



HAL
open science

Assimilation de données variationnelle pour les problèmes de transport des sédiments en rivière

Junqing Yang

► **To cite this version:**

Junqing Yang. Assimilation de données variationnelle pour les problèmes de transport des sédiments en rivière. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1999. Français. NNT : . tel-00004863

HAL Id: tel-00004863

<https://theses.hal.science/tel-00004863>

Submitted on 18 Feb 2004

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

Présentée par

Junqing YANG

Pour obtenir le titre de

Docteur de l'Université Joseph Fourier - Grenoble I
et de l'Université de Wuhan, R. P. Chine

(arrêtés ministériels du 5 juillet 1984 et du 30 mars 1992)

Spécialité: **Mathématiques Appliquées**

Assimilation de données variationnelle pour les problèmes de transport des sédiments en rivière

Date de soutenance: le 26 Novembre 1999

Composition du Jury:

M. Jacques BLUM	Professeur	Président
M. Olivier MONGA	CR, HDR	Rapporteur
M. Bernard SARAMITO	Professeur	Rapporteur
M. François Xavier LE DIMET	Professeur	co-directeur de thèse
M. Chaojiang XU	Professeur	co-directeur de thèse

Thèse préparée au sein du laboratoire LMC (IMAG projet IDOPT)
et du Département de Mathématiques (UW)

Remerciements

Je tiens, tout d'abord, à exprimer ma reconnaissance et ma gratitude à M. François Xavier LE DIMET et M. Chaojiang XU qui ont su me proposer un sujet très enrichissant et ont assuré la direction de cette thèse. Leurs compétences mathématiques, les soutiens et encouragements qu'ils m'ont accordés, leurs conseils qui ont guidé mes recherches au cours de ces années de thèse m'ont permis de mener à bien ces travaux.

Je remercie M. Oliver MONGA et M. Bernard SARAMITO pour l'attention et l'intérêt qu'ils ont porté à mon travail et pour m'avoir fait l'honneur de rapporter sur le manuscrit.

Je témoigner ma sincère reconnaissance à M. Jacques BLUM qui m'a encouragée moralement et matériellement, avec gentillesse, compétence et m'a fait l'honneur de présider le jury.

Je tiens à témoigner ma gratitude à messieurs Patrick WITOMSKI et Jean Claude PAUMIER pour m'avoir accueilli au LMC, dans l'équipe EDP où cette thèse a été réalisée.

Une partie de cette thèse a été fait pendant un stage à l'Institut de Physique de l'Atmosphère (IAP) de l'Académie de Science (Pekin) et il a bénéficié de l'aide de plusieurs membres du laboratoire LASG pendant leurs séjours en France comme M. Qingcun ZENG, M. Dongjian GUO et M. Jiang ZHU, vous trouvez ici mes remerciements.

Je remercie vivement l'ensemble des personnels du département de mathématiques de l'Université de Wuhan, de l'IAP, et du LMC, pour leurs précieux conseils, leur soutien actif et les nombreux services qu'ils m'ont rendus et, avec qui j'ai eu le plaisir de partager de très bons moments pendant ces quatre années de thèse.

Ce mémoire a été réalisé dans le cadre de la coopération Sino-Français, il a bénéficié de l'aide de l'Ambassade de France à Pekin et particulièrement de celle de M. Jean-Paul REBOUILLART que je prie d'accepter mes remerciements.

Enfin, je m'adresse mes sincères remerciements à tous ceux qui ont contribué à rendre l'environnement de cette thèse agréable.

Je ne saurais clore ces remerciements sans en adresser à mon père, ma mère et à ma sœur, qui m'ont rapporté tout leur amour et m'ont encouragée constamment dans les moments difficiles tout au long de ces années de thèse. Je n'oublierai pas ceux qui m'ont soutenue, et aidée de très près, je pense à mes chers amis Chaoyueh, Meiling.

Table des matières

Introduction générale	11
I Introduction à la modélisation du transport des sédiments	15
1.0.1 Introduction	15
1.1 Lois physiques de la sédimentation	17
1.1.1 Deux phases: liquide-solide	17
1.1.2 Échelle de temps	17
1.2 Modèle unidimensionnel du transport des sédiments	18
1.2.1 Description d'une rivière schématisée	18
1.2.2 Équations fondamentales	19
1.3 Equations semi-empiriques	21
1.3.1 Le frottement	21
1.3.2 Loi de chargement	23
1.4 Méthodes numériques	25
II Modèle bidimensionnel de la sédimentation	29
2.1 Description du modèle choisi	29
2.1.1 Equations de Saint-Venant	30
2.1.2 Transport solide	32
2.1.3 Formules semi-empiriques pour le modèle bidimensionnel	33
2.2 Conditions aux limites	35
2.3 L'énergie totale	37
2.4 Changement de la forme du modèle	40
2.4.1 Transformation des variables pour les équations de Saint-Venant	40
2.4.2 Propriétés de conservation des équations de Saint-Venant	44
2.5 Conclusion	46
III Discrétisation du modèle	47
3.1 Méthodes numériques pour résoudre l'équation d'évolution	47
3.1.1 Notations	48

3.1.2	Résultat de stabilité	49
3.1.3	Résultat de convergence	52
3.2	Schémas à pas fractionnaires	54
3.2.1	Résultat de la convergence au cas linéaire	55
3.2.2	Etude de la stabilité des schémas (3.17)	56
3.3	Méthode de prédiction-corrrection	58
3.4	Caractères du modèle direct discret	59
3.4.1	Choix du maillage	59
3.4.2	Discrétisation spatiale	61
3.4.3	Lemmes utiles	64
3.4.4	Discrétisation temporelle	72
3.5	Résolution numérique du problème	74
3.5.1	Application de la méthode de “time-splitting”	74
3.5.2	Description de l’expérimentation et des paramètres	77
3.5.3	Algorithme de la simulation	80
3.5.4	Résultats numériques	81
3.5.5	Conclusion	86
IV	Problèmes d’assimilation de données	89
4.1	Développement du modèle adjoint	89
4.1.1	Introduction	89
4.1.2	Méthodes adjointes	90
4.1.3	Notations	92
4.1.4	Système d’optimalité	93
4.1.5	Méthode de transposition	95
4.1.6	L’algorithme de minimisation	97
4.1.7	Critère d’arrêt de l’algorithme de minimisation	100
4.2	Assimilation de données I – Contrôle des conditions initiales	100
4.2.1	Forme continue du système adjoint	101
4.2.2	Système adjoint des équations de Saint-Venant alternées	105
4.2.3	Pseudo-observations et les expériences jumelles	110
4.2.4	Test de l’exactitude du gradient	111
4.2.5	Expérience numérique	113
4.3	Assimilation de données II – Identification des paramètres	119
4.3.1	Introduction	119
4.3.2	Système de contrôle	120
4.3.3	Choix des paramètres à identifier	122
4.3.4	Validation du gradient	123

4.3.5	Expériences numériques - Identification avec des observations complètes	123
4.3.6	Expériences numériques - Identification à partir des données non complètes	129
4.3.7	Méthode de la pénalisation	135
4.4	Conclusion	141
V	Estimation des erreurs du modèle	143
5.1	Introduction	143
5.2	Détermination des erreurs dans la base canonique	144
5.2.1	Système de contrôle I	144
5.2.2	Un modèle d'application	146
5.2.3	Expériences numériques	147
5.3	Les méthodes de vecteur optimal	155
5.3.1	L'opérateur linéaire tangent	156
5.3.2	Les méthodes de calcul des vecteurs propres dominants	158
5.3.3	La méthode de la puissance itérée	158
5.3.4	La méthode de Lanczos	160
5.3.5	Résultats numériques	162
5.3.6	Critère du choix de la base d'espace composée des vecteurs propres	168
5.3.7	Système de contrôle II	171
5.3.8	Expériences numériques avec réduction de la taille de l'espace de contrôle	172
5.3.9	Conclusion	177
	Conclusion générale	193

Introduction générale

Le transport des matériaux en rivière est l'un des phénomènes physiques les plus importants et les plus compliqués de l'hydraulique des rivières alluvionnaires ; c'est aussi un aspect capital de l'hydraulique fluviale. Ce transport et la sédimentation jouent un rôle fondamental sur le changement de l'environnement naturel comme l'irrigation, l'évolution du lit d'une rivière, la formation et l'évolution d'un delta, l'expansion de la plaine alluviale, autant d'éléments qui doivent être pris en compte dans les projets de travaux hydrauliques ou de navigation.

Pour mieux prévoir les résultats de l'évolution de la sédimentation, il est nécessaire de simuler la circulation des courants fluviaux, le transport et la sédimentation des matériaux. En revanche, du point de vue physique et mathématique, la simulation de l'évolution de la rivière est une tâche très délicate car l'écoulement et le transport des sédiments sont dans un régime non-permanent, non-uniforme et non-équilibré. Soumise aux interventions d'événements naturels (crues exceptionnelles) ou humains (construction d'un barrage ou aménagement), l'évolution du lit d'une rivière est, en effet, le résultat d'interactions à long terme entre l'écoulement et le lit. Le temps nécessaire à une rivière pour atteindre un état relativement équilibré est très long, la simulation, par un modèle physique, en vue de la prévision, est très coûteuse si on veut la réaliser en laboratoire. C'est pour cela que l'emploi d'un modèle numérique est souhaitable. Actuellement les progrès dans le domaine des performances des méthodes numériques permettent un calcul plus précis et plus stable durant une longue période d'intégration numérique.

Les sédiments transportés par l'écoulement ou les matériaux constituant le lit sont naturellement non-uniformes. Le transport de matériaux aux des diamètres différents se manifeste principalement sous deux modes : charriage et suspension. Ces deux types de transport, ajoutés au mouvement liquide, augmentent le degré de complexité du phénomène physique d'évolution du lit. De plus, les deux transports sont en état de déséquilibre, dépendant étroitement des conditions de la phase liquide et des caractéristiques des sédiments. Un modèle de simulation numérique qui résout les équations décrivant les mouvements du fluide et de matériaux est toujours insuffisant pour simuler l'évolution

d'une rivière et pour faire une bonne prévision à cause des simplifications employées :

Par exemple, on impose dans les modèles :

- une condition d'équilibre instantané entre transport potentiel et transport effectif ;
- des sédiments uniformes (un diamètre représentatif des matériaux du lit est choisi pour représenter la granulométrie qui n'est jamais uniforme en réalité) ;
- le mode de transport est unique (par charriage ou par suspension).

Et il manque aussi des connaissances des mécanisme de transport des sédiments en suspension et des paramètres physiques dans le processus d'érosion et de dépôt : la capacité potentielle de transport des sédiments, la condition de démarrage d'une particule immobile du lit surtout en granulométrie étendue, mais aussi les échanges entre les sédiments en suspension et charriés ou les matériaux du lit.

Ces derniers paramètres sont difficiles à évaluer non seulement par mesure expérimentale en laboratoire mais aussi par l'observation. Ils sont souvent décrits par le développement d'une formule analytique empirique. Parmi les paramètres inconnus, l'estimation de la condition initiale et les conditions aux limites du système dynamique évolutif sont difficiles à déterminer.

L'objet de cette thèse est de déterminer les paramètres inconnus et d'améliorer des reconstructions numériques d'un modèle de transport de sédiments en introduisant les données d'observation afin de produire une simulation plus conforme à la réalité. Il s'agit d'étudier la faisabilité de méthodes numériques d'optimisation de type contrôle optimal appliquées aux problèmes d'assimilation des données de transport de sédiments en rivière.

L'assimilation de données variationnelles est une technique dont le but est de déterminer des paramètres inconnus dans un modèle en combinant des informations dans l'observation et le modèle dynamique afin que les paramètres identifiés permettent les comportements du système de s'accorder avec les observations réelles. Le principe de la méthode variationnelle basée sur la théorie du contrôle optimale est d'introduire une fonctionnelle dépendant des paramètres de contrôle et de la minimiser en cherchant des valeurs optimales de ceux-ci dans un espace de contrôle admissible. Dans la plupart des cas, la fonctionnelle (fonction coût) est choisie comme étant l'écart quadratique entre le champs analysé (ou bien la prévision du modèle) et les données observées. Les valeurs optimales des paramètres inconnus de cette fonction coût peuvent être trouvées par un algorithme de minimisation sans contrainte. Pour y parvenir, on a besoin du gradient de la fonctionnelle par rapport à un nombre – qui peut-être très grand – de variables. La méthode classique (par différences finies) de la détermination du gradient n'est pas efficace pour un problème de grande taille à cause du coût de calcul. La méthode adjointe

est plus pratique et plus efficace. Ceci nécessite le développement d'un modèle adjoint qui est le transposé du système linéaire tangent du modèle de simulation.

Beaucoup de techniques d'assimilation de données sont développées en météorologie et en océanographie. Des bibliographies peuvent être trouvées dans Navon [47], Le Dimet et Navon [39], Ghil et Malanotte-Rizzoli [26] et Daley [15]. Ces techniques sont basées sur la théorie du contrôle optimal due à Lions [41] et pour une application générale en météorologie, on peut citer LeDimet et Talagrand [38]. Des succès importants ont été remportés pour des modèles de la prévision du temps en météorologie, par exemple, Derber [17], Lewis et Derber [40], Courtier et Talagrand [10], Talagrand et Courtier [58], et dans les modèles de circulation océanique, comme Moore [46], Thacker et Long [61], Harland et O'Brien [32], Benett [2] etc. Au vu de ces succès, l'attention s'est alors portée à l'application dans les modèles hydrologiques.

Mais les techniques d'assimilation de données sont difficiles à appliquer à un modèle de simulation complet et les données d'observation réelles sont peu nombreuses surtout pour les modèles de transport. Il en résulte donc très peu d'applications. Cette difficulté est liée, d'une part, en pratique, à la densité insuffisante des observations en temps et en espace et aux observations bruitées ou inexactes, d'autre part à l'instabilité numérique de la technique difficile à maîtriser lorsqu'on doit l'appliquer aux modèles non-linéaires. Il s'agit de l'impact des observations incomplètes sur l'existence et l'unicité de la solution et le taux de convergence du processus de minimisation (cf. Zou, Navon et al. [48], [77]). Il faut donc d'abord tester la faisabilité de l'approche de l'assimilation de données variationnelles avant que les données réelles soient introduites. Notre travail a utilisé pour cet objectif un modèle simplifié concernant plus particulièrement le transport solide uniforme dans le cadre bidimensionnel et nous avons utilisé des observations simulées.

Nous donnons dans le chapitre I une brève présentation de la modélisation générale dans le cadre unidimensionnel. Il s'agit d'une connaissance nécessaire des lois physiques du transport et de la sédimentation en rivière et de généralité sur la formulation du modèle mathématique satisfaisant ces lois. Après avoir introduit les équations de base généralement utilisées dans les modèles unidimensionnels décrivant le transport des sédiments en rivière et la variation du fond de la rivière, on rappelle quelques formules semi-empiriques déterminant des variables physiques dans un modèle unidimensionnel afin de compléter les équations de base, ainsi que les méthodes de la résolution numérique.

Le chapitre II se limite à la mise en équation du problème étudié dans le cadre bidimensionnel d'évolution du lit soumis au transport complet par charriage et suspension. Nous utilisons, en fait, le modèle développé par Zeng, Guo et al. [73], qui est un couplage

du transport en liquide et en solide. le transport liquide est une version bidimensionnelle des équations de Saint-Venant. La fermeture du système est effectuée par l'introduction de formules empiriques. Un changement de variables donne une preuve de la conservation de la masse et de l'énergie du fluide dans la couche liquide.

Le chapitre III se concentre sur l'analyse numérique de la stabilité du schéma aux différences finies choisi et est consacré à la mise au point de la résolution numérique du transport des matériaux convectés par l'écoulement. On a établi la stabilité du schéma à pas fractionnaires sous une forme générale caractérisée par un paramètre. On parvient à un algorithme de simulation de type "time-splitting" en vue d'une économie du coût de calcul. On montre également la conservation de l'énergie des équations aux différences de Saint-Venant discrétisées en espace par un maillage alterné. Comme illustration, on fournit des résultats numériques mis en oeuvre dans le cas d'une modélisation de la formation d'un delta dans une rivière.

Le chapitre IV est destiné à la description et à la construction du code de l'adjoint du modèle de simulation et est consacré à l'application au problème d'assimilation de données au transport des matériaux en rivière. On a testé la faisabilité de cette méthode. La validation est testée par la détermination des conditions initiales des champs d'écoulement et du fond du lit. Des résultats encourageants montrent que l'état initial du fond a un rôle dominant pour l'évolution de la rivière. Nous utilisons également ce code adjoint pour l'identification de paramètres empiriquement déterminés par une régression linéaire. Les résultats obtenus montrent l'intérêt de la méthode et les points cruciaux de sa mise en oeuvre.

La réduction de la taille du problème de contrôle optimal est étudiée dans le chapitre V pour le problème d'estimation des erreurs du modèle. L'utilisation des vecteurs singuliers associés aux plus grandes valeurs singulières de la matrice résolvante du système linéaire tangent donnent une idée de reconstituer une nouvelle base d'espace. Ces vecteurs sont déterminés grâce à l'algorithme de Lanczos par l'application itérative des versions linéaire tangente et adjointe du modèle (Laccarra & Talagrand [35], Buizza et al. [5]). Ces vecteurs représentent les directions dans lesquelles les perturbations s'amplifient le plus vite.

Les discussions sur les résultats obtenus et perspectives sont présentées dans la conclusion générale.

Chapitre I

Introduction à la modélisation du transport des sédiments

1.0.1 Introduction

Le transport des sédiments est un aspect fondamental de l'hydraulique des rivières alluviales. Le lit de ces rivières est en effet dans un état d'équilibre asymptotique dépendant étroitement de deux phases transportées, liquide et solide. Cette notion d'équilibre n'a toutefois de sens qu'associée à une certaine échelle de temps. A l'échelle des temps géologiques, il ne peut y avoir équilibre, la planète étant en perpétuelle évolution. En revanche, à l'échelle du temps humain, hormis quelques fluctuations liées au caractère aléatoire des crues, le lit des rivières laissées à l'état naturel est en équilibre.

Les divers aménagements réalisés sur les rivières remettent en cause cet équilibre. Par exemple, les barrages piègent les sédiments transportés, générant un remous solide en amont, tandis qu'une érosion du lit se produit en aval. D'autre part, des mesures très coûteuses doivent être prises pour fixer le lit de la rivière et protéger les ouvrages existants, par exemple en renforçant les berges ou en implantant des seuils au fond de la rivière. Entre deux seuils peut s'établir à terme un nouvel état d'équilibre du bief.

Les phénomènes physiques mis en jeu sont particulièrement complexes et les diverses interventions sur la rivière ont des effets couplés. Les mesures compensatoires permettant de maîtriser l'évolution de la rivière étant particulièrement coûteuses, il est donc nécessaire de chercher à se doter des moyens qui permettent de déterminer les impacts prévisibles d'actions déjà entreprises, en cours ou en projets sur la rivière. Il est en particulier important de pouvoir déterminer quel sera le nouvel équilibre à terme de la rivière suite à un certain nombre d'actions et quelle sera l'échelle de temps de l'évolution du lit.

Il est également de la plus grande importance de chercher à déterminer la durée de

vie d'une retenue, à savoir, l'échelle de temps du comblement de la retenue par les sédiments. Des mesures compensatoires ou limitantes peuvent être envisagées, il importe d'en apprécier l'incidence.

Les moyens nécessaires à l'appréciation de l'incidence de diverses actions sur l'évolution morphodynamique d'une rivière passe d'abord par une bonne connaissance de celle-ci. Ces connaissances générales de la rivière peuvent être prolongées par une approche simple du problème, qui est basée sur les équations classiques de l'hydraulique fluviale appliquées à des sections simplifiées en prenant globalement en compte les phénomènes principaux de la dynamique fluviale.

L'étude de l'évolution des fonds nécessite dans certains cas le recours à la modélisation physique, en particulier pour les problèmes locaux où la prise en compte de la bidimensionnalité ou tridimensionnalité de l'écoulement est nécessaire (par exemple pour la détermination de l'affouillement généré par les piles d'un pont, pour l'évolution d'un delta, l'expansion de la plaine alluviale....). Les modèles numériques déterministes sont en revanche beaucoup mieux adaptés pour résoudre les problèmes à grande échelle d'espace et de temps, d'autant qu'il est alors possible de considérer la rivière comme filaire et d'effectuer une modélisation unidimensionnelle. Les premiers modèles ont été élaborés par De Vries [18], Cunge et Perdreau [14], qui considéraient une granulométrie uniforme et résolvaient les équations de bilan des phases liquides et solides de façon découplée. Ultérieurement des modèles ont été mis au point, qui résolvent ces équations de façon couplée. Plus récemment, des modèles qui prennent en compte l'étendue de la courbe granulométrique ont été élaborés. Il existe de nombreux modèles développés en laboratoire dans un cadre plus général s'appliquant à des cas réels de rivières, mais ils sont toujours limités au transport solide par charriage dans un bief de rivière du fait de la complexité des conditions naturelles.

Dans la suite, nous présenterons le phénomène physique de la sédimentation afin d'introduire les connaissances élémentaires permettant d'établir un modèle mathématique. Après avoir introduit les équations de base généralement utilisées dans les modèles unidimensionnels décrivant le transport des sédiments en rivière et la variation du fond de la rivière, on rappelle quelques formules semi-empiriques déterminant des variables physiques afin de compléter les équations de base. Dans le chapitre suivant, on étudie le système de transport des matériaux dans le cadre bidimensionnel, ainsi que la fermeture du système est effectuée par l'introduction des formules semi-empiriques appliquées dans le cas bidimensionnel.

1.1 Lois physiques de la sédimentation

Il faut naturellement introduire dans un modèle un certain nombre de lois qui serviront d'hypothèses de base susceptibles d'être modifiées par la suite.

1.1.1 Deux phases : liquide-solide

Au cours de son cheminement, la rivière se charge en sédiments divers de sorte que la rivière transporte non seulement de l'eau mais également des graviers, des sables, des limons... Ces sédiments transportés par la rivière tapissent également le fond du lit. Aussi, faut-il considérer à la fois l'écoulement liquide et les sédiments qu'il transporte et les traiter comme un fluide diphasique.

La turbulence de l'écoulement est le moteur du transport solide. Sur le fond du lit, elle est responsable du départ des particules qui se produit lorsque les fluctuations de la force tractrice dépassent un certain seuil. Ces particules peuvent être ensuite charriées par l'écoulement. Elles se déplacent alors en glissant, en roulant ou par petits sauts. Elles peuvent également être entraînées au sein de l'écoulement et pourront y être maintenues en suspension grâce à la turbulence qui contrebalance les effets de gravité. De l'énergie et de la masse sont échangées entre ces deux phases qui interagissent. Il en résulte un problème couplé.

La vitesse de l'écoulement peut être à la fois influencée par les variations de profondeur de l'eau (qui est la conséquence de la variation du lit fluvial) et aussi par le frottement sur le fond qui dépend de la rugosité, de la composition et du diamètre des particules.

1.1.2 Échelle de temps

Le transport des matériaux suspendus au sein de l'écoulement pose un problème difficile car ce transport convecte avec une vitesse caractéristique égale à celle de l'écoulement, mais beaucoup plus vite que celle du charriage et de l'évolution du fond. Donc on peut considérer deux échelles différentes de temps en ce qui concerne l'évolution globale du système :

- une échelle de temps courte en ce qui concerne l'écoulement, l'apparition et la disparition de rides au fond de la rivière ;
- une échelle de temps généralement beaucoup plus grande en ce qui concerne l'évolution de la forme du lit.

Du fait de la turbulence, l'interaction locale liquide-solide a un caractère aléatoire. La mise en mouvement des particules doit probablement être associée au phénomène

turbulent appelé rafale qui, localement, a pour incidence le départ simultané d'un paquet de particules. Des échanges entre les matériaux du lit et la couche de charriage sont possibles, de même qu'entre la couche de charriage et la couche de suspension. Les échanges entre les sédiments suspendus et le lit sont supposés s'effectuer nécessairement à travers cette couche de charriage.

1.2 Modèle unidimensionnel du transport des sédiments

1.2.1 Description d'une rivière schématisée

Le fond réel de la rivière possédant une géométrie complexe et non-homogène, il est difficile de le représenter par une variable continue par rapport au temps et à l'espace. Avant d'introduire les équations de base générales, il nous faut alors en faire une représentation schématique afin d'assurer certaines propriétés mathématiques comme par exemple, la continuité.

Un tronçon filiaire de la rivière, tel que schématisé dans la figure I.1, est décomposé en trois couches :

- La Couche I d'*écoulement liquide* est principalement occupée par de l'eau et des sédiments en *suspension* ou par des substances dissoutes. La vitesse des sédiments est approximativement la vitesse moyenne d'écoulement. Selon le degré de saturation de l'écoulement, on peut distinguer un écoulement en régime non-saturé ou un écoulement en régime saturé.
- La Couche II de *charriage* contient les sédiments charriés. Les particules se déplacent sous l'effet de l'écoulement possédant un mouvement de glissement, de roulement ou de saut à proximité du fond.
- La couche III d'*immobilité* représente le fond de la rivière, lequel comprend le lit à partir d'une côte de référence. Les sédiments dans cette couche ne satisfont pas à la condition critique de début de transport. Cette couche est constituée des sédiments en granulométrie étendue, qui peut aussi être divisée en sous-couches pour mettre la simulation de l'évolution de la composition du fond.

On représente le fond rugueux par une courbe lisse ayant une certaine pente afin d'assurer l'écoulement. Celle-ci représente la surface moyenne du lit. Dans la couche III se trouve la variation du fond.

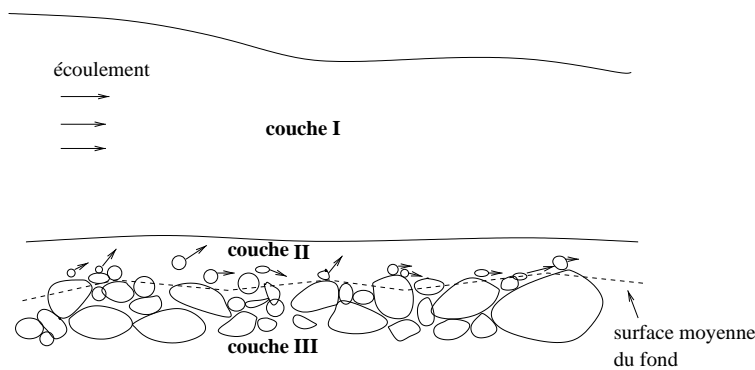


FIG. I.1 – Description de la rivière schématisée

1.2.2 Équations fondamentales

Comme dans toute modélisation, il s'agit de prendre en compte les caractères essentiels du problème physique. Ici, il s'agit de reproduire au mieux l'écoulement liquide, le transport solide et l'évolution du fond de la rivière. Pour cela, la conservation de la masse et de la quantité de mouvement ou de l'énergie sont utilisées pour déduire les équations de base du mouvement de l'écoulement liquide transportant des sédiments.

Ici, on considère le cas d'une granulométrie uniforme (les sédiments ont tous la même taille). Dans ce cas, les lois de conservation suivantes doivent être considérées :

- conservation de l'écoulement liquide ;
- conservation des sédiments ;
- conservation de la quantité de mouvement.

Il est aussi possible de considérer une granulométrie non constante. On discrétise alors la fonction de répartition en classes étendues, mais dans ce cas il faut ajouter à ce système d'équations les lois de conservation de chaque classe granulométrique.

Dans notre cas, la granulométrie du lit est donc caractérisée par une seule valeur de diamètre des sédiments. Les équations d'évolution dans le temps et dans l'espace du fond des rivières comprennent cinq équations : trois correspondent aux lois de conservation et deux expriment les relations entre

- Le débit solide d'équilibre et les paramètres hydrauliques,

- la dissipation par frottement et ces mêmes paramètres hydrauliques.

On décrit ci-dessous les équations unidimensionnelles établies pour le cas simplifié d'un écoulement en canal rectangulaire. Les équations adaptées au cadre bidimensionnel seront discutées au chapitre suivant.

1. Equation de conservation liquide (les phénomènes tels que la pluie peuvent être ajoutés comme terme source)

$$\frac{\partial h}{\partial t} + u \frac{\partial h}{\partial x} + h \frac{\partial u}{\partial x} = p - i; \quad (1.1)$$

2. Equation de la quantité de mouvement

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + g \frac{\partial h}{\partial x} + g \frac{\partial z}{\partial x} + g S_f = 0; \quad (1.2)$$

3. Equation de conservation solide

$$\frac{\partial z}{\partial t} + \frac{1}{1-p} \frac{\partial g_s}{\partial x} = 0; \quad (1.3)$$

4. Formulation du transport solide

$$f_1(g_s, u, h) = 0; \quad (1.4)$$

5. Dissipation par frottement

$$f_2(S_f, u, h) = 0; \quad (1.5)$$

où :

$p - i$ décrit les termes sources de la pluie et l'infiltration, h est la hauteur d'eau,

u est la vitesse de l'écoulement,

z est la cote du fond,

g_s est le transport solide d'équilibre (charriage) par unité de largeur,

S_f est la pente d'énergie,

g est l'accélération de la pesanteur,

p est la porosité du matériau au fond du lit,

f_1, f_2 sont les relations semi-empiriques.

On distingue deux sortes d'équations : les équations (1.1) - (1.3) sont déduites mathématiquement et sont exprimées par des équations aux dérivées partielles, et les équations semi-empiriques (1.4) et (1.5) traduisent une réalité physique complexe et sont déduites d'expérimentation en laboratoire. Ces équations semi-empiriques peuvent être algébriques

ou aux dérivées partielles. La variété des modèles de transport des sédiments s'explique par un grand nombre de ces formules semi-empiriques.

Dans un cadre unidimensionnel, même si on sait qu'il n'existe pas un mouvement unidimensionnel dans la réalité à cause de la géométrie complexe de la rivière, la modélisation est faite en partant de la supposition que toutes les variables physiques de l'écoulement et des sédiments sont homogènes dans la section. Cette modélisation sera adaptée à la résolution des problèmes à grande échelle d'espace et de temps.

1.3 Equations semi-empiriques

1.3.1 Le frottement

Le frottement auquel l'écoulement est soumis dans les rivières alluviales est dû à plusieurs causes : rugosité des parois, rides du sable, force extérieure des constructions artificielles, etc. La dissipation de l'énergie mécanique de l'écoulement est la conséquence de la synthèse de ces frottements ainsi que de la variation des éléments hydrauliques, comme le débit, la hauteur de l'eau et du lit, etc.

La formulation la plus générale du frottement peut s'écrire :

$$f(S_f, u, h, \dots) = 0,$$

où le frottement S_f (ou la pente d'énergie) est une fonction implicite des autres variables du problème. A l'heure actuelle, il reste encore des difficultés pour déterminer les relations entre les variables, cependant, plusieurs formules semi-empiriques ont été établies à partir d'essais en canaux de laboratoire ou en conduites ayant généralement des parois fixes et dans les conditions de régime permanent uniforme. Ces formules sont utilisées pour la simulation des écoulements à fond fixe graduellement variés et donnent des résultats corrects.

La difficulté avec les écoulements à fond mobile réside dans la variabilité de la rugosité du lit de la rivière. Cette variabilité est essentiellement due à deux phénomènes différents qui peuvent d'ailleurs se combiner :

(1) la déformation du fond avec le possible développement de dunes à certains régimes et

(2) l'évolution de la rugosité en fonction des changements de composition granulométrique de la couche de fond au contact direct avec l'écoulement.

La variabilité liée aux ondulations du lit est surtout importante dans les rivières dont le fond est sablonneux. Notant cependant qu'un lit plat n'est qu'approximativement plat car il existe toujours des instabilités tridimensionnelles dans un écoulement. Les rides et les dunes apparaissent à certains régimes subcritiques et aux régimes super-critiques. Il existe donc un fort couplage physique entre la phase solide et la phase liquide à travers la rugosité des fonds. Cet argument plaide en faveur de la résolution couplée des équations des phases liquides et solides.

Chollet [8] tout en considérant un débit constant résout de façon couplée les équations de l'énergie, de la continuité liquide, du débit solide et une équation spécifique pour le frottement qui tient compte de la variabilité des formes du lit, laquelle est basée soit sur la théorie d'Engelund [24], soit sur la théorie d'Einstein [23].

Actuellement, les formules semi-empiriques suivantes sont souvent utilisées à déterminer la pente d'énergie S_f :

- La formule de STRICKLER

$$S_f = \frac{Q^2}{K_s^2 A^2 R^{4/3}}, \quad (1.6)$$

où :

Q est le débit du fluide ;

K_s est le coefficient de STRICKLER ;

A est la surface de la section ;

R est le rayon hydraulique de la section.

- La formule de MANNING

$$S_f = \frac{n^2 Q^2}{A^2 R^{4/3}}, \quad (1.7)$$

où n est le coefficient de MANNING.

On voit que ces deux formules sont équivalentes lorsque les coefficients ont entre eux la relation :

$$K_s = \frac{1}{n},$$

et qu'elles ont les paramètres nécessaires pour représenter globalement une dissipation d'énergie en fonction des variables de l'écoulement, du moins pour des écoulements hydrauliquement rugueux. Le coefficient K_s caractérise une rugosité de paroi et le rayon hydraulique R permet une certaine prise en compte de la forme de la section.

1.3.2 Loi de chargement

Par définition, la capacité de transport de l'écoulement correspond au débit solide à l'équilibre. Quand le lit de la rivière se trouve dans un état d'équilibre, il n'y a ni dépôt ni érosion, et dans le cas de transport en suspension, ceci correspond à une concentration de saturation critique telle que si cette valeur est dépassée (super-saturation) il y a dépôt et dans le cas contraire (sub-saturation) érosion. Dans le cas du transport non-équilibré, le débit solide s'écarte de la valeur à l'équilibre et la loi de chargement a pour le but de préciser cet écart en le reliant à la capacité de transport. Parallèlement, l'équation de continuité lie cette différence à la déformation du lit. Donc, si un écoulement est supposé en régime saturé, la capacité de transport des sédiments peut être exprimée dans l'équation continue, la différence entre le débit des sédiments venant de l'amont et celui des sédiments se déposant au fond de la rivière.

Puisque les moteurs du mouvement de ces deux sortes de sédiments ne fonctionnent pas de la même manière, et puisque les énergies fournies par l'écoulement pour maintenir ces deux mouvements ne sont pas du même type, nous allons préciser séparément les lois de chargement pour la suspension et le charriage. De nombreuses formules semi-empiriques décrivent ces processus, nous nous attacherons ici à celles qui sont utilisées le plus souvent.

1.3.2.1 Capacité de transport en suspension

La formule suivante a été trouvée par l'Institut de l'énergie hydraulique et électrique de Wuhan (Chine) dans les années 1960 :

$$S^* = K \left(\frac{U^3}{gR\omega} \right)^m, \quad (1.8)$$

où :

U est la vitesse de l'écoulement (m/s),

R est le rayon hydraulique de la section (m),

ω est la vitesse de dépôt des sédiments (m/s),

g est l'accélération de la pesanteur (m/s^2),

K est le coefficient de rapport (kg/m^3),

m est une constante empirique.

Cette formule a déjà été utilisée dans de nombreux travaux hydrauliques.

1.3.2.2 La capacité du transport de charriage

Un grand nombre de problèmes fluviaux peuvent être résolus par une approche unidimensionnelle. Dans ce cadre, plusieurs chercheurs ont tenté d'établir des formules semi-empiriques de transport solide comme Meyer-Peter [45] et Einstein [23] qui sont calées

sur des données expérimentales obtenues en canal de laboratoire. Ces formules fournissent une relation univoque entre le débit liquide et le débit solide pour un diamètre des particules donné. Ce sont des relations d'équilibre en ce sens qu'elles ont été établies dans des conditions d'écoulements stationnaires et uniformes et ne sont valables en toute rigueur que pour ces conditions. Le transport solide calculé par ces formules est un transport solide d'équilibre.

La formule de Meyer-Peter et Muller s'écrit :

$$g_s = \sqrt{g \frac{\rho_s - \rho}{\rho}} d_m^3 \times 8 \left[\frac{Q_s}{Q} \left(\frac{K_m}{K_r} \right)^{3/2} \frac{h S_f}{\frac{\rho_s - \rho}{\rho} d_m} - 0,047 \right]^{3/2}, \quad (1.9)$$

où :

g_s est le débit volumique de sédiments charriés par unité de largeur (il s'agit uniquement du volume de la matière, les vides ne sont pas comptés) (m^3/s),

ρ_s est la masse volumique des sédiments (kg/m^3),

ρ est la masse volumique de l'eau (kg/m^3),

d_m est le diamètre moyen de la granulométrie des sédiments (m),

Q est le débit liquide (m^3/s),

Q_s est la portion du débit liquide efficace pour le charriage,

K_m est le coefficient de Strickler du lit ($m^{1/3}s^{-1}$),

K_r est le coefficient de Strickler relatif aux grains ($m^{1/3}s^{-1}$),

h est la hauteur d'eau,

S_f est la pente de l'énergie.

Cette formule d'équilibre a été établie à l'aide de granulométrie uniforme ou peu étendue caractérisée par un diamètre unique. Pour un diamètre donné des particules, le transport solide d'équilibre correspond à une saturation de la capacité de l'écoulement à transporter ces particules. Pour un autre diamètre de particules, le transport solide d'équilibre peut-être tout autre, éventuellement nul.

Mais les granulométries des rivières sont souvent étendues et le fond des rivières n'étant généralement pas homogène, il peut être nécessaire de considérer différentes couches granulométriques, chaque couche granulométrique ou strate de dépôt ayant sa propre granulométrie, caractérisée par les pourcentages de chacune des classes granulométrique. Les travaux se rapportant à ceci concernant peuvent être trouvés dans Rahuel [52] etc.

1.4 Méthodes numériques

Pour résoudre numériquement les équations de base qui gouvernent le mouvement de l'écoulement liquide du transport solide et de la variation du fond de la rivière, il existe trois méthodes possédant des avantages différents suivant les modèles. Ces méthodes ayant des propriétés différentes sont :

- la méthode des caractéristiques ;
- les méthodes d'éléments finis ;
- les méthodes aux différences finies.

1.4.0.3 Méthode des caractéristiques

La méthode des caractéristiques est très bien adaptée à l'analyse d'un système d'équations hyperboliques, et est donc beaucoup utilisée pour résoudre les équations de Saint-Venant unidimensionnelles. Cette méthode est à la base de plusieurs modèles d'écoulement à fond fixe. Elle consiste à calculer la solution à un instant donné en allant chercher l'information par le biais des caractéristiques le long desquelles se propage cette information. Il s'agit d'une méthode générale très performante et bien adaptée au traitement des discontinuités (Vila [66]). Un de ses avantages, pour traiter un modèle unidimensionnel est que l'on peut déterminer où l'on va imposer les conditions aux limites (en amont ou en aval). Cela assure que le problème étudié est bien posé.

Cependant, cette méthode devient rapidement complexe lorsque l'on cherche à l'appliquer à la modélisation des rivières à fond mobile (De Vries [18]).

1.4.0.4 Méthodes d'éléments finis

Les méthodes d'éléments finis se sont avérées particulièrement intéressantes dans la modélisation bidimensionnelle ou tridimensionnelle. Elles ont le gros avantage sur les méthodes aux différences finies d'être beaucoup plus souples d'emploi pour représenter les frontières à géométrie complexe.

En revanche, les résultats obtenus en modélisation unidimensionnelle des écoulements à fond fixe ont été, jusqu'à récemment, sujets à de sérieux problèmes d'oscillation parasites et de stabilité (Cooley et Moin [9]). En raison des difficultés rencontrées pour la modélisation des rivières à fond fixe, la méthode des éléments finis a été jusqu'à présent peu ou pas appliquée à la modélisation des rivières à fond mobile.

1.4.0.5 Méthodes aux différences finies

Dans le cadre des méthodes aux différences finies, des schémas centrés ou décentrés, implicites ou explicites, ont été mises au point aussi bien dans le domaine de l'hydrodynamique que dans celui de l'aérodynamique (Yee [72]). Les méthodes aux différences finies sont actuellement les plus utilisées pour résoudre les équations de base (par exemple, dans le cas unidimensionnel, ce sont les équations (1.1) - (1.5)). Différentes approches ont été tentées avec plus ou moins de succès.

Les méthodes les plus simples résolvent le système d'équations de façon découplée, explicitement et sous l'hypothèse simplificatrice d'un débit liquide constant le long du chenal. Cette approche a été utilisée dans des modèles plus connus tels HEC6 (Thomas et Prasuhn [62]) ou IALLUVIAL (Karim et Kennedy [34]). Elle a l'avantage d'être simple à mettre en oeuvre mais elle nécessite l'utilisation de petits pas de temps pour que la variation du fond de la rivière reste suffisamment faible d'un pas de temps à l'autre. Comme toutes les méthodes qui supposent un débit liquide constant par paliers, elle ne peut être appliquée qu'à des rivières dont la variation du débit d'un palier à l'autre n'est pas trop rapide.

D'autres modèles réalisent également une résolution découplée des équations liquide-solide en considérant un débit liquide variable. C'est le cas de CHARI (Perdreau et Cunge [51]). Dans un premier temps, les deux équations de Saint-Venant sont résolues de façon couplée en utilisant le schéma implicite de Preissmann. Dans une version généralisée et pour une équation hyperbolique de la forme :

$$\frac{\partial u}{\partial t} + \frac{\partial A(u)}{\partial x} = 0,$$

les dérivées spatiales et temporelles sont discrétisées selon :

$$\frac{\partial u}{\partial t} = \frac{\theta_1(u_{j+1}^{n+1} - u_{j+1}^n) + (1 - \theta_1)(u_j^{n+1} - u_j^n)}{\Delta t}$$

et

$$\frac{\partial A}{\partial x} = \frac{\theta_2(A_{j+1}^{n+1} - A_j^{n+1}) + (1 - \theta_2)(A_{j+1}^n - A_j^n)}{\Delta x}.$$

On obtient ainsi un schéma général du premier ordre. Le schéma de Preissmann proprement dit est obtenu pour $\theta_1 = 1/2$ et est du second ordre.

Dans un deuxième temps, l'équation de continuité solide est résolue explicitement en utilisant un schéma décentré :

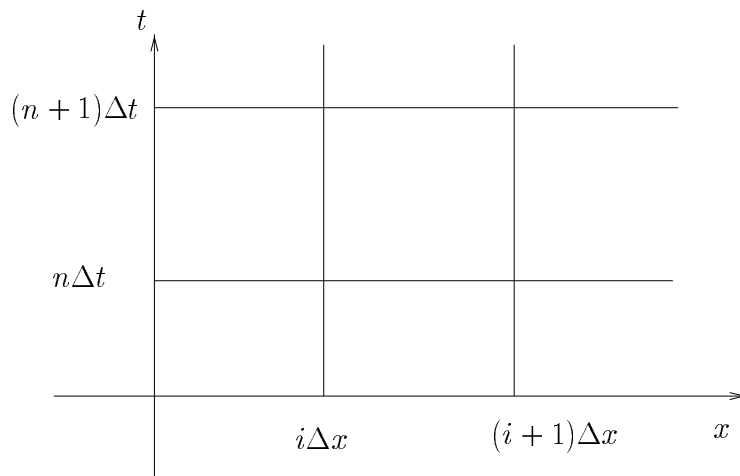


FIG. I.2 –

$$\frac{\partial z}{\partial t} = \frac{z_i^{n+1} - z_i^n}{\Delta t},$$

$$\frac{\partial g_s}{\partial x} = (g_s)_{i+1}^n - (g_s)_i^n.$$

Il en résulte une formule pour z_{i+1}^{n+1} permettant de calculer, de l'amont vers l'aval, l'évolution du lit de la rivière :

$$z_{i+1}^{n+1} = z_{i+1}^n + \frac{\Delta t}{(1-p)\Delta x} ((g_s)_{i+1}^n - (g_s)_i^n).$$

Le schéma étant explicite, sa stabilité numérique n'est que conditionnelle.

Cette approche du problème n'est valable que si les variations du débit liquide sont suffisamment lentes pour que l'on puisse considérer des paliers de débit. Pour les rivières dont la variabilité du débit est forte (comme le cas de crues), il n'y a de transport solide que lors des crues et celles-ci ne durent que quelques heures.

Pour modéliser le transport solide dans ces rivières au régime très variable, Cunge et Perdreau [13] et Lyn [1] utilisent le système d'équation complet de façon couplée en utilisant un schéma implicite de Preissmann à quatre points. Ce schéma est inconditionnellement stable et permet de choisir des pas de temps assez grands, ce qui évite des calculs trop coûteux.

Pour d'autres modèles, on peut citer Cunge et al. [12], dans lequel une analyse comparée de plusieurs modèles est présentée.

Chapitre II

Modèle bidimensionnel de la sédimentation

Contrairement à la modélisation en une dimension, où l'on tient en compte de l'évolution d'une rivière dans sa globalité, la mise en oeuvre de la modélisation dans le cadre bidimensionnel est avantageuse dans le traitement du problème de la sédimentation local sur un tronçon de rivière, comme le réservoir, l'estuaire, la baie, la sortie d'eau, etc. Dans ce contexte, un modèle numérique développé par Zeng [73] en géométrie bidimensionnelle sera introduit dans ce chapitre.

Après une description brève de ce modèle, nous présentons également une forme alternative des équations de Saint-Venant gouvernant le système dynamique par un changement de variables. Nous allons montrer sous cette forme, que notre système possède, sous certaines conditions, de bonnes caractéristiques importantes telles que la conservation de la masse et la conservation de l'énergie, lesquelles nous devons garder après la discrétisation.

2.1 Description du modèle choisi

Le transport solide en eau douce fait référence à un milieu hétérogène que l'on peut diviser en trois parties : le lit, constitué de grains compacts et considérés ici sans cohésion les uns avec les autres, l'écoulement de l'eau situé au-dessus de celui-ci et une couche de grains transportés dans la partie inférieure à la couche fluide. Une telle hétérogénéité combinée avec des échanges de matière et d'énergie spécifique entre les trois milieux suggère un modèle diphasique réalisant le couplage entre les phases liquide et granulaire (ou solide).

Comme nous allons le voir, nous disposons pour la phase liquide d'un modèle ma-

thématique reposant sur les principes de conservation de la masse et de la quantité de mouvement : le système de Saint-Venant, système communément utilisé en hydrologie et qui a fait ses preuves. Un tel modèle devra toutefois être adapté pour tenir compte de la concentration granulaire dans l'écoulement et de la rugosité du lit.

Le transport solide, cependant, ne présente pas les mêmes caractéristiques. Les modèles disponibles reposent en partie seulement sur les principes de conservation (essentiellement conservation de la masse), on utilise par ailleurs des lois empiriques (lois de débit solide ou de mise en charge) englobant différents mécanismes. L'objectif de ce chapitre est donc de s'appuyer le plus possible sur les principes de conservation en vue de réduire cette partie empirique et amener le modèle relatif à la phase solide au niveau de description de l'hydraulique. La phase solide sera décrite par la variable "concentration granulaire" (ou plus précisément fraction volumique solide) moyennée sur la hauteur de la couche fluide, symétriquement à la phase liquide décrit par les variables hauteur d'eau - vitesse du fluide. Une telle approche permettra notamment de mettre en évidence les effets d'inertie du transport solide et l'influence de la concentration solide de la couche de transport sur l'écoulement.

Un modèle bidimensionnel horizontal (sur le plan $x-y$) décrit généralement une valeur moyenne de toutes les variables physiques de l'écoulement et des sédiments intégrées sur la hauteur de l'eau. Il décrit également la variation du fond de la rivière au cours de l'écoulement et le long de la direction transverse. Parmi le grand nombre de modèles existants, on choisit un modèle développé par Zeng, Guo et al. ([73]) qui a déjà été appliqué à certains travaux et qui s'est révélé effectif. Ce modèle est un couplage des trois processus suivants :

- les équations hydrodynamiques régissant l'écoulement fluvial ;
- le transport des sédiments en suspension dans le fleuve ;
- la variation du fond de la rivière du fait d'érosion et dépôt.

2.1.1 Equations de Saint-Venant

Un écoulement à surface libre est un milieu tridimensionnel délimité par une surface rigide, le fond ou surface du lit et la surface libre séparant le liquide de l'air. Moyennant des conditions aux limites adaptées à ces deux types de frontières, un système d'équations de Navier-Stokes tridimensionnel est tout à fait envisageable pour le calcul d'un tel écoulement. Dans ce cas, il faut bien entendu s'attendre à des temps de calcul importants. D'autre part, dans certaines applications, il n'est pas toujours indispensable de connaître le profil de vitesse suivant la profondeur.

Les équations aux dérivées partielles gouvernant un fluide peu profond peuvent être déduites d'une forme limitée des équations générales de la mécanique des fluides lorsque la profondeur de l'eau est assez petite par rapport à la longueur d'onde d'une onde de surface typique. L'intégration de ces équations (en utilisant l'hypothèse hydrostatique) sur la hauteur (suivant l'axe vertical) est alors intéressante : elle supprime l'équation de quantité de mouvement suivant la direction verticale Oz , la composante de la vitesse verticale associée à la surface libre. Le vecteur vitesse ne possède donc plus que deux composantes horizontales dans la direction d'écoulement et de transverse caractérisant la couche fluide.

En tenant compte de la viscosité turbulente du fluide, le système d'équations obtenu est le système de Saint-Venant bidimensionnel, qui décrit la conservation de la masse et la quantité de mouvement dans le fluide. Ce système est souvent appliqué en hydraulique à surface libre bidimensionnelle pour décrire le comportement du fluide. Les équations de Saint-Venant s'écrivent :

$$\begin{cases} \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} - fv = -g \frac{\partial z}{\partial x} + k_1 \Delta u - C_D \frac{|\vec{u}| u}{h}, \\ \frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + fu = -g \frac{\partial z}{\partial y} + k_1 \Delta v - C_D \frac{|\vec{u}| v}{h}, \\ \frac{\partial h}{\partial t} + \frac{\partial(uh)}{\partial x} + \frac{\partial(vh)}{\partial y} = 0, \end{cases} \quad (2.1)$$

où :

\vec{u} est le vecteur représentant le champ de vitesse de l'écoulement, u et v sont les deux composantes de \vec{u} , respectivement sur la direction Ouest-Est et Sud-Nord,

h est la profondeur de la couche liquide, avec la contrainte qu'elle est toujours positive,

z est l'élévation de la surface libre de l'eau ($z = h + Z_b$ avec Z_b la hauteur de la surface du lit fluvial ou bien la cote du fond de la rivière),

k_1 est le coefficient de viscosité turbulente,

C_D est le coefficient de friction,

Δ est l'opérateur laplacien,

f est le paramètre de Coriolis,

et g est l'accélération de la pesanteur.

La pente de l'énergie S_f du modèle mentionnée au chapitre II, lequel décrit le frottement de l'écoulement engendré par la rugosité du fond de la rivière, est en fait un vecteur en deux dimensions. On peut donc noter $S_f = (S_{fx}, S_{fy})$, S_{fx} est la pente d'énergie le long du cours de l'écoulement suivant l'axe des x et S_{fy} est la pente de l'énergie dans la direction transverse suivant l'axe des y . Si on estime que la rugosité est isotrope, c'est à dire ne dépend pas de la direction, et que C_D est le coefficient de friction au fond de la rivière, alors S_{fx} et S_{fy} sont donnés par :

$$\begin{cases} S_{fx} = C_D \frac{|\vec{u}| u}{gh}, \\ S_{fy} = C_D \frac{|\vec{u}| v}{gh}. \end{cases} \quad (2.2)$$

2.1.2 Transport solide

Dans l'hypothèse de transport en régime non-équilibré, c'est-à-dire, il y a le processus de dépôt-érosion dans la couche solide au fond de la rivière qui n'est pas en état d'équilibre, et le taux de saturation ne peut être satisfait simultanément, la concentration des sédiments transportés par le fluide s'écarte de la capacité de l'écoulement à transporter des sédiments et elle atteint seulement une valeur approximative de cette capacité.

Dans ce cas-là, en supposant par ailleurs un diamètre uniforme des sédiments, et à partir de l'équilibre des échanges de sédiments dans la couche proche du lit de la rivière, ou bien à partir de la conservation de la masse solide, on déduit une équation qui décrit le transport des sédiments en suspension. Celle-ci se caractérise par la variable de la concentration des sédiments dans l'écoulement :

Équation de concentration:

$$\frac{\partial(hS)}{\partial t} + \nabla \cdot (\vec{u} hS) = k' \nabla \cdot (h \nabla S) + \alpha \omega (S^* - S), \quad (2.3)$$

où

S est la concentration des sédiments dans l'eau, i.e. la masse des sédiments secs par unité de volume,

k' est le coefficient de diffusion de sédiments,

ω est la vitesse de dépôt de sédiments,

S^* est la capacité de l'écoulement à transporter des sédiments ;

Une autre équation décrit la variation de la cote du fond de la rivière en négligeant l'épaisseur de la couche des sédiments charriés décrite dans le chapitre I ;

Équation de la variation de la hauteur du lit :

$$\rho_1 \frac{\partial Z_b}{\partial t} = \alpha \omega (S - S^*) - T_r, \quad (2.4)$$

avec ρ_1 la densité sèche des sédiments sur le lit de la rivière, $T_r = \nabla \cdot \vec{g}_b$ la masse des matériaux transportés sur la surface du fond de la rivière, avec $\vec{g}_b = (g_{bx}, g_{by})$ le taux de transport des sédiments charriés par unité de largeur, les deux composantes de g_{bx} et g_{by} correspondent aux taux de transport des sédiments charriés par unité de longueur le long de l'axe des x et des y respectivement.

Beaucoup de modèles différents ont été développés pour modéliser le transport des sédiments. Les différences entre ces modèles apparaissent dans des conditions aux limites et dans les termes de dépôt et d'érosion. Dans les équations (2.3) et (2.4), le terme $\alpha \omega (S - S^*)$ représente les processus de dépôt et d'érosion, c'est-à-dire, l'échange entre les sédiments en suspension dans le fluide et ceux à la surface du lit. Les sédiments sont enlevés de la surface du lit par le fluide sous l'effet de la diffusion turbulente. L'agglomération des sédiments sur le lit se fait par l'effet de la gravité. Par ailleurs, on néglige la sédimentation et l'érosion sur les bords latéraux, c'est-à-dire qu'on ne considère pas les échanges entre les sédiments dans le fluide et les parois latérales.

2.1.3 Formules semi-empiriques pour le modèle bidimensionnel

Comme dans le chapitre précédent, il reste à déterminer des variables inconnues afin de compléter les équations de base données ci-dessus. Elles sont toujours fournies par des formules semi-empiriques. Dans le cas bidimensionnel, il n'y a pas beaucoup de résultats de recherche pour déterminer ces variables hydrauliques, la plupart des formules actuelles

étant développées à partir des formules unidimensionnelles en deux dimensions.

Ici, il faut déterminer la capacité du fluide à transporter des sédiments S^* et le taux de transport des sédiments charriés par unité de longueur (g_{bx}, g_{by}) . Les formules que nous allons donner ci-dessous sont aussi choisies les mêmes que Zeng et al. ([74]).

2.1.3.1 La capacité de transport en suspension

On rappelle d'abord la définition de la capacité de l'écoulement à transporter des sédiments qui est déjà donnée dans la section 1.3.2.2, qui est le débit maximum de matériaux transportés vers l'aval.

On applique ici la formule (1.8) au cas bidimensionnel

$$S^* = K \frac{|\vec{u}|^{3m}}{(gh\omega)^m}, \quad (2.5)$$

où

$|\vec{u}| = \sqrt{u^2 + v^2}$ désigne la vitesse du fluide, avec u et v les composantes de la vitesse dans la direction d'écoulement et transverse ;

m et K sont des constantes empiriques.

2.1.3.2 Le taux de transport de charriage par unité de largeur

En supposant une granulométrie uniforme, c'est à dire, un diamètre uniforme des sédiments (on choisit un diamètre moyen), le taux de transport des sédiments par unité de largeur est aussi une représentation de la capacité de l'écoulement à transporter des sédiments charriés. Il existe peu de résultats de recherche sur le taux de transport de charriage dans la direction transverse. On a supposé dans cette formule les transport des matériaux possédant des capacités équivalentes suivant deux directions horizontales.

$$\vec{g}_b = \begin{cases} \Gamma(1 - \frac{\nu_c}{|\vec{u}|}) |\vec{u}|^n h^p d^q \vec{u}, & \text{if } \nu_c \leq |\vec{u}|, \\ 0, & \text{if } \nu_c > |\vec{u}|, \end{cases} \quad (2.6)$$

où ν_c est une constante empirique représentant la vitesse minimum pour que les sédiments commencent à être transportés par le fluide de la surface du fond de la rivière, et où les

autres paramètres $\alpha, K, \Gamma, \nu_c, m, n, p, q$ sont tous des constantes empiriques.

Faute de résultats de recherche concernant l'étude du taux de transport de charriage dans la direction transverse, on suppose implicitement qu'il existe des capacités équivalentes pour le transport des matériaux le long des deux directions horizontales.

