river section is rectangular and is characterized by a slope s (unitless), width W (m), depth H_c (m), length L (m), and a Manning coefficient n (unitless) that quantifies friction as well as channel resistance at the bottom of the river. A more extensive description of these TRIP specific parameters is given below.

The section length is the arc length between the cell center and the direct downstream cell center multiplied by a meandering factor m , equal to 1.4 in order to take the river meanders into account.

The riverbed slope is deduced from the elevation difference between cells ($E - E_{next}$) and based on the Simulated Topological Network (STN-30p) DEM provided at 0.58 3 0.58 resolution by the International Satellite Land Surface Climatology Project, Initiative II (ISLSCP II) database (http://daac.ornl.gov/ISLSCP_II/islscpii.shtml) and the section length as follows:

$$L = 5 \min[10^{25}, \frac{E - E_{next}}{L}]. \quad (4)$$

The river width is globally computed via an empirical geomorphologic relationship between W and the mean annual discharge Q_{yr} ($\text{m}^3 \text{s}^{-21}$; Moody and Troutman 2002):

$$W = 5 \max[20, b (Q_{yr})^{0.5}], \quad (5)$$

with $b = 32$ ($\text{m}^{21/2} \text{s}^{1/2}$) for an equatorial basin such as the Amazon River (Decharme et al. 2012).

The river depth is computed from the river width using the following nonlinear function:

$$H_c = 5 W^{1/3}. \quad (6)$$

The Manning coefficient can be defined as channel resistance to the flow and is generally difficult to estimate. It takes low values (0.025–0.03) for natural streams with deep pools and higher values for small, mountainous streams and floodplains (0.75–0.1) (Maidment 1993). Previous global studies used a global constant value taken equal to 0.035 (Arora et al. 1999). Still, it is commonly assumed that the Manning coefficient should be spatially distributed over an entire catchment. In TRIP, n is therefore spatially distributed according to a simple relationship considering that upstream grid cells contain narrow rivers with high Manning coefficients and that the Manning value

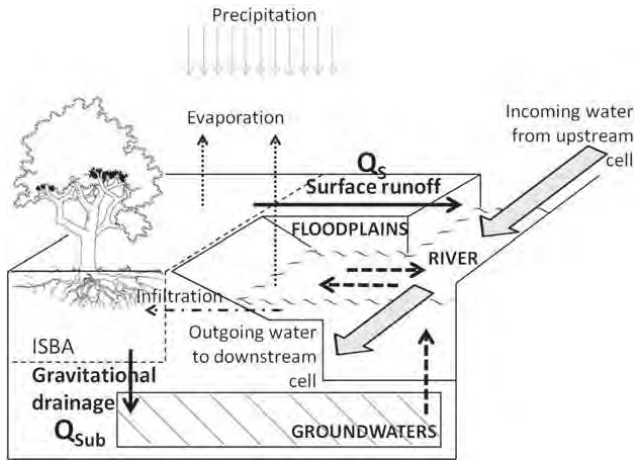


FIG. 1. The ISBA-TRIP system for a given grid cell. ISBA surface runoff (i.e., Q_s) flows into the river/surface reservoir (i.e., S) and ISBA gravitational drainage (i.e., Q_{sub}) feeds the groundwater reservoir (i.e., G). The surface water is transferred from one cell to another following the TRIP river-routing network.

decreases as the cells become closer to the river mouth. Arbitrarily, n was chosen to vary between n_{min} 5 0.04 and n_{max} 5 0.06 according to the following linear relation:

$$n = n_{min} + \frac{(n_{max} - n_{min})}{SO_{max} - SO_{min}} (SO - SO_{min}), \quad (7)$$

with SO being a measure of the relative size of stream in the current cell, SO_{max} being the same measure at the river mouth, and SO_{min} (51) being the measure at source cells.

All these parameters are critical to determine the flow velocity with the Manning formula:

$$y(t) = \frac{1}{n} \left(\frac{Q_s(t)}{W} \right)^{1/2} \frac{r}{2h_s(t)}, \quad (8)$$

where the river water height (m) is estimated from its actual water mass S and river geometry as follows

$$h_s(t) = \frac{S(t)}{r_w L W}, \quad (9)$$

where r_w (kg m^{-3}) is the water density.

The system in Eq. (2) is solved using the fourth-order Runge-Kutta method. Once the water storage (kg) of each reservoir is computed, TRIP determines in each cell a diagnosis of the quantities of the river water height (m), following Eq. (9), and the river discharge ($\text{m}^3 \text{s}^{-1}$), defined as follows:

$$Q_{out}^S(t) = \frac{y(t)}{L} S(t). \quad (10)$$

c. Description of the Amazon basin

In terms of average discharge ($2.3 \times 10^5 \text{ m}^3 \text{s}^{-1}$) and drainage area ($6.15 \times 10^6 \text{ km}^2$), the Amazon is the world's largest river. Its catchment area covers about 40% of South America, and the discharge at its mouth represents 30% of total freshwater inflow to the Atlantic Ocean (Wisser et al. 2010).

As shown in Fig. 2a, the river source is located in the Peruvian Andes. The river flows through the Brazilian tropical rain forest and receives water from several important tributaries: the Japurá River, the Purus River, and the Negro River (16% of the total discharge) at Manaus. At this point, the river has reached 56% of its total discharge. From Manaus to its mouth, it receives water from the Madeira River (17% of the total discharge), the Tapajós River, and the Xingu River (11% of the total discharge; Molinier et al. 1993).

From a geological perspective, the Amazon basin can be divided in three major morphostructural units: the

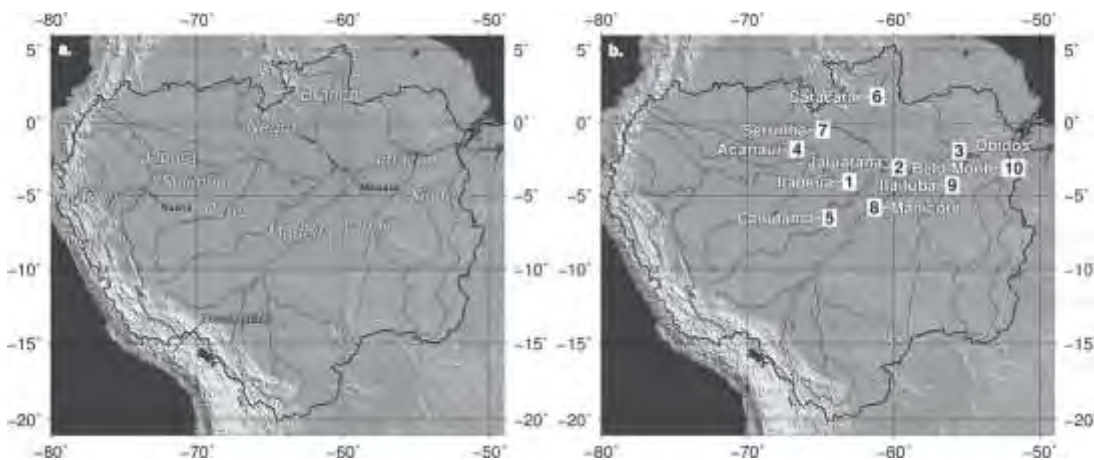


FIG. 2. Maps of the Amazon basin. (a) The Amazon River and its main tributaries. (b) In situ stations along the Amazon River and its main tributaries.

shields at the eastern part of the basin (Guiana shield to the north and Brazilian shield to the south), the central Amazon trough, and the western Andean Cordillera. In terms of vegetation, the basin is covered by tropical rain forest (71%) and savannas (29%; [Sioli 1984](#)).

The whole Amazon basin is located in a humid tropical climatic zone. The central basin is under an equatorial climate zone, implying high surface temperatures, air humidity, and precipitation. Thus, a vast floodplain along the mainstream is filled every year, leading to the damping of discharge extremes. Northern and southern parts of the basin are under a tropical climate with a dry and a wet season. However, because of the seasonal shift of the intertropical convergence zone, the maximum rainfall season occurs at different periods during the year ([Meade et al. 1991](#)). This implies that annual peak discharge in southern tributaries occurs a few months earlier than in northern tributaries.

3. SA with the ANOVA method

This section presents the SA method used in this study.

a. SA generalities

Variance-based SA methods aim to analyze the model output variance against an ensemble of simultaneously modified parameters ([Efron and Stein 1981](#); [Sobol 1993](#)) and are efficient even for nonlinear and nonmonotonic models ([Saltelli and Bolado 1998](#)). The goal is to estimate a set of sensitivity indices (S_i ; one for each parameter) that represent the contribution of the associated parameter in the model output unconditional variance. The two most usual measures of sensitivity are the main effect and the total effect. The main effect is used in factor privatization searching for the parameter that controls the most model output uncertainty. The total effect, on the contrary, is used in factor fixing searching for irrelevant parameters that can be constrained to an arbitrary value ([Saltelli et al. 2008](#)).

There are several methodological variants to calculate partial variances, for example, Sobol's method ([Sobol 1993](#)) or the simple and extended Fourier amplitude sensitivity test (FAST and EFAST; [Cukier et al. 1973](#); [Fang et al. 2003](#)). The main difference between FAST and the Sobol method lies in the approach used to calculate the variances multidimensional integrals. Whereas Sobol's method uses a Monte Carlo integration procedure, FAST uses a pattern search based on sinusoidal functions. However, [Reusser et al. \(2011\)](#) tested the three sampling methods (FAST, EFAST, and Sobol) and found comparable results between each method. Even though these methods are quite popular, they still

have a high computational cost by requiring a large number of model runs to estimate S_i .

Currently, some studies focus on direct calculation of the double-loop multidimensional integrals underlying S_i by trying to reduce computational cost ([Jansen 1999](#); [Saltelli et al. 2010](#); [Sobol 1993](#)). Alternatives have been found to overcome dependence on the problem of dimensionality (linked to the number of parameters) and thereby reducing computational cost, in particular through high-dimensional model representation (HDMR) and metamodeling ([Rabitz et al. 1999](#); [Ratto et al. 2007](#)).

In addition, [Sieber and Uhlenbrook \(2005\)](#) and [Reusser et al. \(2011\)](#) introduced temporal dynamics of parameter sensitivity (TEDPAS) and time series of grouped error in order to perform temporal SA of any model output variable such as river discharge. This kind of analysis aims to identify dominant parameters during a given hydrological process and highlight parameter interactions. For example, [Garambois et al. \(2013\)](#) performed a temporal Sobol's sample ANOVA to identify key parameters of the physically based and spatially distributed Model of Anticipation of Runoff and Inundations for Extreme Events (MARINE) during flash flood events. [Guse et al. \(2014\)](#) used a temporal FAST-ANOVA method on the Soil and Water Assessment Tool (SWAT) model to detect which parameter dominates in poor model performance cases.

In this study, the state-dependent parameter (SDP) metamodeling approach developed by [Ratto et al. \(2007\)](#) is used. Also, a temporal sensitivity analysis following [Garambois et al. \(2013\)](#) and the [Reusser and Zehe \(2011\)](#) TEDPAS methodology is applied. The ANOVA is based on Sobol's sampling method as in [Ratto et al. \(2007\)](#).

b. The ANOVA formulation

The formulation of the ANOVA decomposition is based on [Sobol \(2001\)](#). The model output Y can be defined as a function of k independent uncertain input parameters $p_1, p_2, \dots, p_k$:

$$Y = M(p_1, p_2, \dots, p_k), \quad (11)$$

where M is the model operator and the parameters $p_1, p_2, \dots, p_k$ are defined over V^k , the set of all possible values. It is ensured that the range of variation of these parameters corresponds to the assumed parameter uncertainties and, in the following SA studies, leads to realistic values for these parameters $p_1, p_2, \dots, p_k$.

Now, let $X = (X_1, X_2, \dots, X_k)$ be a possible set of normalized parameters (i.e., $p_1, p_2, \dots, p_k$ scaled between 0 and 1). The relationship that links the normalized model parameters with Y is denoted by

$$Y = f(X) = f(X_1, X_2, \dots, X_k), \quad (12)$$

where f is also the model operator but taking the normalized parameters $X_1, X_2, \dots, X_k$ as inputs.

It is assumed that Y can be decomposed in a sum of real functionals with increasing input dimensionality of the following form:

$$Y = f_0 + \sum_{i=1}^k f_i(X_i) + \sum_{i=1}^k \sum_{j=1}^k f_{ij}(X_i, X_j) + \dots + f_{1,2,\dots,k}(X_1, X_2, \dots, X_k). \quad (13)$$

Under the hypothesis that f is integrable over the unit hypercube of size k , it is assumed that the following integral holds $\int_0^1 \dots \int_0^1 f_{i_1, \dots, i_s}(X_{i_1}, \dots, X_{i_s}) dX_{i_1} \dots dX_{i_s} = 0$.

$$\int_0^1 \dots \int_0^1 f_{i_1, \dots, i_s}(X_{i_1}, \dots, X_{i_s}) dX_{i_1} \dots dX_{i_s} = 0. \quad (14)$$

The variable $f_{i_1, \dots, i_s}(X_{i_1}, \dots, X_{i_s})$ can take any real values, so the above integral implies that the $f_{i_1, \dots, i_s}$ values have a zero mean. It follows from Eq. (14) that the functionals are orthogonal and can be expressed as an integral of $f(X)$. Indeed,

$$\int_0^1 \dots \int_0^1 f(X) dX = f_0 \quad \text{and} \quad (15)$$

$$\int_0^1 \dots \int_0^1 f(X) dX_{\bar{i}} = f_0 + f_i(X_i), \quad (16)$$

where $X_{\bar{i}}$ signifies that the integration is performed over all parameters but X_i . Successively,

$$\int_0^1 \dots \int_0^1 f(X) dX_{\bar{i}j} = f_0 + f_i(X_i) + f_j(X_j) + f_{ij}(X_i, X_j). \quad (17)$$

Moreover, if f is also square integrable over the unit hypercube of size k , so are the summands and

$$\int_0^1 \dots \int_0^1 f^2(X) dX = f_0^2 + \sum_{i=1}^k \sum_{j=1}^k \int_0^1 \dots \int_0^1 f_{ij}^2(X_i, X_j) dX_i dX_j. \quad (18)$$

Now, X is considered as a random variable uniformly distributed over the unit hypercube of size k , meaning that each parameter X_i follows a continuous uniform law between 0 and 1. The model output $f(X)$ and all the summands $f_{i_1, \dots, i_s}$, $\int_0^1 \dots \int_0^1 f_{i_1, \dots, i_s}^2(X_{i_1}, \dots, X_{i_s}) dX_{i_1} \dots dX_{i_s}$ are random variables as well. This implies that integrals of f and $f_{i_1, \dots, i_s}$ represent expectations and integrals of the variances of f^2 and f_{ij}^2 .

For instance, the model output expectation is given by

$$E(Y) = \int_0^1 \dots \int_0^1 f(X) dX = f_0 \quad (19)$$

and the model output conditional expectations given first-order factors are defined as

$$E(Y | X_i) = \int_0^1 \dots \int_0^1 f(X) dX_{\bar{i}} = f_0 + f_i(X_i), \quad (20)$$

meaning that the expectation of the model output Y is computed over all possible values of X keeping X_i fixed.

A similar definition holds for higher-order factors:

$$E(Y | X_i, X_j) = f_0 + f_i(X_i) + f_j(X_j) + f_{ij}(X_i, X_j). \quad (21)$$

The total variance (Var) of the model output Y is then formulated as follows:

$$\begin{aligned} \text{Var}(Y) &= \int_0^1 \dots \int_0^1 f^2(X) dX - f_0^2 \\ &= \sum_{i=1}^k \sum_{j=1}^k \int_0^1 \dots \int_0^1 f_{ij}^2(X_i, X_j) dX_i dX_j. \end{aligned} \quad (22)$$

Rewriting Eq. (22) based on Eq. (13) using Eq. (14), the total variance can be expressed as its so-called ANOVA decomposition:

$$\text{Var}(Y) = \sum_{i=1}^k V_i + \sum_{i=1}^k \sum_{j=1}^k V_{ij} + \dots + V_{1,2,\dots,k}, \quad (23)$$

where

$$V_i = \text{Var}(f_i) = \int_0^1 \int_0^1 f_i^2(X_i) dX_i - f_0^2 \quad (24)$$

is called the main effect of parameter X_i on Y representing the expected reduction of total variance if X_i could be fixed and where

$$V_{ij} = \text{Var}(f_{ij}) = \int_0^1 \int_0^1 f_{ij}^2(X_i, X_j) dX_i dX_j - f_0^2 \quad (25)$$

is the covariance caused by combined effects of parameter X_i and X_j on Y . Two parameters are said to interact when their effects on Y cannot be only explained by V_i and V_j , and so on for higher-order terms.

Sensitivity indices are finally estimated from the partial variances introduced in Eqs. (24) and (25) by dividing them by the total variance, giving

$$S_{i, \dots, j} = \frac{V_{i, \dots, j}}{\text{Var}(Y)}. \quad (26)$$

They represent the part of a given single parameter (or a given subset of parameters) variance in the unconditional variance. They could be seen as an estimate of the

model sensitivity to this particular parameter. From Eq. (23), it appears that

$$\sum_{i=1}^k S_i \mathbf{1} \sum_{j=1}^k S_j \mathbf{1} \cdots \mathbf{1} S_{1,2,\dots,k} \mathbf{1}. \quad (27)$$

In practice, calculating the sensitivity indices requires first the evaluation of the integrals in Eqs. (24) and (25) along with the double-loop integral in Eq. (22). This evaluation is achieved stochastically using an ensemble of model output realizations. First, the ensemble of input parameter sets $X_e = [X^{(1)}, X^{(2)}, \dots, X^{(N_e)}]$, with N_e the size of the ensemble, are generated using a quasi-random sequence generator. As each parameter X_i is independent from others and normalized, each parameter X_i follows a continuous uniform law between 0 and 1. Then, the model output ensemble is established by running the model f with each one of the N_e input parameter sets.

Second, the f_i are calculated using the “SDP meta-model” approach framed by Ratto et al. (2007). It is designed to estimate f_i functionals until second order, that is, f_i and $f_{i,j}$ in Eq. (13). This is done with the Smoothing Spline ANOVA (SS-ANOVA) routine (available at <https://ec.europa.eu/jrc/en/econometric-statistical-software#SS-ANOVA-R>). Once the terms $f_i(X_i)$ are estimated for the entire ensemble, the associated S_i are straightforwardly estimated using the estimator suggested in Doksum and Samarov (1995). The advantage of this method is its adaptability to any sampling method of the input ensemble, as it provides fast, accurate, and unbiased results. The SS-ANOVA procedure provides (besides an estimation of the f_i coefficients) the standard errors of SDP and hence the relative significance of estimated HDMR terms. Finally, a last important routine output is the metamodel correlation coefficient R^2 measuring the likeness between the original and emulated model. In the present study, only first-order S_i are estimated and studied. Model output variance explained only with first-order S_i characterizes a purely additive model and will be confirmed by an R^2 close to 1 and low interaction effects (quantified by $1 - \sum_{i=1}^k S_i$). Otherwise, it will indicate nonnegligible interaction effects. For more details, refer to Ratto et al. (2007).

To sum up, the ANOVA method produces a set of sensitivity indices taking value between 0 and 1 and associated with a single or a subset of input parameters. These indices represent a percentage of the total variance.

Two SAs with different choices of parameters (but all with the same general configuration) have been carried out in this study and are detailed in the following subsection.

c. Simulation description

In this study, the SA is based on a sampling of the parameter space of size $N_e = 1024$ drawings [this value

was chosen following Garambois et al. (2013) and has reasonable computational costs]. To generate the initial ensemble of normalized input parameter sets X_e , the quasi-random sequence LP_t proposed by Sobol (1967) is used. These sequences use a base of two to successively form finer partitions of the unit interval and then reorder the coordinates in each dimension. They spawn the entire unit hypercube for any dimension (the computational routine can generate until $2^{30} - 1$ points of maximum dimension 52); and thereby the entire range of definition for each input parameter is explored.