2.2 Conditions aux limites

Conditions aux bords:

On considère un domaine Ω dont la frontière est composée de trois parties : $\partial\Omega_1, \partial\Omega_2, \partial\Omega_3$, qui représentent respectivement le bord du flux entrant, les parois solides latérales et la sortie du fluide :

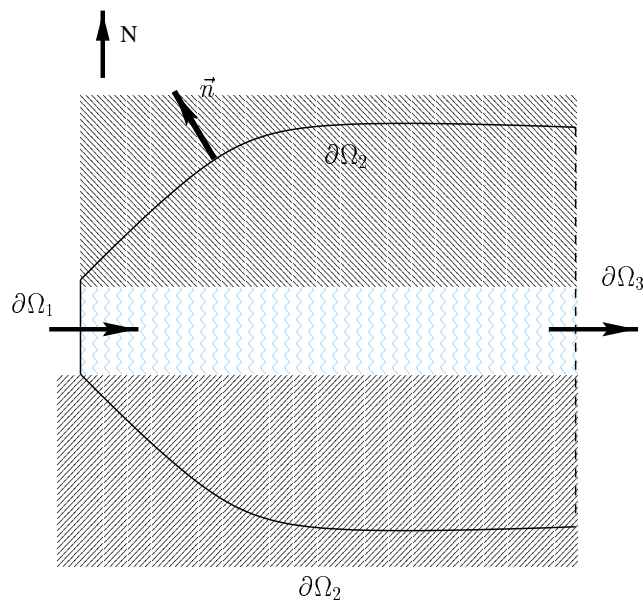


FIG. II.1 – *La composition du domaine .*

On suppose que la rivière s'écoule d'Ouest en Est avec des bords latéraux au Nord et au Sud. On donne à l'entrée (en amont) un flux $\vec{u} \cdot \vec{n}$, $\vec{u} \cdot \vec{l}$, et une hauteur de l'eau h et une concentration des sédiments S qui est prise comme une source des sédiments en amont. \vec{n} est la direction normale orientée vers l'extérieur et \vec{l} la direction tangente.

$$\begin{cases} \vec{u} \cdot \vec{n} = u_1 < 0, \\ \vec{u} \cdot \vec{l} = 0, h = h_1, \\ S = S_1, \\ Z_b = Z_{b1}. \end{cases} \quad (2.7)$$

En vue d'une modélisation sur un domaine plus réel, on donne à la sortie (en aval de la rivière) une condition de type frontière ouverte couramment utilisée pour traiter des problèmes en océanographie. Dans ce cas, la conservation de la masse totale des sédiments servant à déterminer la condition de la concentration de sédiments contenus à la frontière est difficile à obtenir. Pour simplifier le problème, le domaine du modèle ne prend pas en compte le cas de l'estuaire, où il y a peut être de flux entrant sur la frontière ouverte comme la marée.

Les conditions sur ce bord sont données par :

$$\frac{\partial \vec{u}}{\partial \vec{n}} = 0, \quad \frac{\partial h}{\partial \vec{n}} = 0, \quad \frac{\partial S}{\partial \vec{n}} = 0, \quad \frac{\partial Z_b}{\partial \vec{n}} = 0. \quad (2.8)$$

Sur les parois latérales, les conditions aux bords générales sont données par :

$$\left(A \frac{\partial \vec{u}}{\partial \vec{n}} + \mu \vec{u} \right) \cdot \vec{l} = 0 \quad \text{et} \quad -\epsilon \frac{\partial S}{\partial \vec{n}} = \kappa(S - S^{**}) + \mathcal{M}, \quad (2.9)$$

où μ, κ sont deux coefficients non négatifs, κS^{**} et \mathcal{M} sont les taux d'absorption des sédiments et d'éroulement de parois respectivement. $S - S^{**} > 0$ indique que les sédiments sont transportés par l'écoulement et déposés aux parois, tandis que $S - S^{**} < 0$ signifie le cas du processus au contraire. En raison de la simplification, on prend au cas présent

$$\mu = \kappa = \mathcal{M} = 0,$$

ce qui veut dire que l'on ne considère pas d'échanges de sédiments entre le fluide et les parois, et comme il n'y a pas de déformation des parois verticales, on choisit une condition de glissement sans frottement :

$$\vec{u} \cdot \vec{n} = 0, \quad \frac{\partial h}{\partial \vec{n}} = 0, \quad \frac{\partial S}{\partial \vec{n}} = 0, \quad \frac{\partial Z_b}{\partial \vec{n}} = 0. \quad (2.10)$$

Condition initiale:

Au temps initial, on donne des valeurs, sur tout le domaine, de la vitesse de l'écoulement \vec{u} , de l'élévation de la surface de l'eau z , de la concentration de sédiment S et de la hauteur du lit Z_b .

$$\begin{aligned}\vec{u}(t=0) &= \vec{u}_0, & h(t=0) &= h_0, \\ S(t=0) &= S_0, & Z_b(t=0) &= Z_{b0}.\end{aligned}\tag{2.11}$$

Ainsi, on est ramené à un système des équations non linéaire avec conditions initiales et conditions aux limites. Il reste à prouver mathématiquement que le problème est bien posé (existence, unicité, stabilité). Dans le chapitre suivant, nous nous attacherons exclusivement aux solutions données au cas numérique.

2.3 L'énergie totale

Nous voulons maintenant étudier l'énergie totale de notre système de la sédimentation, en prenant en compte les processus de dépôt et d'érosion dans la couche solide au fond de la rivière conséquence du transport de sédiments, on va considérer la partie de l'énergie liée à ces processus. Si le fluide considéré est un mélange d'eau et de sédiments en suspension, si ρ_0 est la densité constante de l'eau douce, et si S est la concentration de sédiments dans le fluide, alors la densité du fluide ρ s'écrit :

$$\rho = \rho_0(1 + S).$$

En fait, on peut écrire les équations (2.1) - (2.4) sous la forme d'un couplage des équations vectorielles suivantes :

$$\frac{\partial \vec{v}}{\partial t} + \vec{v} \cdot (\nabla \cdot \vec{v}) + f(\vec{k} \times \vec{v}) = -g\nabla z + k_1 \Delta \vec{v} - C_D \frac{|\vec{v}| \vec{v}}{h},\tag{2.12}$$

$$\frac{\partial h}{\partial t} + \nabla \cdot (\vec{v} h) = 0,\tag{2.13}$$

$$\frac{\partial (hS)}{\partial t} + \nabla \cdot (\vec{v} hS) = k' \nabla \cdot (h \nabla S) + \alpha \omega (S^* - S),\tag{2.14}$$

$$\rho_1 \frac{\partial Z_b}{\partial t} = \alpha \omega (S - S^*) - \nabla \cdot \vec{g}_b,\tag{2.15}$$

où $\vec{v} = (u, v)$ est le vecteur du champ de vitesse, $\vec{k} = (0, 0, 1)$ désigne le vecteur unité suivant la coordonnée verticale, $\nabla(\cdot) = \left(\frac{\partial(\cdot)}{\partial x}, \frac{\partial(\cdot)}{\partial y}\right)$ et $\Delta(\cdot) = \frac{\partial^2(\cdot)}{\partial x^2} + \frac{\partial^2(\cdot)}{\partial y^2}$.

En multipliant les équations (2.13) et (2.14) par ρ_0 , l'équation concernant la variation de la masse du fluide est donnée par leur combinaison :

$$\frac{\partial(\rho h)}{\partial t} + \nabla \cdot (\vec{v} \rho h) = k' \nabla \cdot (h \nabla \rho) + \alpha \omega (\rho_0 (S^* + 1) - \rho). \quad (2.16)$$

On peut toujours supposer que $z > 0$ et $z^2 > Z_b^2$, qui assure que l'énergie potentielle du fluide $\frac{1}{2} g \rho (z^2 - Z_b^2)$ et l'énergie cinétique $\frac{1}{2} h \rho |\vec{v}|^2$ sont positives. L'énergie totale du système de la sédimentation est la somme de l'énergie cinétique et de l'énergie potentielle. On s'intéresse donc à étudier l'expression de l'énergie totale du système :

$$\frac{\partial}{\partial t} \iint_{\Omega} \left(\rho h \frac{|\vec{v}|^2}{2} + \frac{\rho g}{2} (z^2 - Z_b^2) \right) d\Omega.$$

En multipliant l'équation (2.12) par $\rho h \vec{v}$ et l'équation (2.16) par $\frac{|\vec{v}|^2}{2}$, la somme de ces deux équations nous donne :

$$\begin{aligned} & \frac{|\vec{v}|^2}{2} \frac{\partial \rho h}{\partial t} + \rho h \vec{v} \frac{\partial \vec{v}}{\partial t} \\ = & -\frac{1}{2} \rho h \vec{v} \cdot \nabla (|\vec{v}|^2) - f \rho h \vec{v} \cdot \vec{k} \times \vec{v} - \rho g h \vec{v} \cdot \nabla z \\ & + k_1 \rho h \vec{v} \Delta \vec{v} - C_D \rho |\vec{v}|^3 - \frac{|\vec{v}|^2}{2} \nabla \cdot (\rho h \vec{v}) \\ & + \frac{|\vec{v}|^2}{2} (k' \nabla \cdot (h \nabla \rho) + \alpha \omega (\rho_0 (S^* + 1) - \rho)) \\ = & -\nabla \cdot \left(\rho h \frac{|\vec{v}|^2}{2} \vec{v} \right) - \rho g h \vec{v} \cdot \nabla z \\ & + k_1 \rho h \vec{v} \Delta \vec{v} - C_D \rho |\vec{v}|^3 \\ & + \frac{|\vec{v}|^2}{2} (k' \nabla \cdot (h \nabla \rho) + \alpha \omega (\rho_0 (S^* + 1) - \rho)), \end{aligned}$$

d'autre part, grâce à la relation $z = h + Z_b$:

$$\begin{aligned} \frac{\partial}{\partial t} \frac{\rho g}{2} (z^2 - Z_b^2) &= \frac{\partial}{\partial t} \left(\frac{\rho g}{2} (z + Z_b) h \right) \\ &= g \frac{z + Z_b}{2} \frac{\partial(\rho h)}{\partial t} + \frac{\rho g h}{2} \frac{\partial h}{\partial t} + \rho g h \frac{\partial Z_b}{\partial t}, \end{aligned}$$

mais, due à l'équation (2.16), on obtient

$$\begin{aligned} g \frac{z + Z_b}{2} \frac{\partial(\rho h)}{\partial t} &= g z \frac{\partial(\rho h)}{\partial t} - \rho \frac{g h}{2} \frac{\partial h}{\partial t} - \frac{g h^2}{2} \frac{\partial \rho}{\partial t} \\ &= g z (-\nabla \cdot (\rho h \vec{v}^\top) + k' \nabla \cdot (h \nabla \rho) + \alpha \omega (\rho_0 (S^* + 1) - \rho)) \\ &= -\frac{\rho g h}{2} \frac{\partial h}{\partial t} - \frac{g h^2}{2} \frac{\partial \rho}{\partial t}. \end{aligned}$$

Pour conclure, l'énergie totale du système hydraulique devient :

$$\begin{aligned} &\frac{\partial}{\partial t} \iint_{\Omega} \left[\rho h \frac{|\vec{v}^\top|^2}{2} + \frac{\rho g}{2} (z^2 - z_b^2) \right] d\Omega \\ &= \iint_{\Omega} \left[-\nabla \cdot \left(\rho h \vec{v}^\top \frac{|\vec{v}^\top|^2}{2} \right) - \rho g h \vec{v}^\top \nabla z - z \nabla (\rho g h \vec{v}^\top) \right. \\ &\quad \left. + \left(\frac{|\vec{v}^\top|^2}{2} + g z \right) (k' \nabla (h \nabla \rho) + \alpha \omega (\rho_0 (S^* + 1) - \rho)) \right. \\ &\quad \left. + k_1 \rho h \vec{v}^\top \Delta \vec{v}^\top - \rho g h \frac{\partial Z_b}{\partial t} - \frac{g h^2}{2} \frac{\partial \rho}{\partial t} - C_D \rho h |\vec{v}^\top|^3 \right] d\Omega. \\ &= \iint_{\Omega} \left[\left(\frac{|\vec{v}^\top|^2}{2} + g z \right) (k' \nabla (h \nabla \rho) + \alpha \omega (\rho_0 (S^* + 1) - \rho)) \right. \\ &\quad \left. - \frac{g h^2}{2} \frac{\partial \rho}{\partial t} + \rho g h \frac{\partial Z_b}{\partial t} + k_1 \rho h \vec{v}^\top \Delta \vec{v}^\top - C_D \rho h |\vec{v}^\top|^3 + B_1 + B_2 \right] d\Omega. \quad (2.17) \end{aligned}$$

g où

$$\begin{cases} B_1 = -\nabla \cdot \left(\rho h \vec{v}^\top \frac{|\vec{v}^\top|^2}{2} \right), \\ B_2 = -\nabla \cdot (\rho g h \vec{v}^\top z). \end{cases} \quad (2.18)$$

Après l'intégration des termes B_1 et B_2 sur Ω , on obtient l'intégration sur les bords qui dépend des conditions aux bords choisies.

$$\iint_{\Omega} (B_1 + B_2) d\Omega = \int_{\partial\Omega} \left(\rho h \frac{|\vec{v}^\top|^2}{2} + \rho g h z \right) \vec{v}^\top \cdot \vec{n}^\top dl,$$

dont \vec{n}^\top est la direction normale extérieure.

L'équation d'énergie (2.17) est assez compliquée, et l'interprétation physique de chaque terme n'est pas simple. L'étude concernant la simplification de cette formule nécessite l'introduction de certaines hypothèses afin de déduire des résultats intéressants. En négligeant, par défaut, le terme de viscosité turbulente (le laplacien) et du frottement ou de la force d'extérieur, on considère seulement l'équation s'agissant la dynamique du fluide qui correspond essentiellement aux processus rapides où l'on ne prend pas en compte la variation de la concentration de sédiments donc de la densité du fluide et du lit de la rivière au cours du temps :

$$\frac{\partial Z_b}{\partial t} = 0,$$

l'expression d'énergie (2.17) devient

$$\frac{\partial}{\partial t} \iint_{\Omega} \left[\rho h \frac{|\vec{v}^\top|^2}{2} + \frac{\rho g}{2} (z^2 - z_b^2) \right] d\Omega = \int_{\partial\Omega} \left(\rho h \frac{|\vec{v}^\top|^2}{2} + \rho g h z \right) \vec{v}^\top \cdot \vec{n}^\top dl, \quad (2.19)$$

ce qui, si l'on considère un domaine fermé, implique la propriété de la conservation de l'énergie :

$$\frac{\partial}{\partial t} \iint_{\Omega} \left[\rho h \frac{|\vec{v}^\top|^2}{2} + \frac{\rho g}{2} (z^2 - z_b^2) \right] d\Omega = 0. \quad (2.20)$$

2.4 Changement de la forme du modèle

En pratique, il est important que les équations discrétisées gardent les propriétés des équations continues, notamment, les propriétés de conservation de masse et de quantité de mouvement. Avant de discrétiser les équations envisagées, nous allons d'abord changer la forme de nos équations en introduisant un changement de variables.

2.4.1 Transformation des variables pour les équations de Saint-Venant

Pour étudier les propriétés des équations de Saint-Venant, comme dans Zeng et al. [73], on introduit d'abord un changement de variable suivant en posant :

$$\begin{cases} \Phi = \sqrt{gh} = \sqrt{g(z - Z_b)} = \sqrt{\phi - gZ_b}, \\ U = u\Phi, \\ V = v\Phi. \end{cases} \quad (2.21)$$

Pour les équations de Saint-Venant, on ne considère pas le processus de dépôt ou d'érosion au fond de la rivière, c'est à dire que l'on ne prend pas en compte la variation du lit fluvial au cours du temps :

$$\frac{\partial Z_b}{\partial t} = 0.$$

Donc, les équations (2.1) s'expriment avec ces variables de la manière suivante :

$$\begin{cases} \frac{\partial U}{\partial t} + \left(\frac{\partial uU}{\partial x} - \frac{U}{2} \frac{\partial u}{\partial x} \right) + \left(\frac{\partial vU}{\partial y} - \frac{U}{2} \frac{\partial v}{\partial y} \right) \\ = fV - \Phi \frac{\partial \Phi}{\partial x} + k_1 \Phi \Delta \left(\frac{U}{\Phi} \right) - g^2 C_D \sqrt{U^2 + V^2} \frac{U}{\Phi^3}, \\ \frac{\partial V}{\partial t} + \left(\frac{\partial uV}{\partial x} - \frac{V}{2} \frac{\partial u}{\partial x} \right) + \left(\frac{\partial vV}{\partial y} - \frac{V}{2} \frac{\partial v}{\partial y} \right) \\ = -fU - \Phi \frac{\partial \Phi}{\partial y} + k_1 \Phi \Delta \left(\frac{V}{\Phi} \right) - g^2 C_D \sqrt{U^2 + V^2} \frac{V}{\Phi^3}, \\ \frac{\partial \phi}{\partial t} + \frac{\partial (U\Phi)}{\partial x} + \frac{\partial (V\Phi)}{\partial y} = 0. \end{cases} \quad (2.22)$$

On montre alors que les équations ci-dessus avec le changement de variable peuvent être écrites sous forme vectorielle, en posant

$$\vec{V} = (U, V), \quad \vec{v} = (u, v),$$

nous allons montrer la proposition suivante :

Proposition 1

$$\frac{\partial \vec{V}}{\partial t} + \vec{v} \cdot (\nabla \cdot \vec{V}) + \frac{1}{2} (\nabla \cdot \vec{v}) \vec{V} + f(\vec{k} \times \vec{V}) = -\Phi \nabla \phi + k_1 \Phi \Delta \left(\frac{\vec{V}}{\Phi} \right) - \vec{F}, \quad (2.23)$$

$$\frac{\partial \phi}{\partial t} + \nabla \cdot (\Phi \vec{V}) = 0, \quad (2.24)$$

avec $\vec{F} = g^2 C_D \frac{|\vec{V}| \vec{V}}{\Phi^3} = (g^2 C_D \sqrt{U^2 + V^2} \frac{U}{\Phi^3}, g^2 C_D \sqrt{U^2 + V^2} \frac{V}{\Phi^3})$.

A partir de l'équation (2.13), on obtient l'équation (2.24) par la multiplication du coefficient g , et pour la preuve de (2.23), nous avons besoin du lemme suivant :

Lemme 1

$$\vec{V}^\top(\nabla \cdot \vec{v}^\top) - \vec{v}^\top \frac{\partial \Phi}{\partial t} = \vec{v}^\top(\nabla \cdot \vec{V}^\top) + \frac{1}{2}(\nabla \cdot \vec{v}^\top)\vec{V}^\top. \quad (2.25)$$

Démonstration

Pour un vecteur $\vec{v}^\top = (u, v)$, on note les deux composantes par

$$[\vec{v}^\top]_x = u, \quad [\vec{v}^\top]_y = v,$$

alors on écrit (2.25) composante par composante, c.-à-d. :

$$[\vec{V}^\top(\nabla \cdot \vec{v}^\top)]_x - \left[\vec{v}^\top \frac{\partial \Phi}{\partial t} \right]_x = [\vec{v}^\top(\nabla \cdot \vec{V}^\top)]_x + \left[\frac{1}{2}(\nabla \cdot \vec{v}^\top)\vec{V}^\top \right]_x, \quad (2.26)$$

$$[\vec{V}^\top(\nabla \cdot \vec{v}^\top)]_y - \left[\vec{v}^\top \frac{\partial \Phi}{\partial t} \right]_y = [\vec{v}^\top(\nabla \cdot \vec{V}^\top)]_y + \left[\frac{1}{2}(\nabla \cdot \vec{v}^\top)\vec{V}^\top \right]_y, \quad (2.27)$$

mais comme la démonstration est identique, on ne prouve que la première composante (2.26).

Sous l'hypothèse de la variation nulle de la cote du fond de la rivière, on obtient de (2.21) que

$$\frac{\partial \phi}{\partial t} = 2\Phi \frac{\partial \Phi}{\partial t}, \quad (2.28)$$

et grâce à (2.24), on a

$$\begin{aligned} - \left[\vec{v}^\top \frac{\partial \Phi}{\partial t} \right]_x &= -\vec{v}^\top \cdot \frac{1}{2\Phi} \frac{\partial \phi}{\partial t} = \vec{v}^\top \cdot \frac{1}{2\Phi} \nabla(\Phi \vec{V}^\top) \\ &= \frac{u}{2\Phi} \left(\frac{\partial(\Phi U)}{\partial x} + \frac{\partial(\Phi V)}{\partial y} \right) \\ &= \frac{u}{2\Phi} \left(\Phi \frac{\partial U}{\partial x} + U \frac{\partial \Phi}{\partial x} + \Phi \frac{\partial V}{\partial y} + V \frac{\partial \Phi}{\partial y} \right) \\ &= \frac{u}{2} \frac{\partial U}{\partial x} + \frac{uU}{2\Phi} \frac{\partial \Phi}{\partial x} + \frac{u}{2} \frac{\partial V}{\partial y} + \frac{uV}{2\Phi} \frac{\partial \Phi}{\partial y}, \end{aligned}$$

par suite de l'équation (2.21), il vient :

$$\begin{aligned} \frac{uU}{2\Phi} \frac{\partial \Phi}{\partial x} &= \frac{u}{2} \left(\frac{\partial U}{\partial x} - \Phi \frac{\partial u}{\partial x} \right) \\ &= \frac{1}{2} \left(u \frac{\partial U}{\partial x} - U \frac{\partial u}{\partial x} \right), \end{aligned}$$

de même, on a

$$\frac{uV}{2\Phi} \frac{\partial \Phi}{\partial y} = \frac{1}{2} \left(u \frac{\partial V}{\partial y} - U \frac{\partial v}{\partial y} \right), \quad (2.29)$$

d'où

$$- \left[\vec{v} \frac{\partial \Phi}{\partial t} \right]_x = u \left(\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} \right) - \frac{1}{2} U \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right). \quad (2.30)$$

De plus, grâce à

$$[\vec{V}(\nabla \cdot \vec{v})]_x = \left[\vec{V} \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) \right]_x = U \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right), \quad (2.31)$$

donc

$$[\vec{V}(\nabla \cdot \vec{v})]_x - \left[\vec{v} \frac{\partial \Phi}{\partial t} \right]_x = u \frac{\partial U}{\partial x} + u \frac{\partial V}{\partial y} + \frac{1}{2} U \frac{\partial u}{\partial x} + \frac{1}{2} U \frac{\partial v}{\partial y}. \quad (2.32)$$

Si on considère le terme de droite de (2.26), on voit que :

$$[\vec{v}(\nabla \cdot \vec{V})]_x = \left[\vec{v} \left(\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} \right) \right]_x = u \frac{\partial U}{\partial x} + u \frac{\partial V}{\partial y},$$

et

$$\left[\frac{1}{2} (\nabla \cdot \vec{v}) \vec{V} \right]_x = \left[\frac{1}{2} \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) \vec{V} \right]_x = \frac{1}{2} \left(U \frac{\partial u}{\partial x} + U \frac{\partial v}{\partial y} \right),$$

on obtient donc l'égalité voulue. ■

Pour démontrer l'équation (2.23), nous multiplions l'équation (2.12) par Φ , qui donne

$$\Phi \frac{\partial \vec{v}}{\partial t} + \vec{V}(\nabla \cdot \vec{v}) + f(\vec{k} \times \vec{V}) = -\Phi \nabla \phi + k_1 \Phi \Delta \left(\frac{\vec{V}}{\Phi} \right) - \vec{F}, \quad (2.33)$$

et grâce à

$$\Phi \frac{\partial \vec{v}}{\partial t} = \frac{\partial \vec{V}}{\partial t} - \vec{v} \frac{\partial \Phi}{\partial t} \quad (2.34)$$

et à l'application du lemme 1, nous obtenons l'équation (2.23).

A l'aide de ces équations sous forme vectorielle obtenues par le changement de variable, nous regarderons maintenant les propriétés de conservation du système de Saint-Venant.

2.4.2 Propriétés de conservation des équations de Saint-Venant

2.4.2.1 Conservation de la masse

Si l'on intègre l'équation (2.24) sur un domaine Ω , en supposant implicitement sur ce domaine une vitesse non nulle et une hauteur de la couche liquide positive, on obtient :

$$\begin{aligned} \frac{\partial}{\partial t} \iint_{\Omega} \phi d\Omega &= \iint_{\Omega} \nabla \cdot (\Phi \vec{V}) d\Omega \\ &= \oint_{\partial\Omega} \Phi (\vec{V} \cdot \vec{n}) dl, \end{aligned}$$

où $\oint_{\partial\Omega}$ désigne l'intégration sur $\partial\Omega$ la frontière du domaine Ω , \vec{n} étant la normale unitaire extérieure à Ω . Puisque $\phi = gz$ désigne le géopotentiel, on voit de cette expression que la variation de la masse totale du fluide dans la couche liquide est égale à l'intégrale du débit du fluide sur les bords, en conséquence, si le domaine Ω est supposé fermé, comme $\vec{v} \cdot \vec{n} = 0$ sur S , on obtient l'équation de conservation de la masse du fluide :

$$\iint_{\Omega} \phi d\Omega = \text{constante.} \quad (2.35)$$

2.4.2.2 Conservation de l'énergie

Pour examiner la propriété essentielle des équations (2.22), on néglige, par défaut, le terme de viscosité turbulente (le laplacien) et du frottement ou de la force d'extérieur.

On s'intéresse donc à étudier la variation d'énergie totale du volume du fluide :

$$\frac{\partial}{\partial t} \iint_{\Omega} \frac{1}{2} [\rho h |\vec{v}|^2 + \rho g (z^2 - Z_b^2)] dx dy. \quad (2.36)$$

En appliquant le changement de variable (2.21), l'expression de l'énergie totale est ramenée à étudier l'expression équivalente suivante en négligeant la déformation au fond :

$$\frac{\partial}{\partial t} \iint_{\Omega} \frac{1}{2} (U^2 + V^2 + \phi^2) dx dy. \quad (2.37)$$

En fait, multiplions l'équation (2.23) par \vec{V} , et l'équation (2.24) par ϕ , la somme des intégrations de ces deux termes dans le domaine Ω donne :

$$\begin{aligned} \frac{\partial}{\partial t} \iint_{\Omega} \frac{1}{2}(U^2 + V^2 + \phi^2) dx dy &= \iint_{\Omega} \left[-\vec{V} \cdot (\vec{v} \nabla \cdot \vec{V}) - \frac{1}{2} \vec{V} \cdot ((\nabla \cdot \vec{v}) \vec{V}) \right. \\ &\quad \left. - \vec{V} \cdot f(\vec{k} \times \vec{V}) - \vec{V} \cdot (\Phi \nabla \phi) - \phi \nabla \cdot (\Phi \vec{V}) \right] dx dy. \end{aligned}$$

Pour examiner cette expression, on prouve d'abord une relation suivante :

Lemme 2

$$\vec{V} \cdot (\vec{v} \nabla \cdot \vec{V}) + \frac{1}{2} \vec{V} \cdot ((\nabla \cdot \vec{v}) \vec{V}) = \nabla \cdot \left(\frac{|\vec{V}|^2}{2} \vec{v} \right). \quad (2.38)$$

Démonstration

Considérons tout d'abord le côté gauche de l'égalité (2.38), on a

$$\begin{aligned} \vec{V} \cdot \vec{v} \nabla \cdot \vec{V} &= \vec{v} \cdot \vec{V} \nabla \cdot \vec{V} \\ &= \frac{1}{2} \vec{v} \cdot \nabla |\vec{V}|^2, \end{aligned}$$

et

$$\frac{1}{2} \vec{V} \cdot ((\nabla \cdot \vec{v}) \vec{V}) = \frac{1}{2} |\vec{V}|^2 \nabla \cdot \vec{v}. \quad (2.39)$$

Regardons la somme de ces deux égalités, on en déduit donc le lemme. ■

En plus, f étant constante, il est évident que le terme $\vec{V} \cdot f(\vec{k} \times \vec{V})$ est nul et que

$$\vec{V} \cdot (\Phi \nabla \phi) + \phi \nabla \cdot (\Phi \vec{V}) = \nabla \cdot (\phi \Phi \vec{V}).$$

Par conséquent, l'on obtient l'expression de l'énergie du système de Saint-Venant :

$$\begin{aligned} \frac{\partial}{\partial t} \iint_{\Omega} \frac{1}{2}(U^2 + V^2 + \phi^2) d\Omega &= - \iint_{\Omega} (\nabla \cdot (\phi \Phi \vec{V}) + \nabla \cdot \left(\frac{|\vec{V}|^2}{2} \vec{v} \right)) d\Omega \\ &= - \oint_{\partial\Omega} \left(\frac{|\vec{V}|^2}{2} \vec{v} \cdot \vec{n} + \phi \Phi \vec{V} \cdot \vec{n} \right) dl, \quad (2.40) \end{aligned}$$

ce qui, si l'on considère un domaine fermé, implique la propriété de la conservation de l'énergie. Avec cette expression, on peut considérer le problème dans l'espace de $L^2(\Omega)$, on obtient

$$\frac{d}{dt}(\|U\|^2 + \|V\|^2 + \|\phi\|^2) = 0,$$

à partir de laquelle, on déduit l'unicité de la solution du problème (2.22) avec conditions initiales.

Puisque la stabilité du calcul est liée étroitement à la propriété de conservation d'énergie, il est important de conserver cette propriété de conservation lorsque l'on veut discrétiser les équations (2.22) selon un certain schéma aux différences finies. Nous allons l'envisager dans le chapitre suivant.

2.5 Conclusion

Nous avons donné les équations générales du transport des sédiments pour effectuer la fermeture des systèmes, nous avons également introduit les formules empiriques en précisant la concentration en fonction des paramètres caractéristiques de l'écoulement et des matériaux.

L'introduction d'un changement de variable nous permettent d'étudier les propriétés de conservation de la masse et de l'énergie de l'équation gouvernant l'écoulement dans un cadre d'espace de L^2 , puis dans le prochain chapitre, les équations discrètes adaptées à certain maillage de discrétisation, lesquelles garderont les mêmes propriétés.

Chapitre III

Discrétisation du modèle

Dans ce chapitre, nous présentons des schémas numériques qui permettent le calcul d'une solution approchée. On se concentre sur l'analyse numérique de la stabilité du schéma aux différences finies choisi pour le modèle dans le chapitre précédent, en prenant en compte la mise au point de la résolution numérique du transport des matériaux convectés dans l'écoulement. Après avoir établi un théorème de stabilité du schéma à pas fractionnaires avec une forme générale caractérisée par un paramètre, on introduit un algorithme de simulation de type "time-splitting" permettant d'une économie de coût de calcul à long terme. On montre également la conservation de l'énergie des équations aux différences de Saint-Venant discrétisées en espace par le maillage de type ARAKAWA. En application, on montre des résultats numériques mis en œuvre dans le cas de la modélisation de la formation d'un delta dans une rivière.

3.1 Méthodes numériques pour résoudre l'équation d'évolution

La solution des équations non-linéaires ou quasi-linéaires souvent utilisées dans le domaine de mécanique du fluide et de la prévision du temps ne peut être approchée qu'à l'aide des méthodes numériques. Pour cela, l'étude préalable concernant le bon choix du schéma, surtout sa stabilité de calcul, doit être réalisée avant la mise en pratique de la méthode numérique. Un schéma non stable est sensible à l'erreur d'arrondi propageant dans le calcul numérique et peut provoquer une explosion de la solution numérique qui diffère considérablement de celle du problème différentiel. On s'intéresse aux méthodes de type différences finies pour la résolution des équations utilisées en hydrologie.

3.1.1 Notations

Nous prenons les équations quasi-linéaires dans un cadre plus général en tenant en compte des problèmes d'évolution, il s'agit de trouver une fonction ou une famille de fonctions $X(x, t)$ vérifiant :

$$\begin{cases} \frac{\partial X}{\partial t} + \mathcal{F} \cdot X = \mathcal{G}, & \text{dans } \Omega \times [0, T], \\ X(t=0) = X^0, & \text{dans } \Omega, \end{cases} \quad (3.1)$$

où $X = X(x, t)$, $x \in \Omega$, $t \in [0, T]$, est la variable d'état, k dénote la dimension d'espace, $\mathcal{F} = \mathcal{F}(t, x, X)$ est un opérateur non-linéaire dépendant de la solution X . La solution du problème peut être de plus demandée de vérifier des certaines conditions sur le bord $\partial\Omega$.

Par exemple, pour les équations de Saint-Venant sans terme de diffusion,

$$\begin{cases} \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + \frac{\partial \phi}{\partial x} - fv = 0, \\ \frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + \frac{\partial \phi}{\partial y} + fu = 0, \\ \frac{\partial \phi}{\partial t} + \frac{\partial(u\phi)}{\partial x} + \frac{\partial(v\phi)}{\partial y} = 0 \end{cases} \quad (3.2)$$

on décrit la variable X et l'opérateur \mathcal{F} de la manière suivante :

$$X = \begin{pmatrix} u \\ v \\ \phi \end{pmatrix} \quad (3.3)$$

et

$$\mathcal{F} = \begin{pmatrix} u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} & -f & \frac{\partial}{\partial x} \\ f & u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} & \frac{\partial}{\partial y} \\ \phi \frac{\partial}{\partial x} & \phi \frac{\partial}{\partial y} & u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} \end{pmatrix} \quad (3.4)$$

Soient \mathcal{V} , \mathcal{H} deux espaces de Hilbert avec $\mathcal{V} \subset \mathcal{H}$ (par exemple $\mathcal{H} = L^2(\Omega)$), \mathcal{V} est dense dans \mathcal{H} . Si \mathcal{V}' désigne le dual de \mathcal{V} , on a, en identifiant \mathcal{H} à son espace dual : $\mathcal{V} \subset \mathcal{H} \subset \mathcal{V}'$. Si un opérateur \mathcal{A} défini par :

$$\mathcal{A} : \mathcal{H} \rightarrow \mathcal{V}, \quad (3.5)$$

on dit que \mathcal{A} est **positif** s'il vérifie

$$(\mathcal{A}Y, Y) \geq 0, \quad \forall Y \in \mathcal{V}.$$

La norme de l'opérateur \mathcal{A} est définie par

$$\|\mathcal{A}\|^2 = \sup_{Y \in \mathcal{V}, Y \neq 0} \frac{(\mathcal{A}Y, \mathcal{A}Y)}{(Y, Y)}.$$

Pour résoudre numériquement les équations (3.1) avec une méthode de différences finies, on les discrétise d'abord en espace. Supposons que le domaine d'espace Ω ouvert dans \mathbb{R}^2 , et choisissons les pas d'espace δx et δy , ainsi, le domaine est discrétisé par un maillage rectangulaire. On note M le nombre des unités de maillage. Si le domaine est irrégulier, la frontière sera approchée par une interpolation réalisée sur les points de grille les plus proches. Pour deux fonctions définies sur Ω , on munit du produit scalaire suivant :

$$(X, Y) = \sum_i^M X_i Y_i \Delta_M, ,$$

où $\Delta_M = \delta x \delta y$ désignant la mesure de chaque unité du maillage. La L^2 -norme de X est donnée par $\|X\| = (X, X)^{1/2}$.

3.1.2 Résultat de stabilité

Pour éviter les abus des indices, dans l'étude suivante, on considère l'équation déjà discrétisée en espace, on obtient un système d'équation différentielle ordinaire qui s'écrit de la forme suivante :

$$\begin{cases} \frac{\partial X}{\partial t} + F \cdot X = G, & \text{dans } \mathbb{R}^k \times [0, T], \\ X(t=0) = X^0, & \text{dans } \Omega. \end{cases} \quad (3.6)$$

où F et G sont l'opérateur aux différences correspondant à \mathcal{F} et \mathcal{G} dans l'équation (3.1).

Soit N un entier, on divise l'intervalle du temps $[0, T]$ en N parties, soit $\tau = \frac{T}{N}$ le pas de temps, donc $t_n = n\tau$, $n = 0, \dots, N_T$, sont les instants discrets. On s'intéresse aux schémas numériquement stables et possédant une plus grande précision d'approximation, car il y a naturellement beaucoup d'autres schémas qui ont été proposés. Dans Zeng[75], on a utilisé un schéma aux différences finies selon l'espace, avec une discrétisation en temps dépendant d'un paramètre. Soit $0 \leq \theta \leq 1$, il s'agit de trouver une solution X^n de l'équation discrétisée en temps suivante :

$$\begin{cases} \frac{X^{n+1} - X^n}{\tau} + F^{(n+1,n)} \cdot [\theta X^{n+1} + (1 - \theta)X^n] = G^{(n+1,n)}, \\ X^n|_{n=0} = X^0, \end{cases} \quad (3.7)$$

où X^n dénote l'approximation de l'état au temps t^n , $F^{(n+1,n)}$ désigne l'approximation de l'opérateur \mathcal{F} . Grâce à la dépendance de \mathcal{F} de la variable X , $F^{(n+1,n)}$ dépend une combinaison linéaire des variables discrétisées X^n et X^{n+1} , $G^{(n+1,n)} = \theta G^{n+1} + (1 - \theta)G^n$.

En fait, on constate que le schéma (3.7) possède une forme plus générale que des schémas couramment utilisés. Dans cette partie, on va établir les conditions de stabilité en ce qui concerne le schéma (3.7).

Pour étudier la stabilité du schéma (3.7), on définit le sens de stabilité comme suit :

Définition 1 (1) On dit que un schéma est stable, si quelque soit h et $0 \leq n \leq N$, avec $T (= N\tau)$ fixé, lorsque τ est suffisamment petit, il existe une constante C indépendante de τ , h et X_h^0 telle que la solution numérique vérifiant

$$\|X^{n+1}\| \leq [1 + O(\tau)]\|X^n\| + \tau\|G^{(n+1,n)}\|.$$

(2) Un schéma sera absolument stable s'il est stable pour un choix arbitraire de $\tau > 0$ et $h > 0$. Autrement, s'il y a quelque condition à satisfaire entre τ et h afin que le schéma soit stable, on dit que le schéma est conditionnellement stable.

Théorème 1 (1) Soit $\theta \in [0, 1]$, pourvu que $F = F^{(n+1,n)}$ est un opérateur positif, c'est-à-dire, $(FY, Y) \geq 0$, alors :

(1) Si $\frac{1}{2} \leq \theta \leq 1$, le schéma (3.7) est absolument stable.

(2) Si $\theta = 0$, le schéma est stable si de plus F satisfait la condition suivante :

$$|1 + \|F\|^2 \tau^2 - 2\tau m_h| \leq 1 + O(\tau), \quad (3.8)$$

où

$$m_h = \inf_{Y \in L^2} \frac{(FY, Y)^{1/2}}{\|Y\|}.$$

Preuve :

On décrit le principe de la démonstration du théorème comme ci-dessous :

(1) Si $\frac{1}{2} \leq \theta \leq 1$, posons

$$\overline{X}^\theta = \theta X^{n+1} + (1 - \theta)X^n,$$

son produit scalaire avec l'équation (3.7) donne

$$(\overline{X}^\theta, X^{n+1} - X^n) = -\tau(F\overline{X}^\theta, \overline{X}^\theta) + \tau(\overline{G}, \overline{X}^\theta),$$

on a aussi

$$\begin{aligned} \theta \|X^{n+1}\|^2 - (1 - \theta)\|X^n\|^2 + (1 - 2\theta)(X^{n+1}, X^n) \\ = -\tau(F\overline{X}^\theta, \overline{X}^\theta) + \tau(\overline{G}, \overline{X}^\theta), \end{aligned}$$

car F est positif et

$$\tau(\overline{G}, \overline{X}^\theta) \leq \tau\|\overline{G}\|(\theta\|X^{n+1}\| + (1 - \theta)\|X^n\|),$$

on en déduit que

$$\theta\|X^{n+1}\|^2 - (1 - \theta)\|X^n\|^2 \leq (2\theta - 1)(X^{n+1}, X^n) + \tau\|\overline{G}\|(\theta\|X^{n+1}\| + (1 - \theta)\|X^n\|).$$

Grâce à $\frac{1}{2} \leq \theta \leq 1$, et au fait que

$$(X^{n+1}, X^n) \leq \frac{1}{2}(\|X^{n+1}\|^2 + \|X^n\|^2),$$

on a

$$\|X^{n+1}\|^2 - 2\tau\theta\|\overline{G}\|\|X^{n+1}\| + (\tau\theta\|\overline{G}\|)^2 \leq \|X^n\|^2 + 2\tau\theta\|\overline{G}\|\|X^n\| + (\tau\theta\|\overline{G}\|)^2,$$

c'est-à-dire que

$$\|X^{n+1}\| \leq \|X^n\| + \tau\|\overline{G}\|,$$

dont on peut sûrement déduire la condition de la stabilité d'après la définition 1.

Si $\theta = 0$, on prend en général $F^n = F(X^n)$, l'équation (3.7) revient au schéma explicite:

$$\frac{X^{n+1} - X^n}{\tau} + F^n X^n = G^n,$$

donc

$$X^{n+1} = (I - \tau F^n)X^n + \tau G^n,$$

mais

$$\|(I - \tau F^n)X^n\|^2 = \|X^n\|^2 + \tau^2\|F^n X^n\|^2 - 2\tau(F^n X^n, X^n),$$

d'après (3.8) et $\|F^n X\| \leq \|F^n\|\|X\|$, on a

$$\|(I - \tau F^n)X^n\|^2 \leq (1 + \tau^2\|F^n\| - 2\tau m^2)\|X^n\|^2 \leq (1 + O(\tau))\|X^n\|^2.$$

Quand τ est suffisamment petit, on obtient

$$\sqrt{1 + O(\tau)} \approx 1 + O(\tau),$$

dont on déduit

$$\|X^{n+1}\| \leq [1 + O(\tau)]\|X^n\| + \tau\|G^n\|,$$

ce qui revient à la stabilité quand $\theta = 0$.

Corollaire : Si $G = 0$ et $(FX, X) = 0$, lorsque $\theta = \frac{1}{2}$, le schéma (3.7) conserve l'énergie du système au sens de

$$\|X^n\|^2 = \|X^0\|^2.$$

3.1.3 Résultat de convergence

Quand les indices de discrétisation h, τ tendent vers zéro, on veut savoir si la solution numérique obtenue par l'intégration des équations de différences (3.7) tend vers la solution des équations continues (3.1). Si l'on pose un problème sous la forme suivante :

$$L^{h\tau} X^{h\tau} = f^{h\tau},$$

où l'opérateur $L^{h\tau}$ et les fonctions $X^{h\tau}$ et $f^{h\tau}$ sont discrétisés en temps et en espace. Filippov ([25]) a défini la stabilité comme étant pour l'opérateur $(L^{h\tau})^{-1}$ d'être borné uniformément, c'est-à-dire le fait qu'il existe une constante uniforme C tel que

$$\|(L^{h\tau})^{-1}\| \leq C.$$

et il a prouvé que, dans le cas linéaire, la consistance et la stabilité impliquent la convergence de la solution de l'équation aux différences finies vers celle de l'équation différentielle. Pour un problème d'évolution bien posé, Lax a proposé un système de définitions de consistance et de stabilité avec lesquelles la stabilité est déduite simultanément avec la convergence quand la consistance est établie, ceci est connu comme le théorème de Lax (cf. Richtmyer et Morton [53]).

Maintenant, on veut étudier l'approximation du schéma (3.7) lorsque $\theta = \frac{1}{2}$, on réécrit ici le schéma pour l'équation homogène ci-dessous :

$$\frac{X^{n+1} - X^n}{\tau} + F^{(n+1,n)} \cdot \left(\frac{X^{n+1} + X^n}{2} \right) = 0. \quad (3.9)$$

A propos, on donne une proposition suivante :

Proposition 2 *Le schéma*

$$\frac{X^{n+1} - X^n}{\tau} + F\left(\frac{X^{n+1} + X^n}{2}\right) \cdot \left(\frac{X^{n+1} + X^n}{2}\right) = 0$$

et une approximation d'ordre 2 par rapport au temps.

Démonstration

D'abord, on définit HX par :

$$HX \equiv \frac{\partial X}{\partial t} + F(X) \cdot X, \quad (3.10)$$

et son approximation H_τ^n , dont X^n signifie la projection de la solution exacte de (3.1) sur le maillage discrétisé au temps t_n , par :

$$H_\tau^n X^n = \frac{X^{n+1} - X^n}{\tau} + F\left(\frac{X^{n+1} + X^n}{2}\right) \cdot \left(\frac{X^{n+1} + X^n}{2}\right). \quad (3.11)$$

On définit une norme $\|\cdot\|_{C_\tau}$ qui estime l'approximation de la discrétisation :

$$\|(HX)^n - H_\tau^n X^n\|_{C_\tau} = \|H_\tau^n X^n\|_{C_\tau} = \max_{t_n} \|H_\tau^n X^n\|, \quad (3.12)$$

où $\|\cdot\|$ désigne la norme sur le maillage défini (pour $t = t_n$).

Si on suppose que X , étant la solution de l'équation

$$HX = 0, \quad \text{i.e. } X_t = -F(X) \cdot X, \quad (3.13)$$

est d'ordre 3, noté $X_t = \frac{\partial X}{\partial t}$, on écrit le développement de Taylor :

$$X^{n+1} = X^n + \tau(X_t)^n + \frac{\tau^2}{2}(X_{tt})^n + \dots, \quad (3.14)$$

on a

$$\begin{aligned} X_{tt} &= -\frac{\partial}{\partial t}(F(X) \cdot X), \\ &= -\left(\left[\frac{\partial F}{\partial X}\right] X_t\right) \cdot X - F \cdot X_t, \\ &= -\left(\left[\frac{\partial F}{\partial X}\right] F \cdot X\right) \cdot X + F^2 X, \end{aligned}$$

avec lequel, l'application de (3.14) sur les variables discrètes implique que

$$X^{n+1} = X^n - \tau F^n \cdot X^n + \frac{\tau^2}{2} \left(\left[\frac{\partial F}{\partial X} \right]^n F^n \cdot X^n + (F^n)^2 \right) \cdot X^n + O(\tau^3),$$

où $F^n = F(X^n)$ désigne la valeur d'approximation de F en prenant les coefficients X au temps t_n .

Donc on peut mesurer la norme de la différence entre l'opérateur continu et discret de façon que :

$$\begin{aligned} & \|H_\tau^n X^n - (HX)^n\|_{C_\tau} (= \|H_\tau^n X^n\|_{C_\tau}, \quad \text{car (3.13)}) \\ &= \max_{t_n} \left\| -F^n X^n + \frac{\tau}{2} \left(\left[\frac{\partial F}{\partial X} \right]^n F^n \cdot X^n + (F^n)^2 \right) \cdot X^n \right. \\ & \quad \left. + F\left(\frac{X^{n+1} + X^n}{2}\right) \left(X^n - \frac{\tau}{2} F^n X^n + \frac{\tau^2}{4} \left(\left[\frac{\partial F}{\partial X} \right]^n F^n \cdot X^n + (F^n)^2 \right) \cdot X^n + O(\tau^3) \right) \right\|, \\ &= \max_{t_n} \left\| F\left(\frac{X^{n+1} + X^n}{2}\right) \cdot X^n - F^n \cdot X^n \right. \\ & \quad \left. + \frac{\tau}{2} \left(\left[\frac{\partial F}{\partial X} \right]^n F^n \cdot X^n + (F^n)^2 - F\left(\frac{X^{n+1} + X^n}{2}\right) F^n \right) \cdot X^n + O(\tau^2) \right\|. \end{aligned} \quad (3.15)$$

De plus, si l'on suppose aussi que l'opérateur F est suffisamment régulier par rapport au temps et à X , on peut obtenir

$$\begin{aligned} F\left(\frac{X^{n+1} + X^n}{2}\right) &= F\left(X^n - \frac{\tau}{2} F^n X^n + O(\tau^2)\right), \\ &= F(X^n) - \frac{\tau}{2} \left[\frac{\partial F}{\partial X} \right]^n F^n X^n + O(\tau^2), \end{aligned}$$

en appliquant ceci à (3.15) on en déduit que

$$\|H_\tau^n X^n - (HX)^n\|_{C_\tau} = O(\tau^2).$$

ce qui démontre la proposition. ■

3.2 Schémas à pas fractionnaires

Supposons que dans la section précédente, l'opérateur différentiel \mathcal{F} admet une sorte de décomposition "naturelle" :

$$\mathcal{F}(X) = \sum_{i=1}^m \mathcal{F}_i(X), \quad (3.16)$$

$$\begin{cases} \frac{d}{dt} X^{n+\frac{1}{m}}(t) + \mathcal{F}_1 \cdot X^{n+\frac{1}{m}} = 0, & n\Delta t < t < (n+1)\Delta t, \\ X^{n+\frac{1}{m}}(n\Delta t) = X^n. \end{cases} \quad (3.18)$$

Ce problème admet une solution unique dans $(n\Delta t, (n+1)\Delta t)$; on définit ensuite $X^{n+\frac{2}{m}}$ pour $t > n\Delta t$ comme étant la solution des équations suivantes avec $X^{n+\frac{1}{m}}((n+1)\Delta t)$ comme condition initiale :

$$\begin{cases} \frac{d}{dt} X^{n+\frac{2}{m}}(t) + \mathcal{F}_2 \cdot X^{n+\frac{2}{m}} = 0, & n\Delta t < t < (n+1)\Delta t, \\ X^{n+\frac{2}{m}}(n\Delta t) = X^{n+\frac{1}{m}}((n+1)\Delta t). \end{cases} \quad (3.19)$$

De la même manière, on déduit $\{X^{n+\frac{i}{m}}\}_{1 \leq i \leq m-1}$ par des processus successifs, on obtient

$$\begin{cases} \frac{d}{dt} X^{n+\frac{m}{m}}(t) + \mathcal{F}_m \cdot X^{n+\frac{m}{m}} = 0, & n\Delta t < t < (n+1)\Delta t, \\ X^{n+\frac{m}{m}}(n\Delta t) = X^{n+\frac{m-1}{m}}((n+1)\Delta t), \end{cases} \quad (3.20)$$

et on prend finalement

$$X^{n+1} = X^{n+1}((n+1)\Delta t).$$

Il a été montré (cf. Temam [59]) que la convergence, lorsque $\Delta t \rightarrow 0$, de la fonction X^n dans l'intervalle $(n\Delta t, (n+1)\Delta t)$, vers la solution X du problème initial, a bien lieu.

3.2.1.2 Approximation discrète

Les principes sont les même que précédemment mais les équations aux dérivées partielles sont remplacées par des équation aux différences finies.

Un point doit être bien noté, c'est la manière de bien choisir la forme discrète des opérateurs \mathcal{F}_i décomposés de \mathcal{F} pour assurer la positivité de chaque \mathcal{F}_i . Dans le cas quasi-linéaire ou non-linéaire, les opérateurs de différences de \mathcal{F}_i dépendent du temps et de la forme discrète de la variables X .

3.2.2 Etude de la stabilité des schémas (3.17)

Dans cette partie, on revient au schéma (3.17), on veut étudier la convergence et la condition de la stabilité du schéma pour l'approximation discrète. Si l'on suppose que,

pour $1 \leq i \leq m$, l'opérateur est positif au sens dans la section 3.1, on constate que la i -ème équation :

$$\frac{X^{n+\frac{i}{m}} - X^{n+\frac{i-1}{m}}}{\Delta t} + F_i^{(n+1,n)} \overline{X}_i^\theta = 0, \quad (3.21)$$

avec

$$\overline{X}_i^\theta = \theta X^{n+\frac{i}{m}} + (1-\theta) X^{n+\frac{i-1}{m}}, \quad \text{pour } 0 \leq \theta \leq 1,$$

possède la même forme que l'équation (3.7).

De la même manière, on établit un théorème de stabilité analogue et en même temps la convergence du schéma (3.17) :

Théorème 2 *Si l'on suppose pour chaque i , les opérateurs $F_i^{(n+1,n)}$ sont positifs, et si $\frac{1}{2} \leq \theta \leq 1$, le schéma (3.17) est stable au sens :*

$$\|X^{n+\frac{i}{m}}\| \leq C \|X^0\|, \quad C \text{ étant une constante indépendante de } \tau, h; \quad (3.22)$$

Si $\frac{1}{2} \leq \theta \leq 1$, il est convergent au sens suivant :

lorsque $\Delta t \rightarrow 0$, $X^{n+\frac{i}{m}} \rightarrow X$ dans $L^2(0, T; H)$ faible, où X est la solution du problème initial.

Démonstration

En utilisant la même méthode que pour démontrer le théorème 1, on peut obtenir facilement la stabilité du schéma au sens de (3.22), car on peut déduire la relation suivante de l'équation (3.21) :

$$\|X^{n+\frac{i}{m}}\| \leq \|X^{n+\frac{i-1}{m}}\| \leq \|X^0\|, \quad 1 \leq i \leq m. \quad (3.23)$$

On va démontrer la convergence.

En multipliant l'équation (3.21) par

$$\overline{X}_i^\theta = \theta X^{n+\frac{i}{m}} + (1-\theta) X^{n+\frac{i-1}{m}},$$

on en déduit que

$$\theta \|X^{n+\frac{i}{m}}\|^2 - (1-\theta) \|X^{n+\frac{i-1}{m}}\|^2 - (2\theta-1)(X^{n+\frac{i}{m}}, X^{n+\frac{i-1}{m}}) = -\Delta t (F_i^{(n+1,n)} \overline{X}_i^\theta, \overline{X}_i^\theta).$$

Du fait que $\frac{1}{2} \leq \theta \leq 1$ et $F_i^{(n+1,n)}$ est positif, en utilisant la relation

$$(a, a-b) = \frac{1}{2} (\|a\|^2 - \|b\|^2 + \|a-b\|^2),$$

on a

$$\frac{1}{2}\|X^{n+\frac{i}{m}}\|^2 - \frac{1}{2}\|X^{n+\frac{i-1}{m}}\|^2 + \left(\theta - \frac{1}{2}\right)\|X^{n+\frac{i}{m}} - X^{n+\frac{i-1}{m}}\|^2 \leq 0,$$

d'où

$$\sum_{i=1}^m \sum_{n=0}^N \|X^{n+\frac{i}{m}} - X^{n+\frac{i-1}{m}}\|^2 \Delta t \leq \frac{1}{2\theta - 1} \sum_{i,n} (\|X^{n+\frac{i}{m}}\|^2 - \|X^{n+\frac{i-1}{m}}\|^2) \leq C \Delta t,$$

dont C dépend de $\|X^0\|$.

Par conséquent, lorsque $\Delta t \rightarrow 0$,

$$X^{n+\frac{i}{m}} - X^{n+\frac{i-1}{m}} \rightarrow 0 \quad \text{dans } L^2(0, T; H).$$

Par extraction éventuellement de sous suite, on déduit alors de (3.22) que $X^{n+\frac{i}{m}}$ converge vers la même limite quel que soit i , d'où le théorème. ■

3.3 Méthode de prédiction-correction

La méthode prédicteur-correcteur peut être considérée comme une des multiples formes de la méthode à pas fractionnaires, bien qu'elle ait été appliquée antérieurement et indépendamment à des problèmes divers. Elle peut être appliquée pour la construction de schéma à corrections d'approximation, et dans le cas des équations non-linéaires, elle peut être aussi utilisée pour la construction de schémas conservatifs. L'utilisation d'un prédicteur pour les équations non-linéaires a été proposée récemment dans les travaux de Douglas [20], S.K. Godonov et K. A. Semendjaev [28], et Marchuk [43], N. N. Yanenko [69].

Le schéma prédicteur-correcteur, excepté la variant du choix de la forme, est principalement composé de deux parties, le prédicteur et le correcteur. On propose ici un schéma prédicteur-correcteur pour résoudre les équations (3.1) que l'on traite comme suit :

$$\frac{X^{n+\frac{1}{2}} - X^n}{\tau/2} + F(X^n)X^n = 0, \quad (3.24)$$

$$\frac{X^{n+1} - X^n}{\tau} + F(X^{n+\frac{1}{2}})\frac{X^{n+1} + X^n}{2} = 0. \quad (3.25)$$

L'équation (3.24) constitue le prédicteur donnant X au temps $(n + \frac{1}{2})\tau$ par un schéma explicite; l'équation (3.25) est le correcteur. Normalement la preuve de la stabilité du schéma n'est pas toujours évidente à cause des cas complexes, même pour un problème

linéaire. Nous allons prendre en compte ci-dessus la stabilité de ce schéma et la précision de l'approximation de la solution numérique vers la solution continue.

En ce qui concerne la stabilité, on suppose que $F(X^{n+\frac{1}{2}})$ est positif, vérifiant ici

$$\left(F(X^{n+\frac{1}{2}}) \cdot \left(\frac{X^{n+1} + X^n}{2} \right), \left(\frac{X^{n+1} + X^n}{2} \right) \right) \geq 0,$$

d'après le théorème 1, le schéma (3.25) est inconditionnellement stable au sens de :

$$\|X^{n+1}\| \leq \|X^n\|.$$

De plus, ce schéma a une précision de second ordre par rapport au temps. En fait, grâce à l'équation (3.24), on a

$$X^{n+\frac{1}{2}} = X^n - \frac{\tau}{2} F(X^n) \cdot X^n, \quad (3.26)$$

donc en identifiant

$$F^n = F(X^n),$$

et en supposant que l'opérateur F est assez régulier par rapport à la solution X , on écrit son développement de Taylor

$$\begin{aligned} F(X^{n+\frac{1}{2}}) &= F\left(X^n - \frac{\tau}{2} F^n X^n\right) \\ &= F^n - \frac{\tau}{2} \left[\frac{\partial F}{\partial X} \right]^n F^n X^n + O(\tau^2). \end{aligned} \quad (3.27)$$

De la même façon que la démonstration de la proposition 2, on déduit la précision en remplaçant dans (3.15) la discrétisation de l'opérateur $H_\tau^n X^n$ par

$$H_\tau^n X^n = \frac{X^{n+1} - X^n}{\tau} + F(X^{n+\frac{1}{2}}) \cdot \left(\frac{X^{n+1} + X^n}{2} \right)$$

on obtient ainsi

$$\|(HX)^n - H_\tau^n X^n\|_{C_\tau} = O(\tau^2).$$

3.4 Caractères du modèle direct discret

3.4.1 Choix du maillage

Avant de discrétiser en espace les équations (2.1)-(2.4) aux dérivées partielles qui gouvernent le système hydraulique avec une méthode de différences finies, on définit d'abord les variables physiques sur un maillage entrecroisé rectangulaire de type ARAKAWA -

maillage de “grille-C”. Dans le plan $x - y$, les coordonnées des points de grille sont définis par :

$$\begin{cases} x_I = I\Delta x, I = 0, 1, \dots, M; \\ y_J = J\Delta y, J = 0, 1, \dots, N; \end{cases} \quad (3.28)$$

avec Δx et Δy les pas en espace.

Avec ceci, la valeur d’une certaine variable X en un certain point s’écrit :

$$X(x_I, y_J) = X_{IJ}.$$

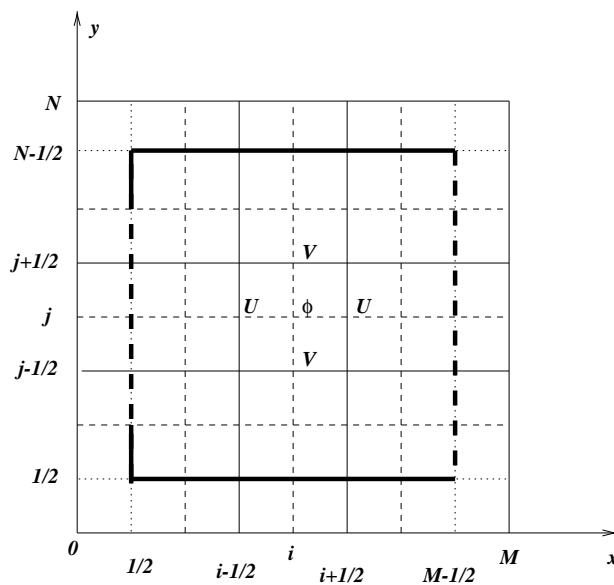


FIG. III.1 – Distribution des variables dans un maillage de type “grille-C”.

Dans ce maillage, on définit trois types de point. On définit un h -point par les coordonnées entières (i, j) , pour $i = 0, 1, \dots, M, j = 0, 1, \dots, N$. Ce sont les points où l’on calcule les approximations des variables h, ϕ, Φ, S et Z_b . Un u -point sera défini par les coordonnées $(i + 1/2, j)$, $i = 0, 1, \dots, M, j = 0, 1, \dots, N$, où les variables u et U seront approximées, et un v -point sera défini par les coordonnées $(i, j + 1/2)$, $i = 0, 1, \dots, M, j = 0, 1, \dots, N$, où les variables v, V seront calculées. La distribution de ces trois types de points est montrée dans la figure Fig.III.1.

Ce type de maillage s’appelle aussi un maillage de type “grille-C”. Il nous permet de choisir la position du domaine réel dans le maillage d’après les conditions que l’on donne

aux bords. Dans notre cas, les parois du Nord et du Sud passent respectivement les v -points de coordonnées $(i, N - 1/2)$ et $(i, 1/2)$, où $i = 0, 1, \dots, M$. Sur ces deux bords, il n'y a que des v -points, d'après les conditions aux bords données: $\vec{u} \cdot \vec{n} = 0$, on a $v = 0$ identiquement. On remarque alors que les approximations des variables u et h n'ont pas besoin d'être calculées sur ces deux frontières. Ça évite aussi le problème lié au comportement irrégulier sur les parois quand la profondeur de l'eau est égale à zéro ($h = 0$). Par contre, on suppose que l'entrée (Ouest) de l'écoulement passe par les u -points avec les coordonnées $(1/2, j)$, et la sortie (Est) passe par $(M - 1/2, j)$, $j = 0, 1, \dots, N$.

3.4.2 Discrétisation spatiale

Avec la définition du maillage de "grille-C", nous discrétisons le système de sédimentation en espace par un schéma de différences finies sur le grille Fig. III.1 avec le même pas d'espace dans les deux directions x et y . Puisque les variables sont situées de manière alternée sur le maillage, pour éviter les abus de langage, soit δh la discrétisation en espace, on exprime la discrétisation des dérivées partielles par rapport aux coordonnées des espaces sans distinguer les indices entiers et demi-entiers.

Si une certaine variable X n'est pas définie sur un certain noeud, on lui fournit une approximation en prenant la valeur moyenne des valeurs de la variable sur les noeuds voisins.

Si l'on pose

$$\begin{cases} I = i + \frac{1}{2} \text{ ou } i, & i = 0, 1, \dots, M, \\ J = j + \frac{1}{2} \text{ ou } j, & j = 0, 1, \dots, N, \end{cases} \quad (3.29)$$

on définit :

$$\begin{cases} \overline{X}_{I,J}^x = \frac{X_{I+\frac{1}{2},J} + X_{I-\frac{1}{2},J}}{2}, \\ \overline{X}_{I,J}^y = \frac{X_{I,J+\frac{1}{2}} + X_{I,J-\frac{1}{2}}}{2}, \\ \overline{X}_{I,J}^{xy} = \frac{X_{I+\frac{1}{2},J+\frac{1}{2}} + X_{I-\frac{1}{2},J+\frac{1}{2}} + X_{I+\frac{1}{2},J-\frac{1}{2}} + X_{I-\frac{1}{2},J-\frac{1}{2}}}{4}. \end{cases} \quad (3.30)$$

Par le même sens de (I, J) , nous posons

$$\left(\frac{\partial X}{\partial x}\right)_{I,J} \approx (X_x)_{I,J} = \frac{X_{I+\frac{1}{2},J} - X_{I-\frac{1}{2},J}}{\delta h}, \quad (3.31)$$

et

$$\left(\frac{\partial X}{\partial y}\right)_{I,J} \approx (X_y)_{I,J} = \frac{X_{I,J+\frac{1}{2}} - X_{I,J-\frac{1}{2}}}{\delta h}. \quad (3.32)$$

L'opérateur laplacien, est approché par le schéma classique à cinq points :

$$(\underline{\Delta}X)_{I,J} = \frac{X_{I+1,J} + X_{I-1,J} + X_{I,J+1} + X_{I,J-1} - 4X_{I,J}}{\delta h^2}. \quad (3.33)$$

On rappelle d'abord les équations gouvernant le système de sédimentation après le changement de variables donné au chapitre II :

$$\left\{ \begin{array}{l} \frac{\partial U}{\partial t} + \left(\frac{\partial(uU)}{\partial x} - \frac{U}{2} \frac{\partial u}{\partial x} \right) + \left(\frac{\partial(vU)}{\partial t} - \frac{U}{2} \frac{\partial v}{\partial y} \right) \\ \quad = fV - \Phi \frac{\partial \phi}{\partial x} + k_1 \Phi \Delta \left(\frac{U}{\Phi} \right) - g^2 C_D \sqrt{U^2 + V^2} \frac{U}{\Phi^3}, \\ \frac{\partial V}{\partial t} + \left(\frac{\partial(uV)}{\partial t} - \frac{V}{2} \frac{\partial u}{\partial x} \right) + \left(\frac{\partial(vV)}{\partial t} - \frac{V}{2} \frac{\partial v}{\partial y} \right) \\ \quad = -fU - \Phi \frac{\partial \phi}{\partial y} + k_1 \Phi \Delta \left(\frac{V}{\Phi} \right) - g^2 C_D \sqrt{U^2 + V^2} \frac{V}{\Phi^3}, \\ \frac{\partial \phi}{\partial t} + \frac{\partial(U\Phi)}{\partial x} + \frac{\partial(V\Phi)}{\partial y} = 0. \end{array} \right. \quad (3.34)$$

$$\frac{\partial(hS)}{\partial t} + \frac{\partial(uhS)}{\partial x} + \frac{\partial(vhS)}{\partial y} = k' \left(\frac{\partial}{\partial x} \left(h \frac{\partial S}{\partial x} \right) + \frac{\partial}{\partial y} \left(h \frac{\partial S}{\partial y} \right) \right) + \alpha \omega (S^* - S), \quad (3.35)$$

$$\rho_1 \frac{\partial Z_b}{\partial t} = \alpha \omega (S - S^*) - \left(\frac{\partial g_{bx}}{\partial x} + \frac{\partial g_{by}}{\partial y} \right). \quad (3.36)$$

Avec les notations (3.29)-(3.33), on discrétise ces équations en espace qui s'écrivent à l'intérieur du domaine :

$$\left\{ \begin{array}{l}
\left[\frac{\partial U}{\partial t} \right]_{i+\frac{1}{2},j} + \left[(\bar{u}^x \bar{U}^x)_x - \frac{1}{2} U \bar{u}_x^x \right]_{i+\frac{1}{2},j} + \left[(\bar{v}^y \bar{U}^y)_y - \frac{1}{2} U \bar{v}_y^y \right]_{i+\frac{1}{2},j} \\
= [f \bar{V}^{xy} - \bar{\Phi}^x \phi_x]_{i+\frac{1}{2},j} + \left[k_1 \bar{\Phi}^x \Delta \left(\frac{U}{\bar{\Phi}^x} \right) - \bar{F}_X^x \right]_{i+\frac{1}{2},j}, \\
\left[\frac{\partial V}{\partial t} \right]_{i,j+\frac{1}{2}} + \left[(\bar{u}^y \bar{V}^x)_x - \frac{1}{2} V \bar{u}_x^y \right]_{i,j+\frac{1}{2}} + \left[(\bar{v}^y \bar{V}^y)_y - \frac{1}{2} V \bar{v}_y^y \right]_{i,j+\frac{1}{2}} \\
= [-f \bar{U}^{xy} - \bar{\Phi}^y \phi_y]_{i,j+\frac{1}{2}} + \left[k_1 \bar{\Phi}^y \Delta \left(\frac{V}{\bar{\Phi}^y} \right) - \bar{F}_Y^y \right]_{i,j+\frac{1}{2}}, \\
\left[\frac{\partial \phi}{\partial t} \right]_{i,j} + [(U \bar{\Phi}^x)_x + (V \bar{\Phi}^y)_y]_{i,j} = 0,
\end{array} \right. \quad (3.37)$$

où

$$\left\{ \begin{array}{l}
\bar{F}_X^x = g^2 C_D \sqrt{U^2 + (\bar{V}^{xy})^2} \frac{U}{(\bar{\Phi}^x)^3}, \\
\bar{F}_Y^y = g^2 C_D \sqrt{(\bar{U}^{xy})^2 + V^2} \frac{V}{(\bar{\Phi}^y)^3}.
\end{array} \right.$$

Si l'on suppose que les variables h , S et Z_b se trouvent sur les h -points avec les coordonnées (i, j) , on peut discrétiser l'équation concernant la concentration de sédiments (3.35) par :

$$\left[\frac{\partial (hS)}{\partial t} \right]_{i,j} + [(\bar{h} S^x u)_x + (\bar{h} S^y v)_y]_{i,j} = k' [(\bar{h}^x S_x)_x + (\bar{h}^y S_y)_y]_{i,j} + [\alpha \omega (S^* - S)]_{i,j}, \quad (3.38)$$

et l'équation de l'évolution du lit de la rivière par (3.36) :

$$\rho_1 \left[\frac{\partial Z_b}{\partial t} \right]_{i,j} = -[\alpha \omega (S^* - S)]_{i,j} - [(g_{bx})_x + (g_{by})_y]_{i,j}, \quad (3.39)$$

où g_{bx} et g_{by} sont deux composantes du vecteur \vec{g}_b , leurs discrétisations dépendent de celles des variables qui les constituent.

Pour étudier les propriétés du schéma aux différences finies (3.37), on néglige les termes de la dissipation et du frottement du fond, en dénotant $X_h = (U_{i+\frac{1}{2},j}, V_{i,j+\frac{1}{2}}, \phi_{i,j})$, le système (3.37) s'écrit aussi de la manière que

$$\frac{\partial X_h}{\partial t} = F_h X_h.$$

En considérant le domaine borné décrit dans la figure Fig. III.1, nous allons montrer que

Théorème 3 *Si le domaine est fermé et les conditions aux bords satisfont :*

$$U_{\frac{1}{2},j} = U_{M-\frac{1}{2},j} = V_{i,\frac{1}{2}} = V_{i,N-\frac{1}{2}} = 0,$$

alors l'opérateur F_h est positif au sens de

$$(F_h X_h, X_h) = 0,$$

c'est-à-dire que, à l'intérieur du domaine, la somme vérifiant

$$\begin{aligned} & (F_h X_h, X_h) \\ = & \sum_{i=1}^{M-2} \sum_{j=1}^{N-1} \left\{ U_{i+\frac{1}{2},j} \left[(\bar{u}^x \bar{U}^x)_x - \frac{1}{2} U \bar{u}_x^x \right]_{i+\frac{1}{2},j} + U_{i+\frac{1}{2},j} \left[(\bar{v}^x \bar{U}^y)_y - \frac{1}{2} U \bar{v}_y^x \right]_{i+\frac{1}{2},j} \right. \\ & \left. - U_{i+\frac{1}{2},j} [f \bar{V}^{xy} - \bar{\Phi}^x \phi_x]_{i+\frac{1}{2},j} \right\} \\ + & \sum_{i=1}^{M-1} \sum_{j=1}^{N-2} \left\{ V_{i,j+\frac{1}{2}} \left[(\bar{u}^y \bar{V}^x)_x - \frac{1}{2} V \bar{u}_x^y \right]_{i,j+\frac{1}{2}} + V_{i,j+\frac{1}{2}} \left[(\bar{v}^y \bar{V}^y)_y - \frac{1}{2} V \bar{v}_y^y \right]_{i,j+\frac{1}{2}} \right. \\ & \left. + V_{i,j+\frac{1}{2}} [f \bar{U}^{xy} + \bar{\Phi}^y \phi_y]_{i,j+\frac{1}{2}} \right\} \\ + & \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \left\{ \phi_{i,j} [(U \bar{\Phi}^x)_x + (V \bar{\Phi}^y)_y]_{i,j} \right\} \\ = & 0. \end{aligned}$$

En vue de la démonstration, nous allons montrer les lemmes suivants.

3.4.3 Lemmes utiles

Lemme 3

$$\sum_{i=1}^{M-2} \sum_{j=1}^{N-1} U_{i+\frac{1}{2},j} \left[(\bar{u}^x \bar{U}^x)_x - \frac{1}{2} U \bar{u}_x^x \right]_{i+\frac{1}{2},j} = \text{les termes aux bords}.$$

Démonstration

Pour obtenir cette égalité, nous avons besoin de deux égalités suivantes :

$$[(\bar{u}^x \bar{U}^x)_x]_{i+\frac{1}{2},j} = [U \bar{u}_x^x]_{i+\frac{1}{2},j} + [\bar{u}^x U_x^x]_{i+\frac{1}{2},j}, \quad (3.40)$$

$$[(\bar{u}^x (\tilde{U}^x)^2)_x]_{i+\frac{1}{2},j} = [U^2 \bar{u}_x^x + 2U(\bar{u}^x U_x^x)]_{i+\frac{1}{2},j}, \quad (3.41)$$

où :

$$[(\tilde{U}^x)^2]_{l,j} = U_{l+\frac{1}{2},j} \cdot U_{l-\frac{1}{2},j}, \quad l = i + \frac{1}{2}, \text{ ou } i.$$

En fait, à partir des notations précédentes, le terme de gauche de l'égalité (3.40) devient,

$$\begin{aligned} & [(\bar{u}^x \bar{U}^x)_x]_{i+\frac{1}{2},j} \\ &= \frac{[\bar{u}^x \bar{U}^x]_{i+1,j} - [\bar{u}^x \bar{U}^x]_{i,j}}{h}, \\ &= \frac{1}{h} \left(\frac{u_{i+\frac{3}{2},j} + u_{i+\frac{1}{2},j}}{2} \frac{U_{i+\frac{3}{2},j} + U_{i+\frac{1}{2},j}}{2} - \frac{u_{i+\frac{1}{2},j} + u_{i-\frac{1}{2},j}}{2} \frac{U_{i+\frac{1}{2},j} + U_{i-\frac{1}{2},j}}{2} \right). \end{aligned}$$

De même, le terme de droite devient,

$$\begin{aligned} & [U \bar{u}_x^x]_{i+\frac{1}{2},j} + [\bar{u}^x U_x^x]_{i+\frac{1}{2},j} \\ &= U_{i+\frac{1}{2},j} \frac{[\bar{u}^x]_{i+1,j} - [\bar{u}^x]_{i,j}}{\delta h} + \frac{[\bar{u}^x U_x^x]_{i+1,j} + [\bar{u}^x U_x^x]_{i,j}}{2}, \\ &= \frac{U_{i+\frac{1}{2},j}}{\delta h} \left(\frac{u_{i+\frac{3}{2},j} + u_{i+\frac{1}{2},j}}{2} - \frac{u_{i+\frac{1}{2},j} + u_{i-\frac{1}{2},j}}{2} \right) \\ &\quad + \frac{1}{2} \left(\frac{u_{i+\frac{3}{2},j} + u_{i+\frac{1}{2},j}}{2} \frac{U_{i+\frac{3}{2},j} - U_{i+\frac{1}{2},j}}{\delta h} + \frac{u_{i+\frac{1}{2},j} + u_{i-\frac{1}{2},j}}{2} \frac{U_{i+\frac{1}{2},j} - U_{i-\frac{1}{2},j}}{\delta h} \right), \\ &= \frac{1}{2} \left(\frac{u_{i+\frac{3}{2},j} + u_{i+\frac{1}{2},j}}{2} \frac{U_{i+\frac{3}{2},j} + U_{i+\frac{1}{2},j}}{\delta h} + \frac{u_{i+\frac{1}{2},j} + u_{i-\frac{1}{2},j}}{2} \frac{U_{i+\frac{1}{2},j} + U_{i-\frac{1}{2},j}}{\delta h} \right), \end{aligned}$$

on en déduit donc l'égalité (3.40).

Regardons maintenant l'égalité (3.41), on a pour le terme de gauche,

$$\begin{aligned}
& \left[(\bar{u}^x (\tilde{U}^x)^2)_x \right]_{i+\frac{1}{2},j} \\
&= \frac{[\bar{u}^x (\tilde{U}^x)^2]_{i+1,j} - [\bar{u}^x (\tilde{U}^x)^2]_{i,j}}{\delta h}, \\
&= \frac{1}{\delta h} \left(\frac{u_{i+\frac{3}{2},j} + u_{i+\frac{1}{2},j}}{2} U_{i+\frac{3}{2},j} U_{i+\frac{1}{2},j} - \frac{u_{i+\frac{1}{2},j} + u_{i-\frac{1}{2},j}}{2} U_{i+\frac{1}{2},j} U_{i-\frac{1}{2},j} \right),
\end{aligned}$$

avec le résultat obtenu au pas précédent, on a pour le terme de droite

$$\begin{aligned}
& [U^2 \bar{u}_x + 2U(\overline{\bar{u}^x U_x^x})]_{i+\frac{1}{2},j}, \\
&= U_{i+\frac{1}{2},j}^2 \frac{u_{i+\frac{3}{2},j} - u_{i-\frac{1}{2},j} + u_{i+\frac{1}{2},j} - u_{i+\frac{1}{2},j}}{2\delta h} \\
&\quad + 2U_{i+\frac{1}{2},j} \frac{1}{2} \left(\frac{u_{i+\frac{3}{2},j} + u_{i+\frac{1}{2},j}}{2} \frac{U_{i+\frac{3}{2},j} - U_{i+\frac{1}{2},j}}{\delta h} \right. \\
&\quad \left. + \frac{u_{i+\frac{1}{2},j} + u_{i-\frac{1}{2},j}}{2} \frac{U_{i+\frac{1}{2},j} - U_{i-\frac{1}{2},j}}{\delta h} \right), \\
&= \frac{u_{i+\frac{3}{2},j} + u_{i+\frac{1}{2},j}}{2\delta h} \left(U_{i+\frac{1}{2},j}^2 + U_{i+\frac{1}{2},j} U_{i+\frac{3}{2},j} - U_{i+\frac{1}{2},j}^2 \right) \\
&\quad + \frac{u_{i+\frac{1}{2},j} + u_{i-\frac{1}{2},j}}{2\delta h} \left(-U_{i+\frac{1}{2},j}^2 + U_{i+\frac{1}{2},j}^2 - U_{i+\frac{1}{2},j} U_{i-\frac{1}{2},j} \right),
\end{aligned}$$

par comparaison de ces deux termes, on a démontré l'égalité (3.41).

Alors :

$$\begin{aligned}
U_{i+\frac{1}{2},j} \left[(\bar{u}^x \bar{U}^x)_x - \frac{1}{2} U \bar{u}_x^x \right]_{i+\frac{1}{2},j} &= \left[U^2 \bar{u}_x + U(\overline{\bar{u}^x U_x^x}) - \frac{1}{2} U^2 \bar{u}_x^x \right]_{i+\frac{1}{2},j} \\
&= \frac{1}{2} [U^2 \bar{u}_x + 2U(\overline{\bar{u}^x U_x^x})]_{i+\frac{1}{2},j} \tag{3.42} \\
&= \frac{1}{2} \sum_i \sum_j \left[(\bar{u}^x (\tilde{U}^x)^2)_x \right]_{i+\frac{1}{2},j}
\end{aligned}$$

Sommons maintenant en tous les indices i et j , avec $i = 1, \dots, M-2, j = 1, \dots, N-1,$

en appliquant ces deux égalités (3.40) et (3.41), nous obtenons :

$$\begin{aligned}
& U_{i+\frac{1}{2},j} \left[(\overline{u^x \overline{U^x}})_x - \frac{1}{2} U \overline{u^x} \right]_{i+\frac{1}{2},j} \\
&= \frac{1}{2\delta h} \sum_{j=1}^{N-1} \sum_{i=1}^{M-2} \left(U_{i+\frac{3}{2},j} U_{i+\frac{1}{2},j} \frac{u_{i+\frac{3}{2},j} + u_{i+\frac{1}{2},j}}{2} - U_{i+\frac{1}{2},j} U_{i-\frac{1}{2},j} \frac{u_{i+\frac{1}{2},j} + u_{i-\frac{1}{2},j}}{2} \right), \\
&= \frac{1}{2\delta h} \sum_{j=1}^{N-1} \left(U_{M-\frac{1}{2},j} U_{M-\frac{3}{2},j} \frac{u_{M-\frac{1}{2},j} + u_{M-\frac{3}{2},j}}{2} - U_{\frac{3}{2},j} U_{\frac{1}{2},j} \frac{u_{\frac{3}{2},j} + u_{\frac{1}{2},j}}{2} \right).
\end{aligned}$$

■

Lemme 4

$$\sum_{i=1}^{M-1} \sum_{j=1}^{N-2} V_{i,j+\frac{1}{2}} \left[(\overline{v^y \overline{V^y}})_y - \frac{1}{2} V \overline{v^y} \right]_{i,j+\frac{1}{2}} = \text{les termes aux bords} .$$

Démonstration

On prouve ce lemme de la même manière que le lemme 3, seulement en changeant les positions de u et v .

■

Lemme 5

$$\sum_{i=1}^{M-2} \sum_{j=1}^{N-1} U_{i+\frac{1}{2},j} \left[(\overline{v^x \overline{U^y}})_y - \frac{1}{2} U \overline{v^x} \right]_{i+\frac{1}{2},j} = \text{les termes aux bords} .$$

Démonstration

Pour prouver ce lemme, on va d'abord établir l'égalité suivante :

$$[(\overline{v^x \overline{U^y}})_y]_{i+\frac{1}{2},j} = [\overline{v^x} U]_{i+\frac{1}{2},j} + [\overline{U^y \overline{v^x}}]_{i+\frac{1}{2},j} . \quad (3.43)$$

En fait, si nous regardons le terme de gauche de cette égalité, on a

$$\begin{aligned} [(\overline{v^x \overline{U^y}})_y]_{i+\frac{1}{2},j} &= \frac{[\overline{v^x \overline{U^y}}]_{i+\frac{1}{2},j+\frac{1}{2}} - [\overline{v^x \overline{U^y}}]_{i+\frac{1}{2},j-\frac{1}{2}}}{\delta h}, \\ &= \frac{1}{\delta h} \left(\frac{v_{i+1,j+\frac{1}{2}} + v_{i,j+\frac{1}{2}}}{2} \frac{U_{i+\frac{1}{2},j+1} + U_{i+\frac{1}{2},j}}{2} \right. \\ &\quad \left. - \frac{v_{i+1,j-\frac{1}{2}} + v_{i,j-\frac{1}{2}}}{2} \frac{U_{i+\frac{1}{2},j} + U_{i+\frac{1}{2},j-1}}{2} \right). \end{aligned}$$

De l'autre coté de l'égalité (3.43), nous voyons que

$$\begin{aligned} [\overline{v_y^x U}]_{i+\frac{1}{2},j} &= U_{i+\frac{1}{2},j} \frac{[\overline{v^x}]_{i+\frac{1}{2},j+\frac{1}{2}} - [\overline{v^x}]_{i+\frac{1}{2},j-\frac{1}{2}}}{\delta h}, \\ &= \frac{U_{i+\frac{1}{2},j}}{\delta h} \left(\frac{v_{i+1,j+\frac{1}{2}} + v_{i,j+\frac{1}{2}}}{2} - \frac{v_{i+1,j-\frac{1}{2}} + v_{i,j-\frac{1}{2}}}{2} \right), \end{aligned}$$

et

$$\begin{aligned} [\overline{U_y \overline{v^{xy}}}]_{i+\frac{1}{2},j} &= \frac{[U_y \overline{v^x}]_{i+\frac{1}{2},j+\frac{1}{2}} + [U_y \overline{v^x}]_{i+\frac{1}{2},j-\frac{1}{2}}}{2}, \\ &= \frac{1}{2} \left(\frac{v_{i+1,j+\frac{1}{2}} + v_{i,j+\frac{1}{2}}}{2} \frac{U_{i+\frac{1}{2},j+1} - U_{i+\frac{1}{2},j}}{\delta h} \right. \\ &\quad \left. + \frac{v_{i+1,j-\frac{1}{2}} + v_{i,j-\frac{1}{2}}}{2} \frac{U_{i+\frac{1}{2},j} - U_{i+\frac{1}{2},j-1}}{\delta h} \right), \end{aligned}$$

leur sommes nous donnent :

$$\begin{aligned} [\overline{v_y^x U}]_{i+\frac{1}{2},j} + [\overline{U_y \overline{v^{xy}}}]_{i+\frac{1}{2},j} &= \frac{1}{\delta h} \left(\frac{v_{i+1,j+\frac{1}{2}} + v_{i,j+\frac{1}{2}}}{2} \frac{U_{i+\frac{1}{2},j+1} + U_{i+\frac{1}{2},j}}{2} \right. \\ &\quad \left. - \frac{v_{i+1,j-\frac{1}{2}} + v_{i,j-\frac{1}{2}}}{2} \frac{U_{i+\frac{1}{2},j} + U_{i+\frac{1}{2},j-1}}{2} \right), \end{aligned}$$

d'où l'égalité (3.43), avec laquelle le terme de gauche dans ce lemme devient

$$\begin{aligned}
U_{i+\frac{1}{2},j} \left[(\overline{v^x \tilde{U}^y})_y - \frac{1}{2} U \overline{v^x} \right]_{i+\frac{1}{2},j} &= \frac{1}{2} [U^2 \overline{v^x} + 2U \overline{U_y v^{xy}}]_{i+\frac{1}{2},j} \\
&= \frac{1}{2} [(\overline{v^x} (\tilde{U}^y)^2)_y]_{i+\frac{1}{2},j},
\end{aligned} \tag{3.44}$$

où $(\tilde{U}^y)_{i+\frac{1}{2},j}^2 = U_{i+\frac{1}{2},j+\frac{1}{2}} \cdot U_{i+\frac{1}{2},j-\frac{1}{2}}$.

La dernière égalité est établie grâce à

$$\begin{aligned}
[U^2 \overline{v^x} + 2U \overline{U_y v^{xy}}]_{i+\frac{1}{2},j} &= \frac{U_{i+\frac{1}{2},j}^2}{\delta h} \left(\frac{v_{i+1,j+\frac{1}{2}} + v_{i,j+\frac{1}{2}}}{2} - \frac{v_{i+1,j-\frac{1}{2}} + v_{i,j-\frac{1}{2}}}{2} \right) \\
&\quad + U_{i+\frac{1}{2},j} \left(\frac{U_{i+\frac{1}{2},j+1} - U_{i+\frac{1}{2},j}}{\delta h} \frac{v_{i+1,j+\frac{1}{2}} + v_{i,j+\frac{1}{2}}}{2} \right. \\
&\quad \left. + \frac{U_{i+\frac{1}{2},j} - U_{i+\frac{1}{2},j-1}}{\delta h} \frac{v_{i+1,j-\frac{1}{2}} + v_{i,j-\frac{1}{2}}}{2} \right), \\
&= \frac{v_{i+1,j+\frac{1}{2}} + v_{i,j+\frac{1}{2}}}{2} \frac{U_{i+\frac{1}{2},j} U_{i+\frac{1}{2},j+1}}{\delta h} \\
&\quad - \frac{v_{i+1,j-\frac{1}{2}} + v_{i,j-\frac{1}{2}}}{2} \frac{U_{i+\frac{1}{2},j} U_{i+\frac{1}{2},j-1}}{\delta h}, \\
&= \frac{[\overline{v^x} (\tilde{U}^y)^2]_{i+\frac{1}{2},j+\frac{1}{2}} - [\overline{v^x} (\tilde{U}^y)^2]_{i+\frac{1}{2},j-\frac{1}{2}}}{\delta h}, \\
&= [(\overline{v^x} (\tilde{U}^y)^2)_y]_{i+\frac{1}{2},j}.
\end{aligned}$$

Donc

$$\begin{aligned}
& \sum_{i=1}^{M-2} \sum_{j=1}^{N-1} U_{i+\frac{1}{2},j} \left[(\bar{v}^x \bar{U}^y)_y - \frac{1}{2} U \bar{v}_y^x \right]_{i+\frac{1}{2},j} \\
&= \frac{1}{2} \sum_{i=1}^{M-2} \sum_{j=1}^{N-1} \left(\frac{v_{i+1,j+\frac{1}{2}} + v_{i,j+\frac{1}{2}}}{2} \frac{U_{i+\frac{1}{2},j} U_{i+\frac{1}{2},j+1}}{\delta h} \right. \\
&\quad \left. - \frac{v_{i+1,j-\frac{1}{2}} + v_{i,j-\frac{1}{2}}}{2} \frac{U_{i+\frac{1}{2},j} U_{i+\frac{1}{2},j-1}}{\delta h} \right) \\
&= \frac{1}{2} \sum_{i=1}^{M-2} \left(\frac{v_{i+1,N-\frac{1}{2}} + v_{i,N-\frac{1}{2}}}{2} \frac{U_{i+\frac{1}{2},N-1} U_{i+\frac{1}{2},N}}{\delta h} \right. \\
&\quad \left. - \frac{v_{i+1,\frac{1}{2}} + v_{i,\frac{1}{2}}}{2} \frac{U_{i+\frac{1}{2},1} U_{i+\frac{1}{2},0}}{\delta h} \right).
\end{aligned}$$

■

Lemme 6

$$\sum_{i=1}^{M-1} \sum_{j=1}^{N-2} V_{i,j+\frac{1}{2}} \left[(\bar{u}^y \bar{V}^x)_x - \frac{1}{2} V \bar{u}_x^y \right]_{i,j+\frac{1}{2}} = \text{les termes aux bords} .$$

Démonstration

La démonstration de ce lemme et du lemme 5 sont similaires à cause de positions symétriques de u et v , U et V .

■

Pour la même raison, on va prouver un des deux lemmes suivants qui sont équivalents :

Lemme 7

$$\sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \phi_{i,j} [(V \bar{\Phi}^y)_y]_{i,j} + \sum_{i=1}^{M-1} \sum_{j=1}^{N-2} V_{i,j+\frac{1}{2}} [\bar{\Phi}^y \phi_y]_{i,j+\frac{1}{2}} = \text{les termes aux bords} .$$

Lemme 8

$$\sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \phi_{i,j} [(U \bar{\Phi}^x)_x]_{i,j} + \sum_{i=1}^{M-2} \sum_{j=1}^{N-1} U_{i+\frac{1}{2},j} [\bar{\Phi}^x \phi_x]_{i+\frac{1}{2},j} = \text{les termes aux bords} .$$

Démonstration

On prouve ici seulement le lemme 8.

En fait,

$$\begin{aligned} \phi_{i,j} [(U\bar{\Phi}^x)_x]_{i,j} &= \phi_{i,j} \frac{[U\bar{\Phi}^x]_{i+\frac{1}{2},j} - [U\bar{\Phi}^x]_{i-\frac{1}{2},j}}{\delta h} \\ &= \frac{\phi_{i,j}}{\delta h} \left[U_{i+\frac{1}{2},j} \frac{\Phi_{i+1,j} + \Phi_{i,j}}{2} - U_{i-\frac{1}{2},j} \frac{\Phi_{i,j} + \Phi_{i-1,j}}{2} \right], \end{aligned} \quad (3.45)$$

et

$$U_{i+\frac{1}{2},j} [\bar{\Phi}^x \phi_x]_{i+\frac{1}{2},j} = U_{i+\frac{1}{2},j} \frac{\Phi_{i+1,j} + \Phi_{i,j}}{2} \frac{\phi_{i+1,j} - \phi_{i,j}}{\delta h}, \quad (3.46)$$

donc

$$\begin{aligned} & \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \left(\phi_{i,j} [(U\bar{\Phi}^x)_x]_{i,j} \right) + \sum_{i=1}^{M-2} \sum_{j=1}^{N-1} \left(U_{i+\frac{1}{2},j} [\bar{\Phi}^x \phi_x]_{i+\frac{1}{2},j} \right) \\ &= \sum_{j=1}^{N-1} \sum_{i=1}^{M-1} \left(\frac{\phi_{i+1,j} U_{i+\frac{1}{2},j}}{\delta h} \frac{\Phi_{i+1,j} + \Phi_{i,j}}{2} - \frac{\phi_{i,j} U_{i-\frac{1}{2},j}}{\delta h} \frac{\Phi_{i,j} + \Phi_{i-1,j}}{2} \right) \\ & \quad - \sum_{j=1}^{N-1} \left(U_{M-\frac{1}{2},j} \frac{\Phi_{M,j} + \Phi_{M-1,j}}{2} \frac{\phi_{M,j} - \phi_{M-1,j}}{\delta h} \right) \\ &= \frac{1}{2\delta h} \sum_{j=1}^{N-1} \left(\phi_{M,j} U_{M-\frac{1}{2},j} \frac{\Phi_{M,j} + \Phi_{M-1,j}}{2} - \phi_{1,j} U_{\frac{1}{2},j} \frac{\Phi_{1,j} + \Phi_{0,j}}{2} \right) \\ &= \text{les termes aux bords.} \end{aligned}$$

■

Lemme 9

$$\sum_{i=1}^{M-2} \sum_{j=1}^{N-1} U_{i+\frac{1}{2},j} [f\bar{V}^{xy}]_{i+\frac{1}{2},j} - \sum_{i=1}^{M-1} \sum_{j=1}^{N-2} V_{i,j+\frac{1}{2}} [f\bar{U}^{xy}]_{i,j+\frac{1}{2}} = \text{les termes aux bords.}$$

Preuve

De même, f étant une constante, on constate que

$$\begin{aligned}
& \sum_{i=1}^{M-2} \sum_{j=1}^{N-1} U_{i+\frac{1}{2},j} [f\bar{V}^{xy}]_{i+\frac{1}{2},j} - \sum_{i=1}^{M-1} \sum_{j=1}^{N-2} V_{i,j+\frac{1}{2}} [f\bar{U}^{xy}]_{i,j+\frac{1}{2}} \\
&= \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \left(U_{i+\frac{1}{2},j} [f\bar{V}^{xy}]_{i+\frac{1}{2},j} - V_{i,j+\frac{1}{2}} [f\bar{U}^{xy}]_{i,j+\frac{1}{2}} \right) \\
& \quad - \sum_{j=1}^{N-1} U_{M-\frac{1}{2},j} [f\bar{V}^{xy}]_{M-\frac{1}{2},j} + \sum_{i=1}^{M-1} V_{i,N-\frac{1}{2}} [f\bar{U}^{xy}]_{i,N-\frac{1}{2}}
\end{aligned}$$

mais

$$\begin{aligned}
& \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \left(U_{i+\frac{1}{2},j} [f\bar{V}^{xy}]_{i+\frac{1}{2},j} - V_{i,j+\frac{1}{2}} [f\bar{U}^{xy}]_{i,j+\frac{1}{2}} \right) \\
&= \frac{f}{4} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \left(U_{i+\frac{1}{2},j} \frac{V_{i+1,j+\frac{1}{2}} + V_{i,j+\frac{1}{2}} + V_{i+1,j-\frac{1}{2}} + V_{i,j-\frac{1}{2}}}{4} \right. \\
& \quad \left. - V_{i,j+\frac{1}{2}} \frac{U_{i+\frac{1}{2},j+1} + U_{i-\frac{1}{2},j+1} + U_{i+\frac{1}{2},j} + U_{i-\frac{1}{2},j}}{4} \right) \\
&= \frac{f}{4} \sum_{j=1}^{N-1} \left[V_{M,j+\frac{1}{2}} (U_{M-\frac{1}{2},j} + U_{M-\frac{1}{2},j+1}) - V_{1,j+\frac{1}{2}} (U_{\frac{1}{2},j} + U_{\frac{1}{2},j+1}) \right] \\
& \quad + \frac{f}{4} \sum_{i=1}^{M-1} \left[U_{i+\frac{1}{2},1} (V_{i,\frac{1}{2}} + V_{i+1,\frac{1}{2}}) - U_{i+\frac{1}{2},N} (V_{i,N-\frac{1}{2}} + V_{i+1,N-\frac{1}{2}}) \right],
\end{aligned}$$

d'où le lemme. ■

En conséquence, le théorème 3, c'est-à-dire la positivité de l'opérateur aux différences finies (3.40) peut être conduite de ces lemmes. Cette propriété nous permet d'appliquer les résultats de stabilité au début de ce chapitre pour analyser la stabilité du schéma (3.37).

3.4.4 Discrétisation temporelle

En définissant les séquences de temps

$$t_n = n\Delta t, n = 0, 1, \dots$$

avec Δt le pas en temps, X^n désignant l'approximation à l'instant t_n .

Nous utilisons un schéma implicite pour la discrétisation en temps :

$$\frac{X^{n+1} - X^n}{\Delta t} + F\left(\frac{X^{n+1} + X^n}{2}\right) \cdot \frac{X^{n+1} + X^n}{2} = 0.$$

c'est-à-dire, si l'on pose

$$X_t = \frac{X^{n+1} - X^n}{\Delta t}, \quad \bar{X} = \frac{X^{n+1} + X^n}{2},$$

on remplace dans les équations discrètes (3.37) U par \bar{U} , V par \bar{V} et ϕ par $\bar{\phi}$. D'après les résultats de stabilité au début de ce chapitre, et grâce au théorème 3, l'opérateur de différences gouvernant le système d'évolution discret est positif, qui entraîne que le schéma choisi est inconditionnellement stable et de second ordre de précision par rapport au temps.

Pour calculer l'énergie du système discret, on multiplie $\bar{U}_{i+\frac{1}{2},j}$, $\bar{V}_{i,j+\frac{1}{2}}$ et $\bar{\phi}_{i,j}$ séparément par chacune des équations (3.37), et grâce aux lemmes 3 - lemme 9, nous obtenons

Théorème 4 *L'expression de la variation temporelle de l'énergie du système discret s'écrit*

$$\begin{aligned} & \sum_{i=1}^{M-2} \sum_{j=1}^{N-1} [\bar{U}U_t]_{i+\frac{1}{2},j} + \sum_{i=1}^{M-1} \sum_{j=1}^{N-2} [\bar{V}V_t]_{i,j+\frac{1}{2}} + \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} [\bar{\phi}\phi_t]_{i,j} \\ = & \sum_{i=1}^{M-2} \sum_{j=1}^{N-1} \bar{U}_{i+\frac{1}{2},j} \left[k_1 \bar{\Phi}^x \Delta \left(\frac{U}{\bar{\Phi}^x} \right) - \bar{F}_X^x \right]_{i+\frac{1}{2},j} + \sum_{i=1}^{M-1} \sum_{j=1}^{N-2} \bar{V}_{i,j+\frac{1}{2}} \left[k_1 \bar{\Phi}^y \Delta \left(\frac{V}{\bar{\Phi}^y} \right) - \bar{F}_Y^y \right]_{i,j+\frac{1}{2}} \end{aligned} \quad (3.47)$$

A partir de cette expression, on trouve que si l'on ne prend pas en compte les termes du frottement au fond et de la dissipation turbulente (le laplacien), on obtient la conservation de l'énergie du système discret de Saint-Venant :

$$\sum_{i=1}^{M-2} \sum_{j=1}^{N-1} (U_{i+\frac{1}{2},j}^n)^2 + \sum_{i=1}^{M-1} \sum_{j=1}^{N-2} (V_{i,j+\frac{1}{2}}^n)^2 + \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} (\phi_{i,j}^n)^2 = \text{constante}. \quad (3.48)$$

Cela signifie que les équations discrétisées possèdent la même propriété de conservation que les équations différentielles continues, et que le schéma de discrétisation a l'avantage d'être conservatif en énergie. Si le domaine n'est pas fermé, le schéma sera aussi conservatif en énergie au sens que : le flux entrant d'énergie est égal au flux sortant. Cela donne un critère pour tester la solution numérique du problème envisagé.

3.5 Résolution numérique du problème

3.5.1 Application de la méthode de “time-splitting”

Grâce au fait que l'évolution du lit d'une rivière est le résultat d'interaction entre l'écoulement diphasique et le lit, et que la vitesse de l'écoulement et la cote de l'eau sont à la fois influencées par la déformation du fond de la rivière, il conduit à beaucoup de difficultés pour la construction des équations fondamentales, la résolution numérique, ainsi le choix des équations complémentaires.

Dans le cadre bidimensionnel, surtout pour les problèmes lorsque la profondeur de l'eau est assez petite et qu'il ne contient pas beaucoup de sédiments dans le fluide, on propose souvent des méthodes non-couplées. Mais dans le cas où la concentration de sédiments de l'écoulement est assez grande, les influences de la variation de la géométrie du lit de la rivière et de la concentration ne peuvent pas être négligées. En plus, d'après l'étude faite au chapitre I, on tient compte du fait que par rapport à la variation du lit de la rivière, les échelles de temps en ce qui concerne l'écoulement et le transports solide sont beaucoup plus courtes. Si l'on prend un petit pas de temps pour résoudre toutes les équations, ça va coûter très cher. Par conséquent, nous allons adopter une méthode de “time-splitting” qui implicitement mettre en œuvre l'interaction entre le solide et le liquide, que nous pouvons réaliser itérativement par trois étapes en prenant trois pas de temps différents pour :

- Dynamique du fluide (avec pas de temps Δt_f);
- Transport des sédiments (avec pas de temps Δt_s);
- Déformation du fond de la rivière (avec pas de temps Δt_z);

dont le pas de temps pour la déformation est beaucoup plus grand que les deux autres :

$$\Delta t_z \gg \Delta t_f \geq \Delta t_s,$$

A chaque date T :

- Dans la première étape, en fixant les variables liées à la concentration de sédiments S et à la cote du fond Z_b , on intègre l'équations de l'écoulement (3.37) par le pas de temps Δt_f jusqu'à ce que le comportement de l'écoulement atteignent un état d'équilibre;
- Dans la deuxième étape, prenant de la première étape les valeurs du champ de vitesse \vec{v} et la hauteur de l'eau h , on intègre l'équation du transport des sédiments (3.38) jusqu'à un état d'équilibre;

- Dans la troisième étape, en utilisant toute les variables hydrauliques obtenues et en intégrant de l'équation de la variation du lit (3.39), on obtient un nouveau fond Z_b à la date prochaine $T + 1$.

3.5.1.1 Application du schéma à pas fractionnaires

En ce qui concerne la résolution de l'équation dynamique (3.37), on va mettre en œuvre l'utilisation de la méthode à pas fractionnaires (voir la section 3.2), qui s'inspire de la méthode de décomposition de opérateurs. On décompose le système dynamique par trois processus :

- Processus géostrophique ;
- Processus d'advection ;
- Processus de dissipation ;

Le schéma est adopté de la manière suivante : dans l'intervalle $[t_n, t_{n+1}]$ ($t_{n+1} - t_n = \Delta t_f$ qui est le pas de temps), supposant $X^n = (U^n, V^n, \phi^n)$ la variable discrète au temps t_n connu, on détermine X^{n+1} par trois processus :

(1) Processus géostrophique

$$\left\{ \begin{array}{l} \left[\frac{\partial U}{\partial t} \right]_{i+\frac{1}{2}, j} = [(f\bar{V}^{xy} - \bar{\Phi}^x \phi_x)]_{i, j+\frac{1}{2}}, \quad t \in]t_n, t_{n+1}[\\ \left[\frac{\partial V}{\partial t} \right]_{i, j+\frac{1}{2}} = [(-f\bar{U}^{xy} - \bar{\Phi}^y \phi_y)]_{i, j+\frac{1}{2}}, \\ \left[\frac{\partial \phi}{\partial t} \right]_{i, j} = [(U\bar{\Phi}^x)_x + (V\bar{\Phi}^y)_y]_{i, j}, \\ U(t_n) = U^n, V(t_n) = V^n, \phi(t_n) = \phi^n, \end{array} \right. \quad (3.49)$$

qui donne $U^I = U(t_{n+1})$, $V^I = V(t_{n+1})$, et $\phi^I = \phi(t_{n+1})$;

(2) Processus d'advection

$$\left\{ \begin{array}{l} \left[\frac{\partial U}{\partial t} \right]_{i+\frac{1}{2},j} = \left[(\bar{u}^x \bar{U}^x)_x - \frac{1}{2} U \bar{u}_x^x \right]_{i,j+\frac{1}{2}} + \left[(\bar{v}^x \bar{U}^y)_y - \frac{1}{2} U \bar{v}_y^x \right]_{i,j+\frac{1}{2}}, \quad t \in]t_n, t_{n+1}] \\ \left[\frac{\partial V}{\partial t} \right]_{i,j+\frac{1}{2}} = \left[(\bar{u}^y \bar{V}^x)_x - \frac{1}{2} V \bar{u}_x^y \right]_{i,j+\frac{1}{2}} + \left[(\bar{v}^y \bar{V}^y)_y - \frac{1}{2} V \bar{v}_y^y \right]_{i,j+\frac{1}{2}}, \\ \left[\frac{\partial \phi}{\partial t} \right]_{i,j} = 0, \\ U(t_n) = U^I, V(t_n) = V^I, \phi(t_n) = \phi^I, \end{array} \right. \quad (3.50)$$

qui donne $U^{II} = U(t_{n+1})$, $V^{II} = V(t_{n+1})$, et $\phi^{II} = \phi(t_{n+1})$, les conditions initiales sont celles données par le résultat du processus géostrophique ;

(3) Processus de dissipation

$$\left\{ \begin{array}{l} \left[\frac{\partial U}{\partial t} \right]_{i+\frac{1}{2},j} = \left[k_1 \bar{\Phi}^x \Delta \left(\frac{U}{\bar{\Phi}^x} \right) - \bar{F}_X^x \right]_{i+\frac{1}{2},j}, \quad t \in]t_n, t_{n+1}] \\ \left[\frac{\partial V}{\partial t} \right]_{i,j+\frac{1}{2}} = \left[k_1 \bar{\Phi}^y \Delta \left(\frac{V}{\bar{\Phi}^y} \right) - \bar{F}_Y^y \right]_{i,j+\frac{1}{2}}, \\ \left[\frac{\partial \phi}{\partial t} \right]_{i,j} = 0, \\ U(t_n) = U^{II}, V(t_n) = V^{II}, \phi(t_n) = \phi^{II}, \end{array} \right. \quad (3.51)$$

qui donne $U^{n+1} = U(t_{n+1})$, $V^{n+1} = V(t_{n+1})$ et $\phi^{n+1} = \phi(t_{n+1})$.

Il doit être remarqué que les valeurs de u , v , et Φ doivent être modifiées à la fin de chaque processus par le changement de variables :

$$\left\{ \begin{array}{l} u_{i+\frac{1}{2},j} = \left[\frac{U}{\bar{\Phi}^x} \right]_{i+\frac{1}{2},j} = \frac{2U_{i+\frac{1}{2},j}}{\Phi_{i+1,j} + \Phi_{i,j}}, \\ v_{i,j+\frac{1}{2}} = \left[\frac{V}{\bar{\Phi}^y} \right]_{i,j+\frac{1}{2}} = \frac{2V_{i,j+\frac{1}{2}}}{\Phi_{i,j+1} + \Phi_{i,j}}, \\ \Phi_{i,j} = [\sqrt{\phi - gZ_b}]_{i,j}. \end{array} \right. \quad (3.52)$$

Sachant qu'il s'agit de trois processus de vitesses différentes, on choisit des pas de temps différents pour chaque processus. En général, le processus de géostrophique varie beaucoup plus vite que les deux autres processus, pour permettre un calcul plus précis, on a adopté un pas de temps plus petit que les autres processus. Par cette méthode, on prend un pas de temps Δt_1 pour le processus d'advection et pour le processus de dissipation, et un pas de temps Δt_2 plus petit pour le processus géostrophique, avec

$$\Delta t_2 = \frac{\Delta t_1}{N_g},$$

où N_g un entier positif.

Comme étant un corollaire du théorème 3 et 4, on peut prouver pour les processus géostrophique et d'advection les propriétés de la conservation de l'énergie.

En résumé, on décompose le système de la sédimentation en trois processus : le dynamique du fluide et le transport du solide correspondant aux processus rapides, la déformation géométrique du fond correspondant au processus très lent, et le dynamique du fluide est elle-même décomposé en 3 processus. Il nous permet de résoudre les équations envisagées par des schémas numériques d'un ordre de précision plus élevé et à la fois d'économiser le temps de calcul. En revanche, la décomposition dépend beaucoup de la connaissance de chaque équation concrète, sinon, il entraînera une mauvaise modélisation.

3.5.2 Description de l'expérimentation et des paramètres

Prenons un exemple analogue à celui dans Zeng [73], on a appliqué le schéma numérique décrit précédemment à la simulation numérique de la sédimentation et à l'évolution d'un delta, nous avons testé la stabilité de ce schéma pour une longue période de calcul.

On fait ici une expérience numérique sur un domaine d'un lac-rivière (comme sur Fig. III.2), qui est de 100 km de long et de 100 km de large. Dans la figure Fig. III.4, on donne une forme initiale pour le lit, où il y a un canal central dans la rivière avec une pente descendante uniforme (voir Fig. III.3). On discrétise ce domaine par des pas d'espace : $\Delta x = \Delta y = 2\text{km}$ et les pas de temps $\Delta t_f = 140\text{sec}$ pour les processus de dissipation et d'advection dans l'écoulement, si $N_g = 5$, donc soit un pas de temps 28sec pour le processus de géostrophique ; $\Delta t_s = 70\text{sec}$ pour le transport des sédiments et $\Delta t_z = 3\text{jour}$ pour la variation du lit fluvial.

A l'entrée du fluide, on donne un flux dont $\vec{v} \cdot \vec{n} = u_0 = 2\text{m} \cdot \text{sec}^{-1}$, avec les valeurs connues de la profondeur h qui est de 10 mètres, et la hauteur du fond Z_b est de 2 mètres, la concentration de sédiments est de $1.2 \text{kg} \cdot \text{m}^{-3}$, ce qui assure un débit de l'écoulement

ainsi qu'une source de sédiments constante.

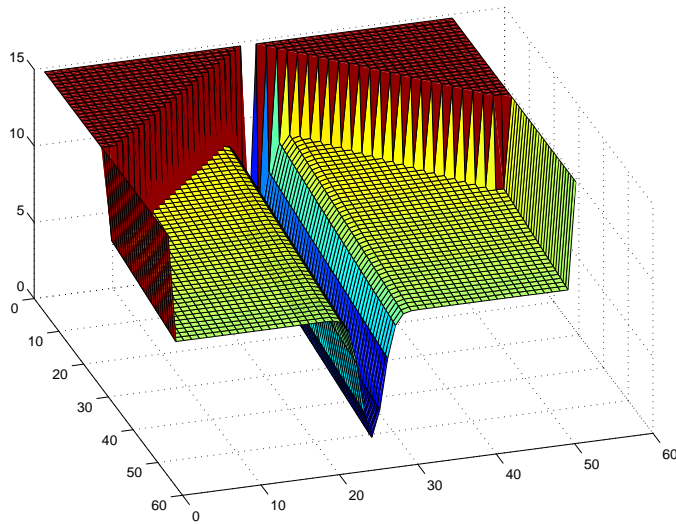


FIG. III.2 – *Le domaine lac-rivière du modèle.*

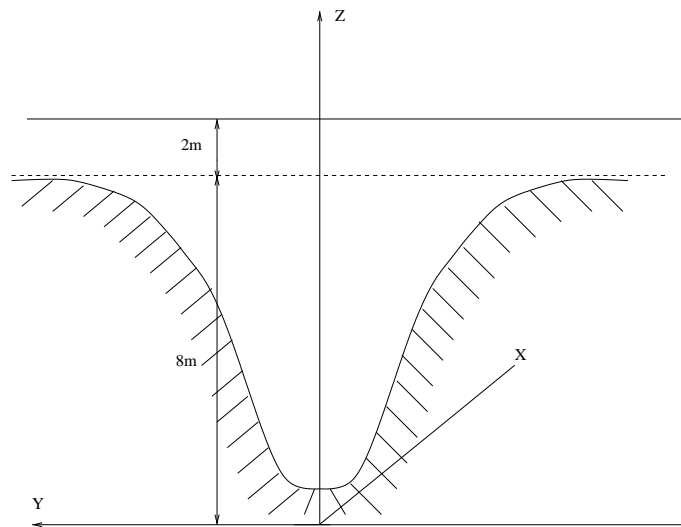


FIG. III.3 – *La géométrie initiale du lit de la rivière dans la section verticale de direction de l'écoulement.*

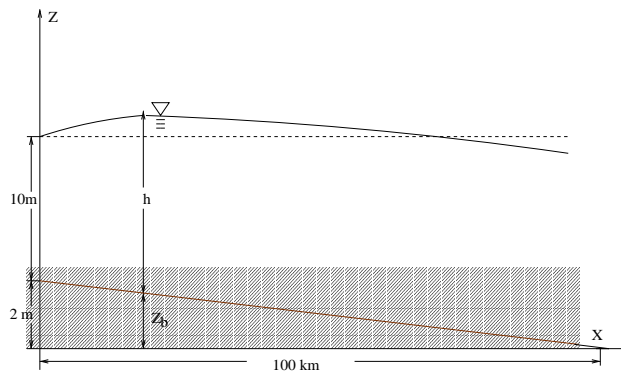


FIG. III.4 – *Le profil initial du canal de navigation au centre du lac .*

On donne ci-dessous un tableau des paramètres (voir Tab. III.1) que l'on a utilisé dans le modèle de simulation. Pour déterminer la capacité du transport solide S^* et \vec{g}_b qui sont données par les formules empiriques mentionnées au Chapitre II, on choisit des paramètres qui proviennent de l'observation dans la section au milieu du fleuve Yangzé [19]:

paramètre	valeur	unité
f	1.0E-4	1/s
g	9.8	m/s ²
k_1	1.25E2	m ² /s
C_D	5.0E-3	-
k'	1.0	m ⁴ /s
ω	1.0E-2	m/s
ρ_1	1.4E3	kg/m ³
α	0.3	-
K	0.35	-
m	0.92	-
v_c	0	m/s
n	3	-
p	-0.25	-
q	0.25	-
d	1.0E-3	m

TAB. III.1 – *Tableau des paramètres*

3.5.3 Algorithme de la simulation

A partir des résultats numériques, on remarque que l'évolution du lit de la rivière est un processus très lent ce qui est en accord avec l'évolution de l'environnement naturel en comparaison de l'évolution du champ de vitesse de l'écoulement et de la concentration de sédiments en suspension dans l'eau. On a testé l'algorithme de simulation de type "time-splitting".

Si l'on fixe d'abord la forme du lit et si l'on calcule l'écoulement à partir d'un état initial au repos, on va voir que, après une période d'environ 16 heures, le champ de vitesse atteint une situation d'équilibre comme il est décrit dans la Fig. III.5, selon le critère que la différence de la valeur maximum du champ de vitesse entre deux pas de temps successif est assez petite :

$$\max_{\Omega} \left| |\vec{u}^{n+1}| - |\vec{u}^n| \right| \leq \epsilon, \quad (3.53)$$

ici, on prend $\epsilon = 5 \times 10^{-4}$. Et $|\vec{u}^n|$ désigne la vitesse du fluide au temps t_n .

De même, on constate que la distribution de la concentration de sédiments en suspension atteint une situation de balance, après à peu près 10 heures (Fig. III.6), selon un critère d'équilibre similaire.

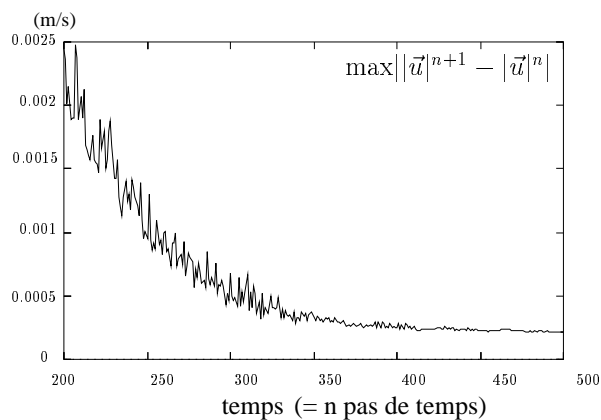


FIG. III.5 – La différence maximale du champ de vitesse entre deux pas de temps successifs après 200 pas de temps.

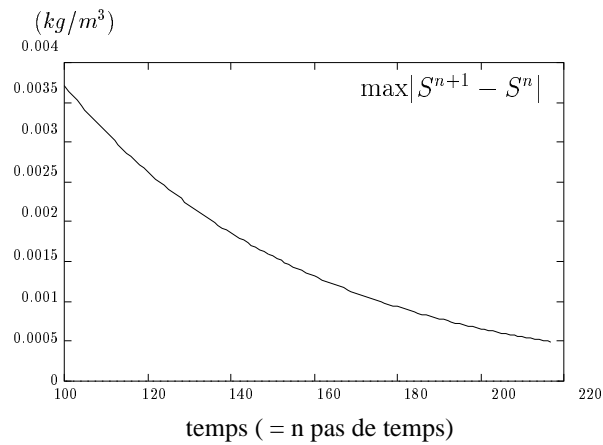


FIG. III.6 – La différence maximum du champ de concentration de sédiments entre deux pas de temps successifs après 100 pas de temps.