A model output ensemble element consists of daily time series over the period 2008–10. The temporal SA is conducted over the three years. The years 2006 and 2007 are used as spinup to avoid effects of the initial condition and transitive states in the model output variance, but no SA is conducted over these two years. The year 2009 was a particularly wet year, as mid-June floods reached their highest level for 50 years. Conversely, in 2010, the Amazon River experienced an extreme drought and reached its lowest level for half a century. The year 2008 is considered an average year. Therefore, studying the model sensitivity for the full 3-yr period from 2008 to 2010 is relevant for all types of hydrological years.

ISBA-TRIP is run in offline mode and uses atmospheric forcing from the Global Soil Wetness Project phase 3 (GSWP3; <http://hydro.iis.u-tokyo.ac.jp/GSWP3>). This project consists of three global-scale experiments with the objective of investigating long-term changes of the energy–water–carbon cycle components and their interactions. The 3-hourly resolution atmospheric boundary conditions used in the present study were generated by dynamically downscaling the global 28-resolution Twentieth Century Reanalysis (Compo et al. 2011).

ISBA-TRIP estimates water levels and discharges at each of the 2028 cells contained in the Amazon basin mesh grid. For reasons of computational cost, the temporal SA is detailed for a selection of cells corresponding to known in situ stations (Fig. 2b). Table 1 briefly describes these evaluation cells.

To carry out this temporal SA, two different sets of parameters have been chosen and presented in the next section.

d. Choice of parameters

1) GLOBAL SA OVER THE AMAZON BASIN

The first step is to analyze TRIP model sensitivity to all parameters over the entire basin. All TRIP parameters are considered: the river and floodplain Manning coefficient n (unitless) and n_F (unitless), the groundwater time constant t_G (days), the riverbed slope s (unitless), the river width W (m), the river bankfull depth H_c (m),

TABLE 1. Description of the in situ stations used for the study (Fig. 2b).

	Station	Lat	Lon	River	Drained area (km ²)
1	Itapéua	5.958S	64.998W	Solimões	1 780 000
2	Jatuarana	4.958S	60.378W	Amazon	2 930 000
3	Óbidos	2.048S	56.508W	Amazon	4 670 000
4	Acanauí	2.098S	67.408W	Japurá	249 000
5	Canutama	7.478S	65.628W	Purus	236 000
6	Caracarai	1.818N	62.898W	Branco	126 000
7	Serrinha	1.548S	65.198W	Negro	293 000
8	Manicoré	6.818S	62.708W	Madeira	1 150 000
9	Itaituba	5.748S	56.048W	Tapajós	458 000
10	Belo Monte	4.908S	52.248W	Xingu	482 000

and the river section length meandering coefficient m (unitless). See Table 2 for an overview of these parameters' nominal range.

Apart from the groundwater time constant and the meandering ratio, which are constant, all other parameters are spatially distributed over the entire basin. To avoid overparameterization, parameter fields are perturbed by applying a constant multiplying factor. Then, sensitivity to a given field is studied via this multiplying factor. Moreover, particular attention is paid to the perturbation of the riverbed slope field. Indeed, following Pappenberger et al. (2010), the riverbed slope is perturbed by first applying a powering constant s_{pow} and then a multiplying factor s_{mult} . Multiplying and powering factors have a nominal value of 1 and are defined in a 6% interval around it. The « value is chosen so that the perturbation range is representative of the parameter's uncertainty (Gatelli et al. 2009). Existing parameter value tables for the Manning coefficient (Maidment 1993), previous studies (Pappenberger et al. 2010; Decharme et al. 2010; Paris et al. 2016), and comparison to a remotely sensed optical-image-based database for river width (Yamazaki et al. 2014) were used to fix « for each parameter, which led to the range of parameter listed in Table 3.

This first SA study is denoted by AMA8—“AMA” for Amazon (though it is planned to extend the ANOVA method to other basins in future work) and “8” since eight perturbed parameters are considered. Note that except for the riverbed slope (for which two coefficients are used), the sensitivity to any of the other parameters presented in Table 3 will be directly referred to as the sensitivity to the corresponding TRIP parameter (see fifth column in Table 3); for example, sensitivity to W_{mult} will be directly mentioned as the sensitivity to W . When discussing slope sensitivity, the use of s_{mult} and s_{pow} will be preserved.

TABLE 2. Name and nominal range or value for each TRIP parameter.

TRIP parameter	Nominal range/value
t_G	30 days
W	20–15 000 m
H_c	3–17 m
n	0.04–0.06
s	From 2.3×10^{24} to 5.3×10^{22}
m	1.4
n_F	0.03–0.09

2) REGIONAL SA OVER THE AMAZON BASIN

In the first SA study, a preliminary insight into TRIP model sensitivities is drawn. Based on these results, a second SA study, referred to as AMA45 in the following, is carried out and aimed at regionalized sensitivities (disregarded in the first SA, since parameter perturbations are made at the entire basin scale for AMA8). A subset of input parameters considered as the most critical ones for understanding TRIP model behavior is then considered per geographical zone. For this purpose, the Amazon basin is divided into nine hydrogeomorphological zones. The interest of such a subdivision is to study the separated impact of the parameters of each zone. These zones were designed according to 1) hydrological arguments as the main course is separated from the tributaries that have their own zones and 2) geological arguments as three major morphostructural units are distinguishable.

The nine zones (Fig. 3) are the following: 1) the upstream Andean part of the basin until the city of Iquitos, Peru; 2) the main stream from Iquitos to Óbidos; 3) the main stream from Óbidos to the river mouth; 4) left-bank tributaries from the Napo River to the Japurá River, including the Japurá River; 5) left-bank tributaries from the Japurá River to Óbidos, including the Negro River and its drainage area; 6) right-bank tributaries from Iquitos to the Purus River confluence at Anamá; 7) right-bank tributaries from Anamá to Óbidos, including the Madeira River; 8) right-bank tributaries exiting in zone 3, including the Tapajós River and the Xingu River; and 9) left-bank tributaries exiting in zone 3.

The procedure of AMA45 is to select a reduced number of important parameters from AMA8 results and rerun the SA by considering each one of these parameters in any of the nine Amazon basin zones. The overall SA configuration will remain identical to the first SA in terms of number of members within the ensembles (N_e 5 1024), using atmospheric forcing (GSWP3) and configuration of the ISBA–TRIP runs (three years from 2008 to 2010 with spinup from 2006 to 2007).

TABLE 3. Input parameter ranges used in the SA referred to as AMA8. The second column gives the nature of the parameter: direct value of the TRIP parameter (dir), multiplying factor (mult), and powering factor (pow); the third column gives the nominal value of the parameter. The sixth column represents the complete range of variation for physical input parameters in the SA study after perturbations given in the fourth column.

SA parameter	Nature	Nominal value	Perturbation 6«%	TRIP parameter	New range
t_G	Dir	30 days	—	t_G	15–60 days
W_{mult}	Mult	1	650%	W	10–22 500
$H_{c,mult}$	Mult	1	650%	H_c	1.5–25.5
n_{mult}	Mult	1	675%	n	0.01–0.105
s_{mult}	Mult	1	620%	s	From 3×10^{25} to 7×10^{22}
s_{pow}	Pow	1	620%	s	Range above obtained by applying both s_{mult} and s_{pow}
m	Dir	1.4	630%	L	1.0–1.8
$n_{F,mult}$	Mult	1	675%	n_F	0.007–0.15

4. Results and discussion for AMA8

a. Water level sensitivity

Figures 4–6 display the resulting first-order sensitivity indices for TRIP water level [see Eq. (9)]. Table 4 shows the time average over the 3-yr study period of the metamodel R^2 and sum of all higher-order effects (i.e., the interaction effects) for each cell. The R^2 close to 1 indicates a good convergence of the method, and interaction effects close to 0 indicate that the main effects (first-order sensitivity indices) are sufficient to explain the water height variance, characterizing purely additive contribution of the parameters.

The simulated water height ensemble for each evaluation cell generally presents a large spread with amplitude up to 20 m for some cells such as Jatuarana and Óbidos (Fig. 4). The ensemble first and ninth deciles (plotted in black solid and dashed lines, respectively) are clearly distant all year long, even during low-flow season (when the ensemble dispersion is at its lowest). It can be already pointed out that water level presents an important sensitivity to TRIP parameters.

Overall, the sensitivity time series show no clear interannual variations, and four parameters are systematically activated: first, the river Manning coefficient (purple line), closely followed by the riverbed slope powering coefficient (green line), then the riverbed width (orange line), and the groundwater time constant (blue line). In some cases, the river section length meandering coefficient is slightly activated (red line in Figs. 5b, 6). In most figures, S_n , $S_{s_{pow}}$, and S_W explain 40%, 35%, and 20%, respectively, of the model output variance; S_{t_G} and S_m explain a small percentage of the sensitivity also during low-flow period. To summarize the sensitive parameters by order of importance are the following:

$$n - s_{pow} - W - t_G \gg m. \quad (28)$$

Three different behaviors are observed. First, evaluation cells along the Amazon mainstream—Itapéua, Jatuarana, and Óbidos (Fig. 4)—show very light inter-seasonal variations, implying that only three parameters' S_i are significant: S_n and $S_{s_{pow}}$ are quasi-constant and S_W presents a smooth increase when the water level is at its lowest. According to Eq. (9), h_s depends on the actual river reservoir storage and on the channel geometry (its width and length). The dominating parameters n , s_{pow} , and W are directly involved in the Manning formula [Eqs. (1) and (8)] that estimates the flow velocity. The flow velocity itself appears in Eq. (2) to determine the river outflow, thereby impacting the surface water storage. Therefore, these parameters represent the contribution of the S to h_s .

Second, the left-bank-tributary cells—Acanauí, Caracarái, and Serrinha (Fig. 5)—present a similar behavior, but sensitivities to n , s_{pow} , and W are noisier, which can be related to precipitation events. Indeed, left-bank tributaries present smaller drainage areas and are

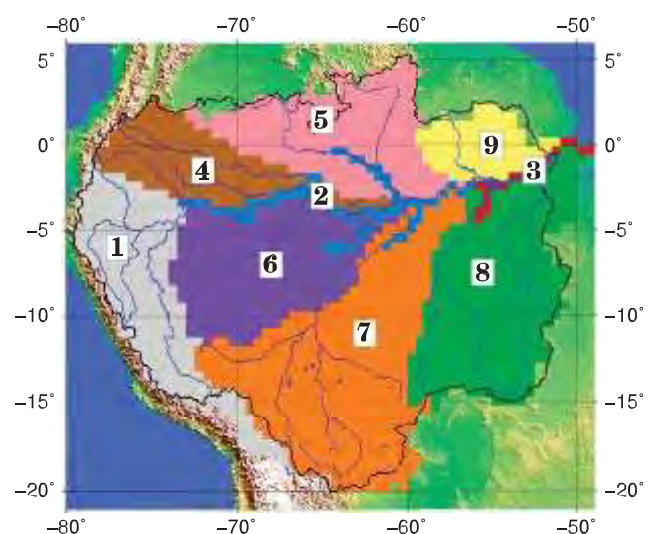


FIG. 3. Hydro-geomorphological areas of the Amazon basin.

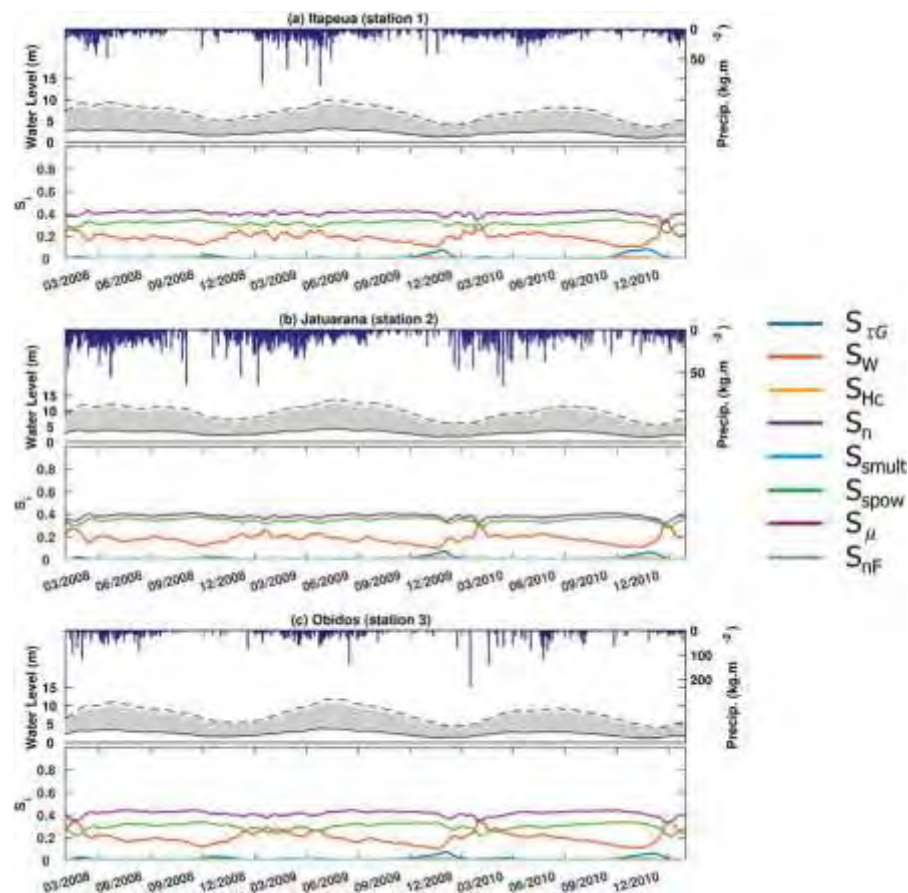


FIG. 4. Results for the AMA8 water level SA for cells along the Amazon mainstream: (a) Itapéua, (b) Jatuarana, and (c) Óbidos. For each panel, the upper graph shows the ensemble of simulated TRIP water heights (gray area) with the ensemble's first (black line) and ninth (black dashed line) deciles on the left y axis and daily precipitation (bar plot) at the observed cell on the reverse right y axis. The lower graph shows first-order daily S_i : S_{t_G} (blue line), S_W (orange line), S_{H_c} (yellow line), S_n (purple line), S_{smult} (light blue line), S_{spow} (green line), S_m (dark red line), and S_{nf} (gray line).

therefore more sensitive to local events, namely precipitation. Besides, S_W peaks are more numerous here, but, as for the general result cells (Fig. 4), they appear when the water level ensemble spread is very narrow and suddenly increases. It is worth noting that for the strongest peaks a smaller but synchronized S_m peak is observed. This implies that during sudden regime changes, the reservoir geometry controls the water level dynamics.

Third, the right-bank-tributary cells—Canutama, Manicoré, Itaituba, and Belo Monte (Fig. 6)—present parameter sensitivity with clear interseasonal patterns: S_n and S_{spow} mainly dominate as observed in previous results but decline slightly in low-flow season. During this period, S_{t_G} distinctly appears and even overpasses other S_i . Eventually, by the end of the low-flow season, S_W and S_m have peaks similar to those observed in Fig. 5, which occur when the water level rises again. The S_{t_G} activates exclusively during low-flow season. This parameter

represents the water inflow from the groundwater reservoir into the river. It is a continuous source of water in the river, but the ANOVA method, as used here, determines sensitivities as a fraction of the overall variance. Then, the groundwater storage contribution takes a higher proportion during the low-water season when other contributions (precipitation, surface runoff, etc.) impacting the water level are lowered. This is undoubtedly due to the very low water level during this season compared to the rest of the year. This seasonal pattern in S_i and the high contribution of t_G is more pronounced for right-bank tributaries than left-bank tributaries and is explained by the large-amplitude difference between the ensemble maximum and minimum values (almost 20 m).

b. Discharge sensitivity

Figures 7–9 display first-order sensitivity for TRIP discharges, and Table 5 presents the time average over

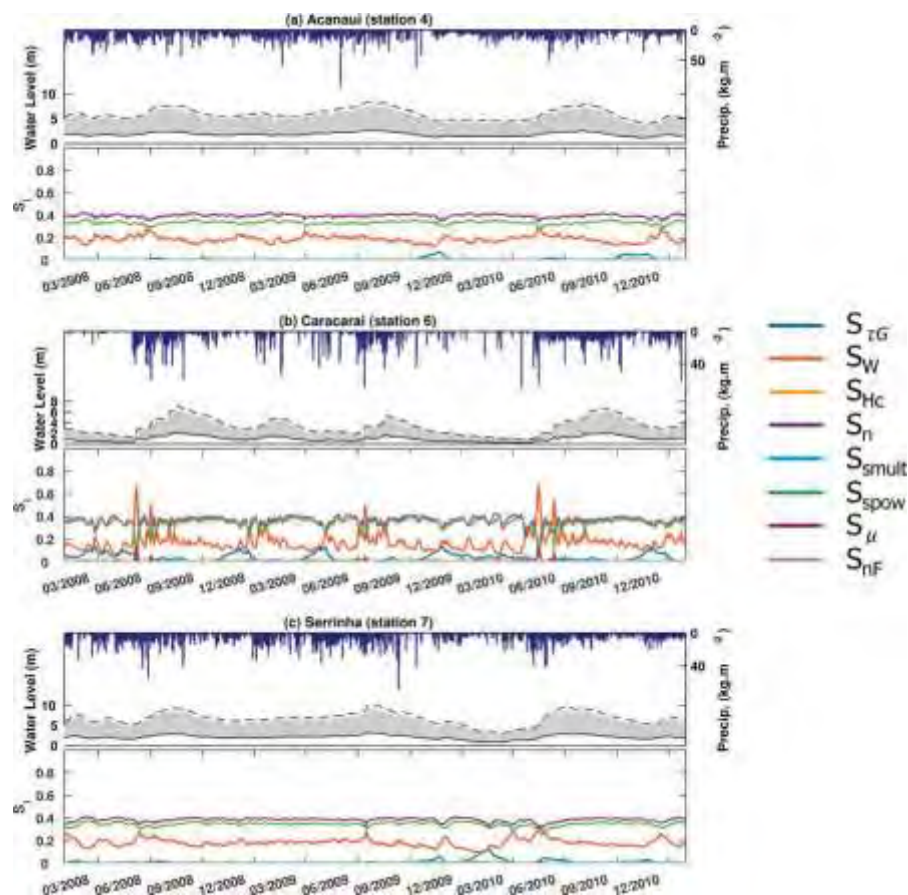


FIG. 5. As in Fig. 4, but for cells along left-bank tributaries: (a) Acanai, (b) Caracará, and (c) Serrinha.

the study period for R^2 and sum of all interaction effects. The metamodel coefficients are slightly lower than those obtained for water levels but generally remain above 0.85, showing again a good convergence of the method. However, discharge interaction effects account for 10%–15% of the unconditional variance (e.g., first-order effects explain 85%–90% of the variance).

Contrary to water level ensemble, the simulated discharge ensembles (gray area) present a quite narrow spread with tightened first and ninth deciles (plotted in black solid and dashed lines, respectively) close to ensemble extrema. The same ensemble of input parameter sets X_e has been used to generate both the water height ensemble and the discharge ensemble (see section 3d).

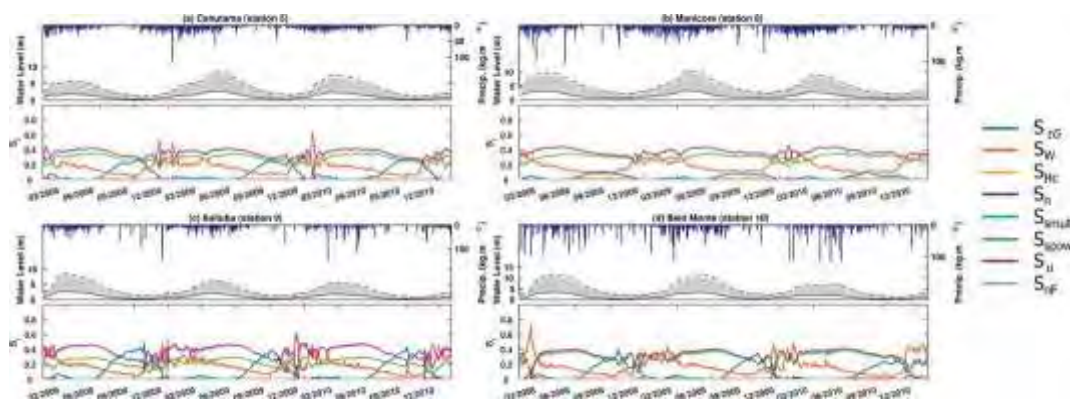


FIG. 6. As in Fig. 4, but for cells along right-bank tributaries: (a) Canutama, (b) Manicoré, (c) Itaituba, and (d) Belo Monte.