Cependant, on constate de nos expériences que une petites variations de la surface du lit (on prend 0.25 mètre dans notre modèle) sont excitées pour une durée approximative de 3 jours. Ces résultats indiquent que la méthode de "time-splitting" assure un calcul assez précis. Il faut remarquer que, pour donner un résultat plus visible, les paramètres sont choisis pour accélérer le processus de la sédimentation.

3.5.4 Résultats numériques

3.5.4.1 Expérience I – simulation de delta

A partir des résultats numériques de simulation sur le domaine que l'on considère, on voit que le problème modélisé correspond à un processus d'évolution de delta, processus qui est très non-linéaire.

Au début, on remarque sur le champ de vitesse (Fig.III.7(a)) que la vitesse est très petite sauf dans la région près de l'entrée de l'écoulement, on voit aussi sur la figure (Fig. III.7(b)) que l'élévation de la surface libre de l'eau a une pente descendante le long de la direction Ouest-Est, ceci pour compenser la dissipation de l'énergie cinétique à cause de la friction sur le fond. La non-symétrie résulte de l'effet de la force de Coriolis.

Les sédiments en suspension dans l'eau se déposent à l'entrée de la rivière, il en résulte l'augmentation de l'élévation du lit dans cette région-là. Avec l'agglomération des sédiments, la profondeur de l'eau devient plus petite, par conséquent, la vitesse de l'eau

devient plus grande. Lorsque la vitesse de l'écoulement a passé une vitesse critique, les sédiments sur le lit seront enlevés et transportés en aval et dans les régions latérales du canal, où les sédiments se déposent très vite.

Après une simulation de 50 années, les résultats montrent qu'un delta se forme avec une île isolée sous marin dans son début et un canal au centre (voir Fig. III.8, aussi Fig. III.9(d)). La hauteur de l'eau sur le delta est très petite, ceci est lié au fait que le champ de vitesse a considérablement augmenté (Fig. III.9(a)) et que le niveau de la surface de l'eau a augmenté à peu près 1 mètre (Fig. III.9(b)). La distribution de la concentration de sédiments est représentée par Fig. III.9(c) où l'on remarque que la concentration est importante près du canal central du delta. Tous ceci correspond à des phénomènes que l'on peut observer dans la nature.

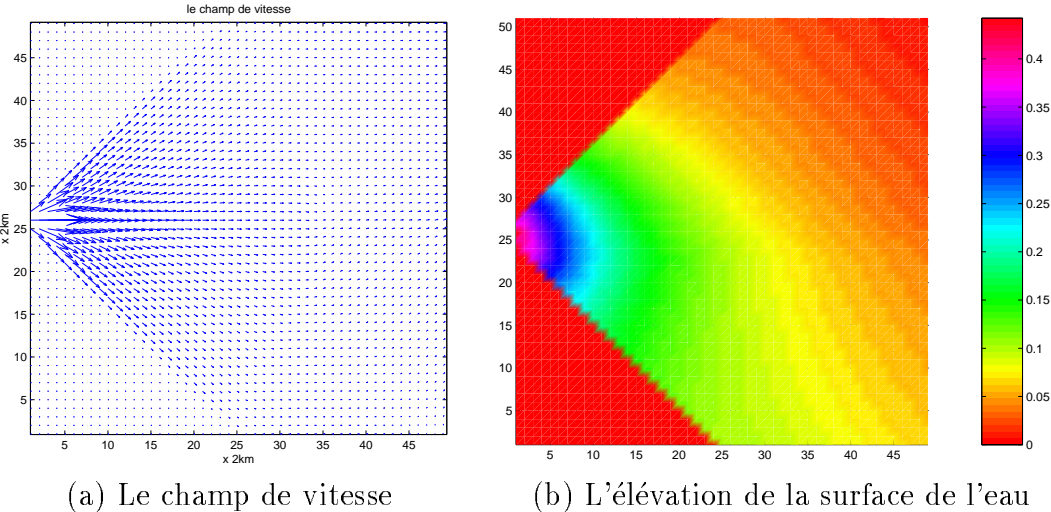


FIG. III.7 – L'état stationnaire du fleuve au début de la simulation.

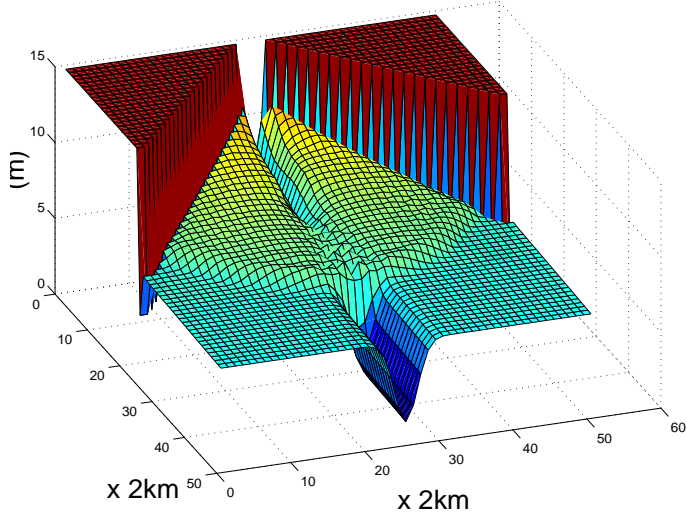


FIG. III.8 – Le fond de la rivière obtenu après une simulation de 50 années .

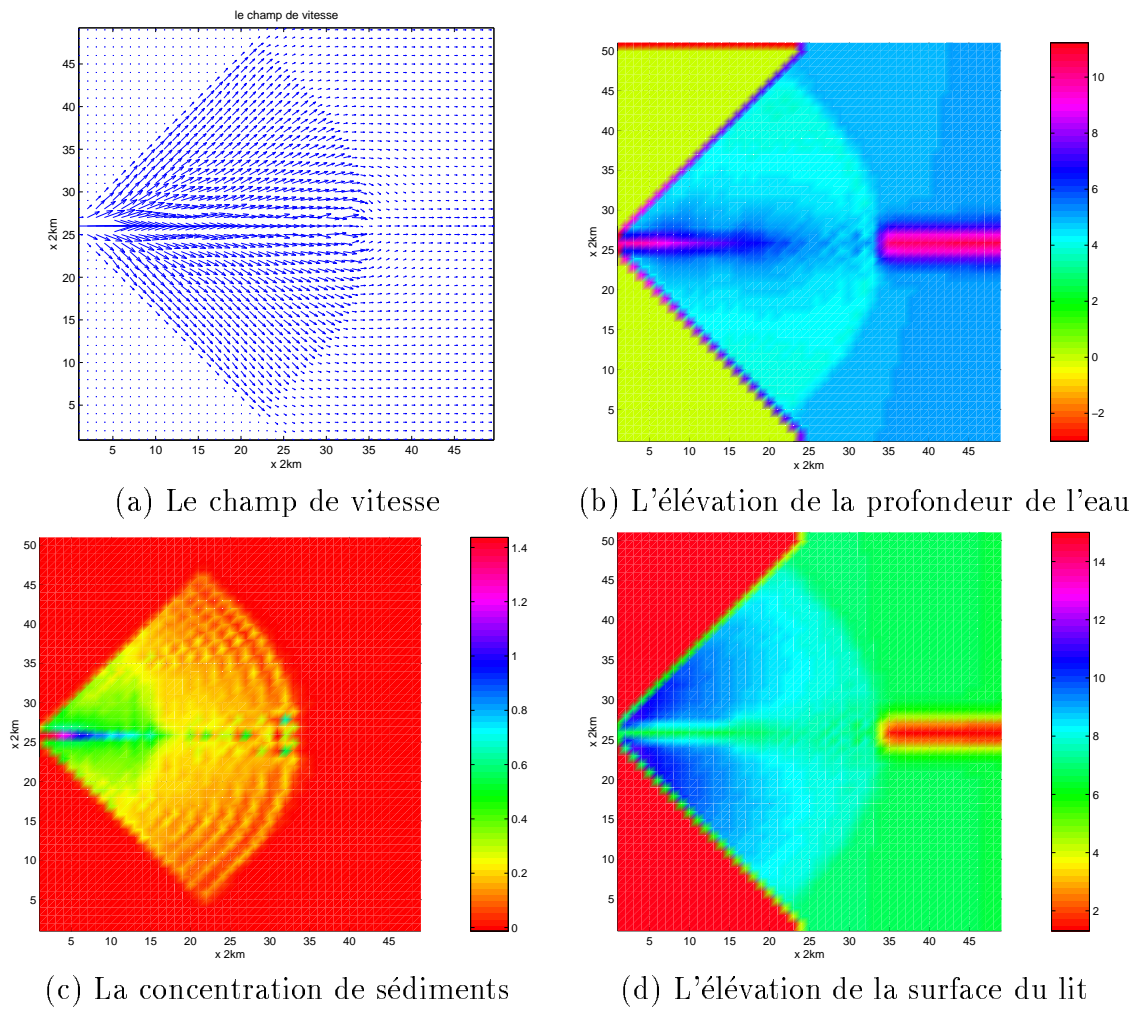


FIG. III.9 – L'état de l'écoulement et la distribution de la concentration de sédiments en rivière après une simulation de 50 années.

3.5.4.2 Expérience II

Puisque la sédimentation en rivière est un très lent processus naturel, nous allons voir qu'après une longue période de temps, c'est l'état initial du fond de la rivière qui joue le rôle dominant dans le processus d'évolution de la rivière, et qu'au contraire, le champ de vitesse et la concentration de sédiments ont seulement très peu de rapport avec leur situations initiales et peuvent être déterminés par la variation du lit obtenue récemment. Ceci est prouvé par les résultats numériques ci-dessous.

Dans les deux expériences ci-dessous, en utilisant les mêmes conditions aux limites que l'expérience I, on a proposé deux conditions initiales différentes. Dans le cas I, les champs

initiaux de l'écoulement et de la distribution des sédiments sont donnés par les résultats de la simulation de 50 années. Dans le cas II, l'état initial du champ de l'écoulement est un état au repos: $\max_{\Omega} |\vec{v}| = \delta, 0 < \delta \ll 1$, et la concentration des sédiments $S = S_c$ est constante. A partir de ces états initiaux et de la même hauteur du fond de la rivière, leurs états d'équilibre sont comparés séparément dans le canal où la sédimentation est la plus importante (voir les Fig III.10 et Fig III.11). Plus précisément, les différences entre ces deux cas sont évaluées par

$$\text{Erreur1} = \max_{1 \leq i \leq 50} \frac{|u1_{i,j_c} - u2_{i,j_c}|}{|u1_{i,j_c}|}$$

et

$$\text{Erreur2} = \max_{1 \leq i \leq 50} \frac{|S1_{i,j_c} - S2_{i,j_c}|}{|S1_{i,j_c}|}$$

où $u1_{i,j_c}, u2_{i,j_c}$ sont respectivement les valeurs absolues de la vitesse de l'écoulement dans le cas I et le cas II aux noeuds (i, j_c) , ici, $j_c=26$; $S1_{i,j_c}$ et $S2_{i,j_c}$ désignent la concentration de sédiments aux mêmes noeuds.

A partir de ces deux figures, on obtient

$$\text{Erreur1} = 0.693 \%,$$

$$\text{Erreur2} = 1.356 \%.$$

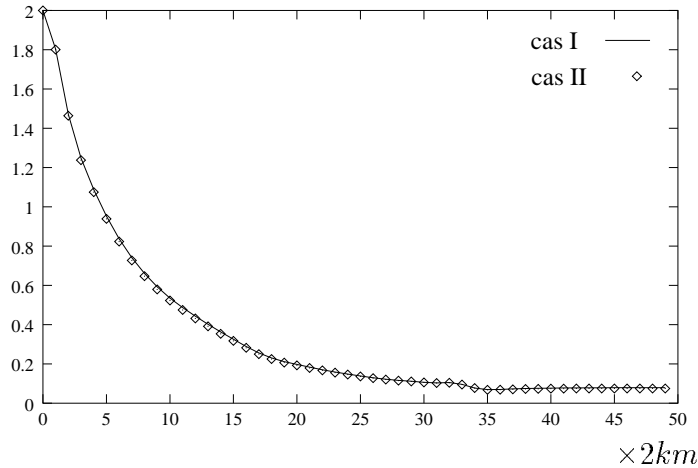


FIG. III.10 – Le résultat comparatif de l'état d'équilibre du champ de vitesse dans le canal à partir de différentes conditions initiales du champ de vitesse (l'unité est de m/s).

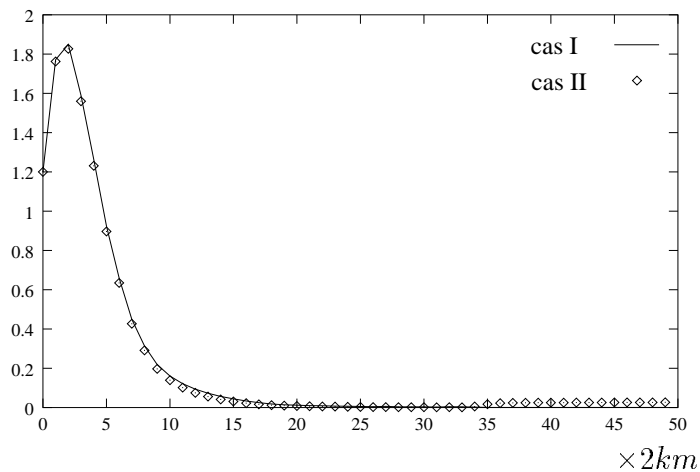


FIG. III.11 – *Le résultat comparatif de l'état d'équilibre de la distribution de concentration de sédiments dans le canal à partir de différentes conditions initiales (l'unité est de kg/m^3).*

Cette expérience est utile notamment pour un problème d'assimilation de données - contrôle des conditions initiales - que nous allons traiter dans le chapitre suivant. D'après cette expérimentation, nous avons remarqué que le contrôle de la condition initiale de la variation du lit de la rivière est plus important que celui du champ de l'écoulement et de la distribution de la concentration de sédiments en rivière.

3.5.5 Conclusion

Dans cette partie, un schéma numérique et son analyse de stabilité sont discutés pour résoudre un système de sédimentation en rivière avec des méthodes de différences finies, ce schéma de discrétisation est utilisé pour mener à bien des expériences de simulation de l'évolution d'un delta dans un domaine analogue à un estuaire. On peut tirer les conclusions suivantes :

1. Le choix du maillage de discrétisation de type "grille-C", où l'on décale le champ de vitesse et le champ potentiel, a fait faire des économies de temps de calcul et de stockage, il est aussi adapté à traiter des frontières irrégulières séparant l'eau et le paroi.
2. Les équations aux différences finies sur ce type d'équation évolutive déterminée par un opérateur positif, permettent une stabilité numérique.

3. En décomposant les équation discrétisées de Saint-Venant en trois parties selon les processus physiques, nous avons fait une grande économie sur la résolution numérique pour un schéma en temps implicite.
4. Les équations discrétisées ont conservé la propriété sur l'énergie du problème original.

Chapitre IV

Problèmes d’assimilation de données

Dans ce chapitre, nous présentons des problèmes d’assimilation variationnelle de données au transport des sédiments en rivière. Dans ces problèmes, on cherche les solutions des problèmes de contrôle optimal. Les méthodes adjointes permettent de déterminer le gradient d’une fonction coût par rapport aux paramètres à assimiler dans le processus d’optimisation. De ce point de vue, nous rappelons et appliquons les techniques de développement du code adjoint du modèle décrit aux chapitres précédents. Pour mettre en valide ce code et tester la faisabilité de la méthode d’optimisation, nous avons reconstruit l’état initial des variables hydrauliques d’une rivière à partir des observations simulées (Yang et Le Dimet [71]). Nous avons également discuté le problème d’identification des paramètres empiriques concernant le transport des sédiments en suspension et charriés.

4.1 Développement du modèle adjoint

4.1.1 Introduction

Pour résoudre numériquement le problème de contrôle optimal, nous utilisons la technique de minimisation de la fonctionnelle de l’écart au sens des moindres carrés par des algorithmes utilisant le vecteur gradient de cette fonctionnelle. La plupart de ces algorithmes de minimisation (sans contrainte) reposent une méthode de descente : en partant d’un point quelconque de l’espace, l’algorithme construit une suite de points qui convergent vers le point de contrôle optimal en faisant diminuer le plus rapidement possible la fonctionnelle calculée en ces points. La direction reliant deux points successifs est appelée “ la direction de descente”. Cette direction est calculée par l’algorithme grâce aux directions du gradient et aux valeurs de la fonctionnelle évaluée en différents points.

L’efficacité d’un algorithme de minimisation dépend surtout de la qualité du calcul de la direction de descente, cela permet à l’algorithme de converger avec moins d’itérations

possibles. Dans les algorithmes de descente, on a besoin, à chaque itération, du vecteur gradient de la fonctionnelle par rapport au paramètre de contrôle (par exemple, la condition initiale). Une idée simple pour calculer ce gradient est de calculer la variation de la fonctionnelle en perturbant le paramètre à contrôler à chaque itération. Évidemment, pour obtenir une variation de la fonctionnelle, il faut intégrer une fois le modèle dans la période de temps analysée, de plus, le vecteur de contrôle est toujours de grande taille, ceci entraîne que cette méthode est impossible à mettre en pratique. Par suite de l'exigence du temps CPU, on a besoin d'une autre manière pour calculer le gradient. On introduit ici la méthode adjointe, dans laquelle à chaque itération de l'algorithme de descente, on a seulement besoin d'une intégration du modèle de prévision et d'une intégration en sens inverse, du temps final au temps initial de l'équation adjointe, qui est l'adjointe de l'équation linéaire tangente. La deuxième intégration nous donne le gradient exigé.

Grâce à cela, ces méthodes ont eu beaucoup d'autres applications, même pour des problèmes en dehors de l'assimilation de données. Pour réaliser ces méthodes numériquement, il n'est pas conseillé de discrétiser séparément les deux équations directe et adjointe. Car en faisant ainsi, il y a très peu de chance que l'équation adjointe discrétisée soit le vrai adjoint de l'équation directe discrétisée. En effet, l'expérience montre qu'il est préférable de prendre directement l'adjoint du système direct discrétisé. Dans le cas contraire, on risque d'observer une divergence de l'algorithme de minimisation.

Le travail peut donc être fait à partir du code direct de l'équation directe pour déduire le code de l'équation adjointe, dont la programmation demande le plus grand soin de la part du développeur afin d'assurer une vraie transposition du code direct (cf. Chao et Chang [7], Rostaing et Dalmas [55]). Ceci demande généralement un investissement humain non négligeable, même pour des codes de calculs de moyenne taille et d'un niveau moyen de complexité. Il existe plusieurs directions de recherche d'études de logiciel (par exemple, Odysée) qui permettent de traiter de façon automatique la transposition des codes en langage fortran ADIFOR, notamment de la part de chercheurs de l'INRIA (Gilbert [27] ou Rostaing et Hassold [54]), et en langage C/C++ ADIC (Griewank [31]).

Dans cette partie, nous allons présenter quelques unes de nos expériences de transposition discrète qui permettent de déduire le système adjoint discret à partir du code direct.

4.1.2 Méthodes adjointes

Les méthodes adjointes sont des méthodes qui ont été récemment beaucoup utilisées dans les problèmes de contrôle optimal et d'assimilation de données. La raison est que l'assimilation de données variationnelle revient à minimiser une fonction coût qui mesure

l'écart entre la prévision du modèle et les observations. Les valeurs optimales des variables de contrôle sont obtenues quand le minimum de la fonction coût est atteint. Donc il est toujours nécessaire de connaître le gradient de la fonction par rapport aux variables de contrôle dans la procédure de minimisation. Les méthodes adjointes sont alors des méthodes pour déterminer ce gradient.

On présente ici l'application pratique à partir d'un code réel. Supposons que $u \in \mathbb{R}^m$ est un vecteur d'entrée d'un système décrit par l'équation directe, soit G l'opérateur, qui représente de manière pratique le total des instructions du programme, et qui transforme u en vecteur de sortie $v \in \mathbb{R}^n$:

$$v = G(u). \quad (4.1)$$

Soit une fonctionnelle J , dont l'expression du gradient par rapport à v : $\nabla_v J$ est facile à déduire, le problème est de calculer le gradient de J par rapport à u . S'il y a une direction de perturbation δu autour de u , cela provoquera une perturbation δv de la sortie, à l'ordre un, on obtient

$$\delta v = \left[\frac{\partial G}{\partial u} \right] \delta u, \quad (4.2)$$

où $\left[\frac{\partial G}{\partial u} \right]$ est la matrice Jacobienne de G par rapport à u . La Gâteaux dérivée de J par rapport à v dans la direction h est égale à

$$J'(v, \delta v) = \langle \nabla_v J, \delta v \rangle, \quad (4.3)$$

grâce à (4.2), il vient

$$J'(v, \delta v) = \langle \nabla_v J, \left[\frac{\partial G}{\partial u} \right] \delta u \rangle = \langle \left[\frac{\partial G}{\partial u} \right]^t \nabla_v J, \delta u \rangle,$$

où $\left[\frac{\partial G}{\partial u} \right]^t$ est la matrice transposée du Jacobien. Par conséquent, le gradient de la fonctionnelle J par rapport à u est donné par :

$$\nabla_u J = \left[\frac{\partial G}{\partial u} \right]^t \nabla_v J. \quad (4.4)$$

Si l'opérateur G se décompose en une suite d'opérations :

$$G = G_N \circ \dots \circ G_2 \circ G_1,$$

son adjoint s'écrit

$$G'^t = G_1'^t \circ G_2'^t \dots \circ G_N'^t,$$

dont G_i^{tt} , $i = 1, \dots, N$, désigne le transposé du opérateur linéarisé de G_i .

Cela montre que pour obtenir le gradient de la fonctionnelle J par rapport à u , il faut d'abord linéariser chacune des opérations puis parcourir en sens inverse la liste des adjoints pour transposer la linéarisation. Voici le principe des méthodes adjointes, mais il reste encore des difficultés techniques pour ses applications, il s'agit notamment des problèmes de stockage et de linéarisation.

4.1.3 Notations

Les problèmes d'assimilation de données sont composés souvent des ingrédients suivants :

Une variable de contrôle : Par exemple, on peut prendre la condition initiale $U = X(t = 0)$ comme variable de contrôle du système, sous l'hypothèse que l'équation a une unique solution sur l'intervalle de temps $[0, T]$. On note \mathcal{U} l'espace des variables de contrôle.

Un modèle : En discrétisant en espace les équations d'évolution d'un système de sédimentation par des méthodes de différences finies, d'éléments finis ou de méthodes spectrales etc. , on représente ces équations sous la forme d'une équation différentielle ordinaire :

$$\begin{cases} \frac{\partial X}{\partial t} + F \cdot X = 0, & \text{sur } \Omega \times [0, T], \\ X(0) = U, \end{cases} \quad (4.5)$$

où F désigne un opérateur, X représente le vecteur des variables physiques aux points de discrétisation du modèle, qui est appelée aussi la **variable d'état** du système appartenant à un espace \mathcal{X} . On suppose que, si U est donnée, on admet une unique solution pour le modèle (4.5).

Des observations : X^{obs} dans le domaine Ω et sur l'intervalle de temps $[0, T]$, considérées continues en temps. En supposant que X^{obs} est dans l'espace d'observations \mathcal{X}^{obs} , qui n'est pas toujours le même que celui de la variable d'état, nous introduisons l'opérateur linéaire $C : \mathcal{X} \rightarrow \mathcal{X}^{obs}$ qui permet de passer de l'espace des variables d'état \mathcal{X} à l'espace des observations \mathcal{X}^{obs} .

Une fonction coût : Nous définissons une fonctionnelle J de l'espace \mathcal{U} de la variable de contrôle U à \mathbb{R} :

$$J(U) = \frac{1}{2} \int_0^T \|C \cdot X(U, t) - X^{obs}(t)\|_{\Omega}^2 dt, \quad (4.6)$$

qui mesure l'écart entre les observations et la solution du modèle déterminée par la variable de contrôle U et le système d'évolution (4.5). Pour chercher le minimum de cette fonctionnelle, on cherche une valeur optimale de la variable de contrôle qui minimise cet écart. $\|\cdot\|_{\Omega}$ est une norme définie sur le domaine Ω déduite d'un produit scalaire $\langle, \rangle_{\Omega}$. On peut prendre par exemple ici la norme de $L^2(\Omega)$ et l'opérateur $C \in \mathcal{L}(L^2(\Omega), L^2(\Omega))$. Si C est l'identité, on observe $X(U, t)$ sur Ω , ou bien, si C est la multiplication par la fonction caractéristique de $\Omega_0 \subset \Omega$, on observe $X(U, t)$ sur Ω_0 .

4.1.4 Système d'optimalité

Pour déduire le système adjoint à partir du système direct, il faut d'abord linéariser le système direct afin de déduire le système linéaire tangent et son adjoint.

On étudie toujours le modèle d'évolution sous la forme abstraite d'un problème de Cauchy :

$$\begin{cases} \frac{\partial X}{\partial t} + F(t, X) \cdot X = 0, & \text{sur } \Omega \times]0, T] \\ X(t=0) = U, & \text{sur } \Omega, \end{cases} \quad (4.7)$$

où F est l'opérateur quasi-linéaire dépendant de t et X .

Soit δU une perturbation sur la condition initiale U , et $U + \delta U$ la nouvelle condition initiale. On définit la variable tangente \hat{X}_U , qui est la Gâteaux-dérivée de X au point U dans la direction δU , par :

$$\hat{X}_U = \lim_{t \rightarrow 0} \frac{X(U + t\delta U) - X(U)}{t}. \quad (4.8)$$

La dérivée de l'opérateur F par rapport à la condition initiale U peut donc être donnée par

$$\nabla_U F = \left[\frac{\partial F}{\partial X} \right] \hat{X}_U.$$

Donc \hat{X}_U est aussi la solution du système linéaire tangent suivant, avec condition

initiale δU :

$$\begin{cases} \frac{\partial \hat{X}_U}{\partial t} + \left[\frac{\partial F}{\partial X} \right] \cdot \hat{X}_U = 0, & \text{sur } \Omega \times]0, T], \\ \hat{X}_U(t=0) = \delta U, & \text{sur } \Omega. \end{cases} \quad (4.9)$$

De plus, $Grad_U J$ étant la dérivée de la fonctionnelle J définie dans (4.6) par rapport à la condition initiale U dans la direction δU est donné par :

$$\langle Grad_U J, \delta U \rangle = \frac{1}{2} \int_0^T \langle CX - X_{obs}, C \hat{X} \rangle_{\Omega} dt.$$

Si l'on suppose qu'une variable P (on l'appelle variable adjointe de X) satisfait les équations suivantes, qui sont appelées également les équations adjointes :

$$\begin{cases} \frac{\partial P}{\partial t} + \left[\frac{\partial F}{\partial X} \right]^t P = C^t(CX - X_{obs}), \\ P(T) = 0, \end{cases} \quad (4.10)$$

où C^t est l'adjoint de l'opérateur C , qui vérifie

$$\langle CX, Y \rangle = \langle X, C^t Y \rangle, \quad \forall X \in \mathcal{X}, Y \in \mathcal{X}^{obs}.$$

Alors le gradient de J par rapport à la condition initiale U peut être donné par :

$$Grad_U J = -P(0), \quad (4.11)$$

et la condition initiale optimale est solution de l'équation :

$$Grad_U J = -P(0) = 0.$$

Pour conclure, la valeur optimale de la condition initiale (la variable de contrôle) est caractérisée par un **système d'optimalité** :

$$\begin{cases} \frac{\partial X}{\partial t} + F(t, X) \cdot X = 0, \\ X(t=0) = U, \\ \frac{\partial P}{\partial t} + \left[\frac{\partial F}{\partial X} \right]^t P = C^t(CX - X_{obs}), \\ P(T) = 0, \\ Grad_U J = -P(0) = 0. \end{cases} \quad (4.12)$$

On constate que par l'intégration des équations adjointes une fois au sens inverse à partir du temps final au temps initial, on peut obtenir le gradient de la fonction coût par rapport à la variable de contrôle, qui est toujours nécessaire pour le processus de minimisation. Les observations sont utilisées ici comme un terme de forçage dans l'équation adjointe.

4.1.5 Méthode de transposition

La discrétisation du système d'optimalité n'est en général pas la même que le système d'optimalité discret, parce que lorsque l'on intègre par partie, pour le modèle continu, on utilise la formule de Green, alors que pour déduire le système d'optimalité discret, on utilise la formule d'Euler. On rencontre toujours la difficulté pour déduire l'écriture du code adjoint qui est la transposition du code linéarisé à partir de celui du modèle de simulation :

- Il demande un travail très précis de la part du développeur afin d'assurer une vraie transposition du code direct, qui permet de calculer le vecteur gradient de la fonctionnelle à minimiser.
- Dans un code de grande taille (même en moyenne taille) ou d'un niveau moyen de complexité, dont le nombre d'instructions exécutables en langage fortran est non-négligable, il nous demande de bien connaître le graphe d'exécution pour le dualiser en exécutant à l'ordre inverse, à cause de celui-ci, le résultat de chaque étape doit être sauvegardé en introduisant des variables intermédiaires, il s'agira donc la stratégie de stockage.

A partir du système linéaire tangent (4.10), on voit que, pour faire la transposition, on a besoin de garder la valeur du Jacobien sur un état de la variable X :

$$\frac{\partial F}{\partial X}.$$

Mais dans un programme, le Jacobien est décomposé et exprimé par une série des instructions, et ces instructions se répartissant en moyenne sur des certaines sous-routines. Par un exemple simple mais typique de Chao et Chang ([7]), on explique un technique principal pour déduire le code adjoint associé à une instruction dans le code du modèle direct.

Supposons une instruction en langage fortran s'écrit :

$$W = AW + BXY, \tag{4.13}$$

où W , X et Y sont les variables intermédiaires, A et B sont les constantes. L'instruction correspondante dans le code linéaire tangent est

$$W' = AW' + BYX' + BXY', \quad (4.14)$$

dont W' , X' et Y' sont les variables perturbées par rapport à la variable de contrôle.

On remarque que nous avons besoin des valeurs de X et Y stockées, si l'on les pris comme étant un opérateur linéaire à transporter, ou bien on doit les recalculer. Donc un problème se pose sur l'estimation du temps d'exécution et le nombre de variables intermédiaires dans le code direct pour pouvoir réaliser un code adjoint avec une négociation entre la place de stockage des variables intermédiaires et le temps nécessaire pour recalculer ces variables dans le code adjoint.

Maintenant on peut traiter chaque instruction dans le code linéaire tangent comme un opérateur linéaire, on peut l'écrire de la manière suivante :

$$W' = A_1W' + B_1X' + C_1Y', \quad (4.15)$$

où A_1 , B_1 et C_1 sont soit les constantes, ou les variables dans le code direct. Supposons le vecteur $\mathcal{W} = (W', X', Y')^t$ et la matrice \mathcal{M} tels que $\mathcal{W} = \mathcal{M}\mathcal{W}$, c.-à-d. ,

$$\begin{pmatrix} W' \\ X' \\ Y' \end{pmatrix} = \begin{pmatrix} A_1 & B_1 & C_1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} W' \\ X' \\ Y' \end{pmatrix}. \quad (4.16)$$

En introduisant le vecteur des variables adjointes $\mathcal{W}^* = (W^*, X^*, Y^*)^t$, par définition de la transposition de \mathcal{M}^t , on a

$$\langle \mathcal{W}^*, \mathcal{W} \rangle = \langle \mathcal{W}^*, \mathcal{M}\mathcal{W} \rangle = \langle \mathcal{M}^t\mathcal{W}^*, \mathcal{W} \rangle,$$

qui donne

$$\mathcal{W}^* = \mathcal{M}^t\mathcal{W}^*,$$

on l'écrit dans le code adjoint par

$$\begin{pmatrix} W^* \\ X^* \\ Y^* \end{pmatrix} = \begin{pmatrix} A_1 & 0 & 0 \\ B_1 & 1 & 0 \\ C_1 & 0 & 1 \end{pmatrix} \begin{pmatrix} W^* \\ X^* \\ Y^* \end{pmatrix} \quad (4.17)$$

c'est-à-dire que

$$\begin{cases} Y^* = C_1W^* + Y^*, \\ X^* = B_1W^* + X^*, \\ W^* = A_1W^*. \end{cases} \quad (4.18)$$

Il faut remarquer que la troisième instruction ne peut pas changer la place avec les autres. En fait, on peut aussi écrire (4.15) de la façon suivante en introduisant une variable Z

$$\begin{cases} Z = A_1 W' + B_1 X' + C_1 Y', \\ W' = Z. \end{cases} \quad (4.19)$$

Pour la même raison, on écrit la transposition de chaque instruction au sens inverse :

$$\begin{cases} Z^* = W^*, \\ Y^* = C_1 Z^* + Y^*, \\ X^* = B_1 Z^* + X^*, \\ W^* = A_1 Z^*. \end{cases} \quad (4.20)$$

L'étape de la linéarisation du modèle direct est simple à mettre en oeuvre sauf quelques fonctions qui ne sont pas différentiables, par exemple

$$\begin{aligned} f(x) &= \sqrt{x}, & \text{quand } x > 0, \\ f(x, y) &= \max(x, y), \\ f(x) &= |x|, & \text{quand } x < 0, \\ & \dots \end{aligned}$$

Dans ces cas, avant de linéariser le code direct, si c'est possible, nous pouvons réécrire le code direct en remplaçant les instructions comprenant cette sorte de fonctions indifférentiables par d'instructions IF THEN-ELSE, avec lesquelles on rassure la linéarisation. En conséquence, les conditions de IF THEN-ELSE doivent être les mêmes dans le code adjoint pour déterminer laquelle séquence des instructions que l'on doit suivre et transposer, et si dans le cas où elles comprennent les variables du code direct, il est nécessaire de stocker ces conditions ou des recalculer. En revanche, l'utilisation d'instructions IF-THEN peut faire obstacle au taux de la convergence de l'optimisation. C'est grâce à ces natures mécaniques et épineuses de l'écriture du code adjoint, le développement de l'opérateur adjoint automatique a attiré beaucoup d'attentions là dessus (Griewank [31]).

4.1.6 L'algorithme de minimisation

Rappelons d'abord que l'algorithme de Newton pour la recherche du zéro d'une fonction au voisinage du point courant est de résoudre successivement des problèmes linéarisés, ce qui donne l'algorithme suivant :

Soit $G : \mathbb{R}^n \rightarrow \mathbb{R}^n$, et soit à résoudre $G(X) = 0$. En partant d'un point $X_0 \in \mathbb{R}^n$ quelconque, on calcule la suite $X_k \in \mathbb{R}^n$, qui converge vers un zéro de G sous certaines conditions, définie par :

$$X_{k+1} = X_k - [JacG(X_k)]^{-1} \cdot G(X_k),$$

où $JacG(X_k)$ est le Jacobien de G calculé au point X_k .

Dans un problème de minimisation, la fonction G dont on veut chercher le zéro est la fonction qui, à chaque point de l'espace associe le vecteur gradient de la fonctionnelle \mathcal{J} à minimiser, calculée en ce point :

$$\begin{aligned} G : \mathbb{R}^n &\rightarrow \mathbb{R}^n \\ X &\rightarrow \vec{\nabla} \mathcal{J}(X). \end{aligned}$$

Le Jacobien $JacG(X_k)$ de G au point X_k dans ce cas est le Hessien de la fonction coût \mathcal{J} calculé au même point. L'algorithme de Newton pour un problème de minimisation peut s'écrire

$$X_{k+1} = X_k - [Hess\mathcal{J}(X_k)]^{-1} \vec{\nabla} \mathcal{J}(X_k),$$

dans le problème d'optimisation, on ajoute un paramètre de pas de descente et l'algorithme s'écrit

$$X_{k+1} = X_k - \rho_k [Hess\mathcal{J}(X_k)]^{-1} \vec{\nabla} \mathcal{J}(X_k),$$

où ρ_k est le pas de descente.

La méthode de minimisation appliquée dans nos tests est un algorithme de type quasi-Newton à mémoire limitée (QNML), la forme utilisée est dite « BFGS » (resp. « BFGS » inverse) due à Broyden, Fletcher et al., qui fournit une façon d'évaluer récursivement une approximation du Hessien (resp. de l'inverse du Hessien) du point de X_k . Cette formule utilise uniquement les vecteurs gradients $G_k = G(X_k)$ et les points X_k auxquels G_k ont été évalués. La méthode de quasi-Newton est en réalité une méthode du type gradient car elle utilise les différentielles seulement jusqu'à l'ordre un de la fonctionnelle \mathcal{J} à minimiser.

Nous avons utilisé plus précisément l'algorithme « M1QN3 » programmé par J. C. Gilbert et C. Lemaréchal de l'INRIA. D'après lequel le programme pour le cas de la double précision a été aussi développé et nommé par « N1QN3 ».

Maintenant nous allons introduire le principe de la méthode QNML. Pour chaque itération de l'algorithme, le point $X_k \in \mathbb{R}^n$ connu, on effectue :

1. Calcul du gradient au point X_k

$$G_k = \vec{\nabla} \mathcal{J}(X_k);$$

2. Calcul de la direction de descente

$$d_k = -H_k G_k,$$

où $H_k = B_k^{-1}$, $B_k \simeq Hess\mathcal{J}(X_k)$;

3. Calcul du pas de descente ρ_k

$$\mathcal{J}(X_k + \rho_k d_k) = \inf_{\rho \in \mathbf{R}} \mathcal{J}(X_k + \rho d_k),$$

i.e. on cherche le minimum de \mathcal{J} dans la direction de descente.

4. Définition du point X_{k+1} :

$$X_{k+1} = X_k + \rho_k d_k$$

A chaque itération B_k ou H_k sont mis à faire par l'une des formules :

1. la formule BFGS

$$[B_{k+1}] = [B_k] + \frac{y_k \otimes y_k}{\langle y_k, s_k \rangle} - \frac{[(B_k \cdot s_k) \otimes (B_k \cdot s_k)]}{\langle s_k, B_k \cdot s_k \rangle};$$

2. la formule BFGS inverse

$$[H_{k+1}] = [H_k] + \frac{[(s_k - H_k y_k) \otimes s_k] + [s_k \otimes (s_k - H_k y_k)]}{\langle y_k, s_k \rangle} - \frac{\langle s_k - H_k \cdot y_k, y_k \rangle}{\langle y_k, s_k \rangle^2} [s_k \otimes s_k]$$

où

$$s_k = X_{k+1} - X_k$$

est la différence entre deux points,

$$y_k = G_{k+1} - G_k$$

est la différence entre deux gradients; $\langle \cdot, \cdot \rangle$ est un produit scalaire choisi dans \mathbb{R}^n ; $[\cdot \otimes \cdot]$ est le produit tensoriel associé à ce produit scalaire, c.-à-d. :

$$\begin{aligned} \otimes : \mathbb{R}^n \times \mathbb{R}^n &\rightarrow \mathcal{L}(\mathbb{R}^n), \\ (U, V) &\rightarrow [U \otimes V] \quad U, V \in \mathbb{R}^n, \\ \text{avec : } [U \otimes V] \cdot X &= \langle V, X \rangle U \quad \forall X \in \mathbb{R}^n. \end{aligned}$$

Remarque 1 : Dans le cas où le produit scalaire choisi est le produit scalaire Euclidien, alors la matrice dans la base canonique $(e_i)_{1 \leq i \leq n}$ du produit tensoriel $[U \otimes V]$ s'écrit $[U \cdot V^t]$.

Remarque 2 Les formules de BFGS et BFGS inverse ont la propriété de conserver le caractère défini positif de B_k et H_k si et seulement si $\langle y_k, s_k \rangle$ est positif, ce qui assure que d_k est une direction de descente.

En pratique, dans un problème de grande taille, il n'est pas possible de calculer et de mémoriser tous les coefficients des matrices B_k ou H_k , mais seulement k couples de vecteurs $(y_i, s_i)_{1 \leq i \leq k}$. Avec un algorithme spécialisé, on peut calculer la direction de descente d_k à la $k^{\text{ième}}$ étape de la minimisation. Par conséquent, la méthode quasi-Newton à mémoire limitée (QNML) consiste à calculer la matrice H_k en ne faisant intervenir qu'un nombre limité de couples de vecteurs (y_i, s_i) les plus récents. La méthode QNML est donc une méthode très efficace pour une capacité de stockage raisonnable.

4.1.7 Critère d'arrêt de l'algorithme de minimisation

Dans le processus de minimisation, il faut toujours choisir un bon critère pour arrêter l'algorithme, surtout pour un système de grand taille. Parce que, soit on arrête l'algorithme trop tôt, et la convergence n'est pas encore arrivée vers le contrôle optimal, soit on a perdu du temps de calcul. Quand l'état de contrôle est proche de l'état optimal dans l'optimisation, la fonction coût et son vecteur gradient par rapport au vecteur de contrôle convergeront vers la valeur nulle d'après la théorie du contrôle optimal. Dans notre cas, nous utilisons un critère d'arrêt qui est très couramment utilisé dans les algorithmes de minimisation et il a été prouvé qu'il peut donner des résultats assez satisfaisants. L'algorithme s'arrête lorsque la condition suivante est réalisée :

$$\frac{\|\vec{\nabla} \mathcal{J}(X_k)\|}{\|\vec{\nabla} \mathcal{J}(X_0)\|} \leq \eta, \quad (4.21)$$

où η est un seuil fixé de l'ordre de 10^{-4} , k désigne le nombre d'itérations.

Le seul point important de ce test est qu'il faut faire attention dans le cas où le vecteur d'initialisation X_0 est proche de l'état optimal car la quantité $\|\vec{\nabla} \mathcal{J}(X_0)\|$ est alors initialement petite.

4.2 Assimilation de données I – Contrôle des conditions initiales

Dans cette partie, nous présentons une application de la méthode adjointe pour le problème de transport des sédiments. On considère à résoudre un problème numérique de type contrôle optimal, où les conditions initiales des variables d'état sont prises comme étant les variables de contrôle. Dans ce problème, on cherche à reconstruire le champ initial du système hydraulique à partir des données observées, afin que l'évolution de la rivière surtout la déformation du fond de la rivière approche le mieux possible les observations.

4.2.1 Forme continue du système adjoint

Posons la variable d'état $X = (u, v, h, S, Z_b)$ comme étant la solution du modèle (4.7), et supposons que l'on observe sur le champ de vitesse $\vec{v}^{\rightarrow}_{obs} = (u_{obs}, v_{obs})$, la concentration de sédiments S_{obs} , la profondeur de l'eau h_{obs} et l'élévation du lit de la rivière $Z_{b_{obs}}$. Pour simplifier, on suppose d'abord que ces observations sont denses sur tout l'espace et à tout l'instant. C'est-à-dire que l'on a des observations sur tous les noeuds de grille, bien que ceci n'est pas toujours possible dans la réalité. On discutera plus tard le cas où les observations ne sont pas denses.

Considérons la fonction coût suivante :

$$\begin{aligned}
 J = \frac{1}{2} \int_D & ((C_u u - u_{obs})^2 + (C_v v - v_{obs})^2 + (C_h h - h_{obs})^2 \\
 & + (C_S S - S_{obs})^2 + (C_{Z_b} Z_b - Z_{b_{obs}})^2) dD,
 \end{aligned} \tag{4.22}$$

dans laquelle C représentant l'opérateur de projection de l'espace d'état dans celui des observations. $D = \Omega \times [0, T]$, avec Ω le domaine discrétisé et $[0, T]$ l'intervalle du temps d'intégration.

Posons $H_X = (H_u, H_v, H_h, H_S, H_{Z_b})$ une perturbation sur la condition initiale $U = X(t = 0)$, $\hat{X} = (\hat{u}, \hat{v}, \hat{h}, \hat{S}, \hat{Z}_b)$ la variable tangente la Gâteaux-dérivée de X vérifiant (4.8), on écrit d'abord le système linéaire tangent à partir des équations (2.1), (2.3) et (2.4) :

$$\left\{ \begin{aligned}
 \frac{\partial \hat{u}}{\partial t} &= -\hat{u} \frac{\partial u}{\partial x} - u \frac{\partial \hat{u}}{\partial x} - \hat{v} \frac{\partial u}{\partial y} - v \frac{\partial \hat{u}}{\partial y} + f \hat{v} - g \frac{\partial \hat{h}}{\partial x} - g \frac{\partial \hat{Z}_b}{\partial x} \\
 &+ k_1 \Delta \hat{u} - C_D \frac{|\vec{v}^{\rightarrow}| \hat{u}}{h} - C_D \frac{(u \hat{u} + v \hat{v}) u}{|\vec{v}^{\rightarrow}| h} + C_D \frac{|\vec{v}^{\rightarrow}| u}{h^2} \hat{h}, \\
 \frac{\partial \hat{v}}{\partial t} &= -\hat{u} \frac{\partial v}{\partial x} - u \frac{\partial \hat{v}}{\partial x} - \hat{v} \frac{\partial v}{\partial y} - v \frac{\partial \hat{v}}{\partial y} - f \hat{u} - g \frac{\partial \hat{h}}{\partial y} - g \frac{\partial \hat{Z}_b}{\partial y} \\
 &+ k_1 \Delta \hat{v} - C_D \frac{|\vec{v}^{\rightarrow}| \hat{v}}{h} - C_D \frac{(u \hat{u} + v \hat{v}) v}{|\vec{v}^{\rightarrow}| h} + C_D \frac{|\vec{v}^{\rightarrow}| v}{h^2} \hat{h}, \\
 \frac{\partial \hat{h}}{\partial t} &= -\frac{\partial(\hat{u} h)}{\partial x} - \frac{\partial(u \hat{h})}{\partial x} - \frac{\partial(\hat{v} h)}{\partial y} - \frac{\partial(v \hat{h})}{\partial y},
 \end{aligned} \right. \tag{4.23}$$

$$\begin{aligned}
\frac{\partial(h\hat{S})}{\partial t} &= -\frac{\partial(\hat{h}S)}{\partial t} - \frac{\partial(\hat{h}uS)}{\partial x} - \frac{\partial(h\hat{u}S)}{\partial x} - \frac{\partial(hu\hat{S})}{\partial x} \\
&- \frac{\partial(\hat{h}vS)}{\partial y} - \frac{\partial(h\hat{v}S)}{\partial y} - \frac{\partial(hv\hat{S})}{\partial y} + \alpha\omega(\hat{S}^* - \hat{S}) \\
&+ k' \left[\frac{\partial}{\partial x} \left(\hat{h} \frac{\partial S}{\partial x} \right) + \frac{\partial}{\partial x} \left(h \frac{\partial \hat{S}}{\partial x} \right) + \frac{\partial}{\partial y} \left(\hat{h} \frac{\partial S}{\partial y} \right) + \frac{\partial}{\partial y} \left(h \frac{\partial \hat{S}}{\partial y} \right) \right],
\end{aligned} \tag{4.24}$$

$$\frac{\partial \hat{Z}_b}{\partial t} = \frac{1}{\rho_1} \alpha \omega (\hat{S} - \hat{S}^*) - \left(\frac{\partial g_{bx}}{\partial x} + \frac{\partial g_{by}}{\partial y} \right). \tag{4.25}$$

D'après les formules (2.5) et (2.6), S^* et (g_{bx}, g_{by}) sont en fonction de \vec{v} et h , on a donc

$$\hat{S}^* = \frac{\partial S^*}{\partial u} \hat{u} + \frac{\partial S^*}{\partial v} \hat{v} + \frac{\partial S^*}{\partial h} \hat{h}, \quad g_{bx} = \frac{\partial g_{bx}}{\partial u} \hat{u} + \frac{\partial g_{bx}}{\partial v} \hat{v} + \frac{\partial g_{bx}}{\partial h} \hat{h}, \quad x = x \text{ ou } y.$$

Les conditions initiales des équations linéaires tangentes sont données par :

$$\hat{u}(0) = H_u; \hat{v}(0) = H_v; \hat{h}(0) = H_h; \hat{S}(0) = H_S; \hat{Z}_b(0) = H_{Z_b},$$

et des conditions aux limites sont homogène de type Dirichlet.

Considérant les variables $P_X = (P_u, P_v, P_h, P_S, P_{Z_b})$, duales de X , en multipliant P_X par les équations (4.23)-(4.25) et ensuite les intégrant sur D , on obtient :

$$\begin{aligned}
& \int_D \left(P_u \frac{\partial \hat{u}}{\partial t} + P_v \frac{\partial \hat{v}}{\partial t} + P_h \frac{\partial \hat{h}}{\partial t} + P_S \frac{\partial (h\hat{S})}{\partial t} + P_{Z_b} \frac{\partial \hat{Z}_b}{\partial t} \right) dD \\
= & \int_D \left\{ -P_u \left(\hat{u} \frac{\partial u}{\partial x} + u \frac{\partial \hat{u}}{\partial x} + \hat{v} \frac{\partial u}{\partial y} + v \frac{\partial \hat{u}}{\partial y} - f\hat{v} \right) \right. \\
& - P_u \left(g \frac{\partial \hat{h}}{\partial x} + g \frac{\partial \hat{Z}_b}{\partial x} - k_1 \Delta \hat{u} + C_D \frac{|\vec{v}| \hat{u}}{h} + C_D \frac{(u\hat{u} + v\hat{v})u}{|\vec{v}|h} - C_D \frac{|\vec{v}|u}{h^2} \hat{h} \right) \\
& - P_v \left(\hat{u} \frac{\partial v}{\partial x} + u \frac{\partial \hat{v}}{\partial x} + \hat{v} \frac{\partial v}{\partial y} + v \frac{\partial \hat{v}}{\partial y} + f\hat{u} \right) \\
& - P_v \left(g \frac{\partial \hat{h}}{\partial y} + g \frac{\partial \hat{Z}_b}{\partial y} - k_1 \Delta \hat{v} + C_D \frac{|\vec{v}| \hat{v}}{h} + C_D \frac{(u\hat{u} + v\hat{v})v}{|\vec{v}|h} - C_D \frac{|\vec{v}|v}{h^2} \hat{h} \right) \\
& - P_h \left(\frac{\partial(\hat{u}h)}{\partial x} + \frac{\partial(u\hat{h})}{\partial x} + \frac{\partial(\hat{v}h)}{\partial y} + \frac{\partial(v\hat{h})}{\partial y} \right) \\
& - P_S \left(\frac{\partial(\hat{h}uS)}{\partial x} + \frac{\partial(h\hat{u}S)}{\partial x} + \frac{\partial(hu\hat{S})}{\partial x} + \frac{\partial(\hat{h}vS)}{\partial y} + \frac{\partial(h\hat{v}S)}{\partial y} + \frac{\partial(hv\hat{S})}{\partial y} \right) \\
& + P_S \left(k' \left[\frac{\partial}{\partial x} \left(\hat{h} \frac{\partial S}{\partial x} \right) + \frac{\partial}{\partial x} \left(h \frac{\partial \hat{S}}{\partial x} \right) + \frac{\partial}{\partial y} \left(\hat{h} \frac{\partial S}{\partial y} \right) + \frac{\partial}{\partial y} \left(h \frac{\partial \hat{S}}{\partial y} \right) \right] \right) \\
& + \left. \alpha\omega \left(P_S (\hat{S}^* - \hat{S}) + \frac{P_{Z_b}}{\rho_1} (\hat{S} - \hat{S}^*) \right) - P_{Z_b} \left(\frac{\partial \hat{g}_{bx}}{\partial x} + \frac{\partial \hat{g}_{by}}{\partial y} \right) \right\} dD.
\end{aligned}$$

Sous l'hypothèse que toutes les variables tangentes sont nulles sur les bords, et avec la formule d'intégration par parties, il en résulte la forme continue suivante du système

adjoint :

$$\left\{ \begin{aligned}
 -\frac{\partial P_u}{\partial t} &= -P_u \frac{\partial u}{\partial x} + \frac{\partial(P_u u)}{\partial x} + \frac{\partial(P_u v)}{\partial y} - P_v \frac{\partial v}{\partial x} - f P_v \\
 &+ h \frac{\partial P_h}{\partial x} + h S \frac{\partial P_S}{\partial x} + k_1 \Delta P_u - P_u C_D \frac{|\vec{v}^\dagger|}{h} - P_u C_D \frac{u^2}{|\vec{v}^\dagger| h} - P_v C_D \frac{uv}{|\vec{v}^\dagger| h} \\
 &+ \alpha \omega \left(P_S \frac{\partial S^*}{\partial u} - P_{Z_b} \frac{1}{\rho_1} \frac{\partial S^*}{\partial u} \right) + \frac{\partial P_{Z_b}}{\partial x} \frac{\partial g_{bx}}{\partial u} + \frac{\partial P_{Z_b}}{\partial y} \frac{\partial g_{by}}{\partial u} + C_u^t (C_u u - u_{obs}), \\
 -\frac{\partial P_v}{\partial t} &= -P_u \frac{\partial u}{\partial y} + \frac{\partial(P_v u)}{\partial x} + \frac{\partial(P_v v)}{\partial y} - P_v \frac{\partial v}{\partial y} + f P_u \\
 &+ h \frac{\partial P_h}{\partial y} + h S \frac{\partial P_S}{\partial y} + k_1 \Delta P_v - P_v C_D \frac{|\vec{v}^\dagger|}{h} - P_v C_D \frac{v^2}{|\vec{v}^\dagger| h} - P_u C_D \frac{uv}{|\vec{v}^\dagger| h} \\
 &+ \alpha \omega \left(P_S \frac{\partial S^*}{\partial v} - P_{Z_b} \frac{1}{\rho_1} \frac{\partial S^*}{\partial v} \right) + \frac{\partial P_{Z_b}}{\partial x} \frac{\partial g_{bx}}{\partial v} + \frac{\partial P_{Z_b}}{\partial y} \frac{\partial g_{by}}{\partial v} + C_v^t (C_v v - v_{obs}), \\
 -\frac{\partial P_h}{\partial t} &= u \frac{\partial P_h}{\partial x} + v \frac{\partial P_h}{\partial y} + \frac{\partial P_S}{\partial t} S + u S \frac{\partial P_h}{\partial x} + v S \frac{\partial P_h}{\partial y} + g \frac{\partial P_u}{\partial x} + g \frac{\partial P_v}{\partial y} \\
 &- k' \left(\frac{\partial P_S}{\partial x} \frac{\partial S}{\partial x} + \frac{\partial P_S}{\partial y} \frac{\partial S}{\partial y} \right) + C_D P_u \frac{|\vec{v}^\dagger| u}{h^2} + C_D P_v \frac{|\vec{v}^\dagger| v}{h^2} \\
 &+ \alpha \omega P_S \frac{\partial S^*}{\partial h} - P_{Z_b} \frac{\alpha \omega}{\rho_1} \frac{\partial S^*}{\partial h} + \frac{\partial P_{Z_b}}{\partial x} \frac{\partial g_{bx}}{\partial h} + \frac{\partial P_{Z_b}}{\partial y} \frac{\partial g_{by}}{\partial h} + C_h^t (C_h h - h_{obs}), \\
 -\frac{\partial P_S}{\partial t} &= u \frac{\partial P_S}{\partial x} + v \frac{\partial P_S}{\partial y} + \frac{k'}{h} \left[\frac{\partial}{\partial x} \left(\frac{\partial P_S}{\partial x} h \right) + \frac{\partial}{\partial y} \left(\frac{\partial P_S}{\partial y} h \right) \right] \\
 &- \frac{1}{h} \alpha \omega P_S + \frac{1}{\rho_1 h} \alpha \omega P_{Z_b} + C_S^t (C_S S - S_{obs}), \\
 -\frac{\partial P_{Z_b}}{\partial t} &= g \frac{\partial P_u}{\partial x} + g \frac{\partial P_v}{\partial y} + C_{Z_b}^t (C_{Z_b} Z_b - Z_{b_{obs}}).
 \end{aligned} \right. \tag{4.26}$$

avec :

$$P_u(T) = 0; \quad P_v(T) = 0; \quad P_h(T) = 0; \quad P_S(T) = 0; \quad P_{Z_b}(T) = 0.$$

4.2.2 Système adjoint des équations de Saint-Venant alternées

4.2.2.1 Adjoint du système avec le changement des variables

Dans le cas où un changement de variables a été utilisé, toutes les variables apparaissant ne sont pas indépendantes, donc il n'est pas nécessaire d'observer en même temps toutes les variables. Par exemple, soient deux variables d'état X et Y vérifiant les équations suivantes :

$$\begin{cases} \frac{dX}{dt} = F(X, Y), & \text{sur } \Omega \times]0, T], \\ G(X, Y) = 0, & \text{sur } \Omega \times]0, T], \\ X(t = 0) = X_0, & \text{sur } \Omega, \end{cases} \quad (4.27)$$

avec F et G deux opérateurs différentiels non-linéaires. Soit Y_0 la condition initiale de Y , elle peut être déterminée par la relation suivante :

$$G(X_0, Y_0) = 0. \quad (4.28)$$

La fonction coût à minimiser supposée la différence entre la solution X du modèle et l'observation sur X est traduite de la manière suivante :

$$J(X_0) = \frac{1}{2} \int_0^T \|CX(X_0, t) - X_{obs}(t)\|_{\Omega}^2 dt, \quad (4.29)$$

où C est un opérateur linéaire de l'espace de la variable d'état X à celle de l'observation X_{obs} , et T est le temps final.

Si l'on impose une perturbation H_X autour de la condition initiale X_0 et X' et Y' sont les Gâteaux dérivées de X et Y par rapport à X_0 respectivement, d'après le système (4.27), on peut déduire que :

$$\begin{cases} \frac{dX'}{dt} = \left[\frac{\partial F}{\partial X} \right] \cdot X' + \left[\frac{\partial F}{\partial Y} \right] \cdot Y', \\ \left[\frac{\partial G}{\partial X} \right] \cdot X' + \left[\frac{\partial G}{\partial Y} \right] \cdot Y' = 0, \\ X'(0) = H_X. \end{cases} \quad (4.30)$$

Posons ensuite X^* et Y^* les variables adjointes de X et Y , qui sont les solutions des équations suivantes :

$$\left\{ \begin{array}{l} \frac{dX^*}{dt} + \left[\frac{\partial F}{\partial X} \right]^t X^* + \left[\frac{\partial G}{\partial X} \right]^t Y^* = C^t(CX - X_{obs}), \\ \left[\frac{\partial F}{\partial Y} \right]^t X^* + \left[\frac{\partial G}{\partial Y} \right]^t Y^* = 0, \\ X^*(t = T) = 0, \end{array} \right. \quad (4.31)$$

dont A^t signifie l'opérateur adjoint s'il vérifie la relation :

$$\langle AX, Y^* \rangle_\Omega = \langle X, A^t Y^* \rangle_\Omega,$$

avec \langle , \rangle le produit scalaire défini dans l'espace Ω , alors le gradient de la fonctionnelle par rapport à X_0 est égal à

$$\nabla_{X_0} J = -X^*(t = 0). \quad (4.32)$$

Pour le montrer, on multiplie la première équation de (4.31) par X' et en intégrant du temps initial au temps $t = T$, on obtient :

$$\int_0^T \left\langle \frac{dX^*}{dt} + \left[\frac{\partial F}{\partial X} \right]^t X^* + \left[\frac{\partial G}{\partial X} \right]^t Y^*, X' \right\rangle = \int_0^T \langle C^t(CX - X_{obs}), X' \rangle,$$

après une intégration par parties, on a

$$\begin{aligned} \int_0^T \left\langle X^*, -\frac{dX'}{dt} + \left[\frac{\partial F}{\partial X} \right] \cdot X' \right\rangle + \int_0^T \left\langle Y^*, \left[\frac{\partial G}{\partial X} \right] X' \right\rangle - \langle X^*(0), H_X \rangle \\ = \int_0^T \langle C^t(CX - X_{obs}), X' \rangle, \end{aligned}$$

d'après (4.30), il vient

$$\begin{aligned} \int_0^T \left\langle X^*, -\left[\frac{\partial F}{\partial Y} \right] Y' \right\rangle - \int_0^T \left\langle Y^*, \left[\frac{\partial G}{\partial Y} \right] Y' \right\rangle - \int_0^T \langle CX - X_{obs}, CX' \rangle \\ = \langle X^*(0), H_X \rangle. \end{aligned} \quad (4.33)$$

Multipliant maintenant la deuxième équation de (4.31) par Y' et l'intégrant sur l'intervalle $[0, T]$, on obtient

$$\int_0^T \left\langle \left[\frac{\partial F}{\partial Y} \right]^t X^*, Y' \right\rangle + \int_0^T \left\langle \left[\frac{\partial G}{\partial Y} \right]^t Y^*, Y' \right\rangle = 0,$$

qui donne de (4.33)

$$- \int_0^T \langle CX - X_{obs}, CX' \rangle = \langle X^*(0), H_X \rangle,$$

et grâce à la Gâteaux dérivée de J par rapport à X_0 sur la direction H_X arbitraire :

$$\hat{J}(X', H_X) = \int_0^T \langle CX - X_{obs}, CX' \rangle$$

on en déduit le gradient (4.32).

4.2.2.2 Adjoint des méthodes de splitting

Dans le chapitre précédent, nous avons donné l'algorithme à pas fractionnaires qui décompose l'opérateur déterminant le modèle d'évolution en trois opérateurs, on présente le système adjoint associé à cet algorithme.

Supposons U_0 est la variable d'entrée, on résout la variable finale U_3 par trois processus,

$$\begin{cases} \text{Processus 1 :} & U_1 = G_1(U_0), \\ \text{Processus 2 :} & U_2 = G_2(U_1), \\ \text{Processus 3 :} & U_3 = G_3(U_2). \end{cases} \quad (4.34)$$

La fonction coût $J = J(U_3)$ est en fonction de U_3 , on cherche le gradient de J par rapport à la variable U_0 .

Soit H une perturbation sur U_0 et $\hat{\cdot}$ la dérivée directionnelle sur la direction H , on obtient le système linéarisé des trois processus ci-dessous :

$$\begin{cases} \hat{U}_1 = \left[\frac{\partial G_1}{\partial U_0} \right] \cdot H, \\ \hat{U}_2 = \left[\frac{\partial G_2}{\partial U_1} \right] \cdot \hat{U}_1, \\ \hat{U}_3 = \left[\frac{\partial G_3}{\partial U_2} \right] \cdot \hat{U}_2, \end{cases} \quad (4.35)$$

et la dérivée directionnelle de J devient

$$\hat{J}(U_3, H) = \langle \nabla_{U_3} J, \hat{U}_3 \rangle. \quad (4.36)$$

Soit P_1, P_2, P_3 les variables adjointes de U_1, U_2, U_3 respectivement, en multipliant trois membres de l'équation (4.35) par P_1, P_2 et P_3 et en les sommant, on a

$$(P_1, \hat{U}_1) + (P_2, \hat{U}_2) + (P_3, \hat{U}_3) =$$

$$(H, \left[\frac{\partial G_1}{\partial U_0} \right]^t \cdot P_1) + (\hat{U}_1, \left[\frac{\partial G_2}{\partial U_1} \right]^t \cdot P_2) + (\hat{U}_2, \left[\frac{\partial G_3}{\partial U_2} \right]^t \cdot P_3).$$

Définissons

$$P_3 = \nabla_{U_3} J, \quad (4.37)$$

alors

$$\begin{aligned} \hat{J}(U_3, H) &= (P_3, \hat{U}_3) \\ &= (\hat{U}_2, \left[\frac{\partial G_3}{\partial U_2} \right]^t \cdot P_3 - P_2) + (\hat{U}_1, \left[\frac{\partial G_2}{\partial U_1} \right]^t \cdot P_2 - P_1) + (H, \left[\frac{\partial G_1}{\partial U_0} \right]^t \cdot P_1). \end{aligned} \quad (4.38)$$

Donc si nous définissons puis

$$P_2 = \left[\frac{\partial G_3}{\partial U_2} \right]^t \cdot P_3, \quad (4.39)$$

et

$$P_1 = \left[\frac{\partial G_2}{\partial U_1} \right]^t \cdot P_2, \quad (4.40)$$

alors il reste

$$\hat{J}(U_3, H) = \left(\left[\frac{\partial G_1}{\partial U_0} \right]^t \cdot P_1, H \right),$$

H étant arbitraire, par identification, il vient

$$\nabla_{U_0} J = \left[\frac{\partial G_1}{\partial U_0} \right]^t \cdot P_1, \quad (4.41)$$

qui est aussi

$$\nabla_{U_0} J = \left[\frac{\partial G_1}{\partial U_0} \right]^t \circ \left[\frac{\partial G_2}{\partial U_1} \right]^t \circ \left[\frac{\partial G_3}{\partial U_2} \right]^t \circ \nabla_{U_3} J. \quad (4.42)$$

La règle est donc d'effectuer les processus adjoints en ordre inverse. Le gradient sera obtenu par l'application du dernier processus à la dernière variable adjointe.

4.2.2.3 Application aux équations de Saint-Venant alternées

En ce qui concerne le système alterné de Saint-Venant (3.34), qui est décomposé en trois équations (3.49) - (3.51), nous appliquons les résultats des deux parties ci-dessus. Supposant respectivement $P_X = (P_U, P_V, P_\phi)$ et $P_Y = (P_u, P_v, P_\Phi)$ les variables adjointes

de $X = (U, V, \phi)$ et $Y = (u, v, \Phi)$. On prend la fonction coût (4.29). On intègre le système adjoint (3.34) en sens inverse à partir du temps final :

$$P_U(T) = 0, \quad P_V(T) = 0, \quad P_\phi(T) = 0.$$

Dans l'intervalle de temps $t_{n+1} \rightarrow t_n$, supposant P_X^{n+1} connu, on détermine P_X^n par trois processus suivants en prenant les conditions initiales données par le résultats du processus précédent :

(1) Adjoint du processus de dissipation :

$$\left\{ \begin{array}{l} -\frac{\partial P_U}{\partial t} = k_2 P_U + P_u + \frac{k_1}{\Phi} \Delta(\Phi P_U) + C_U^t (C_U U - U_{obs}), \\ -\frac{\partial P_V}{\partial t} = \frac{k_1}{\Phi} \Delta(\Phi P_V) - k_2 P_V + P_v + C_V^t (C_V V - V_{obs}), \\ 0 = k_1 \Delta\left(\frac{U}{\Phi}\right) P_U + k_1 \Delta\left(\frac{V}{\Phi}\right) P_V - k_1 \Delta(\Phi P_U) \frac{U}{\Phi^2} - k_1 \Delta(\Phi P_V) \frac{V}{\Phi^2} \\ \quad - u P_u - v P_v + 2\Phi P_\phi + F_x P_U + F_y P_V, \\ 0 = P_\phi, \\ 0 = P_u, \\ 0 = P_v. \end{array} \right. \quad (4.43)$$

(2) Adjoint du processus d'advection :

$$\left\{ \begin{array}{l} -\frac{\partial P_U}{\partial t} = u \frac{\partial P_U}{\partial x} + \frac{1}{2} \frac{\partial u}{\partial x} P_U + v \frac{\partial P_U}{\partial y} + \frac{1}{2} \frac{\partial v}{\partial y} P_U + P_u \\ -\frac{\partial P_V}{\partial t} = u \frac{\partial P_V}{\partial x} + \frac{1}{2} \frac{\partial u}{\partial x} \frac{\partial P_V}{\partial x} - v \frac{\partial P_V}{\partial y} + \frac{1}{2} \frac{\partial v}{\partial y} P_V + P_v, \\ 0 = P_u u + P_v v, \\ 0 = P_\phi, \\ 0 = -P_u \Phi + U \frac{\partial P_U}{\partial x} - \frac{1}{2} \frac{\partial(P_U U)}{\partial x} + V \frac{\partial P_V}{\partial x} - \frac{1}{2} \frac{\partial(V P_V)}{\partial x}, \\ 0 = -P_v \Phi + U \frac{\partial P_U}{\partial y} - \frac{1}{2} \frac{\partial(U P_U)}{\partial y} - \frac{1}{2} \frac{\partial(V P_V)}{\partial y}. \end{array} \right. \quad (4.44)$$

(3) Adjoint du processus géostrophique

$$\left\{ \begin{array}{l} -\frac{\partial P_U}{\partial t} = -fP_V + \Phi \frac{\partial P_\phi}{\partial x} + P_u, \\ -\frac{\partial P_V}{\partial t} = fP_U + \Phi \frac{\partial P_\phi}{\partial y} + P_v, \\ -\frac{\partial P_\phi}{\partial t} = \frac{\partial(\Phi P_U)}{\partial x} + \frac{\partial(\Phi P_V)}{\partial y} - P_\Phi + C_\phi(C_\phi^t \phi - \phi_{obs}) \\ 0 = 2\Phi P_\Phi - P_u u - P_v v - \frac{\partial \phi}{\partial x} P_U - \frac{\partial \phi}{\partial y} P_V + U \frac{\partial P_\phi}{\partial x} - V \frac{\partial P_\phi}{\partial y}, \\ 0 = P_u \Phi, \\ 0 = P_v \Phi. \end{array} \right. \quad (4.45)$$

Les gradients de J par rapport aux composantes de X sont représentés par :

$$\nabla_U J = -P_U(0), \quad \nabla_V J = -P_V(0), \quad \nabla_\phi J = -P_\phi(0). \quad (4.46)$$

4.2.3 Pseudo-observations et les expériences jumelles

Pour déterminer le comportement dynamique qui s'accorde le mieux possible aux données observées dans la méthode de l'assimilation de données, un point à propos de ces observations doit être considéré est qu'il y a quelques caractères décrits ci-dessous qui peuvent faire obstacle à la mise en œuvre de cette méthode pour obtenir des résultats satisfaisants :

- Les données réelles obtenues par des mesures ne sont jamais un champ continuellement distribué, pour cela on introduit les observations dans les formules discrètes ;
- Les données observées réelles n'ont très peu de chance d'être une solutions d'un système dynamique ;
- Les données sont observées avec une erreur qui est due à la précision limitée d'instruments de mesure, cette erreur est amplifiée et se traduit par un très grand niveau d'erreur dans l'estimation du vecteur de contrôle optimal ;
- Les données ne peuvent pas être fournies complètement à tous les instants même discrets ou sur tous les points de maillage qui est défini selon un certain schéma numérique. Car le pas de temps ou le pas d'espace sont choisis d'après les expérimentations. Si les observations ne sont pas satisfaisantes, la méthode d'assimilation peut ne pas donner des résultats ;

- La période de l’assimilation n’est pas arbitraire, surtout pour le système non linéaire, elle n’est déterminée que par des expériences.

Par suite de ces raisons, on évite d’introduire des mesures réelles afin de valider une méthode d’assimilation. Une façon la plus courante est d’utiliser des pseudo-observations et de tester la méthode par des expériences jumelles. Les observations sont engendrées par une simulation numérique du modèle considéré en partant d’un état quelconque du vecteur de contrôle. Par exemple, on choisit U_{ref} un état de référence du vecteur de contrôle dans un système discret :

$$\begin{cases} X^{n+1} = F(X^n), & 0 \leq n \leq N, \\ X^0 = U_{ref}. \end{cases} \quad (4.47)$$

Les solutions $\{X^n\}$ sont prises comme pseudo-observations $\{X_{obs}^n\}$, qui sont complètes en espace et en temps discret, et U_{ref} est l’état de référence du vecteur de contrôle.

Ces observations sont ensuite utilisées dans des expériences jumelles pour la validation du code. A partir d’un état qui est complètement décorrélé à l’état de référence, l’algorithme d’assimilation nous donnera un état optimal U_{opt} , par comparaison entre U_{opt} et U_{ref} suivant un critère, qui porte souvent sur le niveau relatif d’erreur entre ces deux états, nous pouvons donc vérifier l’exactitude de l’algorithme et tester la capacité d’assimilation de notre code.

Une fois la validation de notre système d’optimalité est vérifiée, la dépendance du taux de convergence sur la distribution des données peut être observée, aussi la question concernant la suffisance des données peut être examinée. Si l’on ajoute des erreurs aléatoires sur les observations générées par le modèle ou bien on échantillonne les données dans le champ d’observation, on peut aussi étudier à quel niveau la méthode d’assimilation va fonctionner.

4.2.4 Test de l’exactitude du gradient

Dans le processus d’optimisation, il est très important de calculer correctement le vecteur gradient de la fonction coût par rapport aux variables de contrôle afin de trouver la bonne direction de descente. L’exactitude du calcul du gradient agit sur la vitesse de la convergence de la minimisation et sur la valeur correcte du minimum. Donc avant d’exécuter le processus d’optimisation, on va vérifier le gradient de la fonction coût qui est obtenu par la méthode adjointe.

Pour tester le gradient, Couriter ([11]) a montré que le test fondé sur le développement de Taylor à l’ordre un de la fonctionnelle J est efficace. On écrit le développement de Taylor

aux points $U + \alpha h$ dans le voisinage de U avec la perturbation αh , h une direction fixée, $\alpha \rightarrow 0$, par :

$$J(U + \alpha h) - J(U) = a \langle \nabla J(U), h \rangle + O(\alpha),$$

donc on va tester une fonction $f(\alpha)$ pour laquelle

$$\begin{aligned} \lim_{\alpha \rightarrow 0} f(\alpha) &= \lim_{\alpha \rightarrow 0} \frac{J(U + \alpha h) - J(u)}{\alpha \langle \nabla J(U), h \rangle} \\ &= 1. \end{aligned} \tag{4.48}$$

Pour vérifier chaque composante du vecteur gradient, on utilise ce test non seulement pour J globale mais aussi pour chaque "composante" de la fonction coût J , par exemple, si

$$J(U) = \sum_i J_i(U) = \frac{1}{2} \sum_i \|X(t_i) - X^{obs}(t_i)\|_{\Omega}^2,$$

on va vérifier séparément que chaque J_i satisfait le rapport (4.48). Dans nos tests, on choisit arbitrairement le vecteur d'état U , les observations sont prises seulement au temps final, c.-à-d.

$$J(U) = \frac{1}{2} \|X(T) - X^{obs}(T)\|_{\Omega}^2;$$

la direction perturbée h est choisie comme l'opposé du vecteur gradient calculé au point U : $[\nabla J(U)]^t$.

Du tableau IV.1, on donne l'un des résultats de ce test, on voit que la valeur de la fonction $f(\alpha)$ tend vers 1 quand α varie de 2^0 à 2^{-21} , mais on remarque que si α est proche du zéro de la machine, il ne vérifie plus ce principe.

α	Rapport(4.48)	α	Rapport(4.48)
2^0	1.92878	2^{-11}	0.999987
2^{-1}	1.69950	2^{-12}	0.999394
2^{-2}	1.39879	2^{-13}	0.999987
2^{-3}	1.21369	2^{-14}	0.998404
2^{-4}	1.11040	2^{-15}	0.997084
2^{-5}	1.05578	2^{-16}	0.994445
2^{-6}	1.02771	2^{-17}	0.996556
2^{-7}	1.01348	2^{-18}	1.00500
2^{-8}	1.00632	2^{-19}	0.988111
2^{-9}	1.00269	2^{-20}	1.04723
2^{-10}	1.00081	2^{-21}	0.945884

TAB. IV.1 – Test du gradient de la fonction coût

4.2.5 Expérience numérique

A partir des données simulées, nous allons tester dans cette partie la capacité du système adjoint déduit par la méthode adjointe pour assimiler l'état optimal du vecteur de contrôle. Les expériences sont faites en prenant le même domaine géographique du modèle de simulation au chapitre III, dont l'on a pris 50 pas d'espace pour la longueur et 52 pour la largeur, le pas de temps correspondant à l'ajustement de la forme du fond de la rivière $\Delta t_z = 3$ jours, le nombre de pas de temps pour l'ajustement d'écoulement $N_f = 200\Delta t_f$, avec $\Delta t_f = 140\text{sec}$, et pour l'ajustement du champ de la concentration de sédiments $N_s = 200\Delta t_s$, avec $\Delta t_s = 70\text{sec}$.

D'après l'expérience acquise, la condition initiale du fond de la rivière joue le rôle dominant dans l'évolution de la rivière, donc on ne reconstruit que l'état initial du lit de la rivière. Les expériences sont faites selon le plan suivant :

- On a choisi les résultats de la simulation (cf. Fig. IV.1) comme état de référence, et l'intégration du modèle à partir de cet état comme observations simulées. La période d'intégration est aussi la période d'assimilation ;
- La procédure d'optimisation est initialisée en ajoutant une perturbation aléatoire à l'état de référence (voir Fig. IV.2), où l'état de référence est aussi la solution exacte de la minimisation. Dans ce cas là, nous connaissons la solution exacte du problème de contrôle et la valeur minimale de la fonction coût qui mesure l'écart entre la solution du modèle associée aux conditions initiales et les observation simulées au temps final. Cette valeur minimale est égale à zéro ;

Dans la première expérience, nous faisons d'abord une expérience sur une durée de 15 jours ($= 5\Delta t_z$). On ne dispose des observations qu'au temps final. Le but est de vérifier le bon fonctionnement du code adjoint avant d'envisager son utilisation sur une période de temps d'assimilation plus longue. Les figures Fig. IV.3 et Fig. IV.4 représentent les variations de logarithme décimal de la fonction coût normalisée et du gradient normalisé en fonction des itérations k , où J_0 est la fonction coût et $\vec{\nabla} J_0$ est le vecteur gradient de la fonction coût au début de la procédure d'optimisation. On remarque qu'après 5 itérations, la fonctionnelle normalisée est réduite d'un facteur 5,35 à l'échelle du logarithme décimal tandis que la norme du gradient normalisée l'est de 4,02.

La figure Fig. IV.5 représente l'état assimilé du fond de la rivière. On compare dans la figure Fig. IV.6 la cote du fond dans le canal central avant minimisation, après minimisation et de l'état de référence. Sur le point de l'indice (28,26), on a constaté que l'écart entre l'état de référence et l'état de contrôle diminue de 6,89 mètre à 0,0018 mètre après l'optimisation.

Dans la deuxième expérience, le calcul est effectué sur une période d'assimilation de 30 jours. Sur les figures Fig. IV.7 et Fig. IV.8, la fonction coût normalisée est réduite d'un facteur 5,36 tandis que la norme du gradient l'est de 3,93 après 58 itérations. La figure Fig. IV.9 représente la forme du fond de la rivière assimilée en 30 jours. Sur le même point de grille dans le canal de coordonnée (28,26), on remarque sur la figure Fig. IV.10 que, lors de la minimisation, l'écart entre la cote du lit de la rivière de l'état de référence et l'état de contrôle est réduit de 6,89 mètre à 0,0007 mètre.

A partir de ces résultats obtenues, on peut conclure que la méthode fonctionne aussi bien que pour une durée d'assimilation de 15 jours, mais l'assimilation en 30 jours est beaucoup plus coûteuse. Elle coûte 9,29 heures (33436.33s) de temps CPU tandis que 0.33 heures de temps CPU pour l'assimilation en 15 jours.

Remarque : Dans l'étude de l'impact de la distribution temporelle des observations sur les résultats d'assimilation, on remarque que les résultats sont plus satisfaisants dans le cas où les observations sont disponibles au temps initial et au temps final que dans le cas où les observations sont disponibles à tous les pas de temps.

Conclusion : Les expériences nous montrent que l'assimilation des données est plus précise lorsque la durée diminue. Cela résulte du conditionnement de la matrice Hessienne de la fonction coût par rapport aux variables de contrôle (ici, ce sont les conditions initiales), qui se détériore avec l'allongement de l'intervalle de temps d'intégrer du modèle. Pour améliorer la précision des résultats assimilés, l'utilisation des méthodes de contrôle optimal pour l'assimilation variationnelle de données sur de longues périodes de temps devrait tenir compte de l'étude au second ordre. Dans l'article de Wang, Navon et Le Dimet [68], des études ont déjà été menées pour la détermination du spectre du Hessien de la fonction coût en supposant l'existence d'un intervalle de temps optimal dans lequel les résultats d'assimilation ajusteraient le mieux aux observations. Le calcul du produit de Hessien et vecteur, que l'on a toujours besoin de calculer, peut être déduit par l'analyse du système adjoint au second ordre du modèle direct.

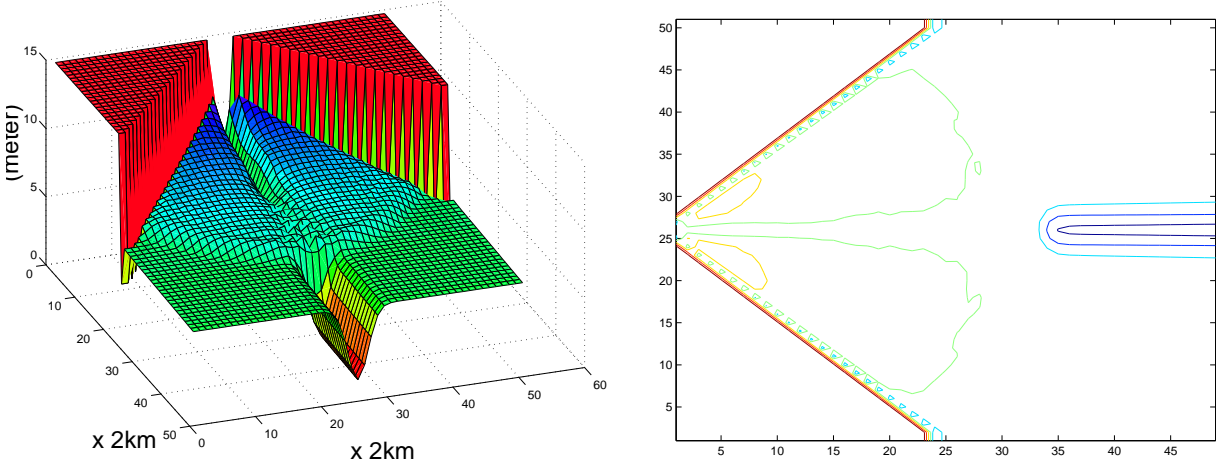


FIG. IV.1 – L'état de reference du fond de la rivière.

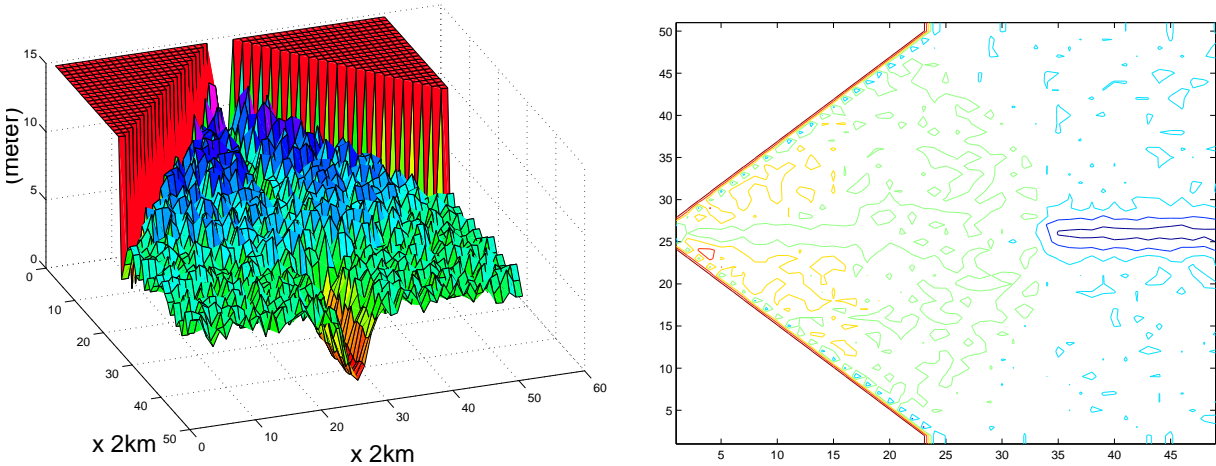


FIG. IV.2 – L'état d'initialisation (first guess) du fond de la rivière.

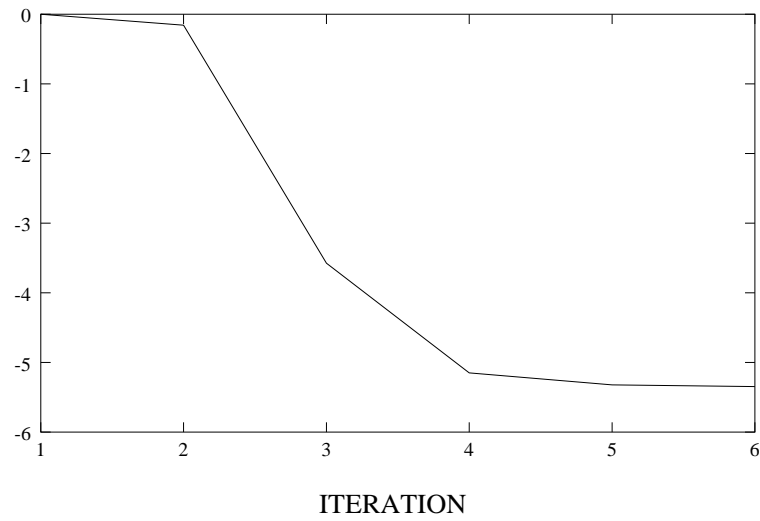


FIG. IV.3 – La variation de $\log(J_k/J_0)$ en fonction des itérations k sur l'assimilation de 15 jours.

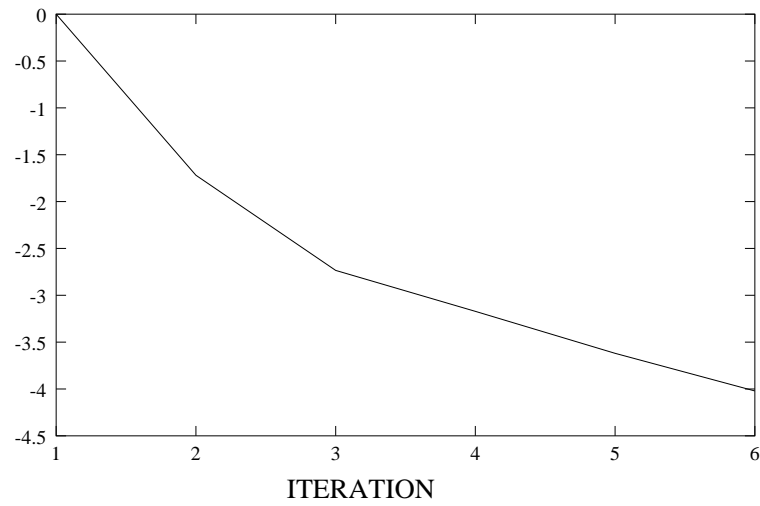


FIG. IV.4 – La variation de $\log(\|\nabla J_k\|/\|\nabla J_0\|)$ en fonction des itérations k sur l'assimilation de 15 jours.

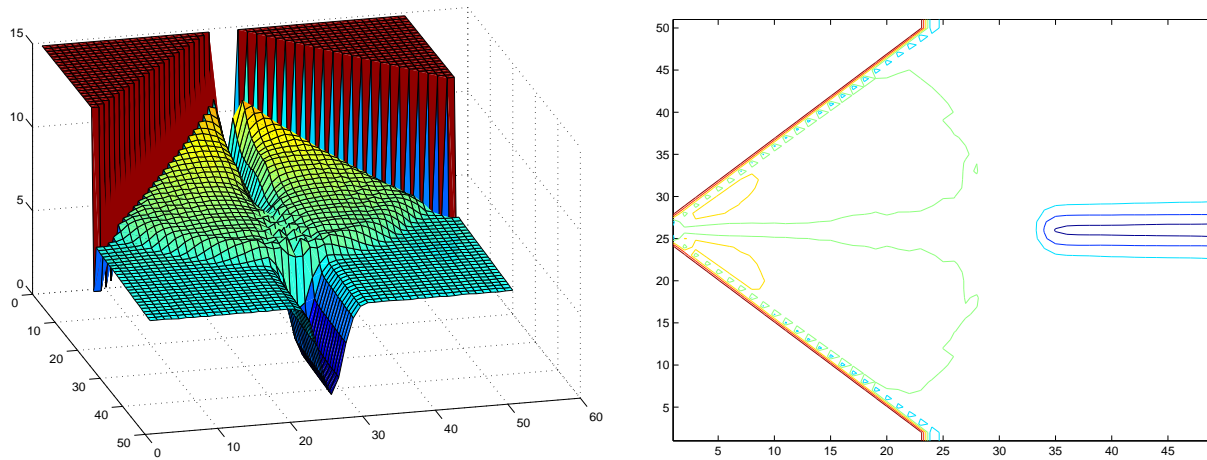


FIG. IV.5 – *L'état optimal du fond initial de la rivière après l'assimilation sur l'assimilation de 15 jours.*

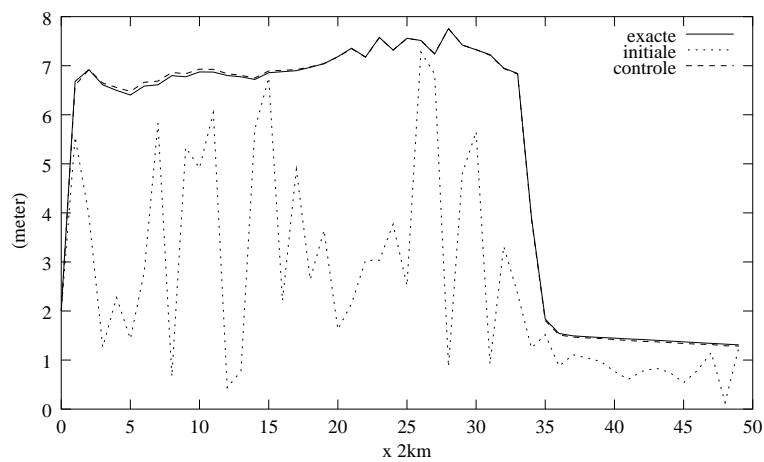


FIG. IV.6 – *Les comparaisons de trois états dans le canal central sur l'assimilation de 15 jours.*

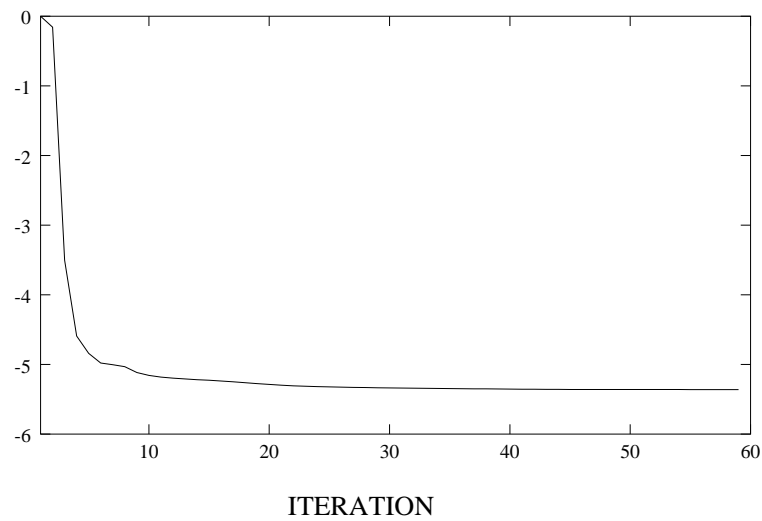


FIG. IV.7 – La variation de $\log(J_k/J_0)$ en fonction des itérations k sur l'assimilation de 30 jours.

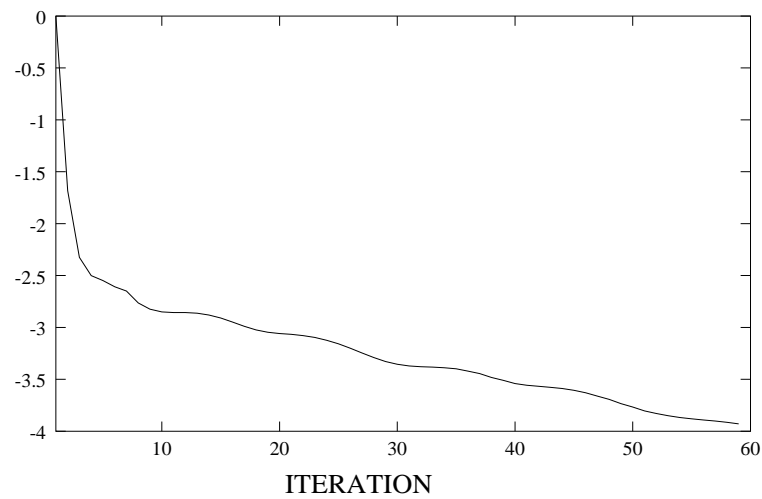


FIG. IV.8 – La variation de $\log(\|\nabla J_k\|/\|\nabla J_0\|)$ en fonction des itérations k sur l'assimilation de 30 jours.

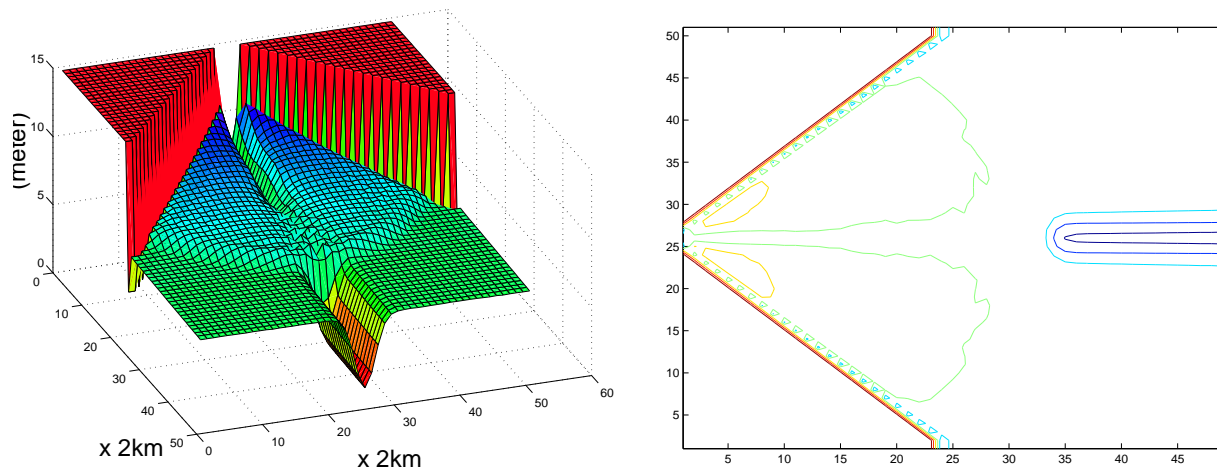


FIG. IV.9 – L'état optimal du fond initial de la rivière après l'assimilation sur 30 jours.

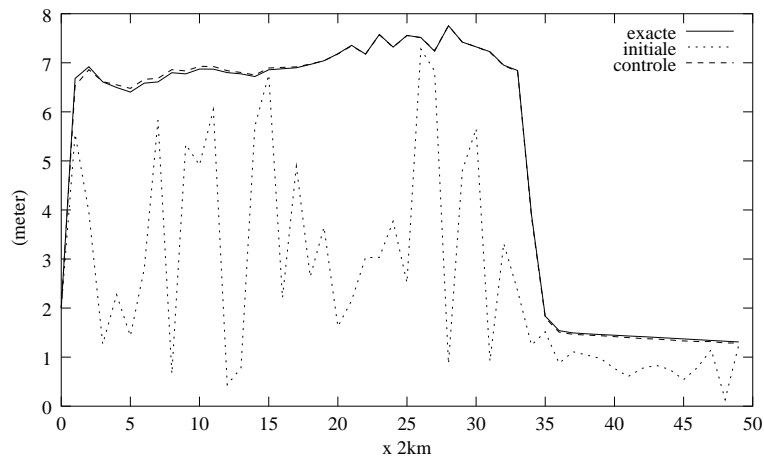


FIG. IV.10 – Les comparaisons de trois états dans le canal central sur l'assimilation de 30 jours.

4.3 Assimilation de données II – Identification des paramètres

4.3.1 Introduction

Pour décrire le transport des sédiments en rivière, qui joue un rôle très important pour l'évolution de son lit, des formules différentes sont proposées pour déterminer les proces-

sus de dépôt et d'érosion, mais il reste encore des difficultés pour mesurer ou identifier des paramètres du modèle pour ces processus que ce soit dans des expériences en laboratoire ou à partir des observations. En plus, il existe plusieurs formules semi-empiriques décrivant les termes de dépôt et d'érosion qui sont établies à partir des données observées aussi des expériences en laboratoire, parmi elles, il y a des coefficients impossibles à ajuster sont généralement déduits des observations ou des expériences empiriques. Tout cela empêche la modélisation numérique du transport des sédiments d'atteindre un bon niveau de précision.

Dans la partie précédente, nous avons remarqué que la méthode d'assimilation des données variationnelles est efficace pour déterminer la condition initiale pour le modèle à partir des observations. Nous sommes intéressés par le problème de contrôle suivant : est-ce qu'on peut utiliser cette méthode pour identifier des paramètres du modèle à partir des données observées afin de trouver un état optimal du système et que la prévision du modèle s'adapte aux données observées? Pour l'instant, il existe déjà quelques travaux en ce sens en océanographie ou en hydrologie (Bennett[3], Panchang[50], Das et Lardner[16], Panchang et Richardon [49]), mais peu sur le problème de transport des sédiments. Nous allons donc utiliser, dans cette partie, la méthode d'assimilation variationnelle de données pour identifier certains paramètres empiriques dans un modèle bidimensionnel sur un domaine représentant un estuaire idéal.

4.3.2 Système de contrôle

L'idée de l'identification des paramètres du modèle avec la méthode variationnelle à partir des observations est qu'en prenant le vecteur des paramètres comme le vecteur de contrôle, on va minimiser une fonction coût qui mesure l'écart entre la prévision du modèle correspondant au vecteur de contrôle et les observations en vue de trouver un vecteur optimal de contrôle.

Posons K le vecteur des paramètres en dimension l , nous considérons l'équation évolutive discrétisée en espace sous une forme générale :

$$\begin{cases} \frac{\partial X}{\partial t} + F(X(K, t), K) = 0, \\ X(K, t = 0) = U, \end{cases} \quad (4.49)$$

et la fonction coût associée, qui est

$$J(K) = \frac{1}{2} \int_0^T \|C \cdot X(K) - X_{obs}\|_{\Omega}^2 dt. \quad (4.50)$$

Ici $X = (u, v, h, S, Z_b)$ est le vecteur d'état.

S'il y a une perturbation H_K sur K , posons \hat{X}_K la Gâteaux dérivée de la variable d'état X par rapport à K dans la direction H_K , qui est la solution du système linéaire tangent :

$$\begin{cases} \frac{\partial \hat{X}_K}{\partial t} + \left[\frac{\partial F}{\partial X} \right] \hat{X}_K + \left[\frac{\partial F}{\partial K} \right] H_K = 0, \\ \hat{X}_K(K, t = 0) = 0. \end{cases} \quad (4.51)$$

La Gâteaux dérivée de J par rapport à K dans la direction H_K vérifie

$$\hat{J}(K, H_K) = \int_0^T \langle C \cdot X(K) - X_{obs}, C \cdot \hat{X}_K \rangle_{\Omega} dt. \quad (4.52)$$

Pour déterminer le gradient de la fonction coût $J(K)$ par rapport à K dans la direction H_K , comme pour le problème de contrôle des conditions initiales, nous introduisons un vecteur P , adjoint du vecteur X , en faisant le produit scalaire $\langle \cdot, \cdot \rangle$ entre P et chacun des membres de la première équation (4.51), et en intégrant sur l'intervalle de temps $[0, T]$, on obtient :

$$\int_0^T \langle P, \frac{\partial \hat{X}_K}{\partial t} \rangle dt + \int_0^T \langle P, \left[\frac{\partial F}{\partial X} \right] \hat{X}_K \rangle dt = - \int_0^T \langle P, \left[\frac{\partial F}{\partial K} \right] H_K \rangle dt,$$

après une intégration par parties, il vient

$$\begin{aligned} \langle P(T), \hat{X}_K(T) \rangle - \langle P(0), \hat{X}_K(0) \rangle + \int_0^T \langle -\frac{\partial P}{\partial t} + \left[\frac{\partial F}{\partial X} \right]^t P, \hat{X}_K \rangle dt \\ = - \int_0^T \langle \left[\frac{\partial F}{\partial K} \right]^t P, H_K \rangle dt. \end{aligned}$$

Si l'on définit le vecteur adjoint P la solution du système :

$$\begin{cases} -\frac{\partial P}{\partial t} + \left[\frac{\partial F}{\partial X} \right]^t P = C^t \cdot (C \cdot X(K) - X_{obs}), \\ P(T) = 0, \end{cases} \quad (4.53)$$

et grâce à (4.51), on a donc

$$\hat{J}(K, H_K) = - \int_0^T \langle \left[\frac{\partial F}{\partial K} \right]^t P, H_K \rangle dt, \quad (4.54)$$

qui donne le gradient de J par rapport au vecteur K :

$$\nabla_K J = - \int_0^T \left[\frac{\partial F}{\partial K} \right]^t P dt. \quad (4.55)$$

En remarquant que le système adjoint (4.53) est le même que celui pour le calcul du gradient de J par rapport à la condition initiale, on trouve que, lorsque l'on a déduit le système adjoint du système direct gouverné par des équations aux dérivées partielles, on peut l'appliquer pour la résolution des autres problèmes de contrôle optimal.

4.3.3 Choix des paramètres à identifier

Dans notre modèle, nous avons utilisé dans le chapitre II des formules semi-empiriques (dans la section 2.1.3) en ce qui concernent les termes hydrauliques dans lesquelles il y a des coefficients de rapport empiriques. Par exemple, la capacité de l'écoulement à transporter des sédiments en suspension est caractérisée par :

$$S^* = b \left(\frac{|\vec{u}|^3}{gh\omega} \right)^m,$$

et le taux de transport des sédiments charriés est décrit par :

$$\vec{g}_b = \begin{cases} \Gamma \left(1 - \frac{\nu_c}{|\vec{u}|} \right) |\vec{u}|^n h^p d^q \vec{u}, & \text{si } \nu_c \leq |\vec{u}|, \\ 0, & \text{si } \nu_c > |\vec{u}|. \end{cases} \quad (4.56)$$

La détermination correcte des paramètres particulièrement des variables S^* et \vec{g}_b est un problème très important grâce à sa influence indispensable sur l'exactitude et la précision du calcul de l'évolution du fond de la rivière et par conséquent sur l'écoulement liquide.

On se place dans un problème simplifié où la taille des particules à transporter sur le fond est considérée uniforme partout. La vitesse de dépôt des sédiments ω correspondant à leur taille peut être mesurée par les outils scientifiques sans beaucoup de difficulté. En revanche, les coefficients b, m, Γ, n, p, q sont empiriques, ainsi que le paramètre α dans le terme représentant les processus de dépôt et d'érosion $\alpha\omega(S - S^*)$ pour les équations de transport des sédiments en suspension et les équations de la variation du fond de la rivière.

Nous voulons alors essayer d'estimer ces paramètres empiriques, en combinant les données observées, par la résolution d'un problème d'assimilation variationnelle de données. Dans ce problème, on prend $K = (\alpha, b, m, \Gamma, n, p, q)$ comme vecteur de contrôle et on cherche sa valeur optimale en minimisant une fonction coût qui mesure l'écart entre la solution du modèle et les données observées.

4.3.4 Validation du gradient

Afin de s’assurer que l’on a une bonne direction de descente, dans l’algorithme d’optimisation, on vérifie numériquement l’exactitude du gradient (4.55) par rapport à chaque composante du vecteur K en prenant la même formule que (4.48). Sur la figure IV.3, on trouve que la fonction $f(\alpha)$ en fonction du logarithme décimal de α tend vers 1 quand α varie entre 2^{-1} et 2^{-20} avec une perturbation dans la direction $(Grad_{K_i} J(K))_{1 \leq i \leq l}$, avec l la dimension de K .

On constate que la précision de la différence entre 1 et $f(\alpha)$ est d’autant meilleure que α est petit, mais lorsque α devient inférieur à 10^{-6} , ce rapport augmente à nouveau en raison des erreurs numériques d’arrondi. Il s’agit donc d’un bon test du gradient lorsque l’on connaît une valeur exacte du gradient, mais cette méthode en elle-même est à déconseiller pour un calcul direct du gradient en raison de ces instabilités numériques.

α	Rapport(4.48)	α	Rapport(4.48)
2^0	22.6728	2^{-11}	1.0011543
2^{-1}	5.744897	2^{-12}	0.999938
2^{-2}	2.599639	2^{-13}	0.999903
2^{-3}	1.659436	2^{-14}	0.999468
2^{-4}	1.299596	2^{-15}	0.997084
2^{-5}	1.142836	2^{-16}	0.983938
2^{-6}	1.069749	2^{-17}	0.983734
2^{-7}	1.034488	2^{-18}	0.989210
2^{-8}	1.017159	2^{-19}	1.122073
2^{-9}	1.008566	2^{-20}	1.126961
2^{-10}	1.004353		

TAB. IV.2 – Validation du gradient

4.3.5 Expériences numériques - Identification avec des observations complètes

L’expérience jumelle est organisée comme au paragraphe 4.2.5, dans lequel, les observations sont générées sur le même domaine par les résultats simulés du modèle dont les paramètres sont pris dans la colonne “valeur exacte” du tableau Tab IV.3. Dans un premier temps, l’identification est effectuée sur la durée d’assimilation de 3 jours, le champs d’écoulement, la distribution des sédiments et le fond de la rivière sont observés au temps final.

Nous connaissons donc la solution exacte du problème de contrôle et la valeur minimale de la fonction coût qui est zéro. Pour initialiser la procédure de minimisation, nous choisissons pour les valeurs d'initialisation par ajoutant une perturbation aléatoire de l'ordre maximal de 50% sur les valeurs exactes, on les figure dans la colonne "valeur initialisée" du Tab IV.3.

La figure FIG.IV.11 montre la décroissance des logarithmes décimaux de la fonction coût normalisée sur 50 itérations. A la 50-ème itération le critère d'arrêt est satisfait. Ici nous prenons $v = 5 \times 10^{-7}$ dans la formule (4.21). La fonction coût normalisée décroît de 6.74 à 3.58×10^{-11} . La figure IV.12 représente la décroissance des logarithmes décimaux de la norme du gradient normalisée de la fonction coût sur 50 itérations, on voit sur cette figure que la norme décroît de 32.48 à 6.84×10^{-5} , ce qui s'explique par le fait que l'optimisateur a convergé vers un minimum local.

La colonne "valeur identifiée" du tableau TAB. IV.3 donne la valeur optimale du vecteur de contrôle par rapport à l'optimisation. Tous les paramètres sauf Γ tendent vers leurs valeurs exactes sur 50 itérations. Sur la figure FIG. IV.13, on voit plus en détail la variation de chaque paramètre durant la minimisation par comparaison à leur valeur exacte. Les valeurs de n, p et q ne varient pas pendant les premières itérations de la procédure d'optimisation, parce que les gradients initiaux de la fonction coût par rapport à n, p et q sont moins importants que ceux de α, b et m . Et la valeur de Γ n'a pratiquement pas varié durant la procédure d'identification, donc dépend de la valeur de départ.

Paramètres	Valeur exacte	Valeur initialisée	Valeur identifiée
b	0.350000	0.385134	0.350000
m	0.920000	1.19991	0.920000
α	0.300000	0.295765	0.300000
Γ	5.0000	5.20477	5.20924
n	3.00000	2.15512	3.00048
p	-0.250000	-0.320287	-0.250718
q	0.250000	0.225900	0.255743

TAB. IV.3 – Résultat de l'identification avec des observations complètes

Les variables hydrauliques n'étant pas indépendants, l'identification de tous les paramètres peut être influencée par l'effet du couplage, la mauvaise identification de Γ peut être due à ce couplage. Pour regarder cette influence sur l'identification, quelques résultats

numériques ont été constatés dans les expériences suivantes :

- On identifie seulement le paramètre Γ en fixant les autres paramètres à leurs valeurs exactes. Nous avons résumé dans le tableau TAB. IV.4, dans trois cas, la valeur de départ Γ_0 , la valeur optimale Γ_{opt} , la fonctionnelle au temps initial et au temps final, ainsi que leurs normes des gradients et le nombre d'itérations. On remarque que dans tous les cas on a bien retrouvé la valeur exacte de Γ qui ne dépend pas de la valeur d'initialisation.

Γ_0	Γ_{opt}	J_0	J_N	$\ \nabla J_0\ $	$\ \nabla J_N\ $	N
0.5	5.0	7.16338×10^{-3}	0	2.65044×10^{-3}	0	3
5.20477	5.0	1.48316×10^{-5}	9.09495×10^{-13}	1.20613×10^{-4}	1.12053×10^{-10}	4
5×10^2	5.0	86.6765	0	0.291548	0	3

TAB. IV.4 – Estimation de Γ avec les observations complètes, où la valeur exacte de Γ est égale à 5.

- On identifie seulement le paramètre q . D'après nos expériences, même la détermination de q est assez délicate. D'abord, le choix de la valeur de départ n'est pas arbitraire, elle dépend de la taille des particules d . Sinon, il entraînera une divergence de l'algorithme de minimisation à cause de la instabilité de l'intégration du modèle direct. Par exemple, dans le cas où $d = 1 \times 10^{-3}m$, la minimisation diverge si la valeur de départ de q ne peut pas dépasser 1. Pour cet exemple, on résume les résultats d'optimisation dans le tableau TAB.IV.5.

q_0	q_{opt}	J_0	J_N	$\ \nabla J_0\ $	$\ \nabla J_N\ $	
2.5×10^{-2}	0.249998	0.123140	3.63798×10^{-12}	1.59432	2.02935×10^{-6}	10
0.225900	0.25	2.90158×10^{-4}	0	2.12658×10^{-2}	0	5
1	0.25	8.74447×10^{-4}	0	2.35391×10^{-2}	0	5

TAB. IV.5 – Estimation de q avec les observations complètes, où la valeur exacte de q est égale à 0.25, la taille des sédiments $d = 1 \times 10^{-3}$.

- Un autre essai est fait, en prenant le paramètre Γ à la valeur exacte, sur l'identification des six autres paramètres. Les résultats au tableau IV.6 nous indique que en 59 itérations, les différences entre les valeurs finales et les valeurs exactes de tous les paramètres sont inférieures à 0.0013%, surtout le paramètre q est beaucoup mieux identifié sans la présence de l'identification du paramètre Γ . En revanche, le contraire n'est pas vrai, c'est-à-dire que la qualité d'identification de Γ ne dépend pas de la présence de q .

Paramètres	Valeur exacte	Valeur initialisée	Valeur identifiée
b	0.35	0.385134	0.349995
m	0.92	1.19991	0.920006
α	0.3	0.295765	0.300002
n	3.0	2.15512	2.99996
p	-0.25	-0.320287	-0.249980
q	0.25	0.225900	0.250002

TAB. IV.6 – Résultats de l'identification en 6 paramètres à partir des observations complètes.

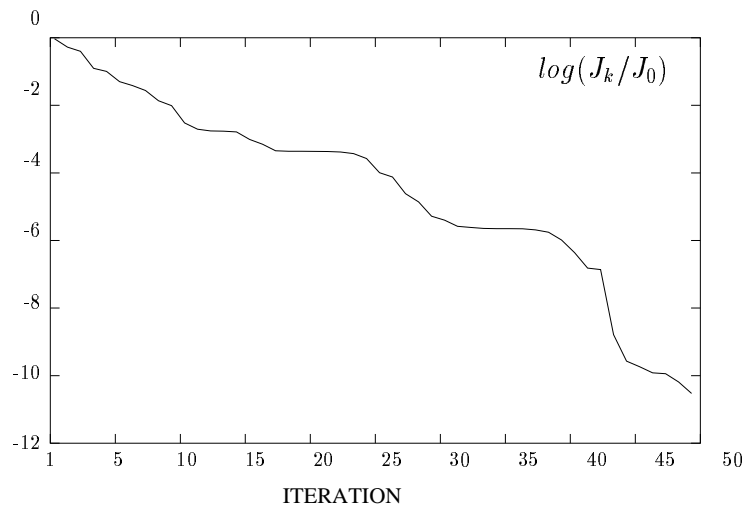


FIG. IV.11 – La variation de $\log(J_k/J_0)$ en fonction des itérations k avec des observations complètes.

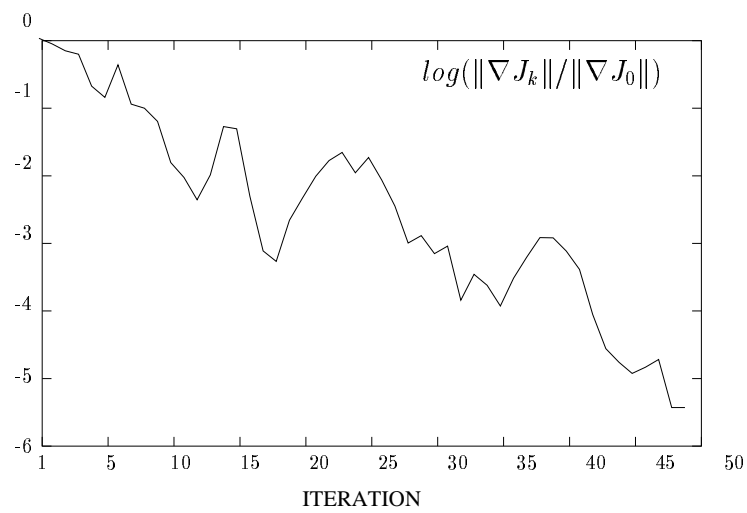


FIG. IV.12 – La variation de $\log(\|\nabla J_k\|/\|\nabla J_0\|)$ en fonction des itérations k avec des observations complètes.

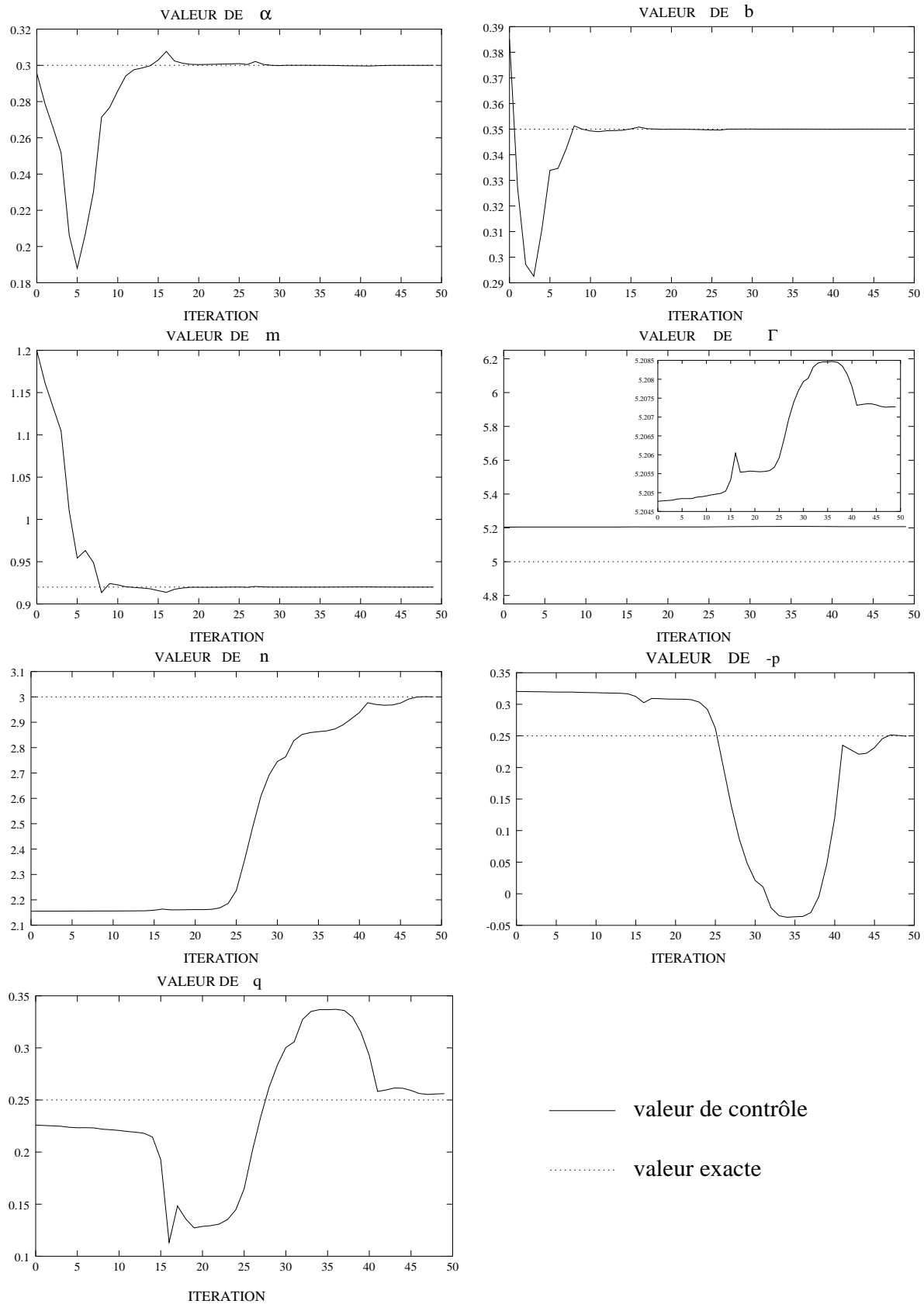


FIG. IV.13 – Résultats de l'identification en 7 paramètres avec des observations complètes.