TABLE 4. AMA8 water level sensitivity. Time-averaged meta-model R^2 and interaction effects for each result cell (see Fig. 2b).

Cell	Time-averaged R^2	$1 \pm S_i $
1	0.95	0.05
2	0.95	0.06
3	0.95	0.06
4	0.95	0.05
5	0.94	0.07
6	0.94	0.07
7	0.95	0.05
8	0.93	0.08
9	0.94	0.07
10	0.93	0.08

But, contrary to the quite large water height ensemble dispersion, the discharge ensemble dispersion is narrower. A first outcome is that discharge is much less sensitive to TRIP parameters.

In similarity with water height results, the discharge sensitivity displays no remarkable interannual sensitivity patterns. The dominating parameters are the same as those for water height but with a different ordering, that is,

$$t_G - n - s_{\text{pow}} - m - W. \quad (29)$$

The parameter t_G is the most dominating parameter, and S_{t_G} may exceed 0.9 (e.g., Fig. 9) and has its highest values during low- and high-flow seasons. Parameter t_G is the parameter for groundwater flow exfiltration to the drainage network (Roux et al. 2011) and represents the mass transfer from the groundwater into the river. This high sensitivity is in accordance with the rainfall–runoff modeling literature and with the conclusions of Pappenberger et al. (2010). In a stable regime, the discharge is mainly driven by water mass transfer from the upstream river and groundwater.

During the transition period between high- and low-flow season, and also in response to precipitation events (Fig. 8 mostly), t_G sensitivity drastically drops and other parameters are sensitive. Figures 7–9 display an evident anti-correlation between S_{t_G} and all other parameter S_i . For example, time- and station-averaged correlation is -0.89 and -0.87 between S_{t_G} and S_n and between S_{t_G} and S_m , respectively. On the contrary, those other parameter sensitivities are highly correlated, for example, 0.94 and 0.97 for the time- and station-averaged correlation between S_n and S_{spow} and between S_m and S_W , respectively.

Between high- and low-flow seasons, or during precipitation events, the discharge is driven by regime changes, that is, by flow velocity changes. Thus, parameters involved

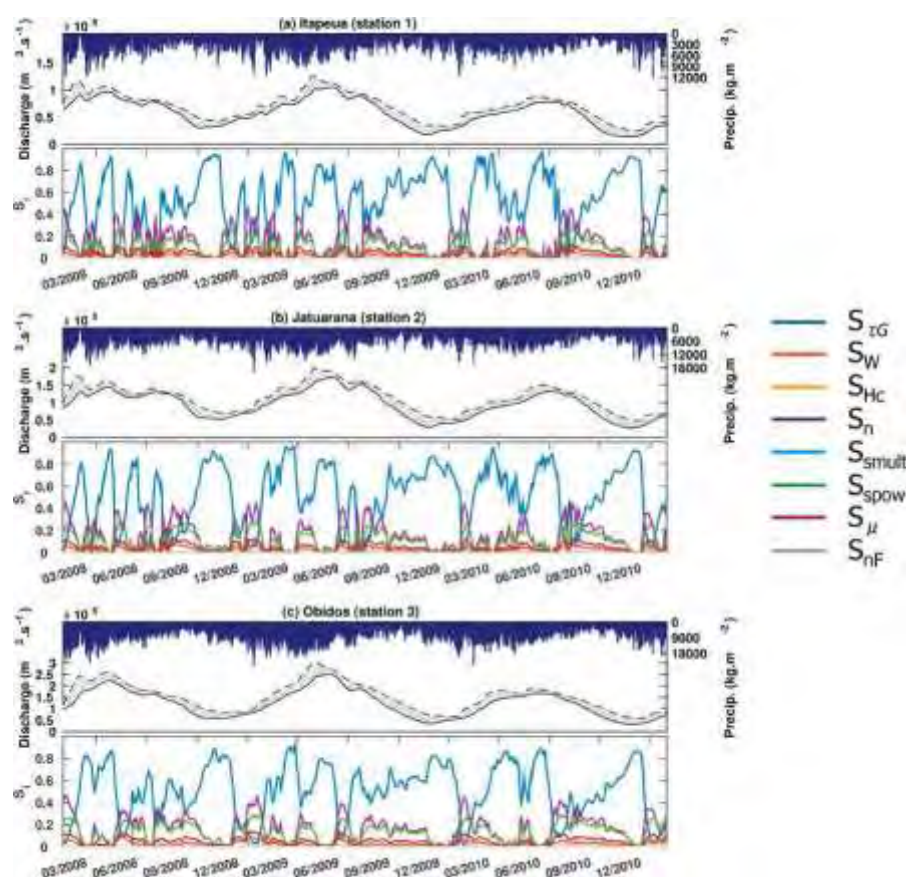


FIG. 7. As in Fig. 4, but for the AMA8 discharge SA.

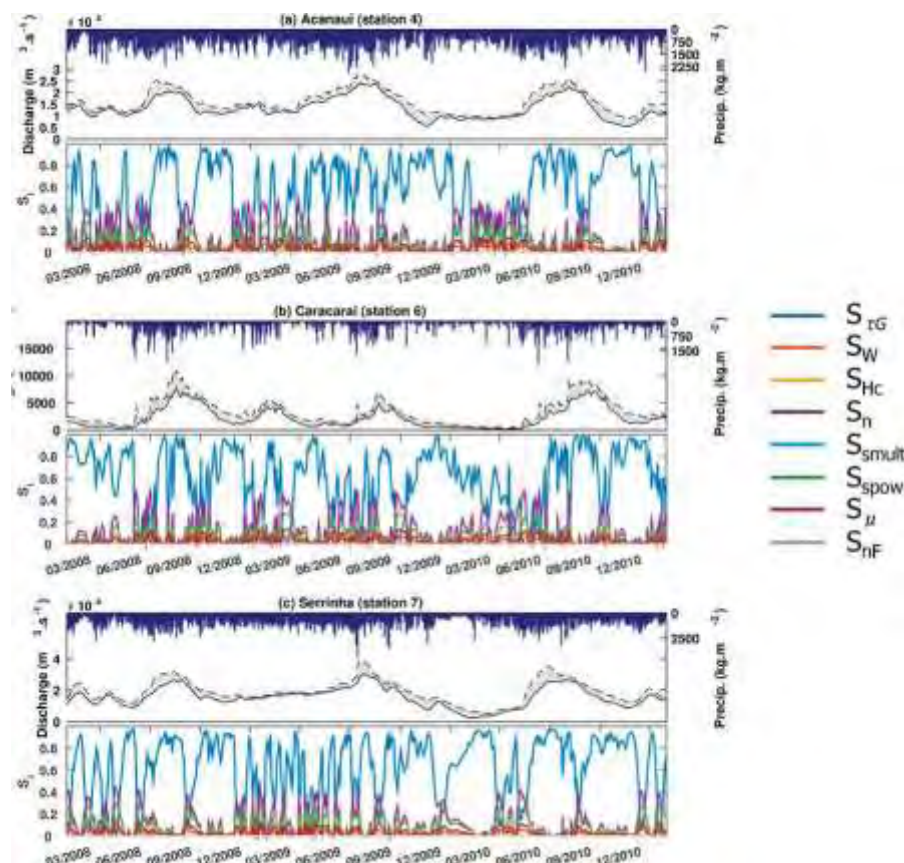


FIG. 8. As in Fig. 5, but for the AMA8 discharge SA.

in the Manning formula activate, respecting the ordering observed in water level sensitivities, that is,

$$n - s_{\text{pow}} - W. \quad (30)$$

In addition, the river section length, represented by m , is also quite an important parameter for the discharge because 1) it is directly used to compile discharge [Eq. (10)] and 2) it is an important river geometry parameter during regime changes.

c. Summary

From this first study, where the parameter sensitivities are studied identically over the entire Amazon basin, five parameters appear to be relevant in explaining TRIP sensitivities: the groundwater time constant, the river Manning coefficient, the riverbed slope powering coefficient, the river width, and the river section length meandering ratio. These parameters translate the predominant impact of both the groundwater reservoir and the use of the Manning formula for flow velocity estimation. Also, it appears that water level is essentially sensitive to parameters describing the geomorphology of the reservoir, while discharge is more sensitive to mass transfer. As discharge seems to be weakly sensitive to TRIP

parameters—deduced from the low dispersion of the simulated ensemble—it is rational to assume that discharge is sensitive to other parts of the ISBA–TRIP system such as precipitation forcing, ISBA parameters, and ISBA outputs.

It is of interest to conserve the five dominating parameters and to study the impact of regionalized parameters over TRIP water height and discharge. The next SA keeps the same configuration, but a different set of five parameters will be taken for each subbasin zone introduced in section 3d (Fig. 3). Giving a total number of 45 parameters (five parameters for nine zones), this simulation will therefore be denoted as AMA45.

5. Results and discussion for AMA45

Concerning the AMA45 experiment, results are shown in Figs. 10–13. The same color code as AMA8 is used to plot parameter sensitivity time series (e.g., purple line for n and orange line for W) and to differentiate two parameters of the same nature but from two different regions; different panels are used for each region.

Results are then discussed for a fewer number of cells than AMA8: Caracará, Belo Monte, Serrinha, Itaituba,

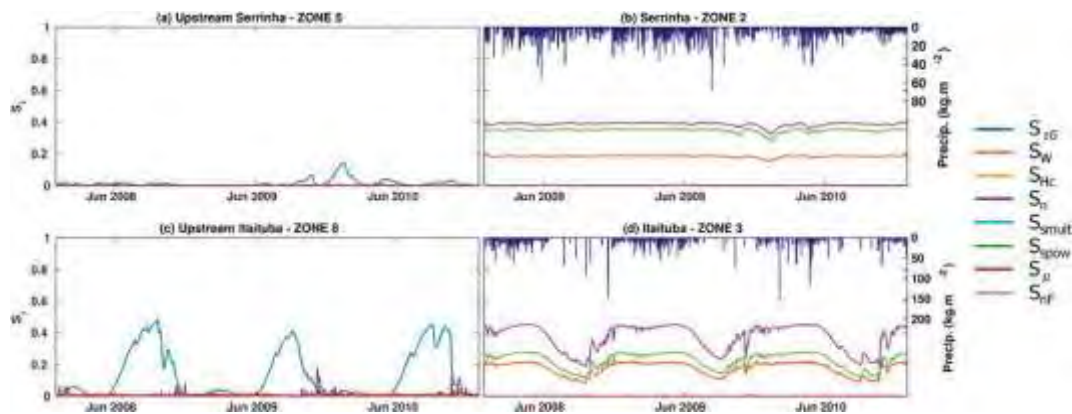


FIG. 10. Results for the AMA45 water level SA at (a),(b) Serrinha and (c),(d) Itaituba. First-order daily S_i for each activated zone— $S_{\tau G}$ (blue line), S_W (orange line), S_n (purple line), S_{spow} (green line), and S_m (dark red line)—with (reverse y axis) local precipitation in the result cell corresponding zone (bar plot). All curves in (a) and (b) [(c) and (d)] correspond to S_i at Serrinha (Itaituba): (a) [(c)] corresponds to the contribution of parameters from the upstream zone of Serrinha (Itaituba) while (b) [(d)] corresponds to the contribution of parameters from the local zone of Serrinha (Itaituba).

to improve water height and discharge diagnostics in ISBA–TRIP. Currently, the river width is defined using an empirical relationship, but it could be directly estimated from existing recent databases such as North American River Width Data Set (NARWidth; Allen and Pavelsky 2015) or Global Width Database for Large Rivers (GWD-LR; Yamazaki et al. 2014). Similarly, other DEMs could replace the currently used DEM in ISBA–TRIP (GTOPO30) to get a better estimate of the riverbed slope, such as Shuttle Radar Topography Mission (SRTM) edited in Hydrological Data and Maps Based on Shuttle Elevation Derivatives at Multiple Scales (HydroSHEDS; Lehner and Grill 2013) for

latitudes under 60°. Additionally, the upcoming Surface Water and Ocean Topography (SWOT; Alsdorf et al. 2007; Fjortoft et al. 2014; Biancamaria et al. 2015) mission will also provide such kinds of information on the river width and the river surface slope, which could be used to set TRIP parameters. Optical images could also be used to spatialize the meandering ratio. A better definition of those parameters would limit the model parameter uncertainty. The most uncertain parameter would therefore be the Manning coefficient and the groundwater time constant, which still remain difficult to estimate.

These SAs are preliminary works in preparation for model reanalysis and dynamic parameter estimation

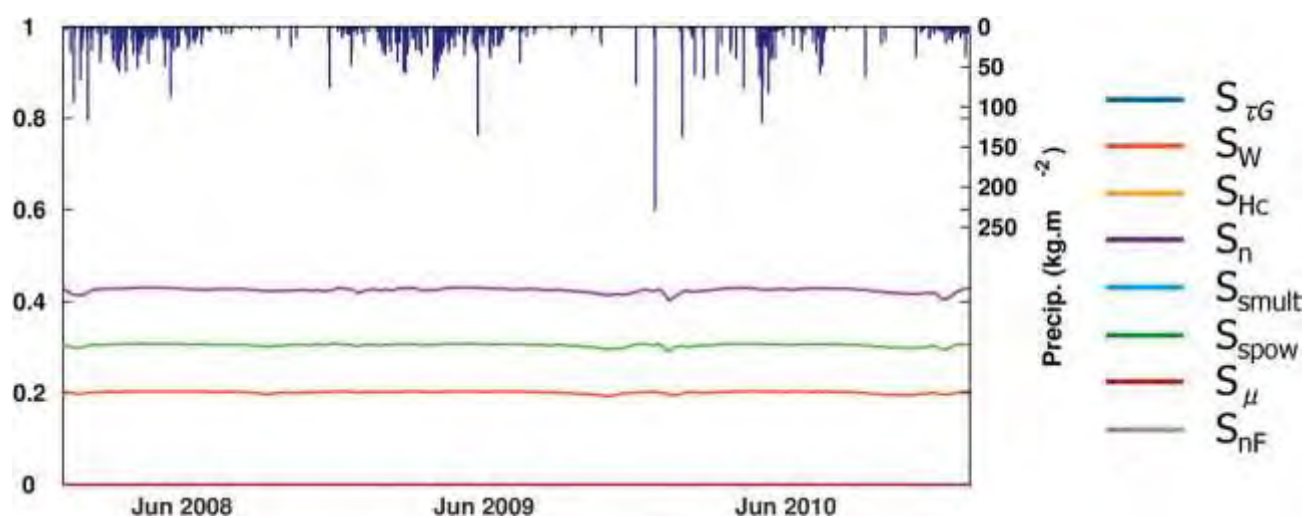


FIG. 11. Results for the AMA45 water level SA at Óbidos. First-order daily sensitivity indices in zone 3: $S_{\tau G}$ (blue line), S_W (orange line), S_n (purple line), S_{spow} (green line), and S_m (dark red line), with (reverse y axis) local precipitation in the result cell corresponding zone (bar plot). The sensitivity to parameters from other zones are all negligible, so the subplots for other zones are not displayed.

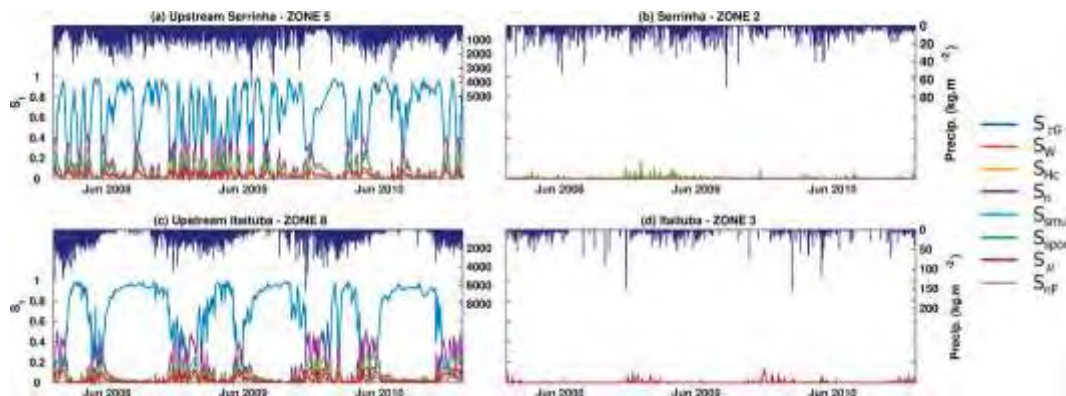
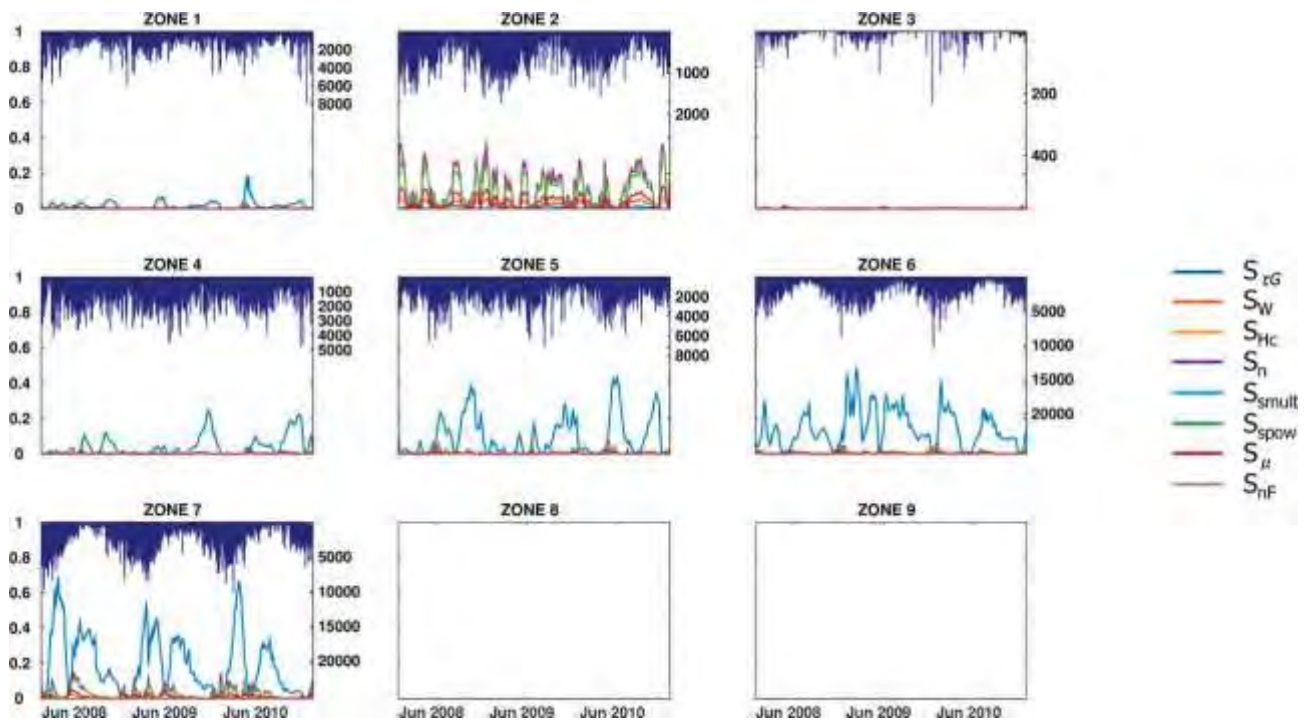


FIG. 12. As in Fig. 10, but for AMA45 discharge SA.

through data assimilation (Moradkhani 2008; Reichle 2008; Pedinotti et al. 2014) that use in situ or remotely sensed water level and discharge data to correct model parameters and/or state. Among them, the incoming SWOT mission offers a valuable potential to improve global-scale hydrological modeling. The idea is to use SWOT satellite products in a parameter estimation configuration to correct the main TRIP parameters highlighted during the SAs.

Additionally, this study focused on the water height exclusively, which is a physical variable of interest, for example, for water depth data assimilation. However, current satellites provide water elevation of top water

body distance to a reference geoid or ellipsoid. Therefore, we quickly investigate the contribution of model parameters with respect to water depth anomaly (i.e., obtained by subtracting the averaged water height over the study time period to the water height). Sensitivities were calculated as previously with the same perturbation ranges. Preliminary results for anomalies are presented at Óbidos in Fig. 14 with the associated S_i for the time-averaged water height. While the overall behavior is quite the same for the Manning coefficient, the riverbed slope, and even the river width (anticorrelated behavior with n and s), t_G clearly activates for short periods when the anomaly ensemble is near to zero. It

FIG. 13. As in Fig. 11, but for AMA45 discharge SA. First-order daily S_i for each zone with (reverse y axis) precipitation from drainage area in per zone (bar plot).

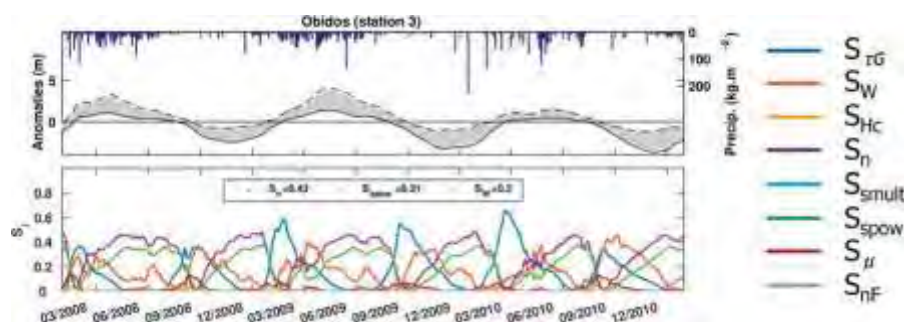


FIG. 14. As in Fig. 4, but for the AMA8 anomaly SA at Óbidos.

appears that the time-averaged water height (used to calculate the anomalies) takes all variance from the parameters n , s , and W , letting t_G dominate when water heights are close to the averaged water height (i.e., when anomalies are close to zero). Those results will be further developed and studied in future works.

The present study focused also only on river-routing model parameters and was applied over a unique river basin—the Amazon—but it offers numerous perspectives. In a global perspective, as SWOT will observe all rivers wider than 100 m between 788S and 788N, the present platform will be extended to other river basins situated in other climatic zones, such as the Mississippi or the Niger. In addition, it will be of great interest to study the impact of the LSM (here the ISBA model), the atmospheric forcing (more precisely precipitation), and even the initial reservoir states on TRIP outputs to improve our understanding of the continental part of the water cycle.

One important final remark about this SA is that the following results are quite dependent on the chosen parameters along with their perturbation range, but also on the chosen model itself. Indeed, the TRIP model considers only the kinematic wave propagation equation for the river reservoir. Other studies include diffusive wave propagation equation (Yamazaki et al. 2011; Winsemius et al. 2013) in their routing models along with a finer description of the topography and the flood dynamics. Even 2D-type finescale hydrodynamic model has been applied at continental scale (Sampson et al. 2015). The ANOVA mathematical formalism can be easily exported to these models and will probably give different results because of the different model physics (e.g., kinematic wave against diffusive wave) and may lead to potentially different sensitivity patterns in space and time.

7. Conclusions

This study aims to analyze the ISBA–TRIP large-scale hydrological model sensitivity over the Amazon River basin. An output model variance decomposition

method was used to identify key river-routing model parameters during a 3-yr period (2008–10).

Two analyses were carried out to evaluate the sensitivity of model parameters at different spatial scales. The first study (AMA8) considered parameters whose uncertainty was defined at the entire catchment scale. The second study (AMA45) used the same parameters but with regionalized uncertainty according to a geological and hydrological division of the Amazon River basin. The objective was to separate the local and upstream impacts of the parameters on both water height and discharge.

For AMA8, the results showed no interannual sensitivity. For both water height and discharge, the importance of river Manning coefficient and riverbed slope was highlighted. These observations are consistent with the use of the Manning formula to estimate the flow velocity. Also, river width and length had a nonnegligible influence. Contrary to water height, the groundwater time constant significantly dominated discharge sensitivity.

The second study used the same parameters but with regionalized uncertainty according to a geological and hydrological division of the Amazon River basin. The aim was to separate contribution from local parameters and upstream parameters.

Water level sensitivity is relatively constant through time, with the Manning coefficient multiplicative constant and the riverbed slope powering coefficient explaining 40% and 35%, respectively, of the unconditional variance. The river width is less present by taking only 20% of the variance. However, when the water level is very low, a nonnegligible impact of the groundwater time constant is observed. Also, in response to sudden regime change and precipitation events, the river width and length may be important. Overall, water level is essentially sensitive to geomorphological parameters. The second regionalized SA indicates that the local geomorphology of the reservoir drives the water level most of the year. In a low-flow period, the inflow from upstream zones is materialized by the activated upstream t_G and turns into an important contributor to water height as local phenomena (e.g., triggered by precipitation events) are minimized.

Discharge sensitivity presents several temporal patterns: seasonal patterns linked to the alternation of high- and low-flow seasons and short-term patterns associated with precipitation events in upstream regions. The groundwater time constant dominates, sometimes at over 90% of the variance. In a permanent regime, this parameter represents the main water inflow into the river. Geomorphological parameters activate during transitional regimes. It is worth noting that, because of the narrow dispersion of simulated discharge ensemble, discharge is actually weakly sensitive to TRIP parameters. A reasonable assumption would be to consider precipitation and even the ISBA configuration (e.g., soil conductivity distribution and vegetation cover) as the main drivers of discharge. The second regionalized study confirms that discharge is mainly driven by mass transfers, in particular upstream mass transfers.

Ultimately, this SA method offers an extensive variety of experiments. The SA results depend on both the studies model and the parameters range. Therefore, the same SA formalism applied with a different set of parameters or another model may give a different behavior of the sensitivity indices. Keeping the ISBA–TRIP model, one could study the impact of the precipitation forcing or the ISBA parameters on TRIP outputs or preserve the same formalism on other river basin. The method could also be applied to other models with a different physics and different parameters.

Acknowledgments. This work was performed using high performance computing resources from Calcul en Midi-Pyrénées (CALMIP; Grant 2015-P1408). The authors are very thankful to M. Ratto and I. M. Sobol for supplying the open-source SS-ANOVA and LP_t routines, respectively. The GSWP3 team is acknowledged for letting the authors use their different forcing fields. The authors acknowledge funding from the CNES Terre–Océan–Surfaces Continentales–Atmosphère (TOSCA) committee. C. M. Emery is supported by a CNES/Conseil Régional de Midi-Pyrénées Ph.D. grant.

REFERENCES

- Alkama, R., M. Kageyama, G. Ramstein, O. Marti, P. Ribstein, and D. Swingedouw, 2008: Impact of a realistic river routing in coupled ocean–atmosphere simulations of the Last Glacial Maximum climate. *Climate Dyn.*, 30, 855–869, doi:10.1007/s00382-007-0330-1.
- , and Coauthors, 2010: Global evaluation of the ISBA–TRIP continental hydrological system. Part I: Comparison to GRACE terrestrial water storage estimates and in situ river discharges. *J. Hydrometeorol.*, 11, 583–600, doi:10.1175/2010JHM1211.1.
- Allen, G., and T. Pavelsky, 2015: Patterns of river width and surface area revealed by the satellite-derived North American River Width data set. *Geophys. Res. Lett.*, 42, 395–402, doi:10.1002/2014GL062764.
- Alsdorf, D., E. Rodriguez, and D. Lettenmaier, 2007: Measuring surface water from space. *Rev. Geophys.*, 45, RG2002, doi:10.1029/2006RG000197.
- Apel, H., A. Thieken, B. Merz, and G. Blöschl, 2004: Flood risk assessment and associated uncertainty. *Nat. Hazards Earth Syst. Sci.*, 4, 295–308, doi:10.5194/nhess-4-295-2004.
- Aronica, G., B. Hankin, and K. Beven, 1998: Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data. *Adv. Water Resour.*, 22, 349–365, doi:10.1016/S0309-1708(98)00017-7.
- Arora, V., F. Chiew, and R. Grayson, 1999: A river flow routing scheme for general circulation models. *J. Geophys. Res.*, 104, 14 347–14 357, doi:10.1029/1999JD900200.
- Balsamo, G., P. Viterbo, A. Beljaars, B. den Hurk, M. Hirchi, A. Betts, and K. Scipal, 2009: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system. *J. Hydrometeorol.*, 10, 623–643, doi:10.1175/2008JHM1068.1.
- Bates, P., M. Horritt, G. Aronica, and K. Beven, 2004: Bayesian updating of flood inundation likelihoods conditioned on flood extent data. *Hydrol. Processes*, 18, 3347–3370, doi:10.1002/hyp.1499.
- Beven, K., and A. Binley, 1992: The future of distributed models: Model calibration and uncertainty prediction. *Hydrol. Processes*, 6, 279–298, doi:10.1002/hyp.3360060305.
- Biancamaria, S., D. P. Lettenmaier, and T. M. Pavelsky, 2015: The SWOT mission and capabilities for land hydrology. *Surv. Geophys.*, 37, 307–337, doi:10.1007/s10712-015-9346-y.
- Blackadar, A.-F., 1976: Modeling the nocturnal boundary layer. Preprints, Third Symp. on Atmospheric Turbulence, Diffusion, and Air Quality, Raleigh, NC, Amer. Meteor. Soc., 46–49.
- Blazkova, S., and K. Beven, 2004: Flood frequency estimation by continuous simulation of subcatchment rainfalls and discharges with the aim of improving dam safety assessment in a large basin in the Czech Republic. *J. Hydrol.*, 292, 153–172, doi:10.1016/j.jhydrol.2003.12.025.
- Boone, A., J.-C. Calvet, and J. Noilhan, 1999: Inclusion of a third soil layer in a land surface scheme using the force–restore method. *J. Hydrometeorol.*, 38, 1611–1630, doi:10.1175/1520-0450(1999)038<1611:IOATSL>2.0.CO;2.
- Coe, M., 1998: A linked global model of terrestrial hydrologic processes: Simulation of modern rivers, lakes, and wetlands. *J. Geophys. Res.*, 103, 8885–8899, doi:10.1029/98JD00347.
- Compo, G., and Coauthors, 2011: The Twentieth Century Re-analysis project. *Quart. J. Roy. Meteor. Soc.*, 137, 1–28, doi:10.1002/qj.776.
- Cukier, R., C. Fortuin, and K. Shuler, 1973: Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I. Theory. *J. Chem. Phys.*, 59, 3873–3878, doi:10.1063/1.1680571.
- Decharme, B., and H. Douville, 2006: Introduction of a subgrid hydrology in the ISBA land surface model. *Climate Dyn.*, 26, 65–78, doi:10.1007/s00382-005-0059-7.
- , —, A. Boone, F. Habets, and J. Noilhan, 2006: Impact of an exponential profile of saturated hydraulic conductivity within the ISBA LSM: Simulations over the Rhône basin. *J. Hydrometeorol.*, 7, 61–80, doi:10.1175/JHM469.1.
- , —, C. Prigent, F. Papa, and F. Aires, 2008: A new river flooding scheme for global climate applications: Off-line evaluation over South America. *J. Geophys. Res.*, 113, D11110, doi:10.1029/2007JD009376.

- , R. Alkama, H. Douville, M. Becker, and A. Cazenave, 2010: Global evaluation of the ISBA–TRIP continental hydrological system. Part II: Uncertainties in river routing simulation related to flow velocity and groundwater storage. *J. Hydrometeor.*, 11, 601–617, doi:10.1175/2010JHM1212.1.
- , —, F. Papa, S. Faroux, H. Douville, and C. Prigent, 2012: Global off-line evaluation of the ISBA–TRIP flood model. *Climate Dyn.*, 38, 1389–1412, doi:10.1007/s00382-011-1054-9.
- Doksum, K., and A. Samarov, 1995: Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. *Ann. Stat.*, 23, 1443–1473, doi:10.1214/aos/1176324307.
- Ducharme, A., C. Golaz, E. Leblois, K. Laval, J. Polcher, E. Ledoux, and G. de Marsily, 2003: Development of a high resolution runoff routing model, calibration and application to assess runoff from the LMD GCM. *J. Hydrol.*, 280, 207–228, doi:10.1016/S0022-1694(03)00230-0.
- Efron, B., and C. Stein, 1981: The jackknife estimate of variance. *Ann. Stat.*, 9, 586–596, doi:10.1214/aos/1176345462.
- Fang, S., G. Gertner, S. Shinkareva, G. Wang, and A. Anderson, 2003: Improved generalized Fourier amplitude sensitivity test (FAST) for model assessment. *Stat. Comput.*, 13, 221–226, doi:10.1023/A:1024266632666.
- Fjortoft, R., and Coauthors, 2014: KaRIn on SWOT: Characteristics of near-nadir Ka-band interferometric SAR imagery. *IEEE Trans. Geosci. Remote Sens.*, 52, 2172–2185, doi:10.1109/TGRS.2013.2258402.
- Francos, A., F. Elorza, F. Bouraoui, G. Bidoglio, and L. Galbiati, 2003: Sensitivity analysis of distributed environmental simulation models: Understanding the model behaviour in hydrological studies at the catchment scale. *Reliab. Eng. Syst. Saf.*, 79, 205–218, doi:10.1016/S0951-8320(02)00231-4.
- Freer, J., K. Beven, and B. Ambroise, 1996: Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the glue approach. *Water Resour. Res.*, 32, 2161–2173, doi:10.1029/95WR03723.
- Gatelli, D., S. Kucherenko, M. Ratto, and S. Tarantola, 2009: Calculating first-order sensitivity measures: A benchmark of some recent methodologies. *Reliab. Eng. Syst. Saf.*, 94, 1212–1219, doi:10.1016/j.res.2008.03.028.
- Garambois, P., H. Roux, K. Larnier, W. Castaigns, and D. Dartus, 2013: Characterization of process-oriented hydrologic model behavior with temporal sensitivity analysis for flash floods in Mediterranean catchments. *Hydrol. Earth Syst. Sci.*, 17, 2305–2322, doi:10.5194/hess-17-2305-2013.
- Gleason, C., and L. Smith, 2014: Toward global mapping of river discharge using satellite images and at-many-stations hydraulic geometry. *Proc. Natl. Acad. Sci. USA*, 111, 4788–4791, doi:10.1073/pnas.1317606111.
- Grayson, R., and G. Blöschl, 2001: *Spatial Patterns in Catchment Hydrology*. Cambridge University Press, 416 pp.
- Guse, B., D. Reusser, and N. Fohrer, 2014: How to improve the representation of hydrological processes in SWAT for a low-land catchment—Temporal analysis of parameter sensitivity and model performance. *Hydrol. Processes*, 28, 2651–2670, doi:10.1002/hyp.9777.
- Hagemann, S., and L. Dümenil, 1997: A parameterization of the lateral waterflow for the global scale. *Climate Dyn.*, 14, 17–31, doi:10.1007/s003820050205.
- Hall, J., S. Tarantola, P. Bates, and M. Horritt, 2005: Distributed sensitivity analysis of flood inundation model calibration. *J. Hydraul. Eng.*, 131, 117–126, doi:10.1061/(ASCE)0733-9429(2005)131:2(117).
- Hornberger, G., and R. Spear, 1981: An approach to the preliminary analysis of environmental systems. *J. Environ. Manage.*, 12, 7–18.
- Jansen, M., 1999: Analysis of variance designs for model output. *Comput. Phys. Commun.*, 117, 35–43, doi:10.1016/S0010-4655(98)00154-4.
- Kavetski, D., G. Kuczera, and S. Franks, 2006: Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resour. Res.*, 42, W03407, doi:10.1029/2005WR004368.
- Lehner, B., and G. Grill, 2013: Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrol. Processes*, 27, 2171–2186, doi:10.1002/hyp.9740.
- Leopold, L., and T. Maddock Jr., 1953: The hydraulic geometry of stream channels and some physiographic implications. USGS Professional Paper 252, 57 pp. [Available online at <http://pubs.usgs.gov/pp/0252/report.pdf>.]
- Lucas-Picher, P., V. Arora, D. Caya, and R. Laprise, 2003: Implementation of a large-scale variable velocity river flow routing algorithm in the Canadian Regional Climate Model (CRCM). *Atmos.–Ocean*, 41, 139–153, doi:10.3137/ao.410203.
- Maidment, D. R., 1993: *Handbook of Hydrology*. McGraw-Hill, 1424 pp.
- Manning, R., 1891: On the flow of water in open channels and pipes. *Inst. Civ. Eng. Irel.*, 20, 161–207.
- Masson, V., and Coauthors, 2013: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. *Geosci. Model Dev.*, 6, 929–960, doi:10.5194/gmd-6-929-2013.
- Meade, R., J. Rayol, S. D. Conceição, and J. Natividade, 1991: Backwater effects in the Amazon River basin of Brazil. *Environ. Geol. Water Sci.*, 18, 105–114, doi:10.1007/BF01704664.
- Molinier, M., J.-L. Guyot, V. Guimarães, and E. de Oliveira, 1993: *Hydrologie du bassin de l'amazone. Colloque Grands Bassins Fluviaux Périalantiques*, Paris, France, PEGI-INSA-CNRS-ORSTOM, 335–345.
- Moody, J., and B. Troutman, 2002: Characterization of the spatial variability of channel morphology. *Earth Surf. Processes Landforms*, 27, 1251–1266, doi:10.1002/esp.403.
- Moradkhani, H., 2008: Hydrological remote sensing and land surface data assimilation. *Sensors*, 8, 2986–3004, doi:10.3390/s8052986.
- Ngo-Duc, T., T. Oki, and S. Kanae, 2007: A variable streamflow velocity method for global river routing model: Model description and preliminary results. *Hydrol. Earth Syst. Sci.*, 4, 4389–4414, doi:10.5194/hessd-4-4389-2007.
- Noilhan, J., and S. Planton, 1989: A simple parameterization of land surface processes for meteorological models. *Mon. Wea. Rev.*, 117, 536–549, doi:10.1175/1520-0493(1989)117<0536:ASPOLS.2.0.CO;2.
- Oki, T., and Y. Sud, 1998: Design of Total Runoff Integrating Pathways (TRIP)—A global river channel network. *Earth Interact.*, 2, doi:10.1175/1087-3562(1998)002<0001:DOTRIP.2.3.CO;2.
- Olivera, F., J. Famiglietti, and K. Asante, 2000: Global-scale flow routing using a source-to-sink algorithm. *Water Resour. Res.*, 36, 2197–2207, doi:10.1029/2000WR900113.
- Pappenberger, F., and K. Beven, 2004: Functional classification and evaluation of hydrographs based on multi-component mapping. *Int. J. River Basin Manage.*, 2, 89–100, doi:10.1080/15715124.2004.9635224.
- , —, A. D. Roo, J. Thielen, and B. Gouweleeuw, 2004: Uncertainty analysis of the rainfall runoff model LisFlood