4.3.6 Expériences numériques - Identification à partir des données non complètes

En réalité, ce n'est pas réaliste que les données soient observées sur tous les points de grille, parce que les données dans une rivière sont observées ou mesurées par des stations hydrauliques qui ne sont pas distribuées de manière homogène dans tout le domaine de la rivière, ni sur tous les noeuds de maillage. Si l'on veut utiliser des observations non complètes pour spécifier les paramètres du modèle avec la méthode variationnelle d'assimilation, c'est important de tester la capacité de la méthode d'assimilation de données pour retrouver la valeur optimale du vecteur de contrôle dans le cas où le nombre de données est inférieur à celui de variables de contrôle. Ce sont des problèmes qui intéressent les mathématiciens et les physiciens.

Dans les premières expériences, nous allons considérer une question : Quel est le nombre minimum de observations dont nous avons besoin pour l'identification des paramètres et quels sont leurs meilleurs positions pour donner des résultats satisfaisants? Ensuite, dans le cas où les données observées sont clairsemés et ne sont pas suffisantes pour spécifier de bons résultats pour les paramètres à identifier, nous allons introduire une méthode de régularisation en ajoutant un terme de pénalité sur la fonctionnelle.

Pour traiter les cas où il y a des observations incomplètes, la formule (4.50) qui définit la fonction coût fait intervenir une matrice de pondération W qui déclare la situation des observations. Dans nos expériences, une matrice diagonale

$$W = \text{diag}(W_U, W_V, W_h, W_S, W_{Z_b})$$

a été choisie pour présenter les poids avec des sous matrices W_U , W_V , W_h , W_S et W_{Z_b} .

Les éléments des sous matrices sont définis par

$$(W_U)_{i,j} = \begin{cases} 1, & \text{s'il y a des données sur le noeud } (i,j), \\ 0, & \text{sinon.} \end{cases} \quad (4.57)$$

La fonction coût devient

$$J(K) = \frac{1}{2} \int_0^T \langle W(C \cdot X - X_{obs}), W(C \cdot X - X_{obs}) \rangle dt. \quad (4.58)$$

De la même manière, le système adjoint correspondant est donné par

$$\begin{cases} -\frac{\partial P}{\partial t} + \left[\frac{\partial F}{\partial t} \right]^t P = C^t W^t W (C \cdot X - X_{obs}) \\ P(T) = 0, \end{cases} \quad (4.59)$$

et le gradient de la fonction coût par rapport au vecteur de contrôle est le même que (4.55).

L'expérience est faite sous la même condition que le paragraphe précédent. Nous supposons que nous n'avons que les données observées au fond de la rivière, notamment dans le canal central où l'action de la sédimentation est plus importante par rapport aux autres régions. Ça veut dire que l'on a,

$$(W_U)_{i,j} = (W_V)_{i,j} = (W_h)_{i,j} = (W_S)_{i,j} = 0, \text{ pour tous } i, j,$$

$$(W_{Z_b})_{i,j} = \begin{cases} 1, & j = j_c, \\ 0, & j \neq j_c, \end{cases} \quad (4.60)$$

où j_c est l'indice correspondant au centre du canal. Pour un domaine de 50×52 points, $j_c=26$.

Comme l'illustrent les figures FIG. IV.14 et FIG. IV.15, on constate que l'algorithme de minimisation s'arrête à la 61-ème itération avec des décroissances remarquables, où le logarithme décimal de la fonction coût normalisée décroît de 0 à -9.63917 et la norme de son gradient normalisé décroît de 0 à -6.10597. En comparaisons avec les valeurs exactes de chaque paramètre sur des graphiques de la figure Fig. IV.16, qui ont montré la variation de chaque composante du vecteur de contrôle au cours de la minimisation. Dans le tableau Tab. IV.7, nous avons trouvé que, comme dans le cas où il y a des observations complètes, les valeurs de contrôle des paramètres α, b, m, n, p peuvent converger vers leurs valeurs exactes à partir d'une valeur perturbée aléatoire, en revanche, la valeur du paramètre Γ n'a pas varié beaucoup au cours de l'optimisation, la valeur finale après la minimisation est égale à 5.21536 quand la valeur de départ est de 5.20477.

En fait, la figure Fig. IV.17 représente la variation du gradient de la fonction coût par rapport à chaque paramètre au cours de la minimisation, à partir de laquelle nous remarquons que le gradient par rapport à Γ n'a augmenté que 5×10^{-3} au cours de la minimisation, ce qui nous demande de chercher un autre algorithme pour l'identification de Γ .

Nous sommes intéressés maintenant à la sensibilité des résultats d'assimilation au zone d'observations. On a réussi à trouver numériquement que, en utilisant l'observation sur la cote du fond au canal central et en seulement 13 nœuds de grille au début du canal, le moindre nombre d'observations, on arrive une bonne identification. Le résultat dans le tableau TAB.IV.7 montre que le résultat est autant satisfaisant que dans le cas où il y a des observations complètes. Sinon, dans les cas où le nombre de sites est inférieur à 13,

ou bien les observations se trouvent plus loin de l'entrée de l'écoulement, on n'arrive pas à continuer la procédure de la minimisation.

Paramètre	valeur exacte	initialisée	identifiée
b	0.35	0.385134	0.350007
m	0.92	1.19991	0.919986
α	0.3	0.295765	0.299985
Γ	5.0	5.20477	5.20992
n	3.0	2.15512	2.99985
p	-0.25	-0.320287	-0.250137
q	0.25	0.225900	0.255875

TAB. IV.7 – Résultats de l'identification avec des observations au fond dans le canal sur les points (i, j) , où $i = 1, \dots, 13$, $j = 26$.

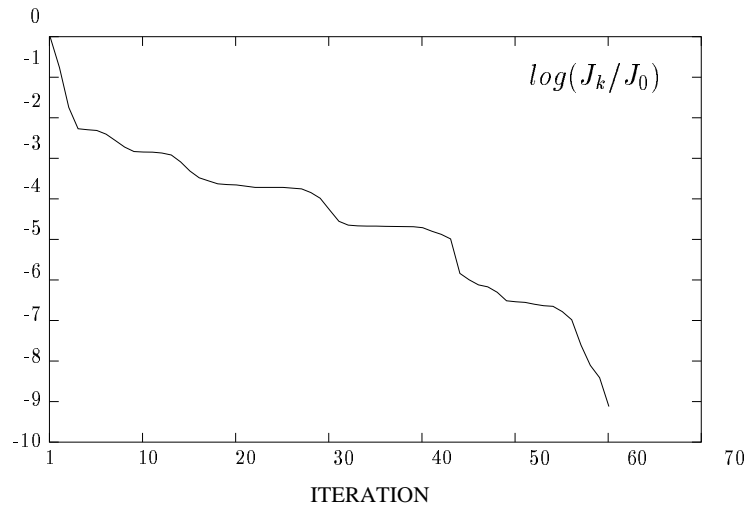


FIG. IV.14 – La variation de $\log(J_k/J_0)$ en fonction des itérations k avec des observations au fond du canal.

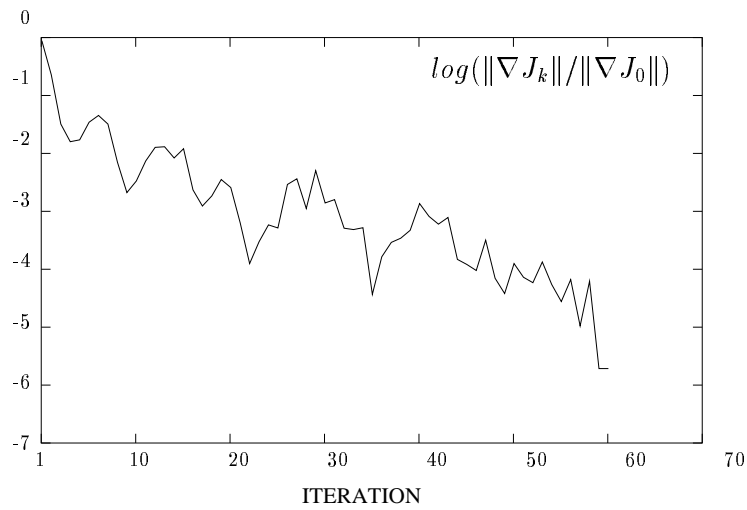


FIG. IV.15 – La variation de $\log(\|\nabla J_k\|/\|\nabla J_0\|)$ en fonction des itérations k avec des observations au fond du canal.

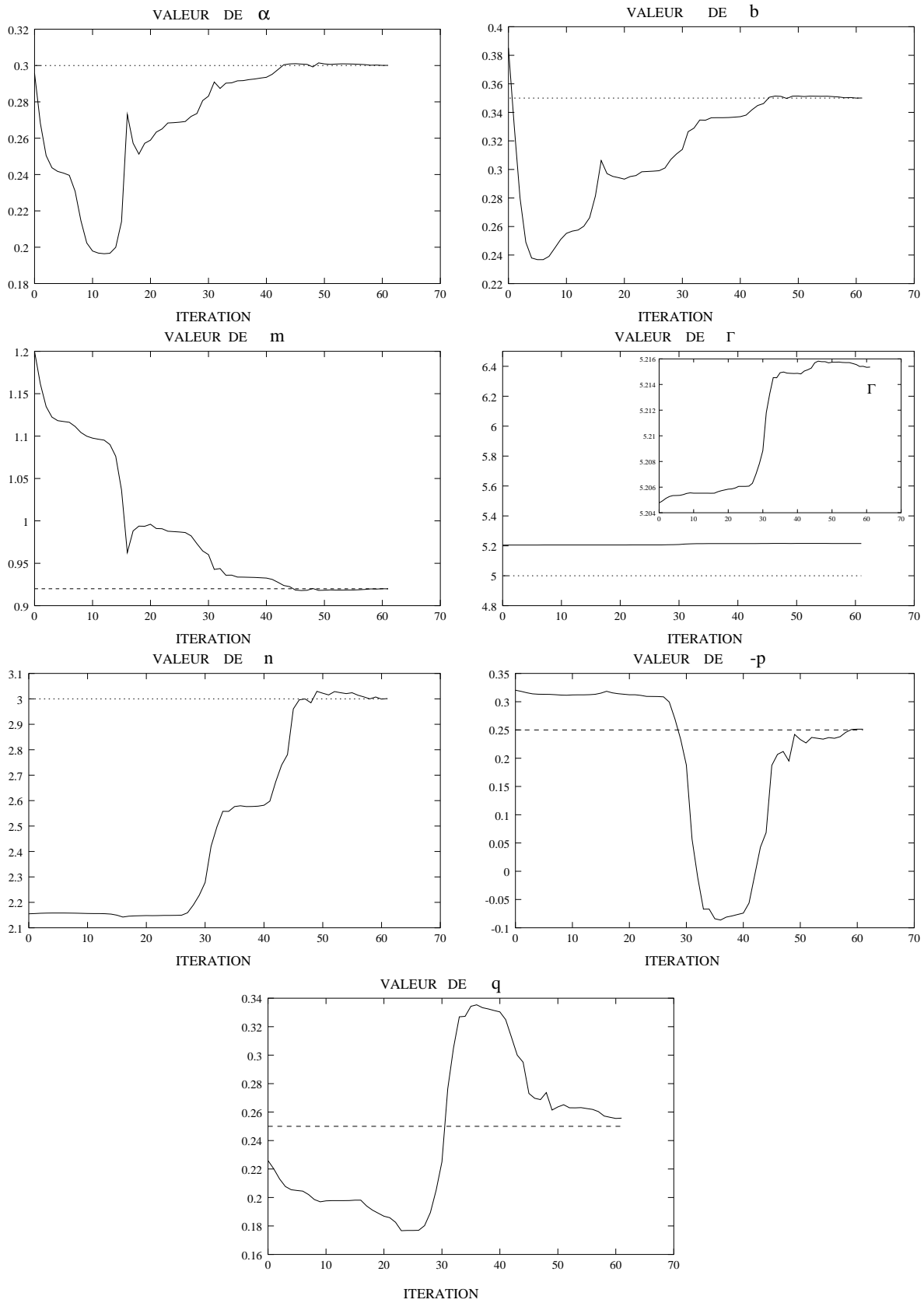


FIG. IV.16 – Résultats de l'identification en 7 paramètres à partir des observations au fond du canal, (—) les valeurs de contrôle en fonction des itérations et (\cdots) les valeurs exactes.

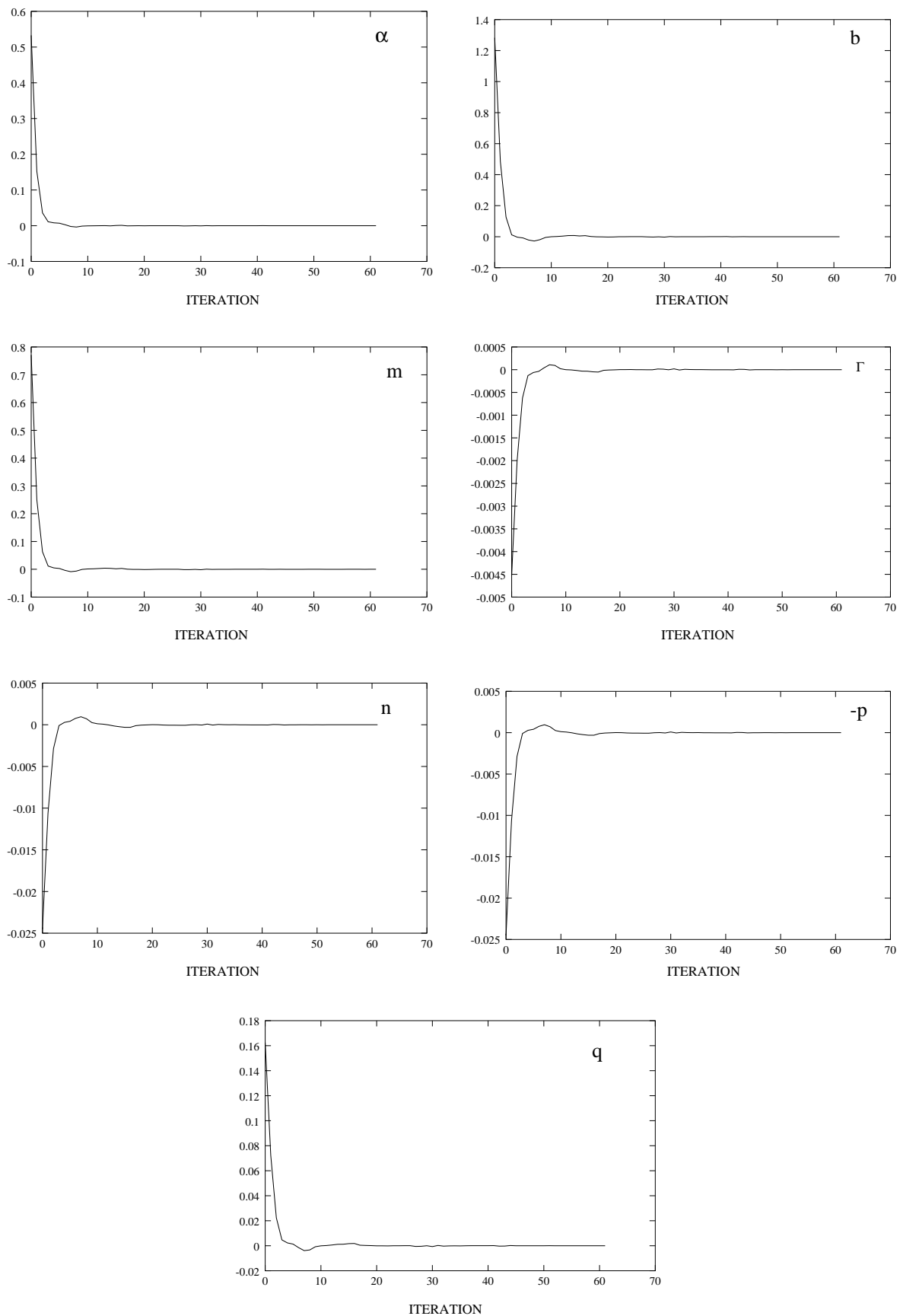


FIG. IV.17 – Le gradient de la fonction coût par rapport à chaque composante du vecteur des paramètres avec des observations au fond du canal central.

4.3.7 Méthode de la pénalisation

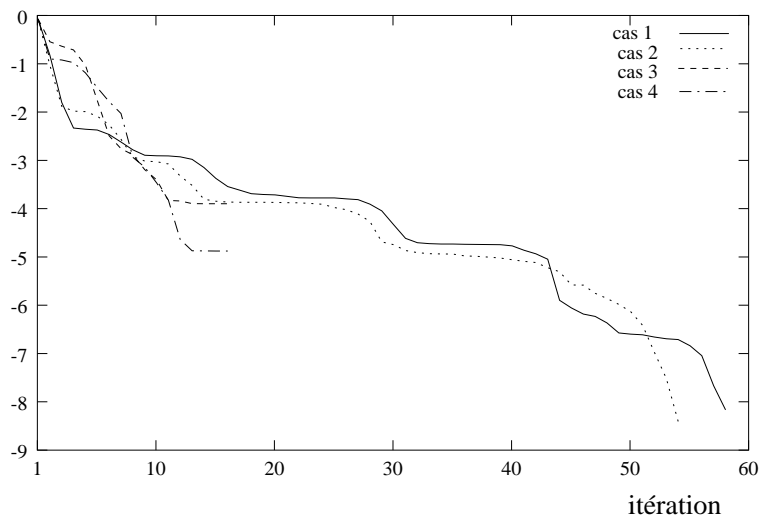
En pratique, l'effet est remarqué que la solution du modèle reconstruite par la méthode variationnelle d'assimilation de données peut provoquer des oscillations spatiales fortes si l'on applique les observations clairsemées ou des observations comprenant du bruit. Une façon à éliminer des oscillations spatiales est d'utiliser de fausses données afin de mettre en vigueur l'aspect lisse de la solution, on peut créer des observations par l'interpolation avec les données aux voisinages des endroits où il manque d'observations, ou bien on peut aussi résoudre ce problème par la méthode de pénalisation en ajoutant un terme de pénalité à l'ancienne fonction coût pour former une fonctionnelle pénalisée. L'idée de la méthode de pénalisation est réduite à beaucoup d'applications (Thacker[60], Zou et Navon [48], Lardner [36]).

Maintenant on diminue le nombre de points de grille où il y a des observations. Puisque l'on a vu que c'est suffisant d'identifier les 7 paramètres avec les données dans le canal de navigation. On diminue le nombre de données observées en prenant dans le centre du canal des mesures de la hauteur du lit de la rivière tous les 2 points, c.-à-d. le nombre de données déçoit de 50 à 25, le résultat de l'identification des paramètres est presque le même que dans le cas où il y a des mesures sur tous les points de grille au canal de navigation. Après l'optimisation selon un critère d'arrêt, les valeurs finales de tous les paramètres sauf Γ sont autant satisfaisant qu'avant la réduction d'observations.

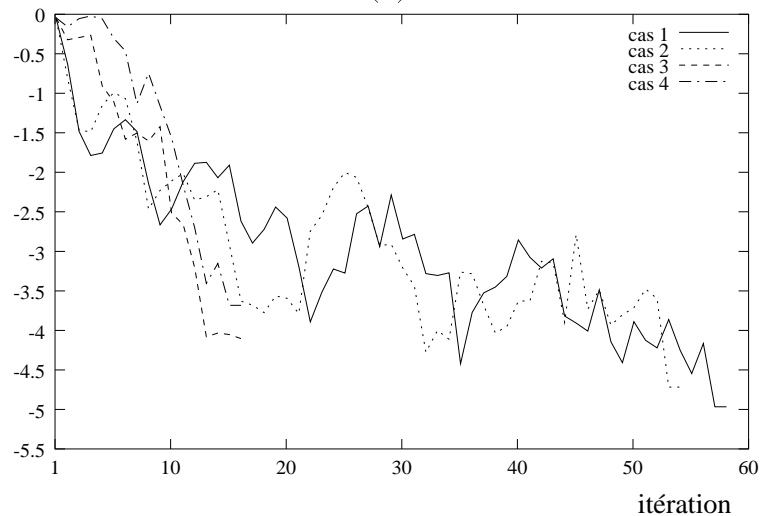
Si l'on diminue encore plus le nombre de données, par exemple on échantillonne l'observation sur tous les 5 points de grille, le processus de la minimisation n'arrive pas à continuer à cause d'un arrêt d'une erreur d'arrondi. Le tableau IV.8 représente le comparaison des résultats de l'identification lorsque le nombre d'observations est diminué. La figure IV.18 représente les variations de la fonction coût et du gradient normalisé par rapport aux paramètres au cours de la minimisation.

Paramètre	Valeur exacte	Valeur de départ	Valeur final avec les observations tous les? points			
			1	2	5	10
b	0.35	0.385134	0.350353	0.350192	0.296336	0.360524
m	0.92	1.19991	0.919622	0.919781	0.959065	0.966842
α	0.3	0.295765	0.300239	0.300132	0.305201	0.294712
Γ	5.0	5.20477	5.21542	5.20598	5.20474	5.20358
n	3.0	2.15512	3.00722	3.00336	2.15096	2.14797
p	-0.25	-0.320287	-0.250798	-0.251268	-0.324778	-0.344579
q	0.25	0.225900	0.256352	0.255625	0.227838	0.309656

TAB. IV.8 – Comparaison des résultats de l'identification avec les différents nombre d'observations.



(a)



(b)

FIG. IV.18 – La variation de $\log(J_k/J_0)$ (a) et de $\log(\|\nabla J_k\|/\|\nabla J_0\|)$ (b) en fonction des itérations k quand les observations sont valables sur tous les un (cas 1), deux (cas 2), cinq (cas 3) et dix (cas 4) points de grille au canal central.

Dans ces expériences, si les données d'observation sont échantillonnées tous les cinq points de maillage au canal de navigation, le tableau Tab.IV.9 montre que nous ne pouvons pas réussir à identifier le vecteur de contrôle, l'écart entre la valeur exacte et celle d'identification reste encore assez grande. Pour cela, nous essayons de voir si l'on peut améliorer les résultats d'identification en utilisant la méthode de pénalisation dans le cas où il y a des observations non complètes ou bruitées.

Cette méthode est réalisée par ajoutant un terme de pénalisation à la fonction coût. Supposons qu'il y a d'estimations a priori valables pour les paramètres à identifier, obtenues à partir des mesures acquises ou des arguments physiques, dénotés par $K_c = (\Gamma_c, n_c, p_c, q_c, m_c, \alpha_c, b_c)$, alors la fonction coût pénalisée à minimiser est donnée par :

$$\begin{aligned} J(K) &= J_0(K) + \frac{1}{2}\beta \|W_K(K - K_c)\|^2 \\ &= \frac{1}{2} \int_0^T \|W(C \cdot X - X_{obs})\|^2 dt + \frac{1}{2}\beta \|W_K(K - K_c)\|^2, \end{aligned} \quad (4.61)$$

où β est le coefficient de pénalité, $W_K = \text{diag}(W_\Gamma, W_n, W_p, W_q, W_m, W_\alpha, W_b)$ désignent les fonctions de poids attachées a priori aux observations.

En prenant toujours le même système adjoint (4.53), le gradient de la fonction coût pénalisée par rapport à chaque paramètre peut être obtenu en ajoutant à (4.55) un terme

$$\beta(W_K)_{jj} K_j, \quad j = 1, \dots, 7,$$

au gradient de la fonction coût non pénalisée.

Si l'on échantillonne les observations au centre du canal tous les 5 points, on ne peut pas obtenu une spécification satisfaisante à cause du fait de manque d'observation. En partant de la même valeur du vecteur de contrôle dans la procédure de la minimisation, on établit une comparaison de l'identification pour les cas sans ou avec application de la méthode de pénalité.

4.3.7.1 Expérience après la pénalisation

L'expérience est faite si l'on suppose que il n'y a l'estimation a priori Γ_c que sur le paramètre Γ , c'est-à-dire que

$$W_\Gamma = 1, \quad W_n = W_p = W_m = W_\alpha = W_b = 0,$$

on ajoute alors un terme de pénalisation $\frac{1}{2}\beta(\Gamma - \Gamma_c)^2$ à la fonction coût $J_0(K)$, où β est bien choisie (ici, $\beta = 10$) pour que la procédure de la minimisation soit effectuée. Dans le tableau Tab. IV.9, on montre que la fonction coût pénalisée (4.61) a trouvé un minimum local en certaines itérations et la qualité de l'identification de paramètres est beaucoup mieux améliorée que dans le cas où il n'y a pas d'application de la méthode de pénalisation.

Paramètre	Valeur exacte	Valeur de départ	Valeur finale	
			sans pénalisation	avec pénalisation $\frac{1}{2}\beta(\Gamma - \Gamma_c)^2$
b	0.35	0.385134	0.296336	0.349975
m	0.92	1.19991	0.959065	0.920010
α	0.3	0.295765	0.305201	0.300036
Γ	5.0	5.20477	5.20474	5.0000
n	3.0	2.15512	2.15096	2.99946
p	-0.25	-0.320287	-0.324778	-0.250050
q	0.25	0.225900	0.227838	0.249946

TAB. IV.9 – Comparaison des résultats de l'identification pour les méthodes sans pénalisation et avec pénalisation en prenant le terme de pénalité $\frac{1}{2}\beta(\Gamma - \Gamma_c)$, où $\beta=10$, $\Gamma_c = 5.0$. Les observations sont valables sur la cote du fond de tous les 5 points dans le canal central.

4.3.7.2 Expériences avec des différents termes de pénalité

Pour comparer l'impact des termes de pénalité sur les résultats de l'identification de paramètre, les expériences sont effectuées avec différents termes de pénalité. Les expériences sont effectuées pour les cas lorsque il n'y a d'estimation a priori que sur Γ , sur q , sur p , ou bien sur tous les paramètres, c'est-à-dire que $(W_K)_{jj} = 1, j=1, \dots, 7$. On illustre les résultats dans le tableau Tab. IV.10, à partir duquel on remarque que l'on obtient le meilleur résultat de spécification lorsque les estimations a priori sont valables sur tous les paramètres à identifier, ou bien seulement sur Γ . Pour les autres cas, en revanche, nous ne pouvons obtenir que de bonnes spécifications du vecteur de paramètre que pour le paramètre sur lequel il y a l'estimation a priori.

Paramètre	Valeur exacte	Valeur de départ	Valeur finale			
			$W_\Gamma = 1$	$W_q = 1$	$W_p = 1$	$(W_K)_j = 1$
b	0.35	0.385134	0.349975	0.298615	0.299921	0.350000
m	0.92	1.19991	0.920010	0.955184	0.966913	0.920000
α	0.3	0.295765	0.300036	0.305146	0.295469	0.30000
Γ	5.0	5.20477	5.0000	5.20574	5.20504	5.0000
n	3.0	2.15512	2.99946	2.15405	2.15185	3.0000
p	-0.25	-0.320287	-0.250050	-0.245738	-0.250001	-0.250000
q	0.25	0.225900	0.249946	0.250000	0.216255	0.250000

TAB. IV.10 – Comparaison de l'identification de paramètre en utilisant différents termes de pénalité, où le coefficient de pénalité $\beta=10$. Les observations sont valables sur la cote du fond de tous les 5 points dans le canal central.

4.3.7.3 Expériences du coefficient de pénalité

Le coefficient de pénalisation β dans l'expression de fonctionnelle pénalisées doit obéir à une loi de compromis, c'est-à-dire, β ne doit pas être trop grand, sinon, on ne tient pas compte suffisamment des observations, car le processus de la minimisation va porter tout son effort pour diminuer seulement le terme de régularisation, si par contre β est trop petit, le problème de l'identification est trop sensible qui va déduire une mauvaise qualité de l'identification.

Les expériences sont réalisées en prenant de différentes valeurs du coefficient de pénalité β . Le tableau Tab.IV.11 nous a montré les résultats de l'identification pour les différents coefficients de pénalité dans les cas où l'estimation a priori est seulement valable sur Γ : $\Gamma_c = 5.0$. Dans tous les cas, la valeur finale de Γ est égale à cette estimation, mais l'identification des autres paramètres peut s'évaluer selon la valeur de β : quand $\beta = 1$, les écarts entre les valeurs identifiées et exactes sont grandes ; si β augmente, par exemple $\beta = 10$, les valeurs finales de tous les paramètres approchent bien leur valeurs exactes ; en revanche, si β est devenu trop grand, on trouve que les résultats sont devenues moins biens.

Si nous faisons un autre test pareil mais nous supposons que nous avons l'estimation a priori sur le paramètre q , donc un terme supplémentaire $\frac{1}{2}\beta(q - q_c)^2$ est ajouté pour former la fonction coût pénalisée. Du tableau Tab. IV.12, nous trouvons que n'importe comment qu'il varie le coefficient β , les valeurs identifiées des paramètres sauf q ne peuvent pas approcher leur valeurs exactes.

Paramètre	Valeur exacte	Valeur de départ	Valeur finale		
			$\beta=1$	$\beta=10$	$\beta=100$
b	0.35	0.385134	0.296610	0.349975	0.349357
m	0.92	1.19991	0.958445	0.920010	0.920654
α	0.3	0.295765	0.305132	0.300036	0.299760
Γ	5.0	5.20477	5.0000	5.0000	5.0000
n	3.0	2.15512	2.15339	2.99946	2.99473
p	-0.25	-0.320287	-0.320286	-0.250050	-0.249275
q	0.25	0.225900	0.223226	0.249946	0.250522

TAB. IV.11 – Comparaisons des résultats d'identification en utilisant différents coefficients de pénalité β pour le terme $\frac{1}{2}\beta(\Gamma - \Gamma_c)^2$. La cote du fond de la rivière est observée sur tous les 5 points du canal central.

Paramètre	Valeur exacte	Valeur de départ	Valeur finale		
			$\beta=1$	$\beta=10$	$\beta=100$
b	0.35	0.385134	0.298544	0.298615	0.295418
m	0.92	1.19991	0.955509	0.955184	0.950209
α	0.3	0.295765	0.305290	0.305146	0.301074
Γ	5.0	5.20477	5.20595	5.20574	5.20516
n	3.0	2.15512	2.15244	2.15405	2.15407
p	-0.25	-0.320287	-0.245896	-0.245738	-0.317570
q	0.25	0.225900	0.250000	0.250000	0.250000

TAB. IV.12 – Comparaison de l'identification en utilisant différents coefficients de pénalité β pour le terme $\frac{1}{2}\beta(q - q_c)^2$. La cote du fond de la rivière est observée sur tous les 5 points du canal central.

Commentaire : Lorsque les observations contiennent toujours des erreurs à cause du bruit ou de la précision de l'appareil de mesure, le résultat de l'assimilation de données sera très sensible à la grandeur d'erreur dans les observations. Pour éliminer la sensibilité du vecteur de contrôle optimal obtenu par la minimisation d'une fonction coût, des méthodes de régularisation par ajout d'un terme de régularisation à la fonction sont toujours utilisées.

4.4 Conclusion

Ce chapitre montre que les méthodes d'assimilation de données peuvent être appliquées aux problèmes d'hydrologie, pour des problèmes tels que la reconstruction de l'évolution du lit d'un fleuve ou pour des problèmes d'identification.

L'étape suivante sera de passer à des données réelles ce qui augmentera d'un ordre de grandeur le coût de calcul. Un des problèmes qui devra être résolu est de savoir quelles données sont pertinentes et quel est le domaine de validité d'un modèle.

Chapitre V

Estimation des erreurs du modèle

5.1 Introduction

Dans les méthodes variationnelles d'assimilation de données, le modèle pris comme contrainte dynamique est toujours supposé parfait, et on cherche à minimiser une fonction coût qui introduit des observations pour que les données observées s'accordent au mieux avec le système dynamique. En revanche, en réalité, même si l'on ne cesse pas d'améliorer le modèle afin qu'il décrit le comportement d'un système dynamique, il reste toujours des difficultés en raison de connaissances limitées et de la limite du coût de calcul. Une question se pose donc : comment estimer ces erreurs afin d'améliorer les résultats de l'assimilation de données?

Il y a plusieurs facteurs qui peuvent entraîner des erreurs du modèle. Tout d'abord en raison du coût de calcul d'un modèle non linéaire, qui est représenté par des équations aux dérivées partielles, on utilise souvent un modèle simplifié en ignorant quelques termes dans les équations, et d'autre part, pour résoudre ces équations numériquement, certaines méthodes d'approximation peuvent être adoptées comme les différences finies ou les éléments finis. Ils interviennent alors une erreur d'approximation quand on utilise un schéma pour discrétiser le modèle en espace et en temps et pour approcher un opérateur différentiel par un opérateur aux différences. On remarque que si le schéma est stable, la croissance des erreurs d'approximation ne peut pas atteindre à une valeur infinie sur une longue période de temps, c'est-à-dire qu'elle a une énergie limitée. Une autre source d'erreur est qu'il y a toujours des écarts entre les valeurs réelles des paramètres du modèle et ses valeurs observées à cause des limites de précision des instruments de mesure ou de la présence du bruit.

Pour toutes de ces raisons, on souhaite estimer l'erreur d'un modèle par la résolution d'un système de contrôle en prenant l'erreur comme un vecteur de contrôle. Dans ce cas, on minimise une fonction coût qui présente l'écart de prévision du modèle avec l'erreur et

les observations.

Si les erreurs sont représentées dans la base euclidienne, la dimension du problème de contrôle sera énorme ainsi que le coût de calcul de l'optimisation, même pour une courte période du temps. Une possibilité de diminution de la taille du vecteur de contrôle est examinée en choisissant une autre base d'espace, qui permet d'estimer les modes de croissance les plus rapides de l'erreur du modèle. Ces modes sont décrites par une série de vecteurs propres trouvés par l'algorithme de Lanczos.

Le but de ce chapitre est de tester des méthodes de réduction de la dimension de l'espace de contrôle. Pour une première approche, il nous a semblé plus adapté d'utiliser un modèle simple d'écoulement de fluide géographique utilisé couramment en météorologie et océanographie. Ce modèle est dérivé des équations d'eau peu profonde (Shallow-water) et décrit dans Grammelvedt ([30]).

La transposition à un problème complexe d'hydrologie demandera d'importants ressources de calcul. Ces études préliminaires se fait donc avec un modèle bien référencé.

5.2 Détermination des erreurs dans la base canonique

5.2.1 Système de contrôle I

On considère un modèle évolutif de forme générale auquel on ajoute une erreur $V(x, t)$ dépendant de l'espace et du temps

$$\begin{cases} \frac{\partial X}{\partial t} = F \cdot X(t) + B \cdot V(x, t), & \text{sur } \Omega \times [0, T], \\ X(0) = U, & \text{sur } \Omega, \end{cases} \quad (5.1)$$

où B est un opérateur linéaire qui permet de transférer V de son espace \mathcal{V} vers celui de la variable d'état.

Notre problème revient à chercher une erreur V qui minimise la fonctionnelle suivante :

$$J(U, V) = \frac{1}{2} \int_0^T (\|CX - X_{obs}\|^2 + \beta \langle \mathcal{N}V, V \rangle) dt, \quad (5.2)$$

où $\mathcal{N} \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ est une matrice définie positive, β est le coefficient de régularisation.

Maintenant on va en déduire la condition de minimisation de cette fonction. S'il y a une perturbation dans la direction h_V autour d'une certaine erreur V et une autre dans la

direction h_U autour de la condition initiale, la dérivée de Gâteaux \hat{J} de J dans la direction (h_U, h_V) est donnée par

$$\hat{J}(U, V, h_U, h_V) = \int_0^T (\langle CX - X_{obs}, C\hat{X} \rangle + \beta \langle \mathcal{N}V, h_V \rangle) dt, \quad (5.3)$$

où \hat{X} est la dérivée directionnelle de X dans la direction (h_U, h_V) , qui est solution du système ci-dessous :

$$\begin{cases} \frac{\partial \hat{X}}{\partial t} = \left[\frac{\partial F}{\partial X} \right] \hat{X}(t) + Bh_V, & \text{sur } \Omega \times]0, T], \\ \hat{X}(0) = h_U, & \text{sur } \Omega. \end{cases} \quad (5.4)$$

On introduit une variable adjointe P qui satisfait

$$\begin{cases} \frac{\partial P}{\partial t} + \left[\frac{\partial F}{\partial X} \right]^t P = C^t(CX - X_{obs}), & \text{sur } \Omega \times [0, T], \\ P(T) = 0, & \text{sur } \Omega. \end{cases} \quad (5.5)$$

Si l'on multiplie la première équation de ce système par \hat{X} et que l'on intègre de 0 à T , on obtient

$$\int_0^T \langle \frac{\partial P}{\partial t}, \hat{X} \rangle dt + \int_0^T \langle \left[\frac{\partial F}{\partial X} \right]^t P, \hat{X} \rangle dt = \int_0^T \langle C^t(CX - X_{obs}), \hat{X} \rangle dt.$$

Par l'intégration par partie, il vient

$$- \langle \hat{X}(0), P(0) \rangle + \int_0^T \langle P, -\frac{\partial \hat{X}}{\partial t} \rangle dt + \int_0^T \langle P, \left[\frac{\partial F}{\partial X} \right] \hat{X} \rangle dt = \int_0^T \langle CX - X_{obs}, C\hat{X} \rangle dt,$$

due au système (5.4), on a

$$- \langle h_U, P(0) \rangle - \int_0^T \langle P, Bh_V \rangle dt = \int_0^T \langle CX - X_{obs}, C\hat{X} \rangle dt,$$

comme h_U et h_V sont arbitraires, on peut déduire de (5.3) que le gradient de la fonctionnelle J par rapport au vecteur (U, V) est égal à :

$$\begin{cases} \nabla_U J = -P(0), \\ \nabla_V J = -B^t P + \beta \mathcal{N}V, \end{cases} \quad (5.6)$$

la caractéristique de la variable optimale de contrôle est donnée par les équations $\nabla_U J = 0$ et $\nabla_V J = 0$; c'est-à-dire que

$$V = \beta^{-1} \mathcal{N}^{-1} B^t P. \quad (5.7)$$

5.2.2 Un modèle d'application

Pour examiner les erreurs du modèle dans un cadre plus simple, on s'intéresse à un cas d'écoulement de flux à partir des équations de Shallow-Water sur une surface limitée qui s'écrivent :

$$\begin{cases} \frac{\partial u}{\partial t} = -u \frac{\partial u}{\partial x} - v \frac{\partial u}{\partial y} + fv - \frac{\partial \phi}{\partial x}, \\ \frac{\partial v}{\partial t} = -u \frac{\partial v}{\partial x} - v \frac{\partial v}{\partial y} - fu - \frac{\partial \phi}{\partial y}, \\ \frac{\partial \phi}{\partial t} = -\frac{\partial(u\phi)}{\partial x} - \frac{\partial(v\phi)}{\partial y}, \end{cases} \quad (5.8)$$

dans lesquelles u et v sont les deux composantes de la vitesse horizontale, $\phi = gh$ est le géopotential, f et g sont respectivement la force de Coriolis et l'accélération de la pesanteur. Puisque il n'y pas de force extérieure, ce système conserve la masse. Ce système est souvent pris comme un exemple pour étudier des problèmes d'assimilation de données en météorologie, ce qui nous permet de comparer facilement les résultats en faisant des expériences jumelles.

Ce système est discrétisé en différences finies sur un domaine rectangulaire de longueur $L = 6000\text{km}$, de largeur $D = 4400\text{km}$, dont les pas de discrétisation sont $\delta x = 300\text{km}$, $\delta y = 220\text{ km}$ en espace, et $\delta t = 600\text{sec}$ en temps. De cette manière on a construit un maillage de 21×21 points. Une condition initiale est choisie pour ce système selon Grammeltvedt ([30]) où le champ d'altitude initial est déterminé de la façon suivante :

$$h = H_0 + H_1 \tanh\left(\frac{y - y_0}{2D}\right) + H_2 \operatorname{sech}\left(\frac{y - y_0}{D}\right) \sin\left(\frac{2\pi x}{L}\right),$$

u et v étant déduits des relations :

$$u = -\frac{g}{f} \frac{\partial h}{\partial y} \quad \text{et} \quad v = \frac{g}{f} \frac{\partial h}{\partial x}.$$

Les valeurs suivantes étant données pour les expériences numériques :

$$\begin{aligned} H_0 &= 2000\text{m}, & H_1 &= -220\text{m}, & H_2 &= 133\text{m}, \\ y_0 &= 2200\text{km}, & f &= 10^{-4}\text{sec}^{-1}, & g &= 10\text{m} \cdot \text{sec}^{-2}. \end{aligned} \quad (5.9)$$

Les conditions périodiques sont choisies sur les bords Ouest-Est, et on suppose que la longueur d'onde du flux est égale à la longueur du domaine. Sur les bords Nord-Sud, les conditions de parois solides sont utilisées, c'est-à-dire que la composante normale de la vitesse est nulle : $v = 0$, et que la dérivée normale du géopotential est nulle aussi : $\partial\phi/\partial t = 0$.

En utilisant un schéma leap-frog en temps et un schéma centré en espace, si $X_{i,j}^n = (u_{i,j}^n, v_{i,j}^n, \phi_{i,j}^n)$ est l'approximation de la solution du système $X = (u, v, \phi)$ au temps $n\delta t$ et au noeud $(i\delta x, j\delta y)$, on pose

$$\bar{u}_{i,j}^n = \frac{u_{i+1,j}^n + u_{i-1,j}^n}{2} \quad \text{et} \quad \bar{v}_{i,j}^n = \frac{v_{i,j+1}^n + v_{i,j-1}^n}{2},$$

on discrétise le système (5.8) par :

$$\begin{aligned} & \frac{u_{i,j}^{n+1} - u_{i,j}^{n-1}}{2\delta t} = \\ & -\bar{u}_{i,j}^n \frac{u_{i+1,j}^n - u_{i-1,j}^n}{2\delta x} - \bar{v}_{i,j}^n \frac{u_{i,j+1}^n - u_{i,j-1}^n}{2\delta y} + f v_{i,j}^n - \frac{\phi_{i+1,j}^n - \phi_{i-1,j}^n}{2\delta x}, \\ & \frac{v_{i,j}^{n+1} - v_{i,j}^{n-1}}{2\delta t} = \\ & -\bar{u}_{i,j}^n \frac{v_{i+1,j}^n - v_{i-1,j}^n}{2\delta x} - \bar{v}_{i,j}^n \frac{v_{i,j+1}^n - v_{i,j-1}^n}{2\delta y} - f u_{i,j}^n - \frac{\phi_{i,j+1}^n - \phi_{i,j-1}^n}{2\delta y}, \\ & \frac{\phi_{i,j}^{n+1} - \phi_{i,j}^{n-1}}{2\delta t} = \\ & -\frac{u_{i+1,j}^n \phi_{i+1,j}^n - u_{i-1,j}^n \phi_{i-1,j}^n}{2\delta x} - \frac{v_{i,j+1}^n \phi_{i,j+1}^n - v_{i,j-1}^n \phi_{i,j-1}^n}{2\delta y}. \end{aligned} \tag{5.10}$$

Si nous notons $X^n = (u^n, v^n, \phi^n)$, le système (5.10) peut s'écrire sous la forme :

$$X^{n+1} = f(X^n) + X^{n-1},$$

en posant $Y^n = (\hat{u}^n, \hat{v}^n, \hat{\phi}^n)$ la variable linéaire tangente de X^n , la linéarisation du système (5.10) devient :

$$Y^{n+1} = J(X^n)Y^n + Y^{n-1}, \tag{5.11}$$

où $J(X^n)$ le Jacobien de f au point X^n . Par la méthode de transposition décrite dans le chapitre précédent, le système adjoint est déduit de l'équation linéarisée (5.11), avec lequel on obtient des équations (5.6) le gradient de la fonctionnelle par rapport aux variables de contrôle.

5.2.3 Expériences numériques

Les expériences ont été effectuées en ajoutant sur le modèle de petites perturbations sur tous les points de grille et à tous les pas de temps. Supposons que ces perturbations

soient dans les mêmes espaces que les solutions du modèle. Elles sont représentées dans la base euclidienne, donc la dimension de l'espace des perturbations est autant que celui des variables d'état. Grâce à l'utilisation des expériences jumelles, les observations sont générées par l'intégration du système Shallow-Water à partir de l'état initial de Grammelvedt (voir Fig. V.1) sur une courte période de temps (48 pas de temps = 8 heures), $N_T = 48$ désigne le nombre des pas de temps.

Posons $V^n = (V_u^n, V_v^n, V_\phi^n)_{1 \leq n \leq N_T}$ variables d'erreur dans le même espace de phase que la solution X^n vérifiant

$$\begin{cases} X^{n+1} = f(X^n) + X^{n-1} + B \cdot V^{n+1}, & n \geq 2, \\ X^2 = f(X^1) + X^1 + B \cdot V^2, \\ X^1 = X^1 + V^1. \end{cases} \quad (5.12)$$

Ici, B est l'opérateur identité. La dimension de l'espace de contrôle est égale à $3 \times 19 \times 19 \times 48 = 51984$. En prenant en compte les degrés de liberté du problème, la dimension des observations doit être supérieure à celle de l'espace de contrôle, c'est pourquoi les observations sont valables à tous les pas de temps. Si on pose $X^n = (u^n, v^n, \phi^n)$ et $X_{obs}^n = (u_{obs}^n, v_{obs}^n, \phi_{obs}^n)$, on veut chercher les erreurs optimales $V = \{V^n\}_{1 \leq n \leq N_T}$ qui minimisent la fonctionnelle :

$$\begin{aligned} J(V) = & \frac{1}{2} \sum_{n=1}^{N_T} W \langle X^n - X_{obs}^n, X^n - X_{obs}^n \rangle \\ & + \frac{1}{2} \beta \sum_{n=1}^{N_T} \mathcal{N} \langle V^n, V^n \rangle, \end{aligned} \quad (5.13)$$

où $W = (W_u, W_v, W_\phi)$ sont les matrices diagonales représentant les poids, et dont les éléments diagonaux sont, dans nos expériences, tels que

$$(W_u)_{jj} = (W_v)_{jj} = 1, (W_\phi)_{jj} = 10^{-2}, \quad 1 \leq j \leq N,$$

avec N le nombre des pas de l'espace, β le coefficient de régularisation et les $\mathcal{N} = (N_u, N_v, N_\phi)$ étant les matrices symétriques définies positives désignant l'inverse de la covariance d'erreur correspondant à chaque variable.

En ce qui concerne l'initialisation du processus de la minimisation, on assigne à chaque composante de $(V^n)_{1 \leq n \leq N_T}$ une valeur aléatoire. En vue d'éviter l'instabilité du système due à la sensibilité du géopotential par rapport aux perturbations, on a choisi une erreur (V_u^n, V_v^n) correspondant au champ de vitesse comprise entre -1 m/s et 1 m/s, alors que l'erreur du géopotential V_ϕ^n se situe entre -100 mètres et 100 mètres (le taux de perturbation soit inférieur à 0.5 ‰ des données observées).

L'algorithme est dit converge quand le critère d'arrêt

$$\frac{\|\nabla J(X^n)\|}{\|\nabla J(X^0)\|} \leq \epsilon$$

est satisfait avec un seuil fixé $\epsilon = 10^{-5}$. Puisque il n'existe pas en général de critère précis sur les choix du coefficient de régularisation β et des matrices $\mathcal{N} = (N_u, N_v, N_\phi)$ qui doit obéir à une loi de compromis entre la stabilité et la qualité de minimisation des résidus par rapport aux observations, en prenant \mathcal{N} une matrice diagonale, des essais ont été faits avec différentes valeurs de β et \mathcal{N} . On présente les valeurs choisies dans les cas suivants :

1. $\beta = 0$;
2. $(N_u)_{ii} = (N_v)_{ii} = (N_\phi)_{ii} = 1$;
 - (a) $\beta = 1$, diverge ;
 - (b) $\beta = 0.1$;
 - (c) $\beta = 0.01$;
3. $(N_u)_{ii} = (N_v)_{ii} = 1, (N_\phi)_{ii} = 0.1$;
 - (a) $\beta = 1$;
 - (b) $\beta = 0.1$;
4. $(N_u)_{ii} = (N_v)_{ii} = 1, (N_\phi)_{ii} = 0.01$;
 - (a) $\beta = 1$;
 - (b) $\beta = 0.1$;

On a remarqué que :

- Si l'on prend β nul, ça veut dire que l'on ne prend pas en compte le terme de régularisation, on constate que le taux de convergence est très lent. Après 4000 itérations, la fonction coût et le gradient n'ont descendu que de l'ordre 10^{-3} sans satisfaire la condition de convergences ;
- Pour comparer le taux de convergence au cours de la minimisation, on présente, pour les cas 2, 3 et 4, sur les graphiques à gauche dans les figures Fig V.2, V.4 et V.6 les variations du logarithme décimal de la fonction coût normalisée tandis que les graphiques à droite décrivent les variations du gradient normalisé de la fonctionnelle par rapport à la variable de contrôle – les erreurs du modèle ;

- Parmi les cas 2, 3 et 4, on a constaté le taux de convergence le plus grand dans le cas où $\beta = 0.1$ et les coefficients de \mathcal{N} sont de même ordre, et plus β est petit, l'algorithme converge moins rapidement.
- On présente dans les figures Fig.V.3, Fig. V.5 et Fig. V.7, pour les derniers trois cas, les normes des erreurs avant le contrôle et après le contrôle. On observe que, bien que les erreurs d'initialisation soient données aléatoirement, leurs valeurs optimales sont décroissantes par rapport au temps.

Remarque: Nous avons observé grâce à ces expériences que si l'on veut traiter un problème d'optimisation en représentant le vecteur de l'erreur dans la base euclidienne, ce sera très coûteux et difficile, notamment pour certaines fonctionnelles. C'est pourquoi nous voulons contrôler l'erreur du modèle en la représentant dans une autre base. Ainsi le problème de contrôle optimal va se traduire par un autre problème de dimension plus petite.

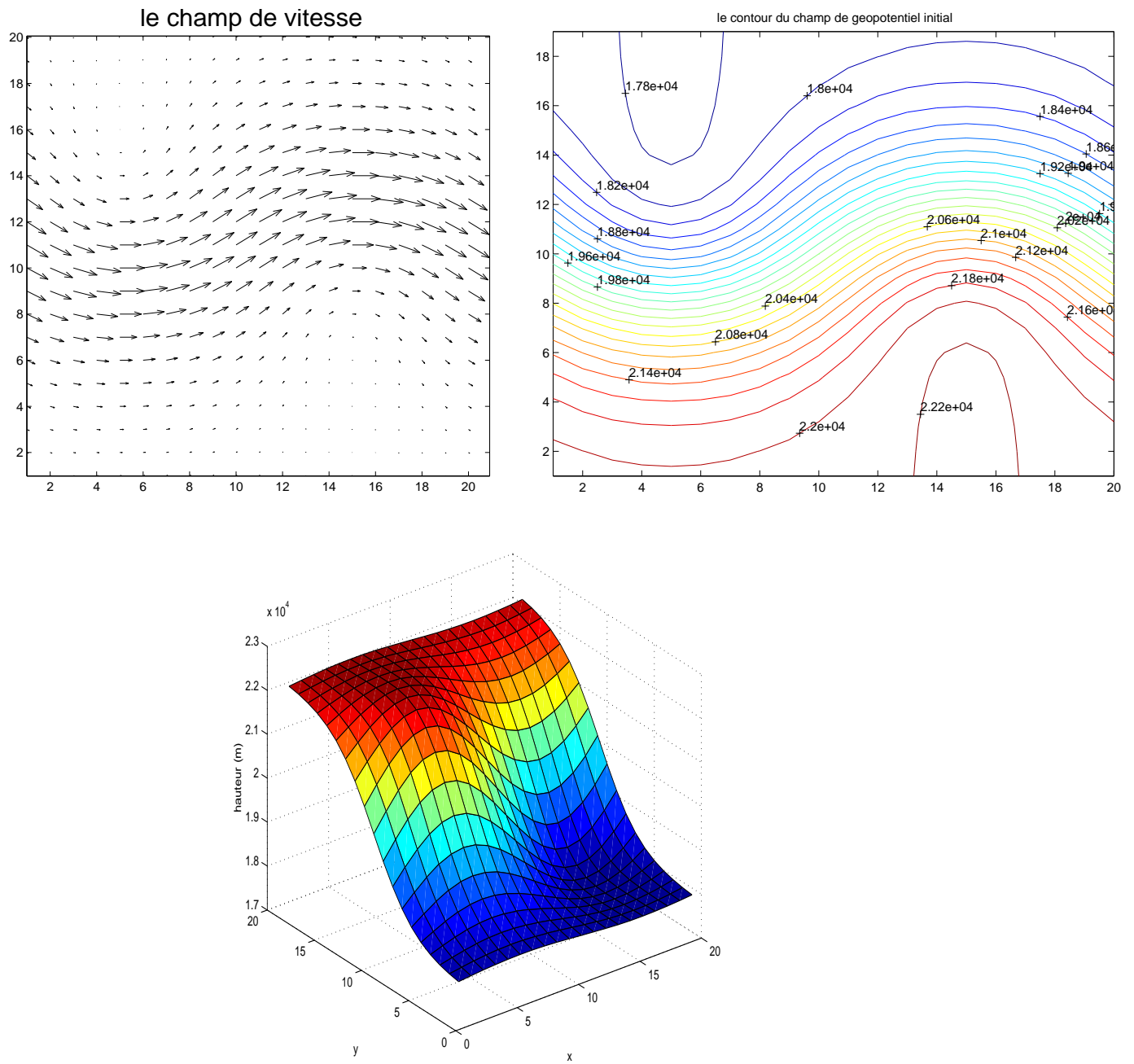
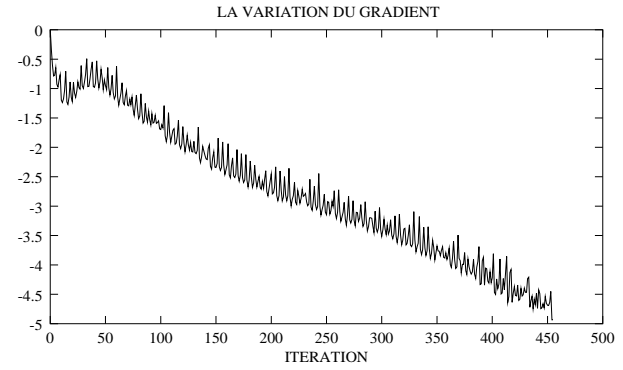
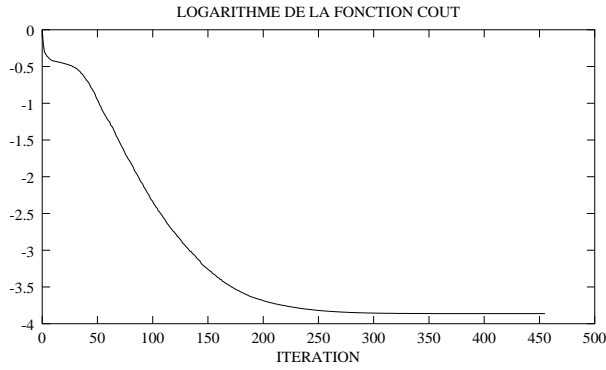


FIG. V.1 – La condition initiale de l'état dynamique : le champ de vitesse (à gauche) et le champ géopotiel (à droite et en bas).

$$(N_u)_{ii} = (N_v)_{ii} = (N_\phi)_{ii} = 1$$

$\beta = 0.1$



$\beta = 0.01$

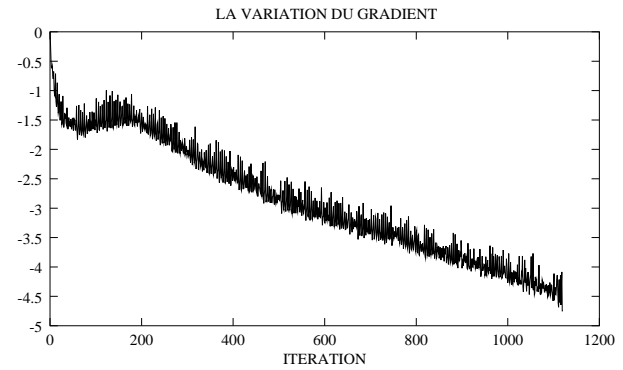
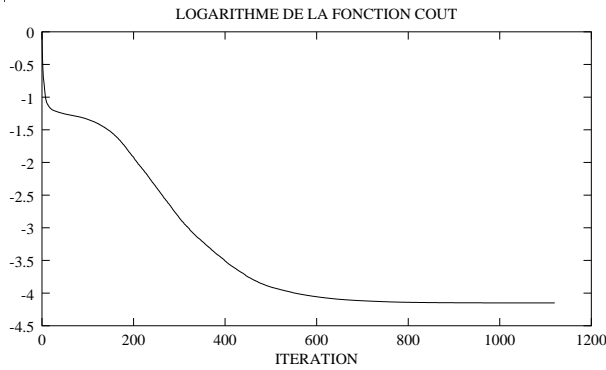


FIG. V.2 – Les figures à gauche décrivent la variation de $\log(J_k/J_0)$ en fonction des itérations k , alors qu'à droite décrivent la variation de $\log(\|\nabla J_k\|/\|\nabla J_0\|)$.

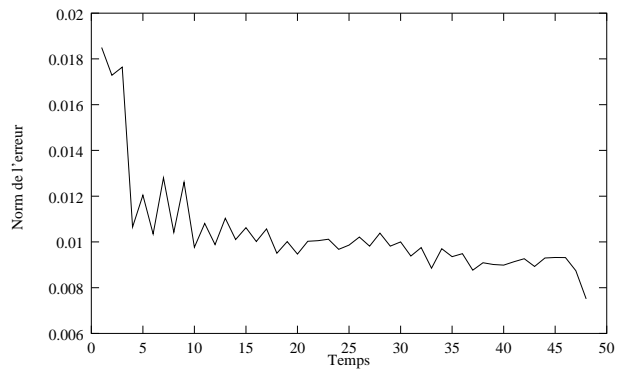
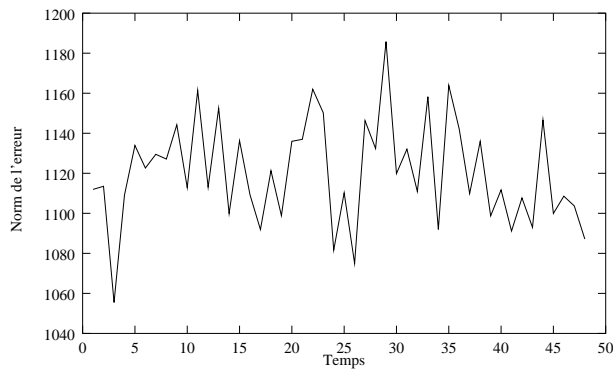


FIG. V.3 – La norme de l'erreur d'initialisation (à gauche) et optimale (à droite) en fonction du temps quand $\beta = 0.1$.

$$(N_u)_{ii} = (N_v)_{ii} = 1, (N_\phi)_{ii} = 0.1$$

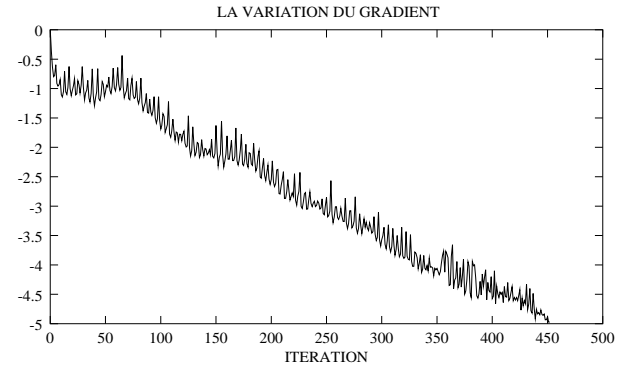
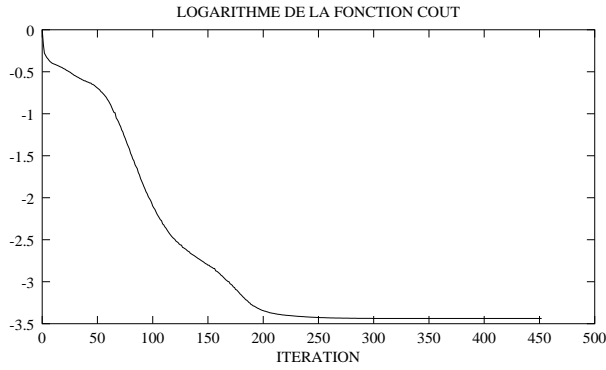
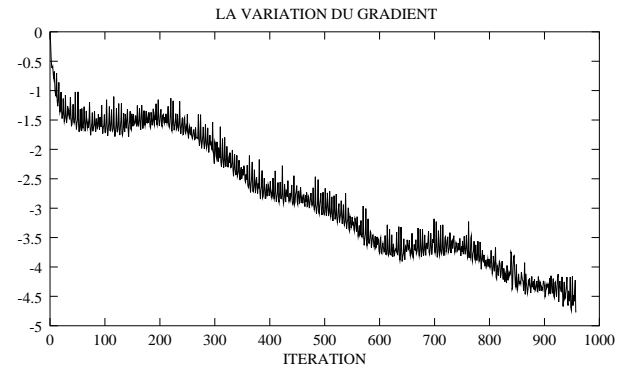
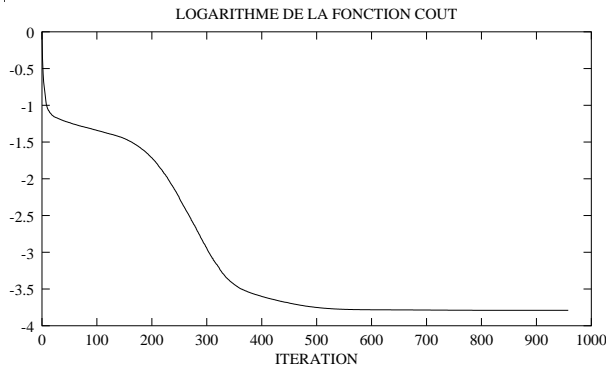
 $\beta = 1$

 $\beta = 0.1$


FIG. V.4 – Les figures à gauche décrivent la variation de $\log(J_k/J_0)$ en fonction des itérations k , alors qu'à droite décrivent la variation de $\log(\|\nabla J_k\|/\|\nabla J_0\|)$.

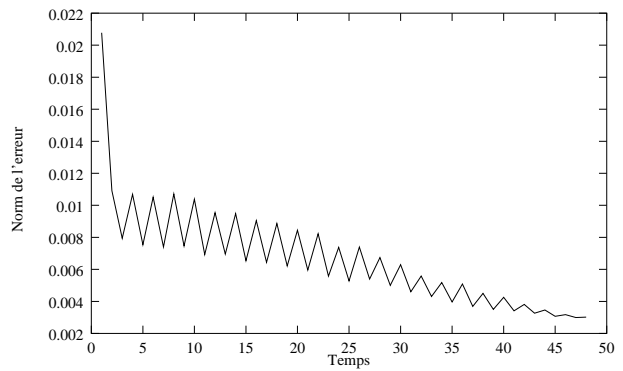
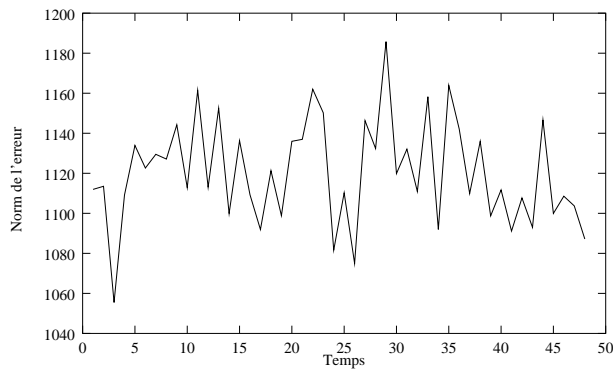


FIG. V.5 – La norme de l'erreur d'initialisation (à gauche) et optimale (à droite) en fonction du temps quand $\beta = 1$.

$$(N_u)_{ii} = (N_v)_{ii} = 1, (N_\phi)_{ii} = 0.01$$

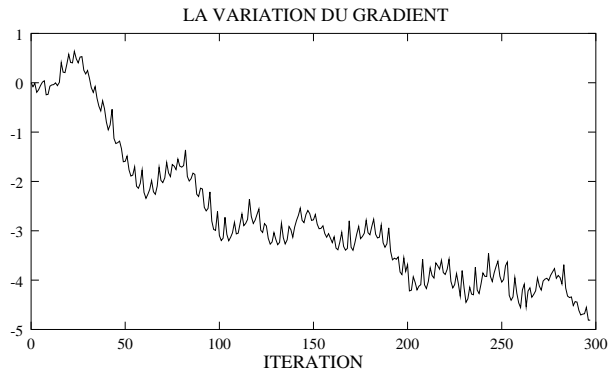
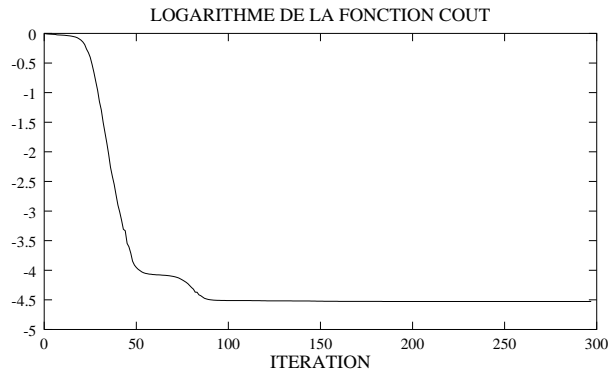
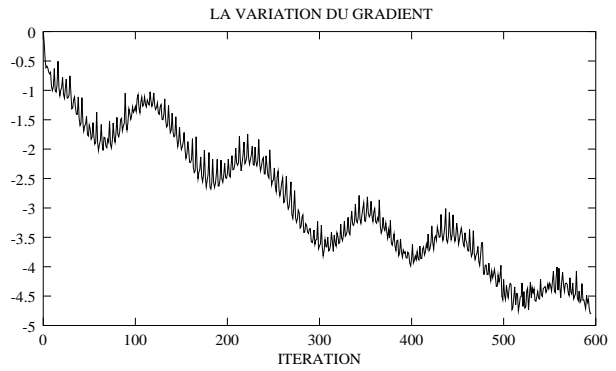
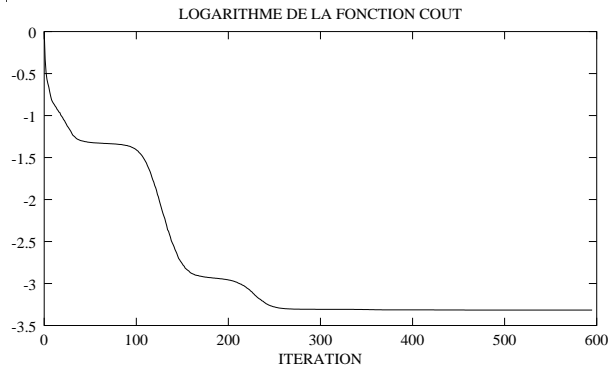
 $\beta = 10$

 $\beta = 1$


FIG. V.6 – Les figures à gauche décrivent la variation de $\log(J_k/J_0)$ en fonction des itérations k , alors qu'à droite décrivent la variation de $\log(\|\nabla J_k\|/\|\nabla J_0\|)$.

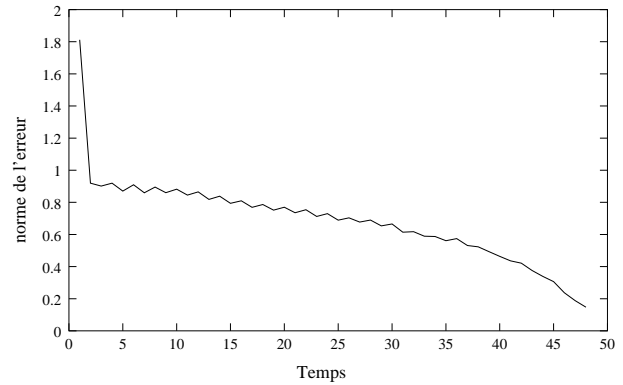
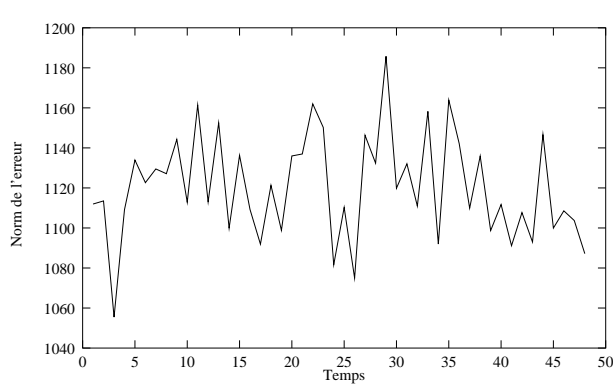


FIG. V.7 – La norme de l'erreur d'initialisation (à gauche) et optimale (à droite) en fonction du temps quand $\beta = 10$.

5.3 Les méthodes de vecteur optimal

Avant d'introduire une autre base d'espace, on va présenter deux méthodes utilisées dans le domaine de prévision de temps qui nous donnent une idée sur comment contrôler la croissance de l'amplitude de l'erreur de modèle.

L'estimation de l'amplitude de la prévision de l'erreur d'un système est toujours un problème intéressant dans le domaine de la prédiction, par exemple en météorologie ou en l'hydrologie. La croissance de l'erreur traduit, du point de vue de la théorie, l'instabilité du modèle. Comme nous le connaissons de la théorie des équations différentielles, une perturbation assez petite sur la solution d'un système dynamique peut se développer exponentiellement dans un intervalle du temps dont la longueur dépend des propriétés du système.

Pour un modèle non linéaire, l'augmentation des erreurs est difficile à déterminer, ce qui oblige à la décrire par une loi empirique extrêmement simple en ignorant la plupart de la dynamique contenu dans le modèle numérique. Pour estimer l'amplitude des erreurs au cours du temps, un problème original qui est primitivement technique est d'abord intéressé dans le domaine de la prévision. Supposons que l'erreur d'analyse de l'état initial du système soit connue grossièrement, l'idée la plus évidente est d'utiliser un générateur aléatoire qui possède la même propriété statistique que celle de l'erreur. Puisque le coût de calcul est très limité, c'est important qu'une petite somme des perturbations choisies puissent représenter l'évolution dominante de l'erreur du modèle.

Il y a deux techniques principales utilisées dans la prévision numérique d'ensemble pour générer des perturbations initiales, ce sont les méthodes BGM (bred grown modes ; Toth & Kalnay [63] [64]) et OV ("Vecteurs Optimaux", Buizza et al. [5], Buizza & Palmer [6]). Par exemple, dans ECMWF (Europe Center for Medium Range Weather Forecasts), une combinaison linéaire des vecteurs singuliers basés sur la norme de l'énergie est utilisée comme analyse perturbée pour la condition initiale. Ces vecteurs singuliers sont des directions suivant lesquelles les erreurs peuvent être amplifiées le plus rapidement possible. Et NCEP (National Centers for Environment Prediction), pour le même but, a adopté la méthode BGM, étant une extension non-linéaire des vecteurs de Lyapunov pour générer les perturbations initiales en groupe des vecteurs de breeding.

Pour diminuer la taille du problème de contrôle optimal, nous allons utiliser l'idée des vecteurs optimaux pour introduire une nouvelle base d'espace en remplaçant l'ancienne base euclidienne. Nous allons présenter ci-dessous comment chercher les vecteurs qui dominent l'amplitude des perturbations par l'algorithme de Lanczos et la méthode

de puissance itérée en utilisant l'intégration du système linéaire tangent et l'intégration rétrograde du système adjoint.

5.3.1 L'opérateur linéaire tangent

On considère toujours le même système

$$\begin{cases} \frac{\partial X}{\partial t} = F(X(t)), & \text{sur } \Omega \times [0, T], \\ X(0) = U, & \text{sur } \Omega, \end{cases} \quad (5.14)$$

soit $X(t) \in \mathbb{R}^n$ la variable d'état après la discrétisation en espace, F l'opérateur dépendant de X et t . Si H est la perturbation sur la condition initiale, la perturbation de la solution du modèle $\hat{X}(t)$ vérifie le système linéarisé :

$$\begin{cases} \frac{\partial \hat{X}}{\partial t} = \left[\frac{\partial F}{\partial X} \right] \hat{X}(t), & \text{sur } \Omega \times [0, T], \\ \hat{X}(0) = H, & \text{sur } \Omega. \end{cases} \quad (5.15)$$

Donc l'amplification de la perturbation au temps T peut être obtenue par l'intégration du système (5.15) sur un intervalle de temps $[0, T]$. Le système étant linéaire par rapport à $\hat{X}(t)$, il existe une matrice résolvante M_T qui vérifie

$$\hat{X}(T) = M_T H.$$

Selon la théorie de la décomposition en valeurs singulières (SVD [29]), il existe deux matrices carrées, orthogonales dans $\mathbb{R}^{n \times n}$:

$$U = [u_1, \dots, u_n], \quad V = [v_1, \dots, v_n]$$

satisfaisant

$$U^t U = I \text{ et } V^t V = I \in \mathbb{R}^{n \times n},$$

telles que

$$U^t M_T V = \Sigma = \text{diag}(\sigma_1, \dots, \sigma_n). \quad (5.16)$$

U et V sont les vecteurs singuliers à gauche et à droite respectivement de la matrice M_T , $\{\sigma_i\}$ sont les valeurs singulières, avec $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$, vérifiant les relations :

$$M_T v_i = \sigma_i u_i, \quad 1 \leq i \leq n. \quad (5.17)$$

En prenant le produit scalaire euclidien et la norme associée, $[u_1, \dots, u_n]$ et $[v_1, \dots, v_n]$ forment deux bases orthogonormées dans \mathbb{R}^n . Pour un vecteur arbitraire H , on peut alors le développer sous la base v_i et l'écrire de la manière suivante :

$$H = \sum_{i=1}^n \langle v_i, H \rangle v_i, \quad (5.18)$$

grâce à (5.17), la perturbation au temps final initialisée par H est égale à

$$\hat{X}(T) = M_T H = \sum_{i=1}^n \langle v_i, H \rangle M_T v_i = \sum_{i=1}^n \langle v_i, H \rangle \sigma_i u_i. \quad (5.19)$$

On en déduit que

$$\|\hat{X}(T)\|^2 = \langle M_T H, M_T H \rangle = \langle M_T^t M_T H, H \rangle = \sum_{i=1}^n |\langle v_i, H \rangle|^2 \sigma_i^2. \quad (5.20)$$

Posons

$$S_T = M_T^t M_T,$$

il est facile de montrer que les vecteurs singuliers droits de la matrice M_T sont les vecteurs propres de la matrice S_T , en fait, grâce à (5.16),

$$S_T v_i = M_T^t M_T v_i = M_T^t \sigma_i u_i = \sigma_i^2 v_i,$$

ce qui implique que $\lambda_i = \sigma_i^2$ sont les valeurs propres de S_T , avec $\lambda_1 \geq \dots \geq \lambda_n \geq 0$.

On remarque que les perturbations aux plus fortes croissances correspondent aux vecteurs propres de la matrice S_T associés à ses plus grandes valeurs propres. En fait, si l'on considère le facteur d'amplification des perturbations de la manière suivante :

$$\alpha^2 = \frac{\|\hat{X}(T)\|^2}{\|H\|^2} = \frac{\sum_{i=1}^n \sigma_i^2 |\langle v_i, H \rangle|^2}{\sum_{i=1}^n |\langle v_i, H \rangle|^2}$$

en fixant la norme de la perturbation initiale $\|H\|$, α atteindra le maximum, si et seulement si $H = v_1$, autrement dit si la perturbation maximale initiale est identique au vecteur propre associé à la plus grande valeur propre.

S_T étant une matrice symétrique, on considère à utiliser l'algorithme de Lanczos pour déterminer les vecteurs propres dominantes de la matrice S_T . Dans cet algorithme, on a

seulement besoin du produit d'une matrice avec un vecteur.

A partir de la construction du système adjoint décrit précédemment, la matrice résolvante du système adjoint est justement M_T^t , donc pour obtenir le produit $S_T H$, nous faisons seulement une intégration du système linéaire tangent dans l'intervalle de temps $[0, T]$, en prenant H comme condition initiale, et une intégration rétrograde du système adjoint de façon à ce que

$$P(T) = \hat{X}(T),$$

la valeur au temps initial de la variable adjointe est le produit que nous demandons.

En résumé, on cherche à résoudre le système suivant :

$$\left\{ \begin{array}{l} \frac{\partial \hat{X}}{\partial t} = \left[\frac{\partial F}{\partial X} \right] \hat{X}, \\ \hat{X}(0) = H, \\ \frac{\partial P}{\partial t} + \left[\frac{\partial F}{\partial X} \right]^t P = 0, \\ P(T) = \hat{X}(T), \\ S_T H = P(0). \end{array} \right. \quad (5.21)$$

5.3.2 Les méthodes de calcul des vecteurs propres dominants

Dans cette partie, on va présenter tout d'abord la méthode de la puissance itérée qui permet le calcul d'une approximation de la valeur propre de plus grand module d'une matrice symétrique ainsi que le vecteur propre associé à cette valeur propre. Dans la seconde partie, on introduit la méthode de Lanczos pour construire plusieurs vecteurs propres.

5.3.3 La méthode de la puissance itérée

Donnons d'abord l'algorithme de la méthode de la puissance itérée, en normalisant à chaque itération par une norme quelconque.

Choisir $x_0 \in \mathbb{C}^N$ tel que $\|x_0\| = 1$.

Pour $k=1, 2, \dots$ calculer :

$$\begin{cases} y_k = Ax_{k-1}, \\ \lambda^{(k)}(j) = \frac{y_k(j)}{x_{k-1}(j)}, & j = 1 \text{ à } N \text{ (facultativement)}, \\ x_k = \frac{y_k}{\|y_k\|}. \end{cases} \quad (5.22)$$

Si le facteur de convergence d'une suite (z_k) qui converge vers z est défini par :

$$\lim_{k \rightarrow +\infty} \frac{\|z_{k+1} - z\|}{\|z_k - z\|},$$

nous avons la convergence des suites (5.22) définies par l'algorithme grâce au théorème suivant (c.f. Lascaux et Théodor [37]) :

Théorème 5 Soit A une matrice carrée d'ordre N , diagonalisable dont la valeur propre de plus grand module λ_1 est unique, c'est-à-dire, que l'on suppose que les valeurs propres sont rangées par ordre de module décroissant :

$$|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_N|$$

si x_0 dans la suite définie dans (5.22) n'est pas orthogonale au sous espace propre à gauche associé à λ_1 , alors on a :

$$\lim_{k \rightarrow +\infty} \left(\frac{\overline{\lambda_1}}{|\lambda_1|} \right)^k x_k$$

est un vecteur propre de norme unité associé à λ_1 ,

$$\lim_{k \rightarrow +\infty} \|Ax_k\| = |\lambda_1|,$$

$$\lim_{k \rightarrow +\infty} \frac{y_{k+1}(j)}{x_k(j)} = \lambda_1, \quad \text{pour } 1 \leq j \leq n, \text{ si } x_k(j) \neq 0.$$

Le facteur de convergence de chacune de cette suite est $\frac{\lambda_{p+1}}{\lambda_1}$, où p est la multiplicité de la valeur propre λ_1 .

Si la matrice n'est pas diagonalisable mais si la valeur propre dominante est unique, il existe aussi un résultat analogue à celui du théorème 5. La seule chose qui diffère est le facteur de convergence qui dépend des propriétés de la matrice A .

Remarque : La méthode de la puissance inverse est une application de la méthode de puissance itérée. C'est la deuxième méthode appliquée soit à la matrice A^{-1} qui permet

d'obtenir la valeur propre de plus petit module et un vecteur propre associé, soit à la matrice $(A - \nu I)^{-1}$ qui permet d'obtenir la valeur propre la plus proche de ν et un vecteur propre associé.

5.3.4 La méthode de Lanczos

La méthode de Lanczos est une méthode de recherche du spectre d'une matrice symétrique qui permet de construire progressivement une matrice de Rayleigh dont l'ordre augmente à chaque itération. Le principe de cette méthode repose sur les propriétés de convergence de l'espace de Krylov engendré à partir d'un vecteur initial. L'idée est la même que celle de la méthode du gradient conjugué.

Soit A une matrice symétrique, la méthode de Lanczos va engendrer une suite de vecteurs orthonormés de \mathbb{R}^n de la manière récurrente suivante :

Soit $q_1 \in \mathbb{R}^n$ tel que $\|q_1\|_2 = 1$, on cherche à construire (q_j) ainsi que les $(\alpha_j, \beta_j) \in \mathbb{R} \times \mathbb{R}$ tels que

$$\begin{cases} Aq_1 = \alpha_1 q_1 + \beta_1 q_2, \\ Aq_j = \beta_{j-1} q_{j-1} + \alpha_j q_j + \beta_j q_{j+1}, \quad j > 1. \end{cases} \quad (5.23)$$

Pour calculer une telle base, l'algorithme suivant est utilisé :

Soit $q_0 = 0, q_1$ non nul tel que $\|q_1\| = 1$, et $\beta = 0$.

Pour $j = 1, 2, \dots$,

$$\left\{ \begin{array}{l} v_0 = q_1 \\ \beta_0 = 1 \\ q_0 = 0 \\ j = 0 \\ \text{FAIRE QUAND } (\beta_j \neq 0) \\ \quad q_{j+1} = v_j / \beta_j \\ \quad j = j + 1 \\ \quad \alpha_j = q_j^t A q_j \\ \quad v_j = (A - \alpha_j I) q_j - \beta_{j-1} q_{j-1} \\ \quad \beta_j = \|v_j\|_2 \\ \text{FIN_FAIRE} \end{array} \right. \quad (5.24)$$

Si l'on les représente en écriture matricielle, au j -ème pas de cet algorithme, on obtient la matrice tridiagonale T_j carrée d'ordre j :

$$T_j = \begin{pmatrix} \alpha_1 & \beta_1 & & 0 \\ \beta_1 & \alpha_2 & \beta_2 & \\ & \ddots & \ddots & \ddots \\ 0 & & \beta_{j-1} & \alpha_j \end{pmatrix}$$

soit

$$Q_j = [q_1, \dots, q_j],$$

on a alors

$$AQ_j^t = T_j Q_j^t.$$

Les valeurs propres et vecteurs propres de la matrice T_j peuvent être trouvés par l'algorithme QR symétrique ou bien par la méthode de bisection. Supposons $(\theta_i, y_i), 1 \leq i \leq j$, les éléments propres de la matrice T_j , c.-à-d.

$$T_j y_i = \theta_i y_i, y_i \in \mathbb{R}^j, \theta_i \in \mathbb{R},$$

qui sont appelés aussi les paires de Ritz de A et donnent une approximation du spectre de A .

En fait, si $\{q_i\} \subset \mathbb{R}^n$ une séquence des vecteurs non nuls vérifiant :

$$[q_1, \dots, q_j] = [q_1, Aq_1, \dots, A^{j-1}q_1],$$

et rappelons que l'espace de Krylov d'ordre j : $\mathcal{K}(A, q_1, j)$ est défini par l'espace vectoriel engendré par $q_1, Aq_1, \dots, A^{j-1}q_1$, c.-à.-d.

$$\mathcal{K}(A, q_1, j) = [q_1, Aq_1, \dots, A^{j-1}q_1],$$

il conduit au problème comment calculer une base orthonormale pour le sous-espace $\mathcal{K}(A, q_1, j)$. Pour cela, rappelons le lien entre la tridiagonalisation de A et la Q-R factorisation de $\mathcal{K}(A, q_1, j)$. On sait que si $Q^t A Q = T$ est diagonale avec $Q e_1 = q_1$, alors on a

$$[q_1, Aq_1, \dots, A^{j-1}q_1] = Q [e_1, T e_1, \dots, T^{j-1} e_1],$$

qui permet de générer q_j en tridiagonalisant A avec une matrice orthogonale.

Théorème 6 *Supposons qu'après j pas d'itération de l'algorithme de Lanczos, il existe $\theta_i, 1 \leq i \leq j$, tels que*

$$S_j^t T_j S_j = \text{diag}(\theta_1, \dots, \theta_j), \quad (\text{Décomposition de Schur})$$

si $Y_j \in \mathbb{R}^{n \times j}$ est définie par

$$Y_j = (y_1, \dots, y_j) = Q_j S_j,$$

où $Q_j = [q_1, \dots, q_j]$, donc pour $i = 1, 2, \dots, j$, nous avons

$$\|Ay_i - \theta_i y_i\|_2 = |\beta_j| |s_{ji}|$$

où $S_j = (s_{pq})$.

La démonstration est dans (Golub et Van Loan [29]). Ce théorème a fourni une estimation de l'erreur de la valeur propre de la matrice T_j :

$$\min_{v \in \lambda(A)} |\theta_i - v| \leq |\beta_j| |s_{ji}|, \quad i = 1, \dots, j.$$

où $\lambda(A)$ désigne le spectre de A .

5.3.5 Résultats numériques

Les expérimentations sont faites en utilisant le modèle de Shallow-Water (5.8) sur un domaine 21×21 , avec les pas d'espace $\delta x = 300$ km et $\delta y = 220$ km et le pas de temps $\delta t = 600$ sec. On a intégré ce modèle sur un intervalle de temps $[0, T]$, avec $T = 288\delta t = 8$ heures, pour obtenir les trajectoires du système. Ces trajectoires seront demandées dans l'algorithme de Lanczos qui exige une intégration du système linéaire tangent et une autre en sens inverse du système adjoint. Les vecteurs propres ainsi acquis par l'algorithme correspondent à la matrice $S_T = M_T^t M_T$, où M_T est la matrice résolvante du système linéarisé dans l'intervalle $[0, T]$.

5.3.5.1 Spectre

On présente dans la figure V.8 le spectre pour les 500 premières valeurs propres correspondant à la période de temps de 8 heures. On a constaté le fait qu'il y a un écart d'ordre 10^3 entre la première et la 361-ème plus grande valeur propre. Ceci correspond aussi au fait constaté en météorologie que le spectre des équations linéarisées du système (5.8) peut se décomposer en deux parties : les fréquences de Rossby de fréquence lente, et les fréquences correspondant aux ondes de gravité. Lorsque l'amplitude des champs devient

grande, le sous-espace vectoriel engendré par les modes de Rossby n'est plus invariant pour les équations d'évolution. C'est ainsi que le concept de variété lente qui traduit le fait que les flux atmosphériques consistent essentiellement en des mouvements de fréquence lente.

Lorsque l'on utilise la méthode de la puissance itérée pour chercher le vecteur propre dominant de la matrice $S_T = M_T^t M_T$, où M_T est la matrice résolvante du système linéaire tangent, le spectre de cette matrice varie en fonction de la longueur de la période de l'intégration du système (5.21). D'après nos expériences, on a trouvé que, plus la période de calcul augmente, la valeur propre dominante devient plus grande. Et effet, selon Szunyogh et al. ([57]), quand le temps tends vers l'infini, le vecteur propre associé à la plus grande valeur propre va tendre vers le plus grand exposant de Lyapunov. Les exposants de Lyapunov globaux d'un système dynamique sont définis par Tsonis [65] :

$$\lambda_i = \lim_{s \rightarrow \infty} \frac{1}{s} \ln(\sigma_i(t_0 + s)),$$

où σ_i sont les valeurs singulières.

En pratique, le plus grand exposant de Lyapunov local peut être estimé par rapport à une période du temps τ de manière à ce que :

$$L(X(t), \tau) = \frac{1}{\tau} \ln A(t, \tau), \quad \text{où } A(t, \tau) = \frac{\|\hat{X}(t + \tau)\|}{\|\hat{X}(t)\|}$$

est le facteur d'amplification local. Les vecteurs singuliers à droite (u_i) sont les vecteurs de Lyapunov locaux.

Remarque: Les caractéristiques de Lyapunov d'un système dynamique décrivent la stabilité du système sur une longue période du temps. Un exposant positif de Lyapunov implique un comportement chaotique et au contraire, un exposant négatif traduit une convergence des trajectoires du système dans la direction associée à ce exposant. De plus, c'est le plus grand exposant de Lyapunov qui joue le rôle le plus important.

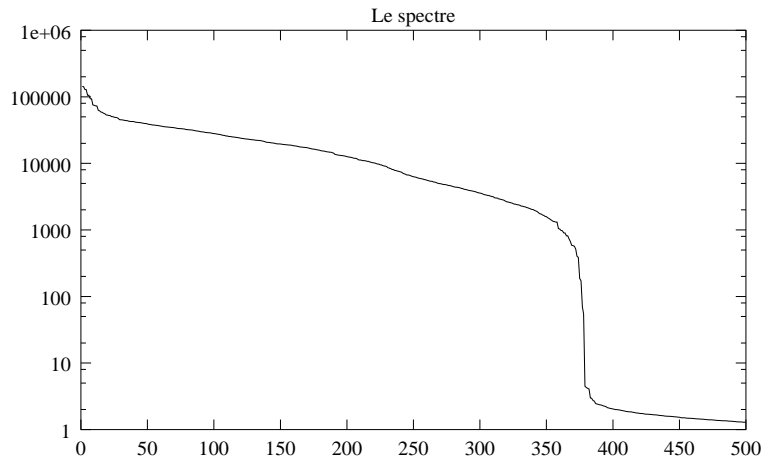


FIG. V.8 – Les valeurs propres en fonction de son numéro.

5.3.5.2 Vecteurs optimaux

Définissons la perturbation optimale V_{opt} par le vecteur initial, par rapport à une norme prédéfinie, qui peut exciter le plus grand taux d'amplification :

$$\tau = \frac{\|\hat{X}(t_0 + T_{opt})\|}{\|\hat{X}(t_0)\|}, \quad \text{dont } \hat{X}(t_0) = V_{opt},$$

sur la période d'optimisation donnée T_{opt} , où \hat{X} est la solution du système linéaire tangent. Pour la norme issue du produit scalaire euclidien, ces perturbations optimales sont les vecteurs propres associés aux plus grandes valeurs propres de la matrice S_t (ici, $t = T_{opt}$). En revanche, les vecteurs obtenus par l'algorithme de Lanczos sont très sensibles à la période du temps d'optimisation T_{opt} et à la norme choisie (cf. Vukicevic [67]). On présente respectivement dans les figures Fig. V.9 et Fig. V.10 les composantes du géopotential des 4 premiers vecteurs propres correspondants aux périodes d'optimisation de 8 heures et 24 heures. On constate qu'il y a des structures analogues entre les vecteurs si les valeurs propres associées sont plus proches. Par contre, il y a de grandes différences entre les vecteurs calculés dans la période sur 8 heures et sur 24 heures.

En appliquant un opérateur scalaire sur les variables du modèle, on peut introduire une autre norme. Par la définition d'une autre norme outre que celle de l'énergie, les perturbations sont limitées au temps initial ou final sur certaines caractéristiques statistiques ou dynamiques. Par exemple, Houtekamer et Derom ([33]) ont utilisé une norme qui a éliminé des structures liées aux caractéristiques statistiques des erreurs d'analyse et Buizza

([4]) a introduit un opérateur de projection locale qui estime les vecteurs optimaux sur une région géographique définie au préalable. Bien que les perturbations optimales dépendent des choix de la période d'optimisation et de la norme, il faut remarquer que tous les perturbations aléatoires, les vecteurs optimaux inclus, tendent vers la direction du premier vecteur de Lyapunov local.

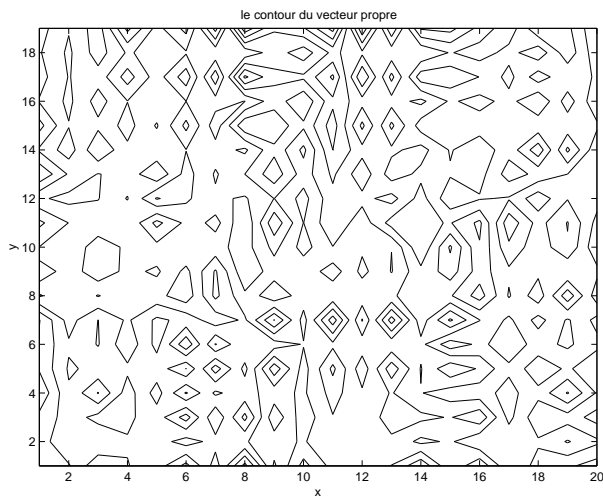
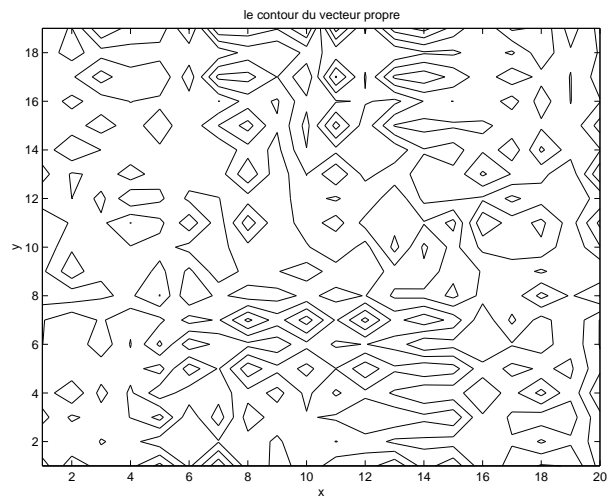
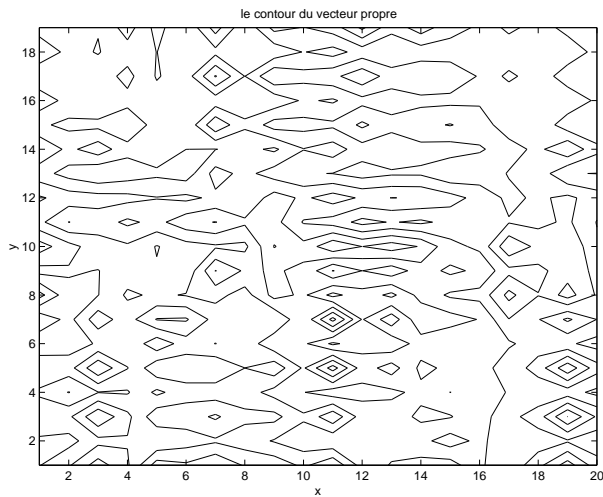
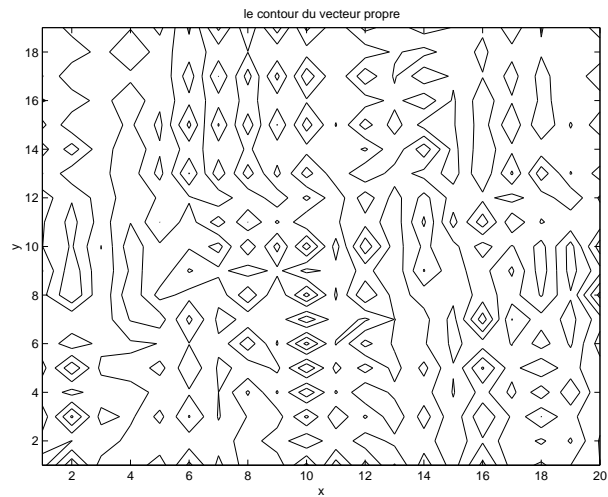
(a) $\lambda = 143285.99010081$ (b) $\lambda = 142097.97379227$ (c) $\lambda = 129350.09345057$ (d) $\lambda = 128191.09532575$ 

FIG. V.9 – Les composantes du potentiel des vecteurs propres associés aux quatre plus grandes valeurs propres qui correspondent à la période d'intégration de 8 heures.

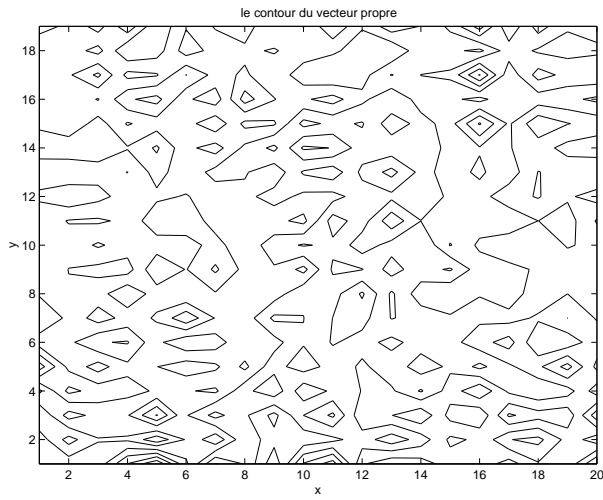
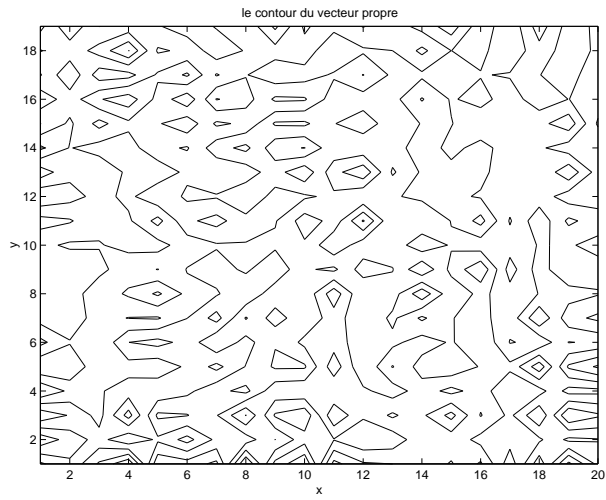
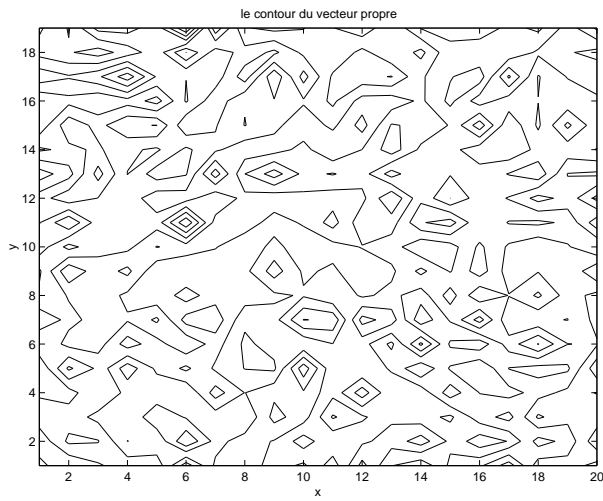
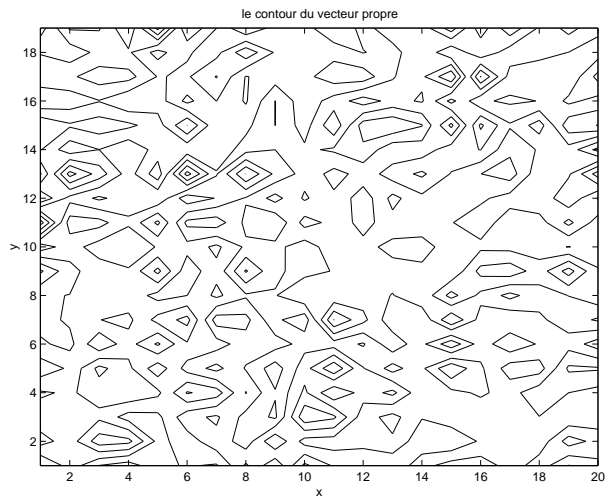
(a) $\lambda = 123541.17385819$ (b) $\lambda = 123316.38735775$ (c) $\lambda = 112585.48404724$ (d) $\lambda = 111324.83106484$ 

FIG. V.10 – Les composantes du potentiel des vecteurs propres associés aux quatre plus grandes valeurs propres qui correspondent à la période d'intégration de 24 heures.

5.3.6 Critère du choix de la base d'espace composée des vecteurs propres

Ayant pour but de diminuer la dimension du problème d'optimisation, nous considérons à choisir les vecteurs propres d'une matrice, dont le nombre est inférieure à la dimension de l'espace de contrôle, comme base d'espace par laquelle l'erreur du modèle sont représentées, une question cruciale concernant est de savoir quels et combien de ces vecteurs sont nécessaires pour que l'approximation d'une erreur réelle soit suffisamment bonne.

Le fait est bien connu que, les vecteurs propres associés aux plus grandes valeurs propres et trouvés par la méthode singulière sont les directions permettant aux erreurs de s'amplifier le plus rapidement possible. Autrement dit, s'il y a des perturbations ont lieu dans ces directions, elles entraînent l'écart le plus remarquable sur la solution du modèle. Donc il faut chercher un critère pour déterminer quel nombre des vecteurs est suffisant pour, en un certain sens, approcher les erreurs du modèle considéré. Pour l'instant, on a étudié la base d'espace spatial, quant à la base d'espace temporel, on adopte la fonction d'indicatrice du temps. Si la période de calcul est très longue, une base d'ondelette pourra être considérée.

Soit H une perturbation arbitraire dans le système linéaire tangent (5.15), et soient les vecteurs propres Y_1, Y_2, \dots associés aux valeurs propres positives $\lambda_1, \lambda_2, \dots$, qui sont rangées par ordre décroissant, i.e. $\lambda_1 \geq \lambda_2 \geq \dots$. On propose deux critères suivants pour chercher le nombre des vecteurs nécessaires :

(i) D'après l'approximation de la norme L^2 :

On cherche m l'entier le plus petit lorsque la condition

$$\|H - \sum_{i=1}^m C_i Y_i\|_{L^2} \leq \epsilon_1 \quad (5.25)$$

soit réalisée, où

$$C_i = \langle H, Y_i \rangle, \quad i = 1, \dots, m,$$

sont les coordonnées de H dans la base orthonormées $\{Y_i\}_{i=1, \dots, m}$.

(ii) D'après l'approximation de l'amplification de l'erreur au temps final :

Soit H_m la projection de H sur l'espace engendré par $\{Y_i\}_{1 \leq i \leq m}$:

$$H_m = \sum_{i=1}^m C_i Y_i,$$

où $C_i = \langle H, Y_i \rangle$, et \hat{X} et \hat{X}_m dénotent respectivement la solutions du système (5.15) et la solution du système initialisé de H_m qui vérifient :

$$\begin{cases} \frac{\partial \hat{X}_m}{\partial t} = \left[\frac{\partial F}{\partial X} \right] \hat{X}_m, & \text{sur } \Omega \times \in [0, T], \\ \hat{X}_m(t=0) = H_m, & \text{sur } \Omega. \end{cases} \quad (5.26)$$

On cherche l'entier m le plus petit pour que la condition

$$\frac{\|\hat{X}(T) - \hat{X}_m(T)\|_{L^2}}{\|\hat{X}(T)\|_{L^2}} \leq \epsilon_2 \quad (5.27)$$

soit réalisée.

Afin de réaliser ces deux critères, on emploie le même modèle de Shallow-Water (5.8), la condition initiale étant toujours donnée par celle de Grammeltvedt, le nombre de pas d'espace étant $N = 1083$ et le pas de temps $\delta t = 600 \text{sec}$. La période d'intégration est de 48 pas de temps.

Dans le premier test, ϵ_1 dans (5.25) étant un seuil fixé de l'ordre 10^{-4} , on voit sur la figure Fig. V.11 que, comme la norme de l'écart entre la perturbation et la projection sur l'espace engendré par la base $\{Y_i\}_{i=1, \dots, m}$ ne diminue pas beaucoup quand m est égal à la moitié de la dimension de l'espace, la condition (5.25) est difficile à réaliser. Autrement dit, ce critère n'est pas utilisable pour notre problème. Par contre, en ce qui concerne le deuxième critère, une expérience a été faite. La figure Fig. V.12 montre que, en fixant toujours le seuil $\epsilon_2 = 10^{-4}$, quand m est supérieur à 371, la condition (5.27) est déjà satisfaite. Cela nous permet d'appliquer ce critère pour déterminer le nombre de vecteurs de base nécessaires pour générer les erreurs du modèle.

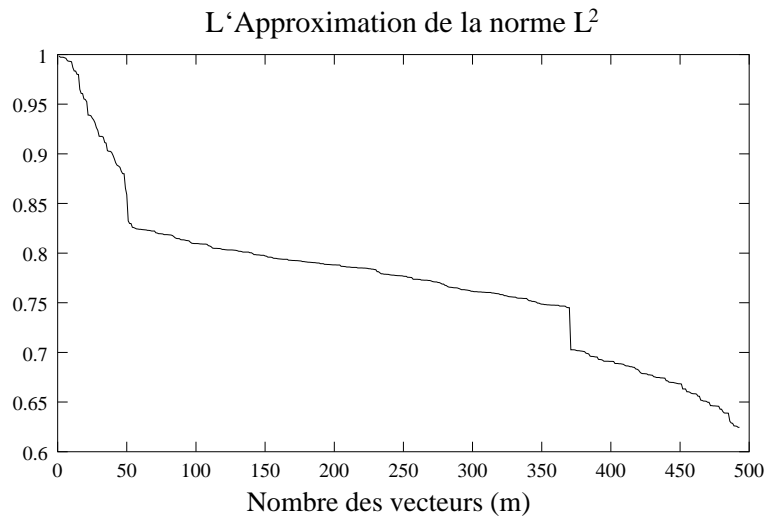


FIG. V.11 – La variation de $\|H - \sum_{i=1}^m C_i Y_i\|_{L^2}$ en fonction du nombre m de vecteurs utilisés.

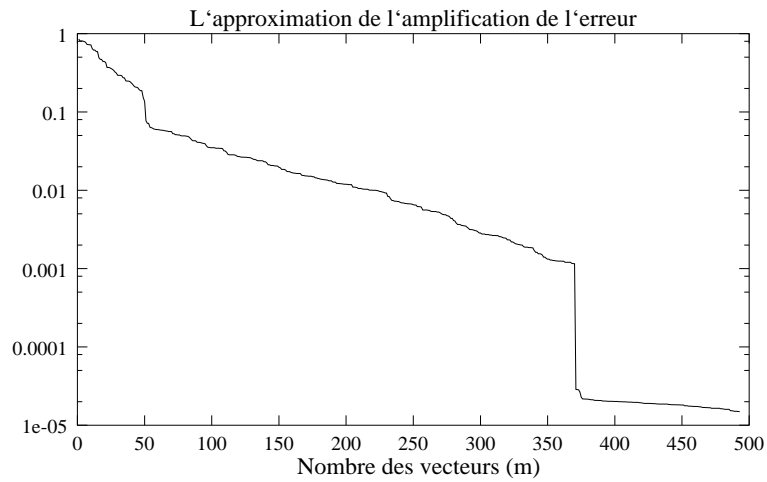


FIG. V.12 – La variation logarithme de $(\|\hat{X}(T) - \hat{X}_m(T)\|) / \|\hat{X}(T)\|$ en fonction du nombre m des vecteurs utilisés.

5.3.7 Système de contrôle II

Par suite du coût de calcul, nous cherchons à déterminer des perturbations engendrées par une autre base canonique que la base euclidienne. Soient $\{Y_i\}_{1 \leq i \leq M}$ les M vecteurs orthonormés dans \mathbb{R}^m (où $m = 3n$, avec n la dimension de l'espace de discrétisation), qui vérifient

$$\langle Y_i, Y_j \rangle = \begin{cases} 1, & \text{si } i = j, \\ 0, & \text{si } i \neq j, \end{cases} \quad (5.28)$$

et soit $\{f_j(t)\}_{1 \leq j \leq N}$ une base de fonctions dépendant du temps, définies sur l'intervalle $t \in [0, T]$, dont le nombre peut être inférieur au nombre de pas de temps N_T . On définit le produit scalaire entre deux vecteurs de cette base comme suit :

$$\int_0^T f_i(t) f_j(t) dt = \begin{cases} 1, & \text{si } i = j, \\ 0, & \text{si } i \neq j. \end{cases} \quad (5.29)$$

Si l'on intéresse toujours le système (5.1), notre problème revient à chercher une condition initiale développée par $\{Y_i\}$:

$$U = \sum_{i=1}^M \theta_i Y_i,$$

ainsi qu'une erreur de modèle développée par $\{Y_i\}$ et $\{f_j\}$:

$$V = \sum_{i=1}^M \sum_{j=1}^N \alpha_{ij} f_j(t) Y_i.$$

Donc l'on ne prend que les coefficients $\alpha = (\alpha_{ij})$ et $\{\theta_i\}$, $1 \leq i \leq M, 1 \leq j \leq N$, comme variables de contrôle, et la dimension du problème d'optimisation devient $M \times 1 + N$.

Considérons une fonction coût sous la forme

$$J(U, V) = \frac{1}{2} \int_0^T (\|CX - X_{obs}\|^2 + \beta \langle \mathcal{N}V, V \rangle) dt, \quad (5.30)$$

où β est le coefficient de régularisation, \mathcal{N} est une matrice symétrique définie positive. S'il y a des perturbations h_θ et h_α sur θ et α respectivement, qui entraîne des perturbations h_U et h_V sur la condition initiale U et sur l'erreur V respectivement par :

$$h_U = \sum_{i=1}^M (h_\theta)_i Y_i, \quad h_V = \sum_{i=1}^M \sum_{j=1}^N (h_\alpha)_{ij} f_j(t) Y_i,$$

donc le système linéaire tangent est présenté de la manière suivante

$$\begin{cases} \frac{\partial \hat{X}}{\partial t} = \left[\frac{\partial F}{\partial X} \right] \hat{X}(t) + B(\sum_{i,j}(h_\alpha)_{ij}f_j(t)Y_i), & t \in [0, T], \\ \hat{X}(0) = \sum_i(h_\theta)_iY_i. \end{cases} \quad (5.31)$$

Introduisant le même système adjoint comme dans la partie précédente :

$$\begin{cases} \frac{\partial P}{\partial t} + \left[\frac{\partial F}{\partial X} \right]^t P = C^t(CX - X_{obs}), & t \in [0, T], \\ P(T) = 0. \end{cases} \quad (5.32)$$

et en multipliant l'équation (5.31) par P , on intègre entre 0 et T et par intégration par partie, c'est facile à déduire que :

$$(\nabla_{\theta}J, h_U) = -(P(0), \sum_i(h_\theta)_iY_i), \quad \text{et}$$

$$(\nabla_{\alpha}J, h_V) = - \int_0^T \left(\left\langle \sum_{i,j}(h_\alpha)_{ij}f_jY_i, B^tP \right\rangle + \beta \left\langle \mathcal{N} \sum_{i,j}\alpha_{ij}f_jY_i, \sum_{i,j}(h_\alpha)_{ij}f_jY_i \right\rangle \right) dt.$$

Ceci fournit le gradient de J par rapport à θ_i et α_{ij} , pour $1 \leq i \leq M, 1 \leq j \leq N$:

$$(\nabla_{\theta}J)_i = - \langle P(0), Y_i \rangle,$$

$$(\nabla_{\alpha}J)_{ij} = \int_0^T \langle (-B^tP + \beta \mathcal{N} \sum_{k,l}\alpha_{kl}f_lY_k), f_jY_i \rangle dt. \quad (5.33)$$

Si la fonction coût atteint le minimum, alors $(\nabla_{\alpha}J)_{ij} = 0$ et on obtient la valeur optimale du coefficient α_{ij} :

$$\alpha_{ij} = \int_0^T \langle \beta^{-1}\mathcal{N}^{-1}B^tP, f_jY_i \rangle dt. \quad (5.34)$$

5.3.8 Expériences numériques avec réduction de la taille de l'espace de contrôle

5.3.8.1 Création des observations

En utilisant la méthode variationnelle d'assimilation de données, on veut estimer les erreurs d'un modèle supposé non parfait à partir des observations. En vue d'une comparaison des résultats de l'estimation sous la même condition, on veut créer d'abord des

observations qui seront utilisées dans nos expériences suivantes. Les observations sont faites selon les étapes ci-dessous :

- Le problème de contrôle optimal est pris en compte sur le domaine rectangulaire (voir la figure Fig. V.1) de longueur 6×10^3 km et de largeur 4.4×10^3 km, qui est discrétisé en grille de 21×21 points, noté G_{21} . On discrétise le domaine d'une manière plus fine en choisissant les pas en espace $\delta x = 150$ km et $\delta y = 110$ km, on obtient une grille de points 41×41 , noté G_{41} . On considère respectivement le modèle correspondant aux grilles G_{41} et G_{21} comme modèles correct et incorrect ;
- Supposons $X_{G_{41}}$ et $X_{G_{21}}$ sont les solutions numériques après l'intégration du modèle G_{41} et du modèle G_{21} sur une période de temps $[0, T]$, où $X_{G_{21}}$ est égale au temps initial la restriction de la condition initiale de $X_{G_{41}}$ sur le domaine G_{21} , c'est-à-dire,

$$X_{G_{21}}(t = 0) = X_{G_{41}}(t = 0)|_{G_{21}}.$$

- Nous choisissons nos pseudo-observations sur le domaine G_{21} au temps $n\delta t$, $1 \leq n \leq N_T$, de façon à ce que :

$$X_{obs}^n = X_{G_{41}}^n|_{G_{21}},$$

et donc en soustrayant deux solutions, les erreurs par rapport aux observations s'écrivent :

$$E_r^n = X_{G_{21}}^n - X_{G_{41}}^n|_{G_{21}}.$$

En fait, soit X^n la solution du modèle correct, la solution au temps $(n+1)\delta t$ est égale à :

$$X^{n+1} = F(X^n),$$

et soit \tilde{X}^n la solution du modèle incorrect, dont la solution au temps $(n+1)\delta t$ satisfait l'équation suivante :

$$\tilde{X}^{n+1} = \tilde{F}(\tilde{X}^n) = F(\tilde{X}^n) + G^{n+1}$$

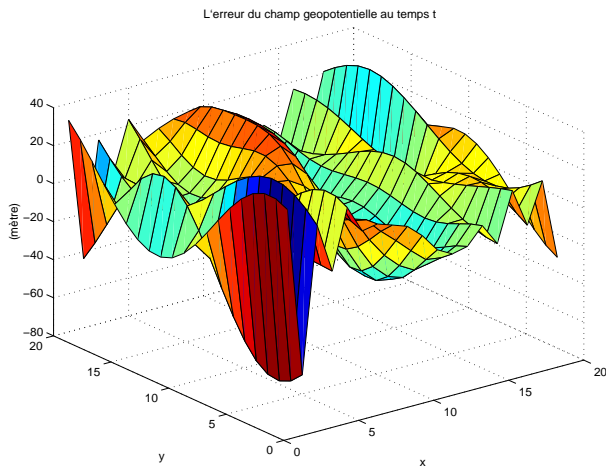
où \tilde{F} correspond au modèle incorrect.