- within the Generalized Likelihood Uncertainty Estimation (GLUE). *Int. J. River Basin Manage.*, 2, 123–133, doi:[10.1080/15715124.2004.9635227](https://doi.org/10.1080/15715124.2004.9635227).
- , H. Cloke, G. Balsamo, T. Ngo-Duc, and T. Oki, 2010: Global runoff routing with the hydrological component of the ECMWF NWP system. *Int. J. Climatol.*, 30, 2155–2174, doi:[10.1002/joc.2028](https://doi.org/10.1002/joc.2028).
- Paris, A., and Coauthors, 2016: Stage–discharge rating curves based on satellite altimetry and modeled discharge in the amazon basin. *Water Resour. Res.*, 52, 3787–3814, doi:[10.1002/2014WR016618](https://doi.org/10.1002/2014WR016618).
- Pedinotti, V., A. Boone, S. Ricci, S. Biancamaria, and N. Mognard, 2014: Assimilation of satellite data to optimize large-scale hydrological model parameters: A case study for the SWOT mission. *Hydrol. Earth Syst. Sci.*, 18, 4485–4507, doi:[10.5194/hess-18-4485-2014](https://doi.org/10.5194/hess-18-4485-2014).
- Rabitz, H., O. Aliş, J. Shorter, and K. Shim, 1999: Efficient input–output model representations. *Comput. Phys. Commun.*, 117, 11–20, doi:[10.1016/S0010-4655\(98\)00152-0](https://doi.org/10.1016/S0010-4655(98)00152-0).
- Ratto, M., A. Pagano, and P. Young, 2007: State dependent parameter metamodelling and sensitivity analysis. *Comput. Phys. Commun.*, 177, 863–876, doi:[10.1016/j.cpc.2007.07.011](https://doi.org/10.1016/j.cpc.2007.07.011).
- Reichle, R. H., 2008: Data assimilation methods in earth sciences. *Adv. Water Resour.*, 31, 1411–1418, doi:[10.1016/j.advwatres.2008.01.001](https://doi.org/10.1016/j.advwatres.2008.01.001).
- Reusser, D., and E. Zehe, 2011: Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity. *Water Resour. Res.*, 47, W07550, doi:[10.1029/2010WR009946](https://doi.org/10.1029/2010WR009946).
- , W. Buytaert, and E. Zehe, 2011: Temporal dynamics of model parameter sensitivity for computationally expensive models with the Fourier amplitude sensitivity test. *Water Resour. Res.*, 47, W07551, doi:[10.1029/2010WR009947](https://doi.org/10.1029/2010WR009947).
- Roux, H., D. Labat, P.-A. Garambois, M.-M. Maubourguet, J. Chorda, and D. Dartus, 2011: A physically-based parsimonious hydrological model for flash floods in Mediterranean catchments. *Nat. Hazards Earth Syst. Sci.*, 11, 2567–2582, doi:[10.5194/nhess-11-2567-2011](https://doi.org/10.5194/nhess-11-2567-2011).
- Saltelli, A., and R. Bolado, 1998: An alternative way to compute Fourier amplitude sensitivity test (FAST). *Comput. Stat. Data Anal.*, 26, 445–460, doi:[10.1016/S0167-9473\(97\)00043-1](https://doi.org/10.1016/S0167-9473(97)00043-1).
- , K. Chan, and E. M. Scott, 2008: Sensitivity Analysis. Wiley, 494 pp.
- , P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola, 2010: Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Comput. Phys. Commun.*, 181, 259–270, doi:[10.1016/j.cpc.2009.09.018](https://doi.org/10.1016/j.cpc.2009.09.018).
- Sampson, C., A. Smith, P. Bates, J. Neal, L. Alfieri, and J. Freer, 2015: A high-resolution global flood hazard model. *Water Resour. Res.*, 51, 7358–7381, doi:[10.1002/2015WR016954](https://doi.org/10.1002/2015WR016954).
- Sieber, A., and S. Uhlenbrook, 2005: Sensitivity analyses of a distributed catchment model to verify the model structure. *J. Hydrol.*, 310, 216–235, doi:[10.1016/j.jhydrol.2005.01.004](https://doi.org/10.1016/j.jhydrol.2005.01.004).
- Sioli, H., 1984: The Amazon and its main affluents: Hydrography, morphology of the river courses, and river types. The Amazon, H. Sioli, Ed., Springer, 127–165, doi:[10.1007/978-94-009-6542-3_5](https://doi.org/10.1007/978-94-009-6542-3_5).
- Sobol, I., 1967: On the distribution of points in a cube and the approximate evaluation of integrals. *URSS Comput. Math. Math. Phys.*, 7 (4), 86–112, doi:[10.1016/0041-5553\(67\)90144-9](https://doi.org/10.1016/0041-5553(67)90144-9).
- , 1993: Sensitivity analysis for non-linear mathematical models. *Math. Model. Comput. Exp.*, 2, 407–414.
- , 2001: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.*, 55, 271–280, doi:[10.1016/S0378-4754\(00\)00270-6](https://doi.org/10.1016/S0378-4754(00)00270-6).
- Stocker, T., and Coauthors, 2013: Technical summary. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 33–115, doi:[10.1017/CBO9781107415324.005](https://doi.org/10.1017/CBO9781107415324.005).
- Trenberth, K., 2011: Changes in precipitation with climate change. *Climate Res.*, 47, 123–138, doi:[10.3354/cr00953](https://doi.org/10.3354/cr00953).
- van Griensven, A., T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, and R. Srinivasan, 2006: A global sensitivity analysis tool for the parameters of multi-variable catchment models. *J. Hydrol.*, 324, 10–23, doi:[10.1016/j.jhydrol.2005.09.008](https://doi.org/10.1016/j.jhydrol.2005.09.008).
- Vörösmarty, C., B. Moore, A. Grace, M. Gildea, J. Melillo, B. Peterson, E. Rastetter, and P. Steudler, 1989: Continental scale models of water balance and fluvial transport: An application to South America. *Global Biogeochem. Cycles*, 3, 241–265, doi:[10.1029/GB003i003p00241](https://doi.org/10.1029/GB003i003p00241).
- Wagener, T., N. McIntyre, M. Lees, H. Wheeler, and H. Gupta, 2003: Towards reduced uncertainty in conceptual rainfall–runoff modelling: Dynamic identifiability analysis. *Hydrol. Processes*, 17, 455–476, doi:[10.1002/hyp.1135](https://doi.org/10.1002/hyp.1135).
- Werner, M., N. Hunter, and P. Bates, 2005: Identifiability of distributed floodplain roughness values in flood extent estimation. *J. Hydrol.*, 314, 139–157, doi:[10.1016/j.jhydrol.2005.03.012](https://doi.org/10.1016/j.jhydrol.2005.03.012).
- Winsemius, H., L. V. Beek, B. Jongman, P. Ward, and A. Bouwman, 2013: A framework for global river flood risk assessments. *Hydrol. Earth Syst. Sci.*, 17, 1871–1892, doi:[10.5194/hess-17-1871-2013](https://doi.org/10.5194/hess-17-1871-2013).
- Wisser, D., B. Feketa, C. J. Vörösmarty, and A. H. Schumann, 2010: Reconstructing 20th century global hydrography: A contribution to the Global Terrestrial Network–Hydrology (GTN-H). *Hydrol. Earth Syst. Sci.*, 14, 1–24, doi:[10.5194/hess-14-1-2010](https://doi.org/10.5194/hess-14-1-2010).
- Yamazaki, D., S. Kanae, H. Kim, and T. Oki, 2011: A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resour. Res.*, 47, W04501, doi:[10.1029/2010WR009726](https://doi.org/10.1029/2010WR009726).
- , F. O’Loughlin, M. Trigg, Z. Miller, T. Pavelsky, and P. D. Bates, 2014: Development of the global width database for large rivers. *Water Resour. Res.*, 50, 3467–3480, doi:[10.1002/2013WR014664](https://doi.org/10.1002/2013WR014664).

Bibliographie

- Albergel, C., de Rosnay, P., Gruhier, C., Muños-Sabater, J., Hasenauer, S., Isaksen, L., Kerr, Y., and Wagner, W. (2012). Evaluation of remotely-sensed and modelled soil moisture products using global ground-based in situ observations. *Remote Sensing Environment*, 118 :215–226.
- Alkama, R., Decharme, B., Douville, H., Becker, M., Cazenave, A., Sheffield, J., Voldoire, A., Tyteca, S., and Moigne, P. L. (2010). Global evaluation of the isba-trip continental hydrological system. part i : Comparison to grace terrestrial water storage estimates and in situ river discharges. *Journal of Hydrometeorology*, 11 :583–600.
- Allen, G. and Pavelsky, T. (2015). Patterns of river width and surface area revealed by the satellite-derived north american river width data set. *geophysical Research Letters*, 42 :395–402.
- Alsdorf, D., Rodríguez, E., Lettenmaier, D., and Famiglietti, J. (2005). Water : The water elevation recovery satellite mission. Technical report, Ohio State University, NASA.
- Alsdorf, D. E. and Lettenmaier, D. P. (2003). Tracking fresh water from space. *Science*, 301 :1485–1488.
- Alsdorf, D. E., Rodríguez, E., and Lettenmaier, D. P. (2007). Measuring surface water from space. *Reviews of Geophysics*, 45(RG2002).
- Alzraiee, A. H., Baú, D., and Elhaddad, A. (2014). Estimation of heterogeneous aquifer parameters using centralized and decentralized fusion of hydraulic tomography data. *Hydrology and Earth System Sciences*, 18 :3207–3223.
- Anderson, J. L. (2001). An ensemble adjustment kalman filter for data assimilation. *Monthly Weather Review*, 129 :2884–2903.
- Anderson, J. L. (2007). Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D : Nonlinear Phenomena*, 230 :99–111.
- Andreadis, K. M., Clark, E. A., Lettenmaier, D. P., and Alsdorf, D. E. (2007). Propects for river discharge and depth estimation through data assimilation of swath-altimetry into a raster-based hydrodynamic model. *Geophysical Research Letters*, 34.
- Andreadis, K. M. and Lettenmaier, D. P. (2006). Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources*, 29 :872–886.
- Apel, H., Thielen, A., Merz, B., and Blochl, G. (2004). Flood risk assessment and associated uncertainty. *Natural Hazards Earth System Sciences*, 4 :295–308.

- Aronica, G., Hankin, B., and Beven, K. (1998). Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data. *Advances in Water Resources*, 22 :349–365.
- Arora, V., Chiew, F., and Grayson, R. (1999). A river flow routing scheme for general circulation models. *Journal of Geophysical Research*, 104 :14347–14357.
- Arora, V. K. and Boer, G. F. (1999). A variable velocity flow routing algorithm for gcms. *Journal of Geophysical Research - Atmospheres*, 104 :30965–30979.
- Balsamo, G., Viterbo, P., Belijaars, A., van den Hurk, B., Hirchi, M., Betts, A. K., and Scipal, K. (2009). A revised hydrology for the ecmwf model : Verification from field site to terrestrial water storage and impact in the integrated forecast system. *Journal of Hydrometeorology*, 10 :623–643.
- Bamber, J. L. (1994). Ice sheet altimeter processing scheme. *International Journal of Remote Sensing*, 15 :925–938.
- Barthélémy, S., Ricci, S., Pannekoucke, O., Thual, O., O., P., and Malaterre (2013). Emulation of an ensemble kalman filter algorithm on a flood wave propagation model. *Hydrology and Earth System Sciences*, 10 :6963–7001.
- Bates, P., Horritt, M., Aronica, G., and Beven, K. (2004). Bayesian updating of flood inundation likelihoods conditioned on flood extent data. *Hydrological Processes*, 18 :3347–3370.
- Bates, P. D. and Roo, A. D. (2000). A simple raster-based model for floodplain inundation. *Journal of Hydrology*, 236 :54–77.
- Bazile, E., Haiti, M. E., Bogatchev, A., and Spiridonov, V. (2002). Improvement of the snow parameterization in arpege/aladin. In *Proceedings of SRNWP/HIRLAM Workshop on Surface Processes, Turbulence and Mountain Effects*, pages 22–24.
- Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., and Ziese, M. (2013). A description of the global land-surface precipitation data products of the global precipitation climatology centre with sample applications including centennial (trend) analysis from 1901-present. *Earth System Science Data*, 5 :71–99.
- Bennet, A. F. (2002). *Inverse modeling of the ocean and atmosphere*. Cambridge University Press.
- Best, M., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., C. B. Ménard, J. M. E., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J. (2011). The joint uk land environment simulator (jules), model description - part 1 : Energy and water fluxes. *Geoscientific Model Development*, 4 :677–699.
- Beven, K. and Binley, A. (1992). The future of distributed models : Model calibration and uncertainty prediction. *Hydrological Processes*, 6 :279–298.
- Beven, K., Cloke, H., Pappenberger, F., Lamb, R., and Hunter, N. (2015). Hyperresolution information and hyperresolution ignorance in modelling the hydrology of the land surface. *Science China Earth Sciences*, 58 :25–35.
- Biancamaria, S., Durant, M., Andreadis, K. M., Bates, P. D., Boone, A., Mognard, N. M., Rodriguez, E., Alsdorf, D. E., Lettenmaier, D. P., and Clark, E. A. (2011). Assimilation of virtual wide swath altimetry to improve arctic river modeling. *Remote Sensing of Environment*, 115 :373–381.

- Biancamaria, S., Lettenmaier, D. P., and Pavelsky, T. M. (2016). The swot mission and its capabilities for land hydrology. *Surveys in Geophysics*, 37 :307–337.
- Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., van de Giesen, N., Winsemius, H., and Wood, E. F. (2015). Hyper-resolution global hydrological modelling : what is next ? *Hydrological Processes*, 29 :310–320.
- Birkett, C. M. (1995). The contribution of topex/poseidon to the global monitoring of climatically sensitive lakes. *Journal of Geophysical Research*, 100 :25179–25204.
- Birkett, C. M. (1998). Contribution of the topex nasa altimeter to the global monitoring of large rivers and wetlands. *Water Resources Research*, 34 :1223–1239.
- Birkett, C. M., Mertes, L. A. K., Dunne, T., Costa, M. H., and Jasinski, M. J. (2002). Surface water dynamics in the amazon basin : Application of satellite radar altimetry. *Journal of Geophysical Research*, 107.
- Birkinshaw, S. J., Moore, P., Kilsby, C. G., O'Donnell, G. M., Hardy, A. J., and Berry, P. A. M. (2014). Daily discharge estimation at ungauged river sites using remote sensing. *Hydrological Processes*, 28 :1043–1054.
- Bishop, C. H., Etherton, B. J., and Majumbar, S. J. (2001). Adaptive sampling with the ensemble transform kalman filter. part i : Theoretical aspects. *Monthly Weather Review*, 129 :420–436.
- Bjerklie, D. M. (2007). Estimating the bankfull velocity and discharge for rivers using remotely sensed river morphology information. *Journal of Hydrology*, 341 :144–55.
- Blackadar, A. K. (1976). Modeling the nocturnal boundary layer. In *Proceedings of 3rd Symposium on Atmospheric Turbulence, Diffusion and Air Quality*, pages 46–49. American Meteorological Society.
- Blazkova, S. and Beven, K. (2004). Flood frequency estimation by continuous simulation of sub-catchment rainfalls and discharges with the aim of improving dam safety assessment in a large basin in the czech republic. *Journal of Hydrology*, 292 :153–172.
- Blösch, G. and Sivapalan, M. (1995). Scale issues in hydrological modeling : a review. *Hydrological Processes*, 9 :251–290.
- Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M. (2014). The concept of essential climate variables in support of climate research, applications, and policy. *Bulletin of the American Meteorological Society*, 95 :1431–1443.
- Boone, A. (2000). *Modélisation des processus hydrologiques dans le schéma de surface ISBA : inclusion d'un réservoir hydrologique, du gel et modélisation de la neige*. PhD thesis, Université Paul Sabatier.
- Boone, A., Calvet, J.-C., and Noilhan, J. (1999). Inclusion of a third layer in a land surface scheme using force-restore method. *Journal of Applied Meteorology*, 28 :1611–1630.
- Boone, A., Calvet, J.-C., and Noilhan, J. (2000). The influence of the inclusion of soil freezing on simulations by a soil-vegetation-atmosphere transfer scheme. *Journal of Applied Meteorology*, 39 :1544–1569.

- Boone, A. and Etchevers, P. (2001). An inter-comparison of three snow schemes of varying complexity coupled to the same land-surface and macroscale hydrologic models. *Journal of Hydrometeorology*, 2 :374–394.
- Bouttier, F. and Courtier, P. (1999). Data assimilation concepts and methods. Technical report, ECMWF.
- Brakenridge, G. R., Nghiem, S. V., Anderson, E., and Chien, S. (2005). Space-based measurement of river runoff. *Eos Transactions, AGU*, 86(19) :185–188.
- Branstetter, M. L. and Famiglietti, J. S. (1999). Testing the sensitivity of gcm-simulated runoff to climate model resolution using a parallel river transport algorithm. In *14th Conference on Hydrology*, pages 391–392. American Meteorological Society.
- Brooks, R. H. and Corey, A. T. (1966). Properties of porous media affecting fluid flow. *Journal of Irrigation and Drainage Engineering*, 2 :61–88.
- Brun, E., Martin, E., Simon, V., Gendre, C., and Coleou, C. (1989). An energy and mass balance model of snow cover suitable for operational avalanche forecasting. *Journal of Glaciology*, 35 :333–342.
- Buis, S., Piacentini, A., and Declat, D. (2006). Palm : a computational framework for assembling high-performance computing applications. *Concurrency and Computation : Practice and Experience*, 18 :247–262.
- Bulygina, N. (2011). Review of land surface models. Slides - Imperial College, London. Disponible sur : www.imperial.ac.uk/media/imperial-college/research-centres-and-groups/environmental-and-water-resource-engineering/NBulygina-LSM-review-14Feb2011.pdf.
- Burgers, G., Leeuwen, P. J. V., and Evensen, G. (1998). Analysis scheme in the ensemble kalman filter. *Monthly Weather Review*, 126 :1719–1724.
- Calmant, S. and Seyler, F. (2006). Continental surface waters from satellite altimetry. *Comptes Rendus Geoscience*, 338 :1113–1122.
- Calvet, J.-C., Noilhan, J., Roujean, J. L., Bessemoulin, P., Cabelguenne, M., Olioso, A., and Wigneron, J. P. (1998). An interactive vegetation svat model tested against data from 6 contrasting sites. *Agricultural and Forest Meteorology*, 92 :73–95.
- Cassé, C. (2015). *Impact du forçage pluviométrique sur les inondations du fleuve Niger à Niamey : Etude à partir de données satellitaires et in-situ*. PhD thesis, Université Paul Sabatier, Toulouse.
- Chambon, P., Jobard, I., Roca, R., and Viltard, N. (2012). An investigation of the error budget of tropical rainfall accumulation derived from merged passive microwave and infrared satellite measurements. *Quarterly Journal of the Royal Meteorological Society*, 139 :879–893.
- Chapnik, B., Desroziers, G., and Rabier, F. (2006). Diagnosis and tuning of observational error in a quasi-operational data assimilation setting. *Quarterly Journal of the Royal Meteorological Society*, 132 :543–565.
- Chertok, D. L. and Lardner, R. W. (1996). Variational data assimilation for a nonlinear hydraulic model. *Applied Mathematical Modelling*, 20 :675–682.
- Clapp, R. and Hornbeger, G. (1978). Empirical equations for some soil hydraulic properties. *Water Resources Researches*, 14 :601–604.

- Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., and Uddstrom, M. J. (2008). Hydrological data assimilation with the ensemble kalman filter : Use of streamflow observations to update states in a distributed hydrological model. *Advances in Water Resources*, 31 :1309–1324.
- Coe, M. (1998). A linked global model of terrestrial hydrologic processes : Simulation of modern rivers, lakes, and wetlands. *Journal of Geophysical Research*, 103 :8885–8899.
- Coe, M. T. and Birkett, C. M. (2004). Calculation of river discharge and prediction of lake height from satellite radar altimetry : Example for the lake chad basin. *Water Resource Research*, 40(W10205).
- Cohn, S. (1997). An introduction to estimation theory. Technical report, Office Note Series on Global Modeling and Data Assimilation, DAO Office Note 97-01, Goddard Space Flight Center, Greenbelt, Maryland.
- Collischon, W., Allasia, D., da Silva, B. C., and Tucci, C. E. M. (2007). The mgb-iph model for large-scale rainfall-runoff modelling. *Hydrological Sciences Journal*, 52 :878–895.
- Compo, G., Whitaker, J., Sardeshmukh, P., Matsui, N., Allan, R., Yin, X., Jr., B. G., Vose, R., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R., Grant, A., Groisman, P., Jones, P., Kruk, M., Kruger, A., Marshall, G., Maugeri, M., Mok, H., Nordli, O., Ross, T., Trigo, R., Wang, X., Woodruff, S., and Worley, S. (2011). The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, 137 :1–28.
- Courtier, P., Thépaut, J.-N., and Hollingsworth, A. (1994). A strategy for operational implementation of 4d-var using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120 :1367–1387.
- Crow, W. T. and Wood, E. F. (2003). The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble kalman filtering : a case study based on estar measurements during sgp97. *Advances in Water Resources*, 26 :137–149.
- Cudlip, W., Ridley, J. K., and Rapley, C. G. (1990). The use of satellite radar altimetry for monitoring wetlands. In Coulson, M. G., editor, *Remote Sensing and Global Change, Proceedings of the 16th Annual Conference of the Remote Sensing Society*, pages 207–216. Nottingham, England.
- Cukier, R., Fortuin, C., and Shuler, K. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. i. theory. *The Journal of Chemical Physics*, 59(8) :3873–3878.
- da Silva, J. S., Calmant, S., Seyler, F., Filho, O. C. R., Cochonneau, G., and Mansur, W. J. (2010). Water levels in the amazon basin derived from the ers 2 and envisat radar altimetry missions. *Remote Sensing of Environment*, 114 :2160–2181.
- da Silva, J. S., Calmant, S., Seyler, F., Lee, H., and Shum, C. K. (2012). Mapping of extreme stage variations using envisat altimetry in the amazon basin rivers. *International Water Technology Journal*, 2 :14–25.
- da Silva, J. S., Calmant, S., Seyler, F., Moreira, D. M., Oliveira, D., and Monteiro, A. (2014). Radar altimetry aids managing gauge networks. *Water Resources Management*, 28 :587–603.
- de Paiva, R. D., Buarque, D. C., Collischonn, W., Bonnet, M.-P., Frappart, F., Calmant, S., and Mendes, C. A. B. (2013). Large scale hydrological and hydrodynamic modeling of the amazon river basin. *Water Resources Research*, 49 :1226–1243.

- de Paiva, R. D., Collischonn, W., and Tucci, C. E. M. (2011). Large scale hydrologic and hydrodynamic modelling using limited data and a gis based approach. *Journal of Hydrology*, 406 :170–181.
- de Rosnay, P., Drusch, M., Vasiljevic, D., Balsamo, G., Albergel, C., and Isaksen, L. (2013). A simplified extended kalman filter for the global operational soil moisture analysis at ecmwf. *Quarterly Journal of the Royal Meteorological Society*, 139(674) :1199–1213.
- Deardoff, J. W. (1977). A parameterization of ground surface moisture content for use in atmospheric prediction models. *Journal of Applied Meteorology*, 16 :1182–1185.
- Decharme, B., Alkama, R., Douville, H., Becker, M., and Cazenave, A. (2010). Global evaluation of the isba-trip continental hydrological system. part ii : Uncertainties in river routing simulation related to flow velocity and groundwater storage. *Journal of Hydrometeorology*, 11 :601–617.
- Decharme, B., Alkama, R., Papa, F., Faroux, S., Douville, H., and Prigent, C. (2012). Global off-line evaluation of the isba-trip flood model. *Climate Dynamics*, 38 :1389–1412.
- Decharme, B., Boone, A., Delire, C., and Noilhan, J. (2011). Local evaluation of the interaction between soil biosphere atmosphere soil multilayer diffusion scheme using four pedotransfer functions. *Journal of Geophysical Research : Atmospheres*, 116 :D20126.
- Decharme, B., Brun, E., Boone, A., Delire, C., Moigne, P. L., and Morin, S. (2016). Impacts of snow and organic soils parameterization on northern eurasian soil temperature profiles simulated by the isba land surface model. *The Cryosphere*, 10(2) :853–877.
- Decharme, B. and Douville, H. (2005). Introduction of a sub-grid hydrology in the isba lsm. *Climate Dynamics*, 26 :65–78.
- Decharme, B. and Douville, H. (2006). Uncertainties in the gswp-2 precipitation forcing and their impacts on regional and global hydrological simulations. *Climate Dynamics*, 27 :695–713.
- Decharme, B., Douville, H., Boone, A., Habets, F., and Noilhan, J. (2006). Impact of an exponential profile of saturated hydraulic conductivity within the isba lsm : simulations over the rhône basin. *Journal of Hydrometeorology*, 7 :61–70.
- Decharme, B., Douville, H., Prigent, C., Papa, F., and Aires, F. (2008). A new river flooding scheme for global climate applications : Off-line evaluation over south america. *Journal of Geophysical Research - Atmospheres*, 113(D11110).
- Delijani, E. B., Pishvaie, M. R., and Boozarjomehry, R. B. (2014). Subsurface characterization with localized ensemble kalman filter employing adaptative thresholding. *Advances in Water Resources*, 69 :181–196.
- Desroziers, G., Berre, L., Chapnik, B., and Poli, P. (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131 :3385–3396.
- Devegowda, D., Arroyo-Negrete, E., and Datta-Gupta, A. (2010). Flow relevant covariance localization during dynamic data assimilation using enfk. *Advances in Water Resources*, 33 :129–145.
- Dickinson, R. E., Henderson-Sellers, A., Kennedy, P. J., and Wilson, M. F. (1986). *Biosphere-Atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model*. NCAR Tech. Note NT-275 + STR, 69 pp.

- Dickinson, R. E., Shaikh, M., Bryant, R., and Graumlich, L. (1998). Interactive canopies for a climate model. *Journal of Climatology*, 11 :2823–2836.
- Doksum, K. and Samarov, A. (1995). Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. *Annals of Statistics*, 23(5) :1443–1473.
- Döll, P., Douville, H., Güntner, A., Schmied, H. M., and Wada, Y. (2016). Modelling freshwater resources at the global scale : Challenges and prospects. *Surveys in Geophysics*, 37 :195–221.
- Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xavier, A., Gruber, A., Drusch, M., Mecklenburg, S., van Oevelen, P., Robock, A., and Jackson, T. (2011). The international soil moisture network : a data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Science*, 15 :1675–1698.
- Dorigo, W. A., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiová, A., Sanchis-Dufau, A. D., Zamojski, D., Cordes, C., Wagner, W., and Drusch, M. (2013). Global automated quality control of in situ soil moisture data from the international soil moisture network. *Vadose Zone Journal*.
- Doucet, A., de Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Springer, New York.
- Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10 :197–208.
- Douville, H., Royer, J. F., and Mahfouf, J. F. (1995). A new snow parameterization for the meteo-france climate. part 1 : validation in stand-alone experiments. *Climate Dynamics*, 12 :21–35.
- Drolon, V. (2016). *Suivi du bilan de masse saisonnier des glaciers alpins grâce aux images satellites optiques basse résolution*. PhD thesis, Université Paul Sabatier, Toulouse.
- Duan, Q., Sorooshian, Q., and Gupta, V. K. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28 :1015–1031.
- Ducharne, A., Golaz, C., Leblois, E., Laval, K., Polcher, J., Ledoux, E., and de Marsily, G. (2003). Development of a high resolution runoff routing model, calibration and application to assess runoff from the lmd gcm. *Journal of Hydrology*, 280 :207–228.
- Ducharne, A. and Laval, K. (2015). *L'eau à découvert*, chapter Première Partie : les grandes thématiques et les enjeux d'aujourd'hui. Le cycle de l'eau, pages 26–27. CNRS Edition.
- Dumenil, H. and Todini, E. (1992). A rainfall-runoff scheme for use in the hamburg climate model. *Advances in Theoretical Hydrology*, 9 :129–157.
- Dunne, S. and Entekhabi, D. (2005). An ensemble-based reanalysis approach to land data assimilation. *Water Resources Research*, 43(W02013).
- Dunne, T. and Black, R.-D. (1970). Partial area contributions to storm runoff in a small new england watershed. *Water Resources Research*, 7 :1160–1172.
- Durand, M., Andreadis, K., Alsdorf, D., Lettenmaier, D., Moller, D., and Wilson, M. (2008). Estimation of bathymetric depth and slope from data assimilation of swath altimetry into a hydrodynamic model. *Geophysical Research Letters*, 35(L20401).

- Durand, M., Gleason, C. J., Garambois, P. A., Bjerklie, D., Smith, L. C., Roux, H., Rodríguez, E., Bates, P. D., Pavelsky, T. M., Monnier, J., Chen, X., Baldassarre, G. D., Fiset, J.-M., Flipo, N., d. M. Frasson, R. P., Fulton, J., Goutal, N., Hossain, F., Humphries, E., Minear, J. T., Mukolwe, M. M., Neal, J. C., Ricci, S., Sanders, B. F., Schumann, G., Schubert, J. E., and Vilmin, L. (2016). An intercomparison of remote sensing river discharge estimation algorithms from measurements of river height, width, and slope. *Water Resources Research*, 52 :4527–4549.
- Durand, M., Neal, J., Rodríguez, E., Andreadis, K., Smith, L., and Yoon, Y. (2014). Estimating reach-averaged discharge for the river severn from measurements of river water surface elevation and slope. *Journal of Hydrology*, 511 :92–104.
- Efron, B. and Stein, C. (1981). The jackknife estimate of variance. *The Annals of Statistics*, 9(3) :586–596.
- Enjolras, V. M. and Rodríguez, E. (2009). Assessment of a ka-band radar interferometer mission accuracy over eurasian rivers. *IEEE Transactions on Geoscience and Remote Sensing*, 47 :1752–1765.
- Evensen, G. (1994a). Inverse methods and data assimilation in nonlinear ocean models. *Physica D*, 77 :108–129.
- Evensen, G. (1994b). Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5) :10143–10162.
- Evensen, G. (1997). Advanced data assimilation for strongly nonlinear dynamics. *Monthly Weather Review*, 125 :1342–1354.
- Evensen, G. (2003). The ensemble kalman filter : theoretical formulation and practical implementation. *Ocean Dynamics*, 53 :343–367.
- Evensen, G. (2004). Sampling strategies and square root analysis schemes for the enfk. *Ocean Dynamics*, 54 :539–560.
- Evensen, G. (2009). *Data assimilation : the Ensemble Kalman Filter*. Springer.
- Evensen, G. and Leeuwen, P. J. V. (1996). Assimilation of geosat altimeter data for the agulhas current using the ensemble kalman filter with a quasi-geostrophic model. *Monthly Weather Review*, 124 :85–96.
- Evensen, G. and Leeuwen, P. V. (2000). An ensemble kalman smoother for nonlinear dynamics. *Monthly Weather Review*, 128 :1852–1867.
- Faroux, S., Tchuente, A. T. K., Roujean, J.-L., Masson, V., Martin, E., and Moigne, P. L. (2013). Ecoclimap-ii/europe : a twofold database of ecosystems and surface parameters at 1km resolution based on satellite information for use in land surface, meteorological and climate models. *Geoscientific Model Development*, 6 :563–582.
- Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Werner, J. U. M., Oskin, M., Burbank, D., and Alsdorf, D. (2007). The shuttle radar topography mission. *Reviews of Geophysics*, 45(RG2004).
- Faures, J. M., Goodrich, D. C., Woolhiser, D. A., and Sorooshian, S. (1995). Impact of small-scale spatial variability on runoff modeling. *Journal of Hydrology*, 173 :309–326.

- Fekete, B. M., Vörösmarty, C. J., Roads, J. O., and Willmott, C. J. (2003). Uncertainties in precipitation and their impact on runoff estimates. *Journal of Climate*, 17 :294–304.
- Fernandez, D. E. and Stenou, N., editors (2016). *Mission Performance*. Presented at the SWOT Science Team Meeting, 13 June 2016, Pasadena, Californie, États-Unis.
- Ferrya, N., Rémya, E., Brasseur, P., and Maesc, C. (2007). The mercator global ocean operational analysis system : Assessment and validation of an 11-year reanalysis. *Journal of Marine Systems*, 65 :540–560.
- Fisher, M. (1998). Minimization algorithms for variational data assimilation. In *Recent Developments in Numerical Methods for Atmospheric Modelling*, pages 364–385. ECMWF.
- Fjörtoft, R., Gaudin, J.-M., Pourthie, N., Lalaurie, J.-C., Mallet, A., Nouvel, J.-F., Martinot-Lagarde, J., Oriot, H., Borderies, P., Ruiz, C., and Daniel, S. (2014). Karin on swot : Characteristics of near-nadir ka-band interferometric sar imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 52 :2172–2185.
- Fouilloux, A. and Piacentini, A. (1990). The palm project : Mpmc paradigm for an oceanic data assimilation software. *Lecture Notes in Computer Sciences*, 1685 :1423–1430.
- Fox, D. (2008). La dégradation des sols dans le monde. Université de Nice Sophia Antipolis. unt.unice.fr/uoh/degsol.php.
- Francos, A., Elorza, F., Bouraoui, F., Bidoglio, G., and Galbiati, L. (2003). Sensitivity analysis of distributed environmental simulation models : understanding the model behavior in a hydrological studies at the catchment scale. *Reliability Engineering & System Safety*, 79 :205–218.
- Frappart, F., Calmant, S., Cauhope, M., Seyler, F., and Cazenave, A. (2006). Preliminary results of envisat ra-2-derived water levels validation over the amazon basin. *Remote Sensing of Environment*, 100 :252–264.
- Freer, J., Beven, K., and Ambrose, B. (1996). Bayesian estimation of uncertainty in runoff prediction and the value of data : An application of the glue approach. *Water Resources Research*, 32(7) :2161–2173.
- Fu, L.-L., Alsdorf, D., Morrow, R., and Rodríguez, E. (2012a). *SWOT : The Surface Water and Ocean Topography Mission. Wide-Swath Altimetric Measurement of Water Elevation on Earth*. disponible à : swot.jpl.nasa.gov/files/SWOT_MSD_final-3-26-12.pdf.
- Fu, L.-L., Alsdorf, D., Rodríguez, E., Morrow, R., Mognard, N., Lambin, J., Vaze, P., and Lafon, T. (2009). The swot (surface water and ocean topography) mission : Spaceborne radar interferometry for oceanographic and hydrological applications. In *OCEANOBS'09*.
- Fu, L. L., Alsdorf, D. E., and Mognard, R. M. A. R. A. (2012b). Swot : the surface water and ocean topography mission. JPL Publications.
- Fu, L.-L. and Rodríguez, E. (2004). High-resolution measurement of ocean surface topography by radar interferometry for oceanographic and geophysical applications. *Geophysical Monograph Series*, 19 :209–224.
- Galtelli, D., Kucherenko, S., Ratto, M., and Tarantola, S. (2009). Calculating first-order sensitivity measures : a benchmark of some recent methodologies. *Reliability Engineering and System Safety*, 94 :1212–1219.

- Garambois, P.-A. (2012). *Étude régionale des crues éclair de l'arc méditerranéen français ; élaboration de méthodologies de transfert à des bassins versants non jaugés*. PhD thesis, Institut National Polytechnique de Toulouse.
- Garambois, P.-A. and Monnier, J. (2015). Inference of effective river properties from remotely sensed observations of water surface. *Advances in Water Resources*, 79 :103–120.
- Garambois, P.-A., Roux, H., Larnier, K., Castaings, W., and Dartus, D. (2013). Characterization of process-oriented hydrologic model behavior with temporal sensitivity analysis for flash floods in mediterranean catchments. *Hydrology and Earth Systems Sciences*, 17 :2305–2322.
- Gauthier, P., Buehner, M., and Fillion, L. (1999). Background-error statistics modelling in a 3d variational data assimilation scheme : Estimation and impact on the analyses. In *Proceedings of the ECMWF Workshop on Diagnosis of Data Assimilation Systems*, pages 131–145.
- Gelb, A. (1974). *Applied optimal estimation*. Cambridge MIT Press, Massachusetts.
- GEOSS (2014). The geoss water strategy : From observations to decisions.
- Getirana, A. C. V., Boone, A., Yamazaki, D., Decharme, B., Papa, F., and Mognard, N. (2012). The hydrological modeling and analysis platform (hymap) : Evaluation in the amazon basin. *Journal of Hydrometeorology*, 13 :1641–1665.
- Getirana, A. C. V., Boone, A., Yamazaki, D., and Mognard, N. (2013). Automatic parameterization of a flow routing scheme driven by radar altimetry data : Evaluation in the amazon basin. *Water Resources Research*, 49 :614–629.
- Gibelin, A. L., Calvet, J.-C., and Viovy, N. (2008). Modeling energy and CO₂ fluxes with an interactive vegetation land surface model. evaluation at high and middle latitudes. *Agricultural and Forest Meteorology*, 148 :1611–1628.
- Gilks, W. R. and Berzuini, C. (2001). Following a moving target - monte carlo inference for dynamic bayesian models. *Journal of the Royal Statistical Society*, 63 :127–146.
- Gland, F. L. (2003). Filtrage particulaire. *Actes du 19ème colloque GRETSI sur le traitement du signal et des images*, 2 :1–8.
- Gleason, C. J. and Smith, L. C. (2014). Toward global mapping of river discharge using satellite images and at-many-stations hydraulic geometry. *Proceedings of the National Academy of Sciences*, 111 :4788–4791.
- Gleason, C. J., Smith, L. C., and Lee, J. (2014). Retrieval of river discharge solely from satellite imagery and at-many-stations hydraulic geometry : sensitivity to river form and optimization parameters. *Water Resources Research*, 50 :9604–9619.
- Gordon, N., Salmond, D., and Smith, A. (1993). Novel approach to non-linear and non-gaussian bayesian state estimation. *IEEE Proceedings, Part F*, 140 :107–113.
- Gottschalck, J., Meng, J., Rodell, M., and Houser, P. (2005). Analysis of multiple precipitation products and preliminary assessment of their impact on global land data assimilation system land surface states. *Journal of Hydrometeorology*, 6 :573–598.
- Goutal, N., Parisot, M., and Zaoui, F. (2014). A 2d reconstruction for the transverse coupling of shallow water models. *International Journal for Numerical Methods in Fluids*, 75.

- Grayson, R. and Blöschl, G. (2001). *Spatial Patterns in Catchment Hydrology*. Cambridge University Press.
- Gruber, A., Dorigo, W. A., Zwieback, S., Xavier, A., and Wagner, W. (2013). Characterizing coarse-scale representativeness of in situ soil moisture measurements from the international soil moisture network. *Vadose Zone Journal*, 12.
- Gupta, V. K., Sorooshian, S., and Yapo, P. O. (1998). Towards improved calibration of hydrological models : multiple and noncommensurable measures of information. *Water Resources Research*, 34 :751–763.
- Guse, B., Reusser, D., and Fohrer, N. (2014). How to improve the representation of hydrological processes in swat for a lowland catchment-temporal analysis of parameter sensitivity and model performance. *Hydrological Processes*, 28 :2651–2670.
- Gustaffson, N., Huang, X. Y., Yang, X. H., Mogensen, K., Lindskog, M., Vignes, O., Wilhelmsson, T., and Thorsteinsson, S. (2012). Dynamic meteorology and oceanography. *Tellus A*, 64.
- Habets, F., Boone, A., Champeaux, J. L., Etchevers, P., Franchistéguy, L., Leblois, E., Ledoux, E., Moigne, P. L., Martin, E., Morel, S., Noilhan, J., Seguí, P. Q., Rousset-Regimbeau, F., and Viennot, P. (2008). The safran-isba-modcou hydrometeorological model applied over france. *Journal of Geophysical Research*, 113.
- Habets, F., Etchevers, P., Golaz, C., Leblois, E., Ledoux, E., Martin, E., Noilhan, J., and Ottle, C. (1999). Simulation of the water budget and the river flows of the rhone basin. *Journal of Geophysical Research : Atmospheres*, 104 :31145–31172.
- Habets, F. and Saulnier, G.-M. (2001). Subgrid runoff parameterization. *Physics and Chemistry of the Earth, Part B : Hydrology, Oceans and Atmosphere*, 26 :455–459.
- Hagemann, S. and Dümenil, L. (1998). A parameterization of the lateral waterflow for the global scale. *Climate Dynamics*, 14 :17–31.
- Hall, J., Tarantola, S., Bates, P., and Horritt, M. (2005). Distributed sensitivity analysis of flood inundation model calibration. *Journal of Hydraulic Engineering*, 131(2) :117–126.
- Hamill, T. M., Whitaker, J. S., and Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter. *Monthly Weather Review*, 129 :2776–2790.
- Harader, E., Borrell-Estupina, V., Ricci, S., Cousteau, M., Thual, O., Piacentini, A., and Bouvier, C. (2012). Correcting the radar rainfall forcing of a hydrological model with data assimilation : application to flood forecasting in the lez catchment in southern france. *Hydrology and Earth System Sciences*, 16 :4247–4264.
- Hervouet, J.-M. (2007). *Hydrodynamics of Free Surface Flows : Modelling with the finite element method*. John Wiley & Sons.
- Hornberger, G. and Spear, R. (1981). An approach to the preliminary analysis of environmental systems. *Journal of Environmental Management*, 12 :7–18.
- Houser, P. R., Lannoy, G. J. M. D., and Walker, J. P. (2010). Land surface data assimilation. In Lahoz, W. A., Khattatov, B., and Ménard, R., editors, *Data assimilation : making sense of observations*, pages 549–597. Springer, Berlin.

- Houtekamer, P. L. and Mitchell, H. L. (2001). A sequential ensemble kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129 :124–137.
- Houtekamer, P. L. and Mitchell, H. L. (2005). Ensemble kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 131 :3269–3289.
- Houtemak, P. and Mitchell, H. (1998). Data assimilation using an ensemble kalman filter technique. *Monthly Weather Review*, 126 :796–811.
- Hunt, B., Kalnay, E., Kostelich, E. J., Ott, E., Patil, D. T., Sauer, T., Szunyogh, I., Yorke, J. A., and Zimin, A. V. (2004). Four-dimensional ensemble kalman filtering. *Tellus*, 56 :273–277.
- Hunt, B. R., Kostelich, E. J., and Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos : A local ensemble transform kalman filter. *Physica D*, 230 :112–126.
- Hwang, C., Peng, M.-F., Ning, J., Luo, J., and Sui, C.-H. (2005). Lake level variations in china from topex/poseidon altimetry : Data quality assessment and links to precipitation and enso. *Geophysical Journal International*, 161 :1–11.
- Häfliger, V., Martin, E., Boone, A., Habets, F., David, C., Garambois, P.-A., Roux, H., Ricci, S., Berthon, L., Thevenin, A., and Biancamaria, S. (2015). Evaluation of regional-scale river depth simulations using various routing schemes within a hydrometeorological modeling framework for the preparation of the swot mission. *Journal of Hydrometeorology*, 16 :1821–1842.
- Hürzeler, M. and Künsch, H. R. (1998). Monte carlo approximations for general state space models. *Journal of Computational and Graphical Statistics*, 7 :175–193.
- Jansen, M. (1999). Analysis of variance designs for model output. *Computer Physics Communications*, 117 :35–43.
- Jazwinski, A. (1970). *Stochastic processes and filtering theory*. Academic Press, San Diego, California, USA.
- Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Transactions of the American Society of the Mechanical Engineers*, 82 :35–45.
- Kavetski, D., Kuczera, G., and Franks, S. (2006). Bayesian analysis of input uncertainty in hydrological modeling : 1. theory. *Water Resources Research*, 42. W03407.
- Kerr, Y. H. (2007). Soil moisture from space : Where are we ? *Hydrogeology Journal*, 15 :117–120.
- Kidd, C., Matsui, T., Chern, J., Mohr, K., Kummerow, C., and Randel, D. (2016). Global precipitation estimates from cross-track passive microwave observations using a physically based retrieval scheme. *Journal of Hydrometeorology*, 17 :383–400.
- Kiel, B., Alsdorf, D., and LeFavour, G. (2006). Capability of srtm c- and x-band dem data to measure water elevations in ohio and the amazon. *Photogrammetric Engineering and Remote Sensing*, 72.
- Kitagawa, G. (1996). Monte carlo filter and smoother for non-gaussian non-linear state space models. *Journal of Computational and Graphical Statistics*, 5 :1–25.
- Knighton, E. (1998). *Fluvial Forms and Processes : A New Perspective*. Hodder Arnold.
- Koblinsky, C. J., Clarke, R. T., Brenner, A. C., and Frey, H. (1993). Measurement of river level variations with satellite altimetry. *Water Resources Research*, 29(6) :1839–1848.

- Koirala, S., Yeh, P. J.-F., Hirabayashi, Y., Kanae, S., and Oki, T. (2014). Global-scale land surface hydrologic modeling with the representation of water table dynamics. *Journal of Geophysical Research - Atmospheres*, 119 :75–89.
- Kouraev, A. V., Zakharovab, E. A., Samainc, O., Mognard, N. M., and Cazenave, A. (2004). Ob' river discharge from topex/poseidon satellite altimetry (1992-2002). *Remote Sensing of Environment*, 93 :238–245.
- Krinner, G., Viovy, N., de Noblet-Ducouré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C. (2005). A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochemical Cycles*, 19(GB1015).
- Lagarde, T., Piacentini, A., and Declat, D. (2001). A new representation of data assimilation methods : the palm flow charting approach. *Quarterly Journal of the Meteorological Society*, 127 :189–207.
- Lahoz, W. A. and Lannoy, G. L. M. D. (2014). Closing the gaps in our knowledge of the hydrological cycle over land : Conceptual problems. *Survey of Geophysics*, 35 :623–660.
- Lawrence, D. M., Olsen, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., Bonan, G. B., and Slater, A. G. (2011). Parameterization improvements and functional and structural advances in version 4 of the community land model. *Journal of Advances in Modelling Earth Systems*, 3.
- Leeuwen, P. V. and Evensen, G. (1996). Data assimilation and inverse methods in terms of a probabilistic formulation. *Monthly Weather Review*, 124(12) :2898–2913.
- LeFavour, G. and Alsdorf, D. (2005). Water slope and discharge in the amazon river estimated using the shuttle radar topography mission digital elevation model. *Geophysical Research Letters*, 32.
- Lehner, B. and Grill, G. (2013). Global river hydrography and network routing : baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, 27 :2171–2186.
- Lehner, B., Verdin, K., and Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *EOS Transactions American Geophysical Union*, 89 :93–94.
- LeMoigne, P., Colin, J., and Decharme, B. (2016). Impact of lake surface temperatures simulated by the flake scheme in the cnrm-cm5 climate model. *Tellus A*, 68 :31274.
- Li, J. and Xiu, D. (2008). On numerical properties of the ensemble kalman filter for data assimilation. *Computer Methods in Applied Mechanics and Engineering*, 197 :3574–3583.
- Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for gsms. *Journal of Geophysical Research*, 99 :14415–14428.
- Lion, C. (2012). *Simulation des données SWOT haute résolution et applications à l'étude de l'estuaire de l'Amazone*. PhD thesis, Université Paul Sabatier - Toulouse III.
- Lohmann, D., Nolte-Holube, R., and Raschke, E. (1996). A large-scale horizontal routing model to be coupled to land surface parametrization schemes. *Tellus*, 48 :708–721.
- Lorenc, A. (1986). Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112 :1177–1194.

- Lucas-Picher, P., Arora, V., Caya, D., and Laprise, R. (2010). Implementation of a large-scale variable velocity river flow routing algorithm in the canadian regional climate model (crcm). *Atmosphere-Ocean*, 41 :139–153.
- MacArthur, J. L. (1980). Altimeter designs - seasat-1 and future missions. *Marine Geodesy*, 3 :39–61.
- Mahfouf, J.-F. and Noilhan, J. (1991). Comparative study of various formulations of evaporation from bare soil using insitu data. *Journal of Applied Meteorology*, 9 :351–362.
- Mahfouf, J.-F. and Noilhan, J. (1996). Inclusion of gravitational drainage in a land surface scheme based on the force restore method. *Journal of Applied Meteorology*, 35 :987–992.
- Maidment, D. R. (1993). *Handbook of hydrology*. McGraw Hill Professional.
- Manabe, S. (1969). Climate and ocean circulation. i. the atmospheric circulation and the hydrology of the earth's surface. *Monthly Weather Review*, 97 :739–774.
- Manning, R. (1891). On the flow of water in open channels and pipes. *Institution of Civil Engineers of Ireland*, 20 :161–207.
- Massart, S., Clerbaux, C., Cariolle, D., Piacentini, A., Turquety, S., and Hadji-Lazaro, J. (2009). First steps towards the assimilation of iasi data into the moca-palm system. *Atmospheric Chemistry and Physics*, 9 :5073–5091.
- Matgen, P., Schumann, G., Henry, J.-B., Hoffmann, L., and Pfister, L. (2007). Integration of sar-derived river inundation areas, high-precision topographic data and a river flow model toward near real-time flood management. *International Journal of Applied Earth Observation and Geoinformation*, 9 :247–263.
- Meade, R., Rayol, J., Conceição, S. D., and Natividade, J. (1991). Backwater effects in the amazon river basin of brazil. *Environmental Geology and Water Sci.*, 18 :105–114.
- Mechri, R. (2014). *Désagrégation spatiale de températures Météosat par une méthode d'assimilation de données (lisseur particulière) dans un modèle de surface continentale*. PhD thesis, Université de Versailles.
- Michailovsky, C. I., Milzow, C., and Bauer-Gottwein, P. (2013). Assimilation of radar altimetry to a routing model of the brahmaputra river. *Water Ressources Research*, 49 :4807–4816.
- Michel, Y. (2013a). Estimating deformations of random processes for correlation modelling in a limited area model. *Quarterly Journal of the Royal Meteorological Society*, 139(671) :534–547.
- Michel, Y. (2013b). Estimating deformations of random processes for correlation modelling : Methodology and the one-dimensional case. *Quarterly Journal of the Royal Meteorological Society*, 139(672) :771–783.
- Milliman, J. D. and Farnsworth, K. L. (2011). *River Discharge to the Coastal Ocean*, chapter Runoff, erosion, and delivery to the coastal ocean, pages 13–69. Cambridge University Press.
- Milly, P. C., Malyshev, S., Shevliakova, E., Dunne, K. A., Findell, K. L., Gleeson, T., Liang, Z., Philipps, P., Stouffer, R. J., and Swenson, S. C. (2014). An enhanced model of land water and energy for global hydrologic and earth-system studies. *Journal of Hydrometeorology*, 15 :1739–1761.

- Milly, P. C. and Shmakin, A. B. (2002). Global modeling of land water and energy balances. part i : The land dynamics (lad) model. *Journal of Hydrometeorology*, 3 :283–299.
- Mirouze, I. (2010). *Régularisation de problèmes inverses à l'aide de l'équation de diffusion, avec application à l'assimilation variationnelle de données océaniques*. PhD thesis, Université de Toulouse, France.
- Misirli, F., Gupta, H. V., Sorooshian, S., and Thiemann, M. (2003). Bayesian estimation of parameter and output uncertainty for watershed models. In *Calibration of watershed models*. American Geophysical Union.
- Molinier, M., Guyot, J.-L., Orstom, B., Guimarães, V., de Oliveira, E., and Dnaee, B. (1993). Hydrologie du bassin de l'amazone. In Editions, O., editor, *Grands Bassins Fluviaux Périalantiques*, pages 335–345, Paris. PEGI-INSA-CNRS-ORSTOM.
- Moller, D. and Esteban-Fernandez, D. (2015). Near-nadir ka-band field observation of fresh water bodies. In *Remote Sensing of the Water Cycle*, pages 143–155. Wiley.
- Montaldo, N. and Albertson, J. D. (2001). On the use of the force-restore svat model formulation for stratified soils. *Journal of Hydrology*, 10 :282–290.
- Montmerle, T. and Berre, L. (2010). Diagnosis and formulation of heterogeneous background-error covariances at the mesoscale. *Quarterly Journal of the Royal Meteorological Society*, 136 :1408–1420.
- Montzka, C., Moradkhani, H., Weihermüller, L., Franssen, H.-J. H., Canty, M., and Vereecken, H. (2011). Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter. *Journal of Hydrology*, 399 :410–421.
- Moody, J. and Troutman, B. (2002). Characterization of the spatial variability of channel morphology. *Earth Surface Processes and Landforms*, 27 :1251–1266.
- Moradkhani, H. (2008). Hydrological remote sensing and land surface data assimilation. *Sensors*, 8 :2986–3004.
- Moradkhani, H., Hsu, K.-L., Gupta, H., and Sorooshian, S. (2005). Uncertainty assessment of hydrologic model states and parameters : Sequential data assimilation using particle filter. *Water Resources Research*, 41(W05012).
- Musso, C., Oudjane, N., and Gland, F. L. (2001). Improving regularized particle filters. In *Sequential Monte Carlo Methods in Practice*, pages 247–271. Springer.
- Neal, J., Schumann, G., Bates, P., Buytaert, W., Matgen, P., and Pappenberger, F. (2009). A data assimilation approach to discharge estimation from space. *Hydrological Processes*, 23 :3641–3649.
- Neal, R. M. (2001). Annealing importance sampling. *Statistics and Computing*, 11 :125–139.
- Ngo-Duc, T., Oki, T., and Kanae, S. (2007). A variable streamflow velocity method for global river routing model : model description and preliminary results. *Hydrology and Earth System Sciences*, 4 :4389–4414.
- Nijssen, B. and Lettenmaier, D. P. (2004). Effect of precipitation sampling error on simulated hydrological fluxes and states : Anticipating the global precipitation measurement satellites. *Journal of Geophysical Research*, 109.

- Noilhan, J. and Mahfouf, J.-F. (1996). The isba land surface parameterization scheme. *Global and Planetary Change*, 13 :145–159.
- Noilhan, J. and Planton, S. (1989). A simple parameterization of land surface processes for meteorological models. *Monthly Weather Review*, 117 :536–549.
- Nykamen, D. K., Foufoula-Georgiou, E., and Lapenta, W. M. (2001). Impact of small-scale precipitation variability on larger-scale spatial organization of land atmosphere fluxes. *Journal of Hydrometeorology*, 2 :105–121.
- Ogden, F. L. and Julien, P. Y. (1993). Runoff sensitivity to temporal and spatial precipitation variability at runoff plane and small basin scale. *Water Resources Research*, 29 :2589–2597.
- Ogden, F. L. and Julien, P. Y. (1994). Runoff model sensitivity to radar precipitation resolution. *Journal of Hydrology*, 158 :1–18.
- Oki, T. and Kanae, S. (2006). Global hydrological cycles and world water resources. *Science*, 313 :1068–1072.
- Oki, T., Nishimura, T., and Dirmeyer, P. (1999). Assessment of annual runoff from land surface models using total runoff integrating pathways (trip). *Journal of the Meteorological Society of Japan*, 77 :235–255.
- Oki, T. and Sud, Y. C. (1998). Design of total integrating pathways (trip)-a global river channel network. *Earth Interactions*, 2 :1–36.
- Ott, E., Hunt, B. R., Szunyogh, I., Kostelich, A. V. Z. A. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A. (2004). A local ensemble kalman filter for atmospheric data assimilation. *Tellus A*, 56 :415–428.
- Ottlé, C. and Mahfouf, J.-F. (2016). Assimilation de données spatiales. In *Système Terre et Environnement*, E. I., editor, *Téledétection pour l'observation des surfaces continentales, Tome 1*. Elsevier. ISBN :978-1-78405-157-0.
- Paiva, R. C. D., Collischonn, W., and Buarque, D. C. (2013). Validation of a full hydrodynamic model for large-scale hydrological modelling in the amazon. *Hydrological Processes*, 27 :333–346.
- Palvis, N. K., Holmes, S. A., Kenyon, S. C., and Factor, J. K. (2012). The development and evaluation of the earth gravitational model 2008 (egm2008). *Journal of Geophysical Research*, 117(B04406).
- Pan, M. and Wood, E. F. (2013). Inverse streamflow routing. *Hydrology and Earth System Sciences*, 17 :4577–4588.
- Pannekoucke, O., Berre, L., and Desroziers, G. (2008). Background error correlation length-scale estimates and their sampling statistics. *Quarterly Journal of the Royal Meteorological Society*, 134 :497–511.
- Papa, F., Prigent, C., and Rossow, W. B. (2007). Ob' river flood inundations from satellite observations : A relationship with winter snow parameters and river runoff. *Journal of Geophysical Research*, 112(D18103).
- Pappenberger, F. and Beven, K. (2004). Functional classification and evaluation of hydrographs based on multi-component mapping. *International Journal of River Basin Management*, 2(2) :89–100.

- Pappenberger, F., Beven, K., Roo, A. D., Thielen, J., and Gouweleeuw, B. (2004). Uncertainty analysis of the rainfall-runoff model lisflood within the generalized likelihood uncertainty estimation (glue). *International Journal of River Basin Management*, 2(2) :123–133.
- Pappenberger, F., Cloke, H., Balsamo, G., Ngo-Duc, T., and Oki, T. (2010). Global runoff routing with the hydrological component of the ecmwf nwp system. *International Journal of Climatology*, 30 :2155–2174.
- Paris, A. (2015). *Utilisation conjointe de données d'altimétrie satellitaire et de modélisation pour le calcul des débits distribués dans le bassin amazonien*. PhD thesis, Université Paul Sabatier, Toulouse III.
- Paris, A., de Paiva, R. D., da Silva, J. S., Moreira, D. M., Calmant, S., Garambois, P.-A., Collischonn, W., Bonnet, M.-P., and Seyler, F. (2016). Stage-discharge rating curves based on satellite altimetry and modeled discharge in the amazon basin. *Water Resources Research*, 52.
- Parrish, D. and Derber, J. (1992). The national meteorological center's spectral statistical interpolation analysis system. *Monthly Weather Review*, 120 :1747–1763.
- Pauwels, V., Balenzano, A., Satalino, G., Skriver, H., Verhoest, N., and Mattia, F. (2009). Optimization of soil hydraulic model parameters using synthetic aperture radar : An integrated multidisciplinary approach. *IEEE Transactions on Geoscience and Remote Sensing*, 47 :455–457.
- Pavelsky, T. M. and Durand, M. T. (2012). Developing new algorithms for estimating river discharge from space. *EOS Transactions American Geophysical Union*, 93.
- Pavelsky, T. M., Durand, M. T., Andreadis, K. M., Beighley, R. E., Paiva, R. C. D., Allen, G. H., and Miller, Z. F. (2014). Assessing the potential global extent of swot river discharge observations. *Journal of Hydrology*, 519 :1516–1525.
- Pedinotti, V., Boone, A., Decharme, B., Crétaux, J. F., Mognard, N., Panthou, G., Papa, F., and Tanimoun, B. A. (2012). Evaluation of the isba-trip continental hydrologic system over the niger basin using in situ and satellite derived datasets. *Hydrology and Earth System Sciences*, 16 :1745–1773.
- Pedinotti, V., Boone, A., Ricci, S., Biancamaria, S., and Mognard, N. (2014). Assimilation of satellite data to optimize large-scale hydrological model parameters : a case study for the swot mission. *Hydrology and Earth System Science*, 18 :4485–4507.
- Pitt, M. K. and Shepard, N. (1999). Filtering via simulations : auxiliary particle filter. *Journal of the American Statistical Association*, 94 :590–599.
- Poitou, D., Amaya, J., El-Hafi, M., and Cuenot, B. (2011). Analysis of the interaction between turbulent combustion : development of a methodology to couple turbulent combustion and radiative heat transfer in les. *Journal of Heat Transfer*, 133.
- Polcher, J. (2003). *Les processus de surfaces à l'échelle globale et leurs interactions avec l'atmosphère*. Habilitation à diriger des recherches, Université Paris VI.
- Puri, S. and Aureli, A. (2009). *Atlas of Transboundary Aquifers - Global Maps, Regional Cooperation and Local Inventories*. UNESCO-IHP ISARM Programme, Paris.
- Rabitz, H., Aliş, O., Shorter, J., and Shim, K. (1999). Efficient input-output model representations. *Computer Physics Communications*, 117 :11–20.

- Rakovec, O., Weerts, A. H., Sumihar, J., and Uijlenhoet, R. (2015). Operational aspects of asynchronous filtering for flood forecasting. *Hydrology and Earth System Sciences*, 19 :2911–2924.
- Rapley, C. G. (1990). Satellite radar altimeters. In Vaughan, R. A., editor, *Microwave Remote Sensing for Oceanographic and Marine Weather-Forecast Models*, pages 45–63. Kluwer Academy, Norwell.
- Rapley, C. G., Guzkowaska, M. A. J., Cudlip, W., and Mason, I. M. (1987). *An Exploratory Study of Inland Water and Land Altimetry Using Seasat Data*. European Space Agency.
- Rasmussen, J., Madsen, H., Jensen, K. H., and Refsgaard, J. C. (2015). Data assimilation in integrated hydrological modeling using ensemble kalman filtering : evaluating the effect of ensemble size and localisation on filter performance. *Hydrology and Earth System Sciences*, 19 :2999–3013.
- Ratto, M., Pagano, A., and Young, P. (2007). State dependent parameter metamodelling and sensitivity analysis. *Computer Physics Communications*, 177 :863–876.
- Reichle, R. H. (2008). Data assimilation methods in the earth science. *Advances in Water Resources*, 31 :1411–1418.
- Reichle, R. H., DeLannoy, G. J. M., Draper, B. A. F. C. S., and Liu, Q. (2014). Connecting satellite observations with water cycle variables through land data assimilation : Examples using the nasa geos-5 ldas. *Surveys in Geophysics*, 35 :577–606.
- Reichle, R. H., Lannoy, G. J. M. D., Forman, B., Draper, C., and Liu, Q. (2013). Connecting satellite observations with water cycle variables through data assimilation : examples using the nasa geos-5 ldas. *Survey of Geophysics*, 46 :577–606.
- Reichle, R. H., McLaughlin, D. B., and Entekhani, D. (2002). Hydrological data assimilation with the ensemble kalman filter. *Monthly Weather Review*, 130 :103–114.
- Reusser, D., Buytaert, W., and Zehe, E. (2011). Temporal dynamics of model parameter sensitivity for computationally expensive models with fast (fourier amplitude sensitivity test). *Water Resources Research*, 47. W07551.
- Reusser, D. and Zehe, E. (2011). Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity. *Water Resources Research*, 47. W07550.
- Ricci, S., Piacentini, A., Thual, O., Pape, E. L., and Joncille, G. (2011). Correction of upstream flow and hydraulic state with the data assimilation in the context of flood forecasting. *Hydrology and Earth Systems Science*, 15 :3555–3575.
- Riccio, M. (2014). Performances du système swot en hydrologie. Master's thesis, Conservatoire National des Arts et Métiers - Ecole Supérieure des Géomètres et Topographes.
- Robock, A., Vinnikov, K. Y., Srinivasan, G., Entin, J. K., Hollinger, S. E., Speranskaya, N. A., Liu, S., and Namkhai, A. (2000). The global soil moisture data bank. *Bulletin of the American Meteorological Society*, 81 :1281–1299.
- Rochoux, M. C. (2014). *Towards a more comprehensive monitoring of wildfire spread : contributions of model evaluation and data assimilation strategies*. PhD thesis, Ecole Centrale Paris.
- Rochoux, M. C., Emery, C., Cuenot, S. R. A., and Trouvé, A. (2015). Towards predictive data-driven simulations of wildfire spread - part ii : Ensemble kalman filter for the state estimation of a front-tracking simulator of wildfire spread. *Natural Hazards and Earth System Sciences*, 15 :1721–1739.

- Rodell, M., Chen, J. L., Kato, H., Famiglietti, J. S., Nigro, J., and Wilson, C. R. (2007). Estimating groundwater storage changes in the mississippi river basin (usa) using grace. *Hydrogeology Journal*, 15 :159–166.
- Rodell, M. and Famiglietti, J. (2002). The potential for satellite-based monitoring of groundwater storage changes using grace : The high plains aquifer. *Journal of Hydrology*, 263 :245–256.
- Rodríguez, E. (2016). *Surface Water and Ocean Topography mission (SWOT) Science Requirements Document*. Jet Propulsion Laboratory, Californian Institute of Technology, Pasadena, CA, USA. Available : http://swot.jpl.nasa.gov/files/swot/D-61923_SRD_RevA_20160318_with_signatures.pdf.
- Rodríguez, E., Moller, D., Smith, L. C., Pavelsky, T. M., and Alsdorf, D. E. (2010). Airswot : an airborne monitoring platform for surface water monitoring. In *AGU Fall Meeting Abstract*. #H32D-06.
- Rodríguez, E. and Pollard, B. D. (2001). The measurement capabilities of wide-swath ocean altimeters. Technical report, Jet Propulsion Laboratory (JPL).
- Ros, D. and Borga, M. (1997). Adaptive use of a conceptual model for real time flood forecasting. *Nordic Hydrology*, 28 :169–188.
- Roux, H. (2004). *Estimation de paramètres en hydraulique fluviale, à partir de données caractéristiques de l'imagerie aérienne*. PhD thesis, Institut National Polytechnique de Toulouse.
- Roux, H., Labat, D., Garambois, P.-A., Maubourguet, M.-M., Chorda, J., and Dartus, D. (2011). A physically-based parsimonious hydrological model for flash floods in mediterranean catchments. *Natural Hazards and Earth System Sciences*, 11 :2567–2582.
- Sakov, P., Evensen, G., and Bertino, L. (2010). Asynchronous data assimilation with the enf. *Tellus*, 62 :24–29.
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. (2010). Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer Physics Communications*, 181 :259–270.
- Saltelli, A. and Bolado, R. (1996). An alternative way to compute fourier amplitude sensitivity test (fast). *Computational Statistics & Data Analysis*, 26 :445–460.
- Saltelli, A., Chan, K., and Scott, E. M. (2008). *Sensitivity Analysis*. Wiley. ISBN : 978-0-470-74382-9.
- Sampson, C., Smith, A., Bates, P., Neal, J., Alfieri, L., and Freer, J. (2015a). A high-resolution global flood hazard model. *Water Resources Research*, 2015 :7358–7381.
- Sampson, C., Smith, A., Bates, P., Neal, J., Alfieri, L., and Freer, J. (2015b). A high-resolution global flood hazard model. *Water Resources Research*, 51 :7358–7381.
- Sanoo, A. K., Pan, M., Troy, T. J., Vinukollu, R. K., Sheffield, J., and Wood, E. F. (2011). Reconciling the global terrestrial water budget using satellite remote sensing. *Remote Sensing of Environment*, 115 :1850–1865.
- Santini, M. and di Paola, A. (2015). Changes in the world rivers' discharge projected from an updated high resolution dataset of current and future climate zones. *Journal of Hydrology*, 531 :768–780.
- Sasaki, Y. (1970). Some basic formalisms in numerical variational analysis. *Monthly Weather Review*, 98 :875–883.

- Schmugge, T. J. (1998). Application of passive microwave observations of surface soil moisture. *Journal of Hydrology*, 212 :188–197.
- Schumann, G., Hostache, R., Puech, C., Hoffmann, L., Matgen, P., Pappenberger, F., and Pfister, L. (2007). High-resolution 3-d flood information from radar imagery for flood hazard management. *IEEE Transactions on Geoscience and Remote Sensing*, 45 :1715–1725.
- Schumann, G., Matgen, P., Cutler, M. E. J., Black, A., Hoffmann, L., and Pfister, L. (2008). Comparison of remotely sensed water stages from lidar, topographic contours and srtm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63 :283–296.
- Schumann, G., Neal, J. C., Mason, D. C., and Bates, P. D. (2011). The accuracy of sequential aerial photography and sar data for observing urban flood dynamics, a case study of the uk summer 2007 floods. *Remote Sensing of Environment*, 115 :2536–2546.
- Schumann, G. J.-P. and Domeneghetti, A. (2016). Exploiting the proliferation of current and future satellite observations of rivers. *Hydrological Processes*, 30 :2891–2896.
- Sellers, P. J., Mintz, Y., Sud, Y. C., and Dalcher, A. (1986). A simple biosphere model (sib) for use within general circulation models. *Journal of the Atmospheric Sciences*, 43 :505–531.
- Sieber, A. and Uhlenbrook, S. (2005). Sensitivity analyses of a distributed catchment model to verify the model structure. *Journal of Hydrology*, 310 :216–235.
- Simpson, J., Adler, R. F., and North, G. R. (1988). A proposed tropical rainfall measuring mission (trmm) satellite. *Bulletin of the American Meteorological Society*, 69 :278–295.
- Sioli, H. (1984). *The Amazon*, volume 56, chapter The Amazon and its main affluents : Hydrography, morphology of the river courses, and river types, pages 127–165. Springer Netherlands.
- Skøien, J. O., Blöschl, G., and Western, A. W. (2003). Characteristic space scales and timescales in hydrology. *Water Resources Research*, 39.
- Smith, E. A., Asrar, G., Furuhashi, Y., Ginati, A., Mugnai, A., Nakamura, K., and Coauthors (2007). International global precipitation measurement (gpm) program and missions : An overview. In Levizzani, V., Bauer, P., and Turk, F. J., editors, *Measuring Precipitation from Space : EURAINSAT and the Future*, pages 611–654. Springer.
- Smith, L. C. and Alsdorf, D. E. (1998). Control on sediment and organic carbon delivery to the arctic ocean revealed with space-borne synthetic aperture radar : Ob' river, siberia. *Geology*, 26 :395–398.
- Smith, L. C. and Pavelsky, T. M. (2008). Estimation of river discharge, propagation speed, and hydraulic geometry from space : Lena river, siberia. *Water Resources Research*, 44(W03427).
- Sobol, I. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4) :86–112.
- Sobol, I. (1993). Sensitivity analysis for non-linear mathematical models. *Mathematical Modeling and Computational Experiment*, 2 :407–414.
- Sobol, I. (2001). Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and Computers in Simulation*, 55 :271–280.
- Sorooshian, S. and Dracup, J. A. (1980). Stochastic parameter estimation procedures for hydrologic rainfall-runoff models : Correlated and heteroscedastic error cases. *Water Resources Research*, 16 :430–442.

- Stallart, R. and Edmont, J. (1983). Geochemistry of the amazon, 2. the influence of geology and weathering environment on the dissolved load. *Journal of Geophysical Research*, 88 :9671–9688.
- Steunou, N., Desjonquères, J. D., Picot, N., Sengenès, P., Noubel, J., and Poisson, J. C. (2015). AltiKa altimeter : instrument description and in flight performance. *Marine Geodesy*, 38 :22–42.
- Strassberg, G., Scanlon, B. R., and Rodell, M. (2007). Comparison of seasonal terrestrial water storage variations from grace with groundwater-level measurements from the high plains aquifer (usa). *Geophysical Research Letters*, 34.
- Sun, A. Y., Morris, A. P., and Mohanty, S. (2009). Sequential updating of multimodal hydrogeologic parameter fields using localisation and clustering techniques. *Water Resources Research*, 45(W07424).
- Talagrand, O. and Courtier, P. (1987). Variational assimilation of meteorological observations with the adjoint vorticity equation, part 1 : Theory. *Quarterly Journal of the Royal Meteorological Society*, 113 :1311–1328.
- Tanajura, C. A. . S., Lima, L. N., and Belyav, K. P. (2015). Assimilation of satellite surface-height anomalies data into a hybrid coordinate ocean model (hycom) over the atlantic ocean. *Marine Physics*, 55.
- Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., and Watkins, M. M. (2004). Grace measurements of mass variability in the earth system. *Science*, 305 :503–505.
- Tarpanelli, A., Brocca, L., Barbetta, S., Faruolo, M., Lacava, T., and Moramarco, T. (2015). Coupling modis and radar altimetry data for discharge estimation in poorly gauged river basins. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8 :141–148.
- Tarpanelli, A., Brocca, L., Lacava, T., Melone, F., Moramarco, T., Faruolo, M., Pergola, N., and Tramutoli, V. (2013). Toward the estimation of river discharge variations using modis data in ungauged basins. *Remote Sensing of Environment*, 136 :47–55.
- Tedesco, M. and Narvekar, P. S. (2010). Assessment of the nasa amsr-e swe product. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3 :141–159.
- Thiemann, M., Trosset, M., Gupta, H., and Sorooshian, S. (2001). Bayesian recursive parameter estimation for hydrologic models. *Water Resources Research*, 37 :2521–2535.
- Thirel, G., Martin, E., Mahfouf, J.-F., Massart, S., Ricci, S., and Habets, F. (2010). A past discharges assimilation system for ensemble streamflow forecasts over france. *Hydrology and Earth System Sciences*, 14 :1623–1637.
- Trenberth, K. E. and Asrar, G. R. (2014). Challenges and opportunities in water cycle research : Wcrp contributions. *Survey in Geophysics*, 35 :515–532.
- Trenberth, K. E., Fasullo, J. T., and Mackaro, J. (2011). Atmospheric moisture transports from ocean to land and global energy flows in reanalyses. *Journal of Climate*, 24 :4907–4924.
- Trenberth, K. E., Smith, L., Qian, T., Dai, A., and Fasullo, J. (2007). Estimates of the global water budget and its annual cycle using observational and model data. *Journal of Hydrometeorology*, 8 :758–769.

- Trudel, M., Leconte, R., and Paniconi, C. (2014). Analysis of the hydrological response of a distributed physically-based model using post-assimilation (enkf) diagnostics of streamflow and in situ soil moisture observations. *Journal of Hydrology*, 514(192-201).
- van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., and Srinivasan, R. (2006). A global sensitivity analysis tool for the parameters of multi-variable catchment models. *Journal of Hydrology*, 324 :10-23.
- Verdin, K. L. and Greenlee, S. K. (1998). Hydro1k documentation. US Geological Survey. <https://lta.cr.usgs.gov/HYDRO1KReadMe>.
- Vergnes, J.-P. (2012). *Développement d'une modélisation hydrologique incluant la représentation des aquifères : évaluation sur la France et à l'échelle globale*. PhD thesis, Université de Toulouse.
- Vergnes, J.-P., Decharme, B., Alkama, R., Martin, E., Habets, F., and Douville, H. (2012). A simple groundwater scheme for hydrological and climate applications : Description and offline evaluation over france. *Journal of Hydrometeorology*, 13 :1149-1171.
- Vergnes, J.-P., Decharme, B., and Habets, F. (2014). Introduction of groundwater capillary rises using subgrid spatial variability of topography into the isba land surface model. *Journal of Geophysical Research - Atmosphere*, 119 :11065-11086.
- Verseghy, D. L. (1991). Class - a canadian land surface scheme for gcms. i. soil model. *International Journal of Climatology*, 11 :111-133.
- Vicente, G. A., Scofield, R. A., and Menzel, W. P. (1998). The operational goes infrared rainfall estimation technique. *Bulletin of the American Meteorological Society*, 79(9) :1883-1898.
- Vinukollu, R. K., Meynadier, R., Sheffield, J., and Wood, E. F. (2011). Multi-model, multi-sensor estimates of global evapotranspiration : climatology, uncertainties and trends. *Hydrological Processes*, 25 :3993-4010.
- Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Moigne, P. L., Martin, E., and Willemet, J.-M. (2012). The detailed snowpack scheme crocus and its implementation in surfex v7.2. *Geoscientific Model Development*, 5 :773-791.
- Vörösmarty, C., Moore, B., Grace, A., Gildea, M., Melillo, J., Peterson, B., Rastetter, E., and Steudler, P. (1989). Continental scale models of water balance and fluvial transport : An application to south america. *Global Biogeochemical Cycles*, 3 :241-265.
- Vrugt, J., Braak, C. T., Diks, C., and Schoups, G. (2013). Hydrologic data assimilation using particle markov chain monte carlo simulation : theory, concepts and applications. *Advances in Water Resources*, 51 :457-478.
- Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S. (2003). A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, 39.
- Wagener, T., McIntyre, N., Lees, M., Wheeler, H., and Gupta, H. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling : dynamic identifiability analysis. *Hydrological Processes*, 17 :455-476.
- Wagner, W., Blochl, G., Pampaloni, P., Calvet, J. C., Bizzarri, B., Wigneron, J. P., and Kerr, Y. (2007). Operational readiness of microwave remote sensing of soil moisture for hydrological applications. *Nordic Hydrology*, 38(1) :1-20.

- Wahr, J., Molenaar, M., and Bryan, F. (1998). Time variability of the earth's gravity field : Hydrological and oceanic effects and their possible detection using grace. *Journal of Geophysical Research-Solid Earth*, 103 :205–230.
- Wahr, J., Swenson, S., Zlotnicki, V., and Velicogna, I. (2004). Time-variable gravity from grace : First results. *Geophysical Research Letters*, 31.
- Weaver, A. and Courtier, P. (2001). Correlation modelling on the sphere using a generalized diffusion equation. *Quarterly Journal of the Royal Meteorological Society*, 127 :1815–1846.
- Werner, M., Hunter, N., and Bates, P. (2005). Identifiability of distributed floodplain roughness values in flood extent estimation. *Journal of Hydrology*, 314 :139–157.
- Williams, B., Rodriguez, E., Biancamaria, S., Durand, M., and Blumstein, D. (2016). Surface water and ocean topography (swot) mission : Hydrology simulator. In *Présenté à la réunion de la Science Mission Team à Pasadena, Californie*.
- Winsemius, H., Beek, L. V., Jongman, B., Ward, P., and Bouwman, A. (2013). A framework for global river flood risk assessments. *Hydrology and Earth System Science*, 17 :1871–1892.
- Wisser, D., Feketa, B., Vörösmarty, C. J., and Schumann, A. H. (2010). Reconstructing 20th century global hydrography : a contribution to the global terrestrial network- hydrology (gtn-h). *Hydrology and Earth System Sciences*, 14 :1–24.
- Wood, E., Roundy, J. K., Troy, T. J., van Beek, R., Bierkens, M., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffe, P., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P. (2011). Hyper-resolution global land surface modeling : meeting a grand challenge for monitoring earth's terrestrial water. *Water Resources Research*, 47 :W05301.
- Wood, E. F., Lettenmaier, D. P., Liang, X., Lohmann, D., Boone, A., Chang, S., Chen, F., Dai, Y. J., Dickinson, R. E., Duan, Q. Y., Ek, M., Gusev, Y. M., Habets, F., Irannejad, P., Koster, R., Mitchel, K. E., Nasonova, O. N., Noilhan, J., Schaake, J., Schlosser, A., Shao, Y. P., Shmakin, A. B., Verseghy, D., Warrach, K., Wetzel, P., Xue, Y. K., Yang, Z. L., and Zeng, Q. C. (1998). The project for intercomparison of land-surface parameterization schemes (pilps) phase 2(c) red-arkansas river basin experiment : 1. experiment description and summary intercomparisons. *Global and Planetary Change*, 19 :115–135.
- WWDR2 (2006). *The United Nations World Water Development Report 2 : Water, a shared responsibility*. UNESCO.
- WWDR4 (2012). *The United Nations World Water Development Report 4 : Managing Water Under Uncertainty and Risk*. UNESCO.
- Yamazaki, D., Kanae, S., Kim, H., and Oki, T. (2011). A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resources Research*, 47(W04501).
- Yamazaki, D., O'Loughlin, F., Trigg, M., Miller, Z., Pavelsky, T., and P.D.Bates (2014). Development of the global width database for large rivers. *Water Resources Research*, 50 :3467–3480.
- Yeh, P., Swenson, S., Famiglietti, J., and Rodell, M. (2006). Remote sensing of groundwater storage changes in illinois using the gravity recovery and climate experiment (grace). *Water Resources Research*, 42.

- Yoon, Y., Durand, M., Merry, C. J., Clark, E. A., Andreadis, K. M., and Alsdorf, D. E. (2012). Estimating river bathymetry from data assimilation of synthetic swot measurements. *Journal of Hydrology*, 464-465 :363-375.
- Zhao, R. J. (1992). The xinanjiang model applied in china. *Journal of Hydrology*, 134 :317-381.
- Zwenzner, H. and Voigt, S. (2009). Improved estimation of flood parameters by combining space based sar data with very high resolution digital elevation model. *Hydrology and Earth System Sciences*, 13 :567-576.