On a donc

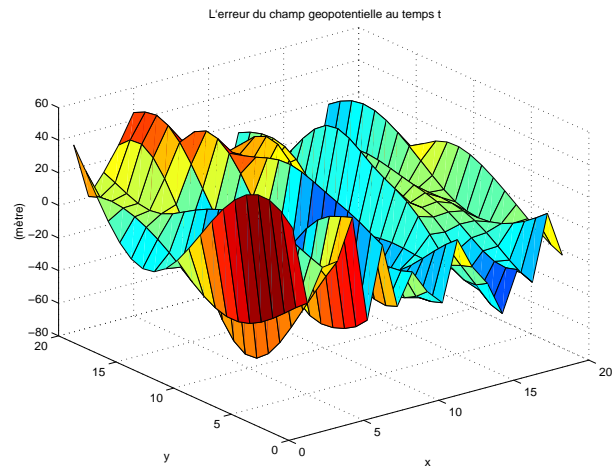
$$E_r^{n+1} = \tilde{X}^{n+1} - X^{n+1} = (F(\tilde{X}^n) - F(X^n)) + G^{n+1},$$

qui contient la propagation des erreurs aux temps précédents : $F(\tilde{X}^n) - F(X^n)$ et l'erreur d'approximation due du schéma aux différences G^{n+1} . On présente sur la figure Fig. V.13 l'évolution de l'écart du champ géopotential et l'observation durant l'intervalle $[0, T]$, ici $T = 8$ heures, soit $48 \delta t$. La croissance de l'amplitude de l'erreur E_r^n est observée.

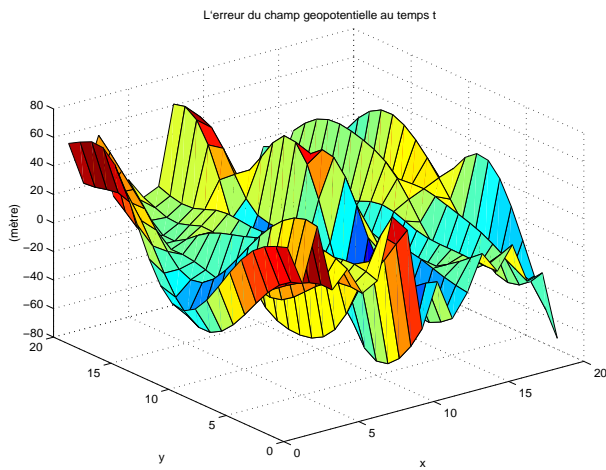
$$(a) t = \frac{1}{4}T$$



$$(b) t = \frac{1}{2}T$$



$$(c) t = \frac{3}{4}T$$



$$(d) t = T$$

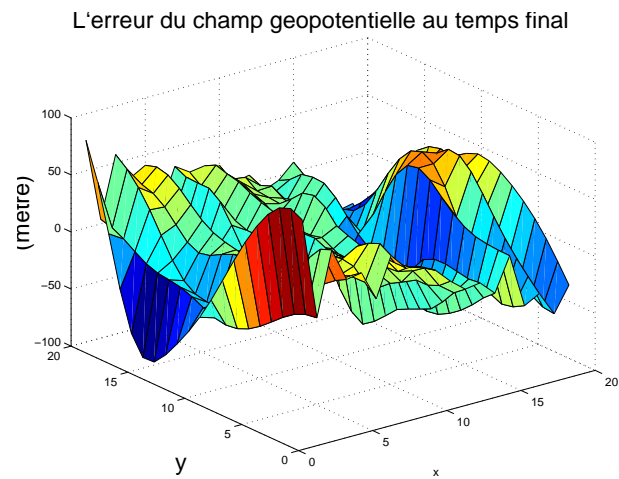


FIG. V.13 – L'évolution de l'erreur du champ géopotentiel entre les solutions du modèle en grille 41×41 et en grille 21×21 .

5.3.8.2 Expérience I - Choix de base d'espace

D'après la section 5.3.7, l'estimation de l'erreur du modèle se ramène à la détermination des coefficients d'espace.

Dans les premiers expériences, on estime d'abord les erreurs sur tous les pas de temps contenant l'erreur de la condition initiale. Pour $i = 1, 2, \dots, N_T + 1$, posant les instants

discrets

$$t_i = (i - 1)\delta t,$$

avec δt le pas de temps, on définit donc une suite de fonctions de base en temps $f_i(t)$ par la fonction d'indicatrice de l'intervalle $[t_i, t_{i+1}[$. Dans ce cas là, le nombre de fonctions de base est égal à celui de pas de temps.

On choisit l'espace de phase engendré par une suite de vecteurs orthonormés $\{Y_i\}$ qui sont trouvés par l'algorithme de Lanczos. Ces vecteurs sont les vecteurs propres de la matrice $M_T^t M_T$, où M_T est la matrice résolvante du système linéaire tangent dans l'intervalle du temps $[0, T]$, on les numérote par ordre décroissant de leur valeurs propres associées.

Pour étudier l'impact du choix et du nombre des vecteurs formant une base d'espace sur le résultat, on va comparer les résultats dans les 4 cas différents en utilisant des bases différentes :

1. Sans réduction : l'espace de dimension 51984 avec la base euclidienne ;
2. Avec réduction : l'espace de dimension 17280 avec les vecteurs de base $\{Y_{N_m}\}$: $N_m = 1 - 360$;
3. Avec réduction : l'espace de dimension 9600 avec les vecteurs de base $\{Y_{N_m}\}$: $N_m = 1 - 200$;
4. Avec réduction : l'espace de dimension 9600 avec les vecteurs de base $\{Y_{N_m}\}$: $N_m = 300 - 500$.

Les expériences sont effectuées avec les mêmes observations créées au début de cette partie, et le même coefficient de régularisation $\beta = 1.0$ et la même matrice \mathcal{N} (5.30). A la première itération du processus d'optimisation, les valeurs de départ des coefficients d'espace sont données par des valeurs aléatoires de module inférieure à 1. Si le critère d'arrêt de la minimisation est satisfait, on dit que l'on a trouvé des coefficients optimaux et les erreurs du modèle sont reconstruites avec ces coefficients optimaux.

Les différences entre les résultats sont comparées et remarquées dans trois aspects suivants :

- On présente dans la figure Fig. V.14 les variations du logarithme décimal de la fonction coût et de la norme du gradient normalisé en fonction des itérations dans les cas avec réduction de taille et sans réduction. On a remarqué que, en utilisant l'espace engendré par les vecteurs propres, l'algorithme converge en beaucoup moins d'itérations et le gradient a descendu presque de même ordre. Parmi les trois cas où l'on a utilisé une autre base d'espace, le résultat avec les 360 premiers vecteurs

est le plus proche du cas où l'on utilise la base canonique de l'espace de phases. Par rapport au même nombre de vecteurs de base, le résultat dans le cas où le numéro est de 1 à 200 est meilleur que dans le cas de 300 à 500. Ceci correspond à l'analyse des propriétés des vecteurs propres associés aux plus grandes valeurs propres.

- On présente dans les figures Fig. V.19 - Fig. V.22 les normes L^2 des coefficients en fonction du temps et aussi celles des erreurs obtenues avec la base orthonormée. A gauche et à droite, on représente respectivement les deux termes : (A) les valeurs au départ de la minimisation et (B) les valeurs identifiées. On a remarqué une tendance de descente de la norme des erreurs par rapport au temps. Parce que l'erreur due à la condition initiale tend à être "oublié".
- Avec les erreurs reconstruites avec les coefficients optimaux, on a intégré de nouveau le modèle dont la prévision doit approcher l'observation. On a comparé dans les figures Fig. V.15 - Fig. V.18 l'écart par rapport à l'observation du champ géopotentiel au temps final avant le contrôle et après l'estimation. On constate que, après l'estimation, s'il n'y a pas de réduction (la dimension de l'espace de phase est de 1083), l'écart de la hauteur maximum du champ du potentiel est inférieur à 10 mètres et si l'on prend les 360 premiers vecteurs singuliers comme base qui permettent d'approcher au mieux les observations à tous les instants. Mais lorsque l'on utilise moins de vecteurs de base, l'écart au temps final reste encore plus grand.

Remarque : Malgré que l'on ait utilisé une base pour réduire la dimension de l'espace de contrôle, le tableau Tab. V.1 montre que le calcul coûte assez cher en raison à l'appel des vecteurs aux base pour construire l'erreur à chaque étape de la simulation sur tous les points et à tous les moments et aussi pour chaque itération de l'optimisation. Ceci nous demande d'optimiser l'algorithme de la programmation. On peut aussi essayer de réduire le nombre en fonction du temps. Pour cela, les expériences ont été faites et seront montrées dans la partie suivante.

	sans réduction	avec réduction		
	base euclidienne	$N_m=1 - 360$	$N_m = 1 - 200$	$N_m = 300 - 500$
Temps CPU (<i>min</i>)	3,28	26,42	22,02	15,54

TAB. V.1 – Comparaison du temps de CPU

5.3.8.3 Expérience II – Diminution du nombre de fonctions de base en temps

On veut diminuer en plus, dans cette partie, le nombre des paramètres à identifier avec réduction de nombre de fonctions de base en temps. Soit M_s un entier et $\left[\frac{N_T}{M_s} \right]$ la

partie entière de N_T/M_s , pour $1 \leq k \leq \left\lceil \frac{N_T}{M_s} \right\rceil$, on prend $f_k(t)$ la fonction d'indicatrice de l'intervalle $[t_{1+(k-1)M_s}, t_{1+kM_s}[$, i.e.

$$f_k(t) = \begin{cases} 1, & t \in [t_{1+(k-1)M_s}, t_{1+kM_s}[\\ 0, & t \notin [t_{1+(k-1)M_s}, t_{1+kM_s}[\end{cases} \quad (5.35)$$

Donc la détermination des erreurs du modèle se ramène à identifier les coefficients α_{ij} de l'erreur V , qui est décrite par :

$$V = \sum_{i=1}^N \sum_{j=1}^{\left\lceil \frac{N_T}{M_s} \right\rceil} \alpha_{ij} f_j(t) Y_i$$

avec Y_i les vecteurs mentionnés dans la partie précédente. On prend ici $N = 360$.

Avec les mêmes conditions que l'expérience I, on a comparé les résultats avec les 3 critères quand $M_s = 2, 3, 6$ et 8. Ce qui sera à estimer l'erreur du modèle tous les 2, 3, 6 ou 8 pas de temps. On a représenté sur la figure Fig. V.23 les taux de convergence de la fonction coût et du gradient. Ce qui est intéressant est que le résultat avec $M_s = 3$, où le nombre de la base de l'espace en temps est de $48/3 = 16$, est plus proche du résultat sans réduction que dans le cas où $M_s = 2$. Les figures Fig. V.24 - Fig. V.27 représentent l'écart par rapport à l'observation du champ du potentiel au temps final quand les erreurs sont données par les coefficients de départ et les coefficients identifiés. Et on a représenté sur les figures Fig. V.28 - Fig. V.31 les normes de l'erreur au départ et après l'estimation en fonction du pas de temps sur M_s .

On a remarqué que avec la réduction, on a beaucoup diminué le temps de calcul (dans le tableau Tab. V.2) et en même temps, cela a assuré le contrôle de l'amplification de l'erreur. Par exemple, quand $M_s = 6$, la dimension du problème de contrôle devient $360 \times 8 = 2880$, qui est beaucoup plus petite que dans le problème originel soit $1083 \times 48 = 51984$, mais la croissance des erreurs est contrôlée de façon assez satisfaisante.

	contrôle de l'erreur de tous les M_s pas de temps			
	$M_s = 2$	$M_s = 3$	$M_s = 6$	$M_s = 8$
Temps CPU (<i>min</i>)	10,38	5,38	2,34	1,40

TAB. V.2 – Comparaison du temps de CPU

5.3.9 Conclusion

Les modèles hydrologique ont des incertitudes qu'il faut tenir compte l'erreur de la modélisation. Le contrôle de l'erreur du modèle nous permet d'estimer des incertitudes

dans le modèle. Pour déterminer les ingrédients incertains d'un modèle, l'utilisation de l'espace engendrée par des vecteurs singuliers de la matrice résolvante du système linéaire tangent permet de réduire la dimension du problème d'optimisation, surtout dans l'aspect de contrôle de la propagation des erreurs liées au modèle auxquelles la prévision est très sensibles. Donc la réduction de l'espace de contrôle rend l'estimation de l'erreur du modèle possible.

Dans le futur travail, on peut aussi essayer d'introduire d'autres sortes de fonctions de base en temps, ainsi de choisir d'autres bases d'espace avec les vecteurs utiles d'après leurs propriétés et les rôles joués différemment dans l'analyse de la propagation des erreurs. Par exemple,

- les vecteurs de Lyapunov ;
- les vecteurs propres du Hessienne, et pour calculer le produit du Hessienne et un vecteur, on a besoin de développer le système adjoint de second ordre ;
- les directions de descente dans l'algorithme d'optimisation.

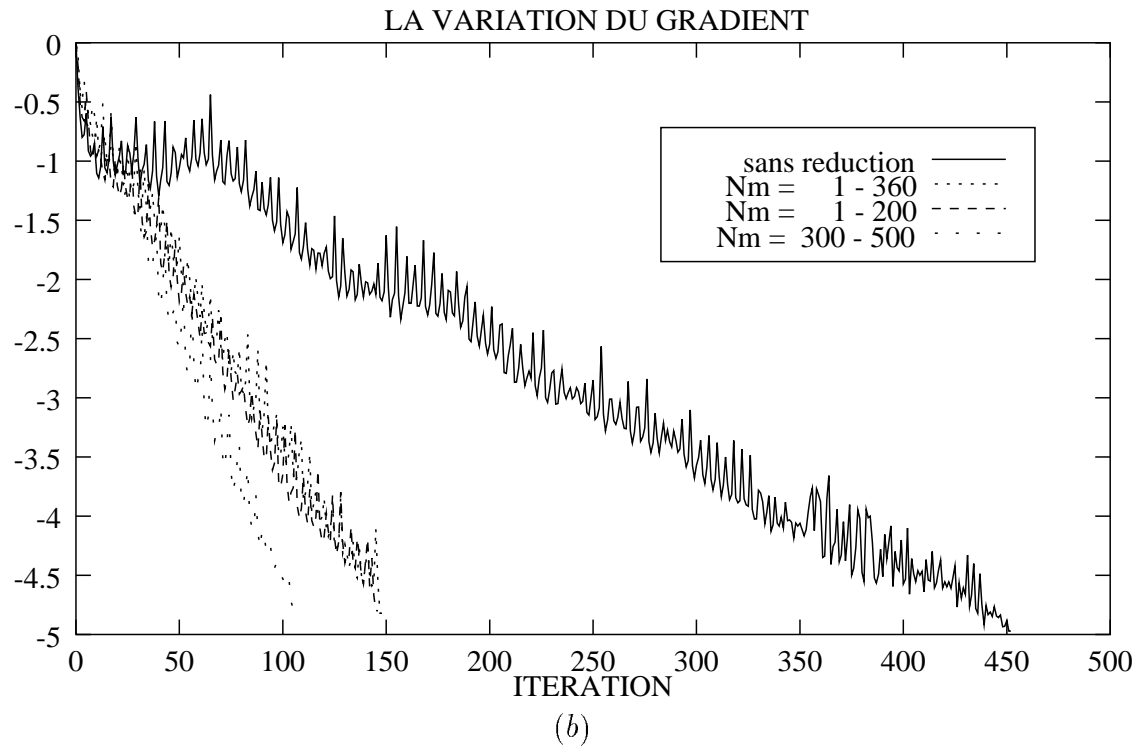
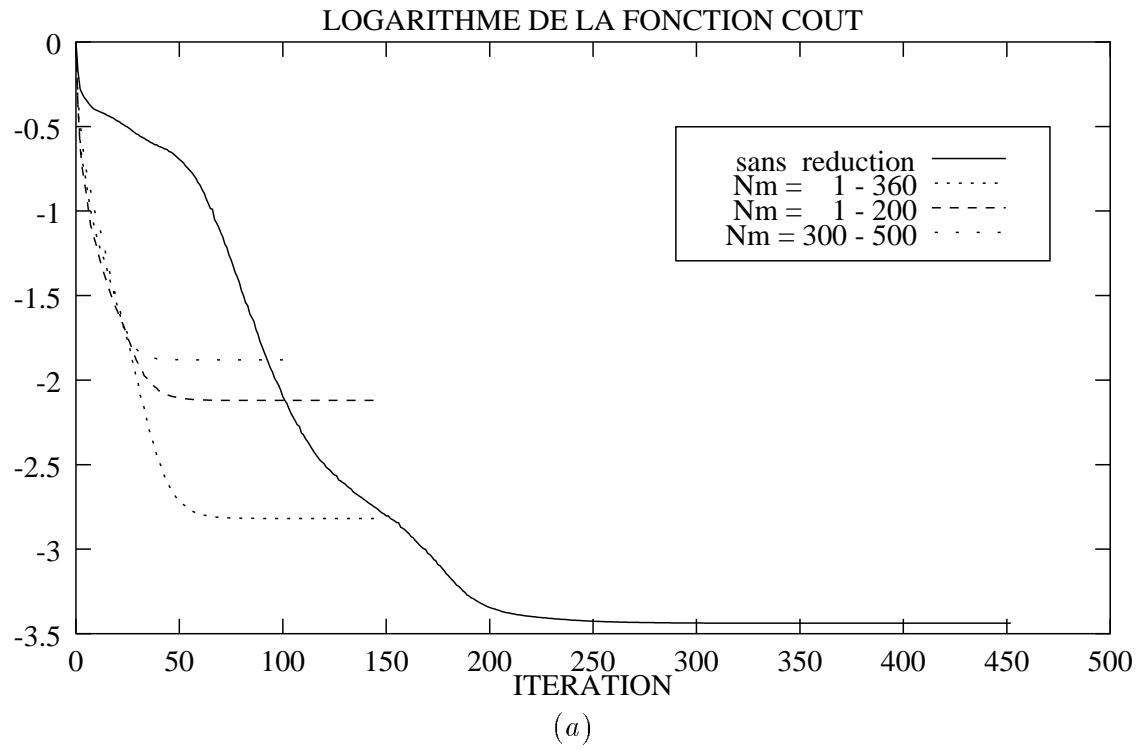
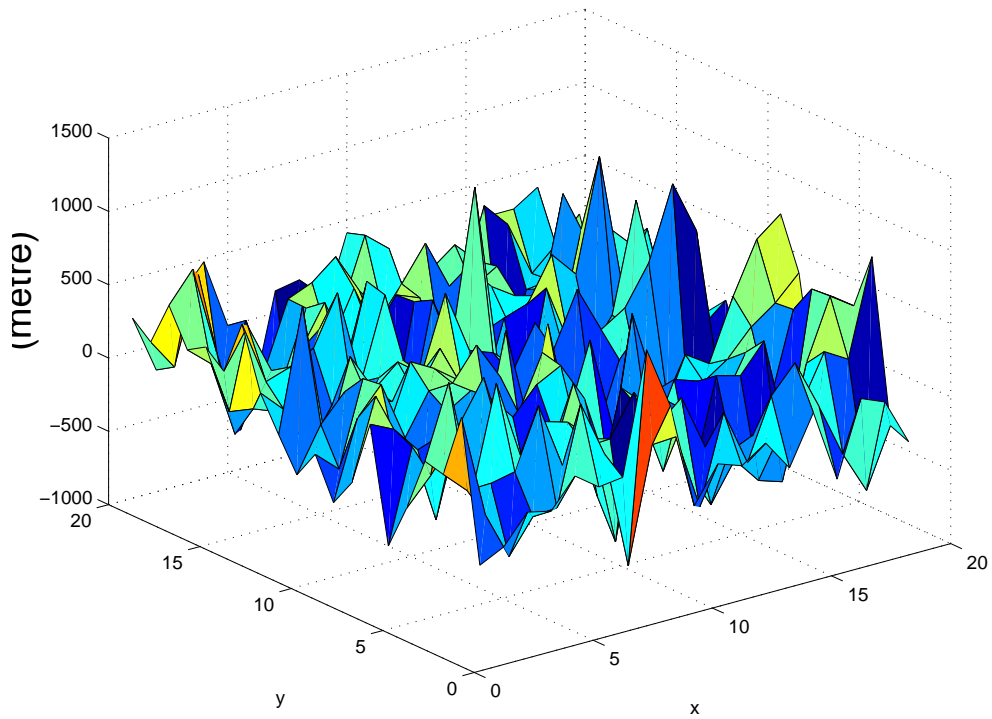
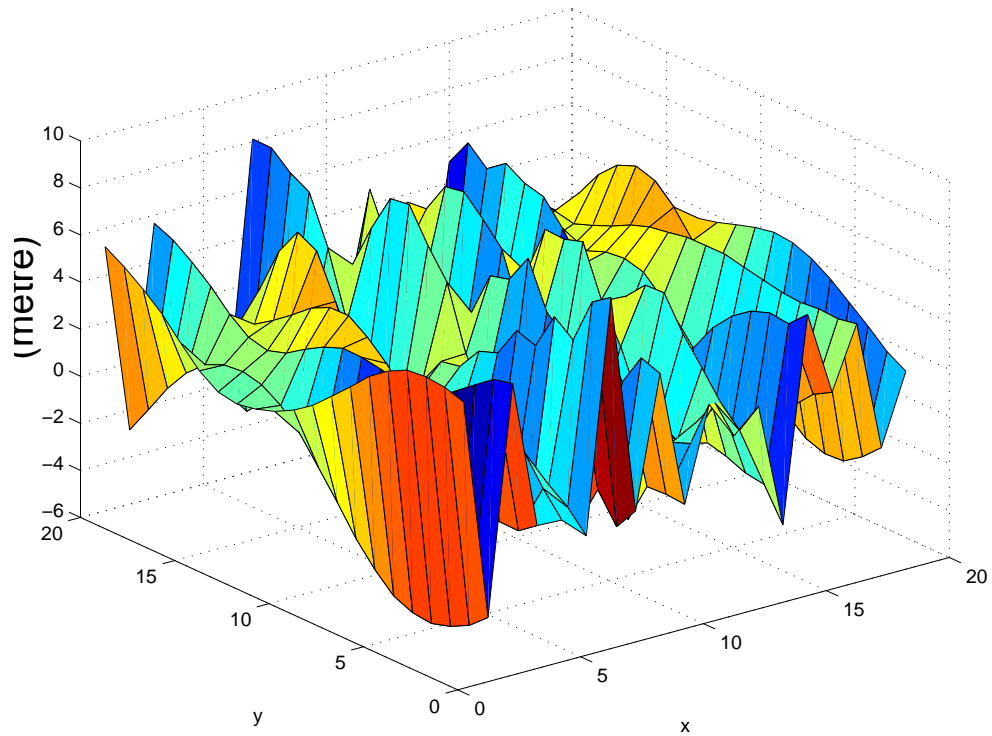


FIG. V.14 – Les comparaisons des variations de $\log(J_k/J_0)$ (a) et des variations de $\log(\|\nabla J_k\|/\|\nabla J_0\|)$ (b) en fonction des itérations k sans la réduction de taille et avec des bases différentes

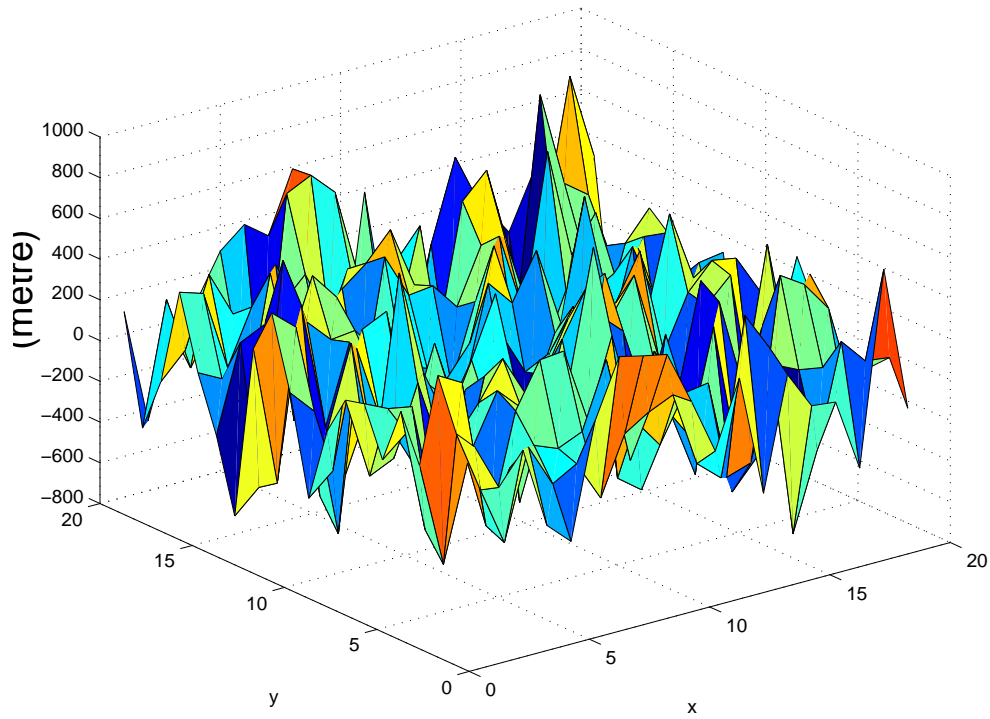


L'écart avant contrôle par rapport à l'observation du champ du potentiel au temps final

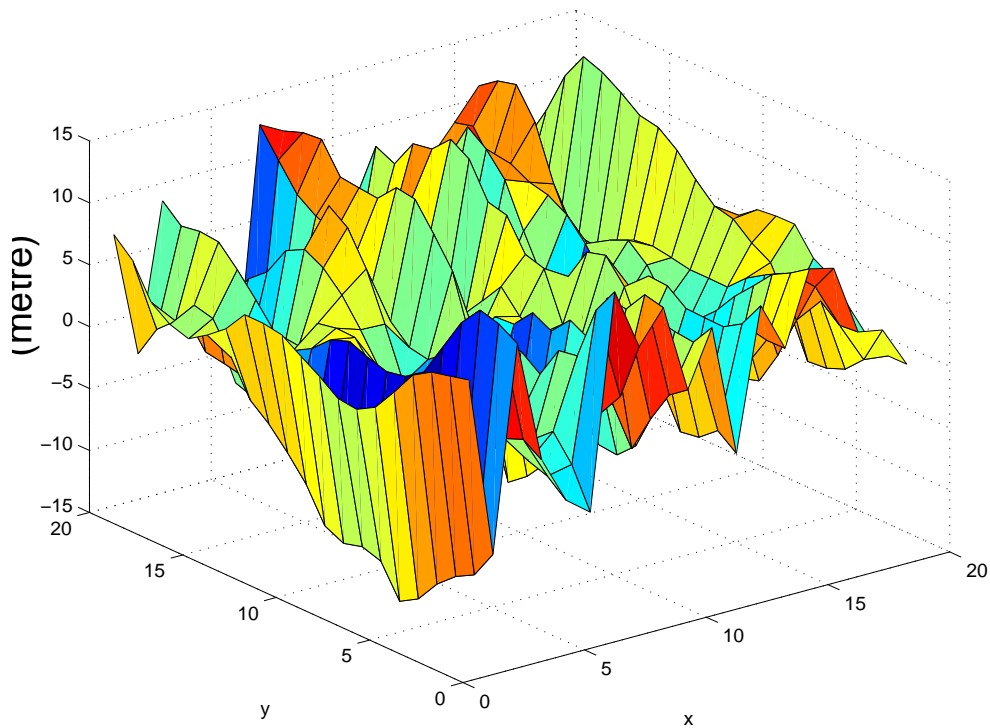


L'écart après contrôle par rapport à l'observation du champ du potentiel au temps final

FIG. V.15 – *Expérience sans réduction de dimension*

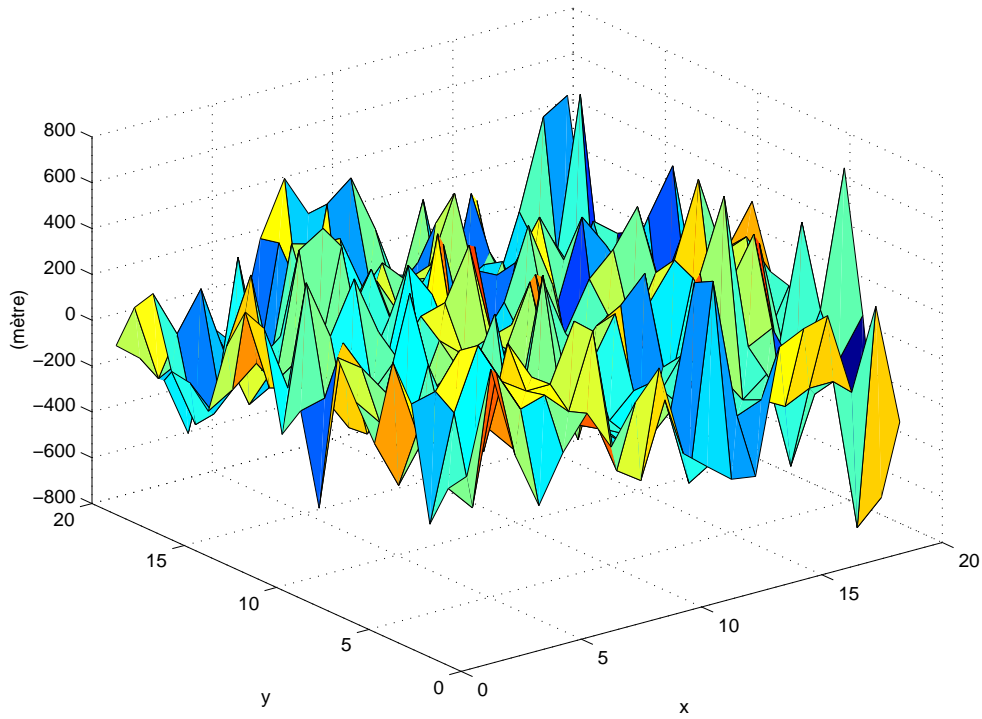


L'écart avant contrôle par rapport à l'observation du champ du potentiel au temps final

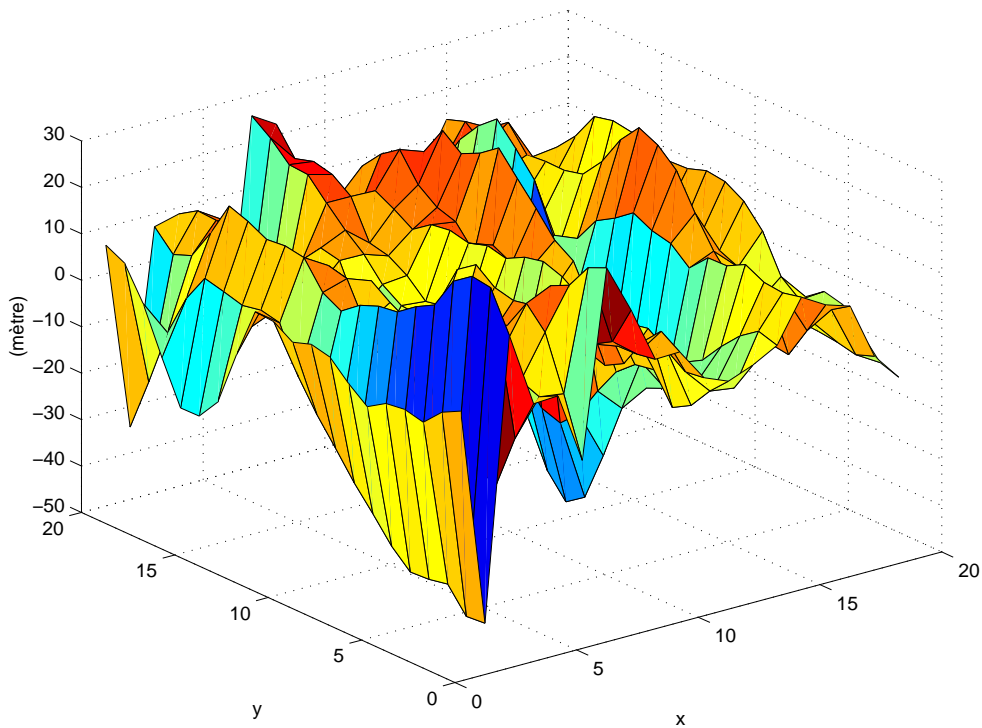


L'écart après contrôle par rapport à l'observation du champ du potentiel au temps final

FIG. V.16 – *Expérience avec les premiers vecteurs $N_m = 1 - 360$*

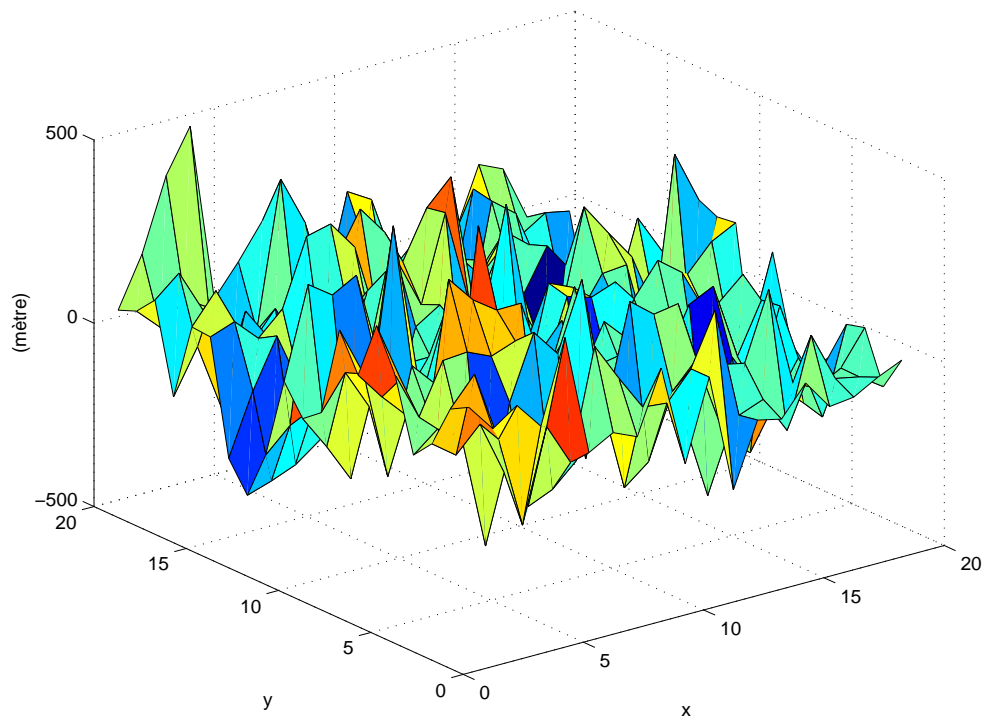


L'écart avant contrôle par rapport à l'observation du champ du potentiel au temps final

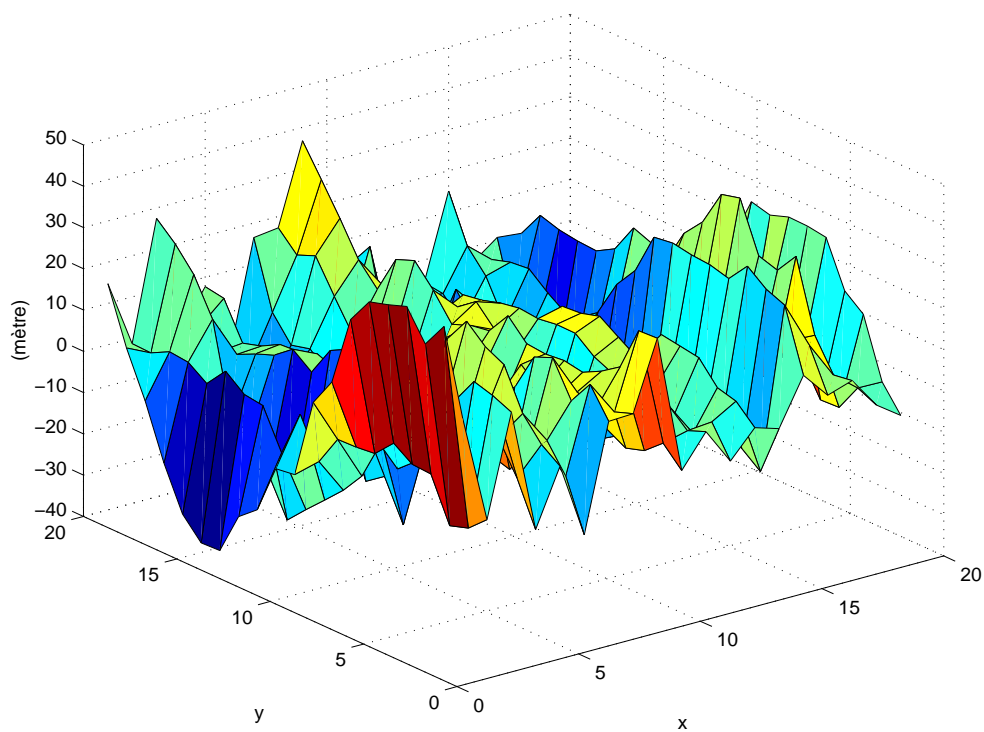


L'écart après contrôle par rapport à l'observation du champ du potentiel au temps final

FIG. V.17 – *Expérience avec les premiers vecteurs $N_m = 1 - 200$*



L'écart avant contrôle par rapport à l'observation du champ du potentiel au temps final



L'écart après contrôle par rapport à l'observation du champ du potentiel au temps final

FIG. V.18 – *Expérience avec les numéros des vecteurs $N_m = 300 - 500$*

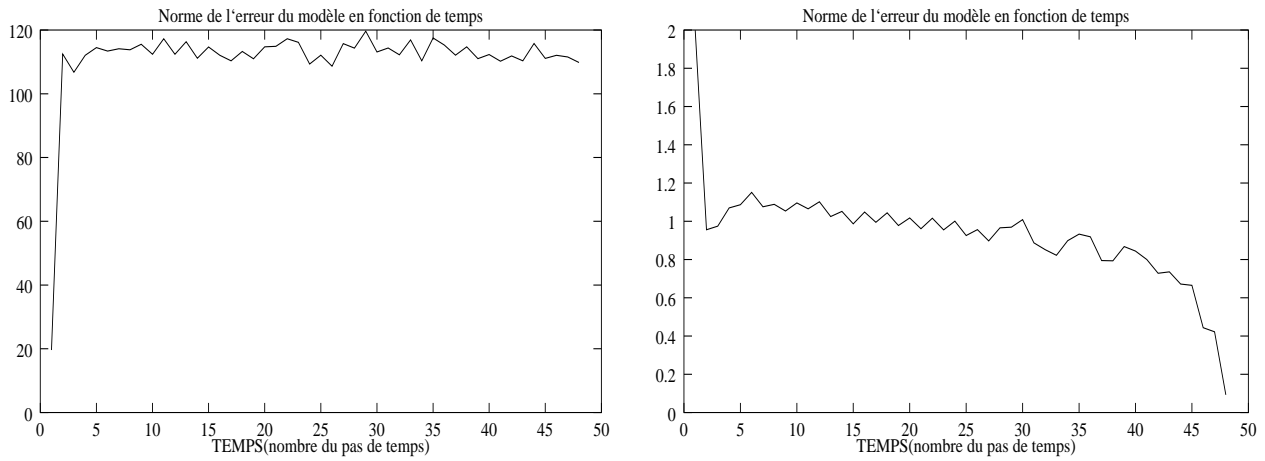


FIG. V.19 – *Expérience sans réduction de la taille de l'espace de contrôle. La norme de l'erreur avant contrôle (à gauche) et de l'erreur optimale (à droite) en fonction du temps.*

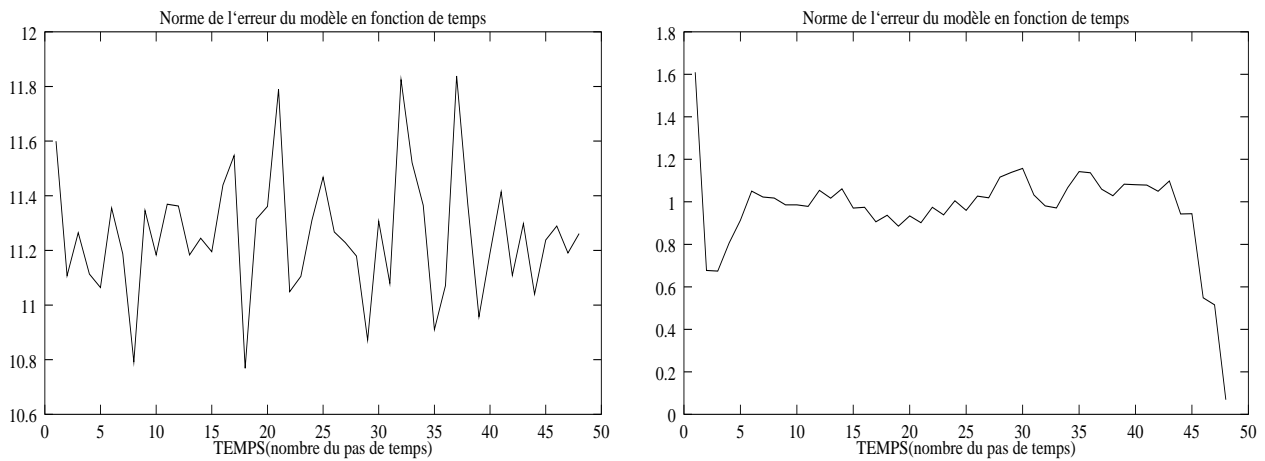


FIG. V.20 – *Expérience avec la base où le numéro du vecteur $N_m = 1 - 360$. La norme de l'erreur avant contrôle (à gauche) et de l'erreur optimale (à droite) en fonction du temps.*

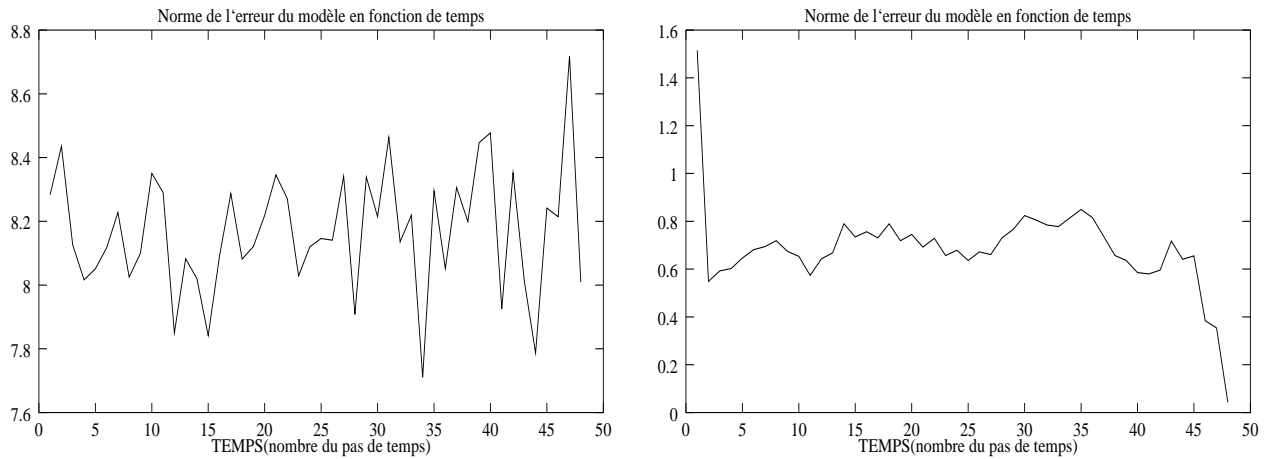


FIG. V.21 – *Expérience avec la base où le numéro du vecteur $N_m = 1 - 200$. La norme de l'erreur avant contrôle (à gauche) et de l'erreur optimale (à droite) en fonction du temps.*

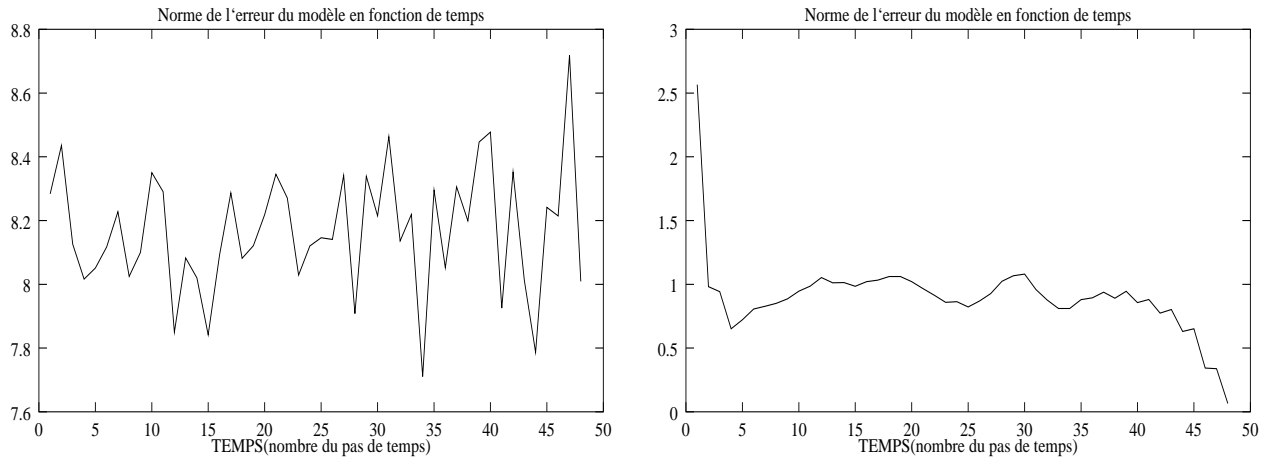
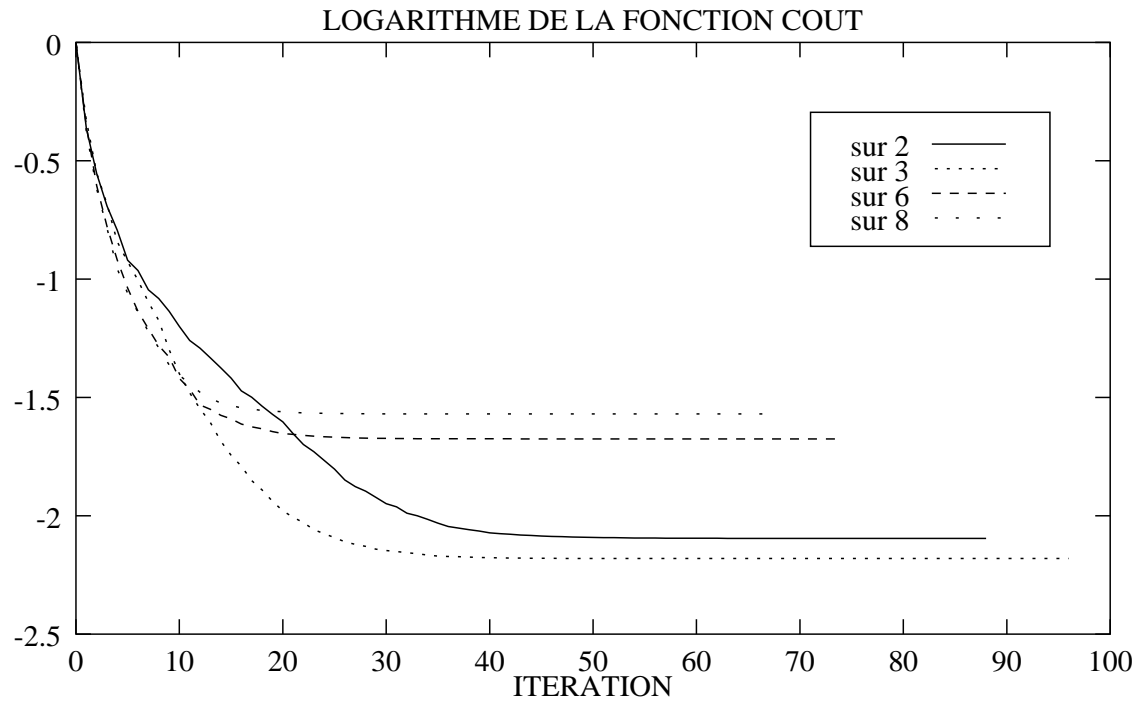
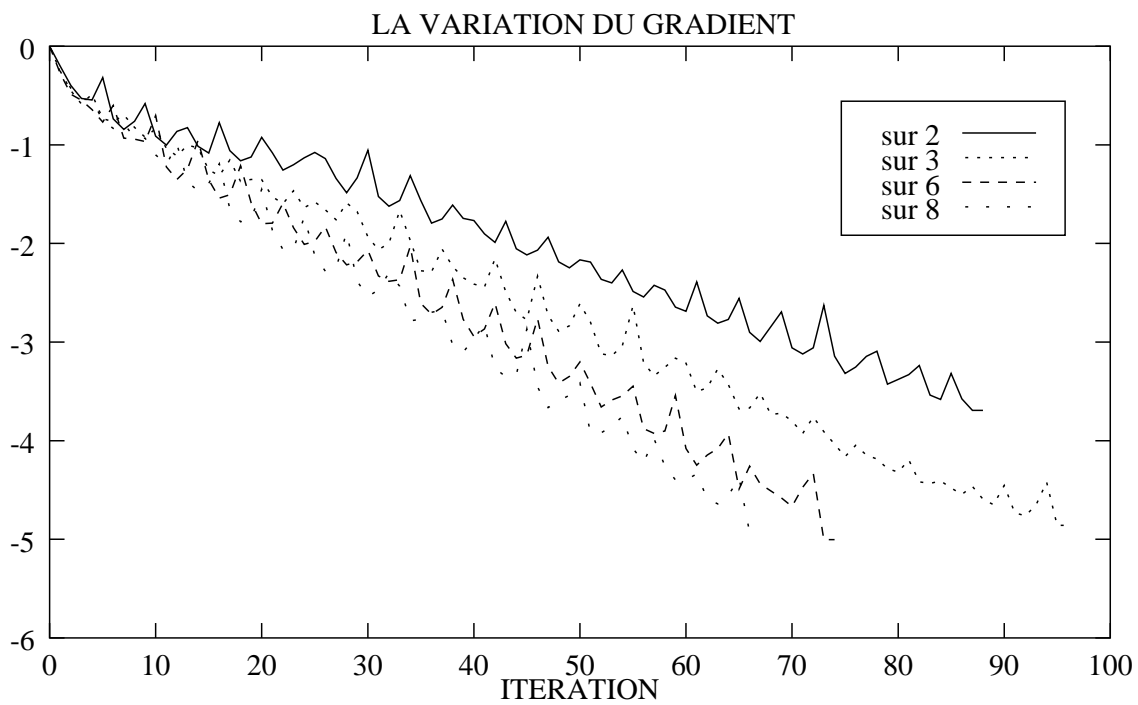


FIG. V.22 – *Expérience avec la base où le numéro du vecteur $N_m = 300 - 500$. La norme de l'erreur avant contrôle (à gauche) et de l'erreur optimale (à droite) en fonction du temps.*

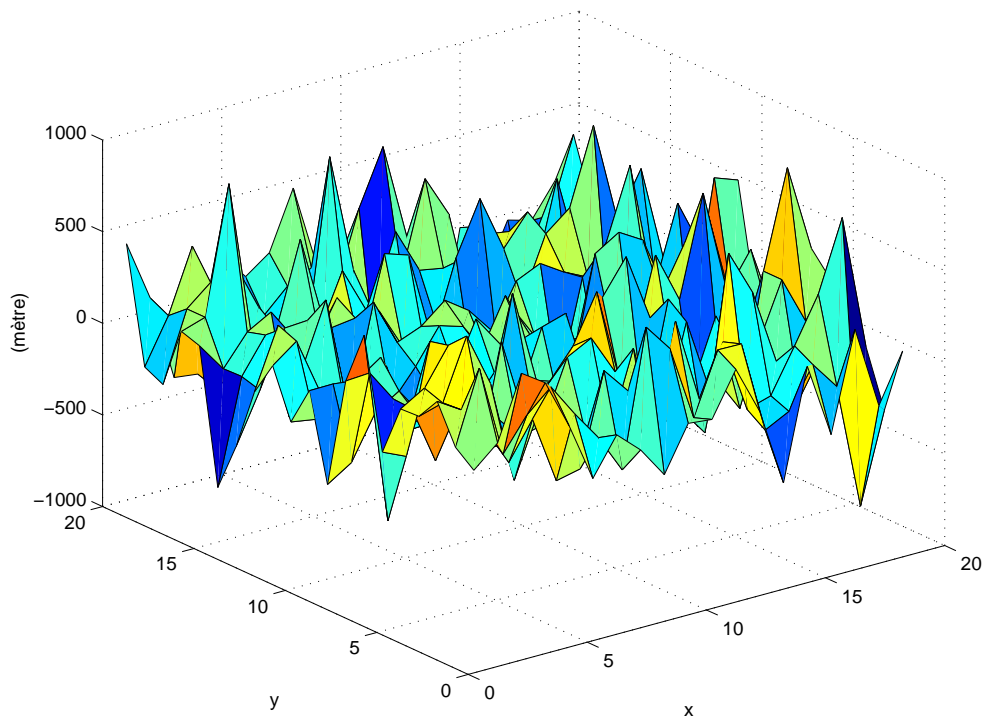


(a)

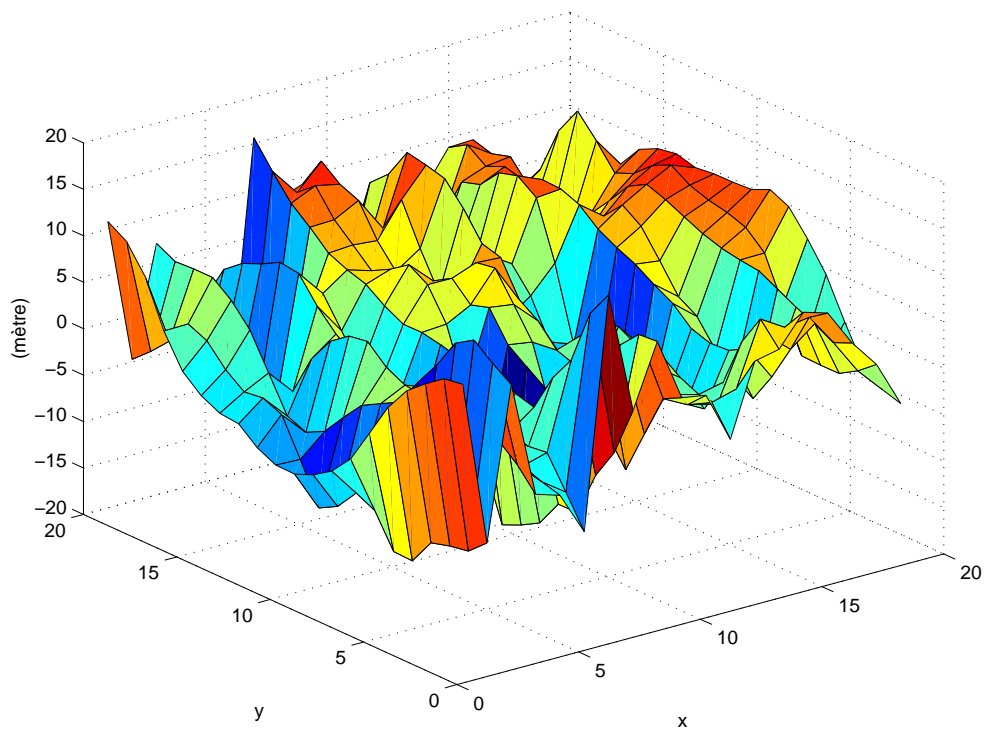


(b)

FIG. V.23 – Les comparaisons des variations de $\log(J_k/J_0)$ (a) et des variations de $\log(\|\nabla J_k\|/\|\nabla J_0\|)$ (b) en fonction des itérations k sans la réduction de taille et avec des bases différentes

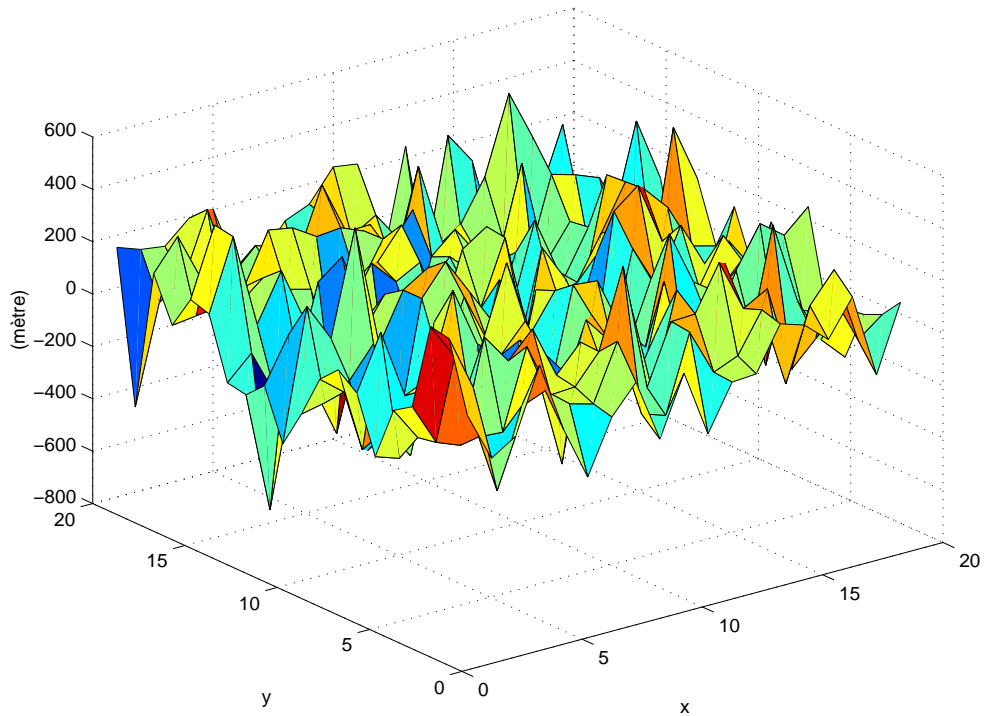


L'écart avant contrôle par rapport à l'observation du champ du potentiel au temps final

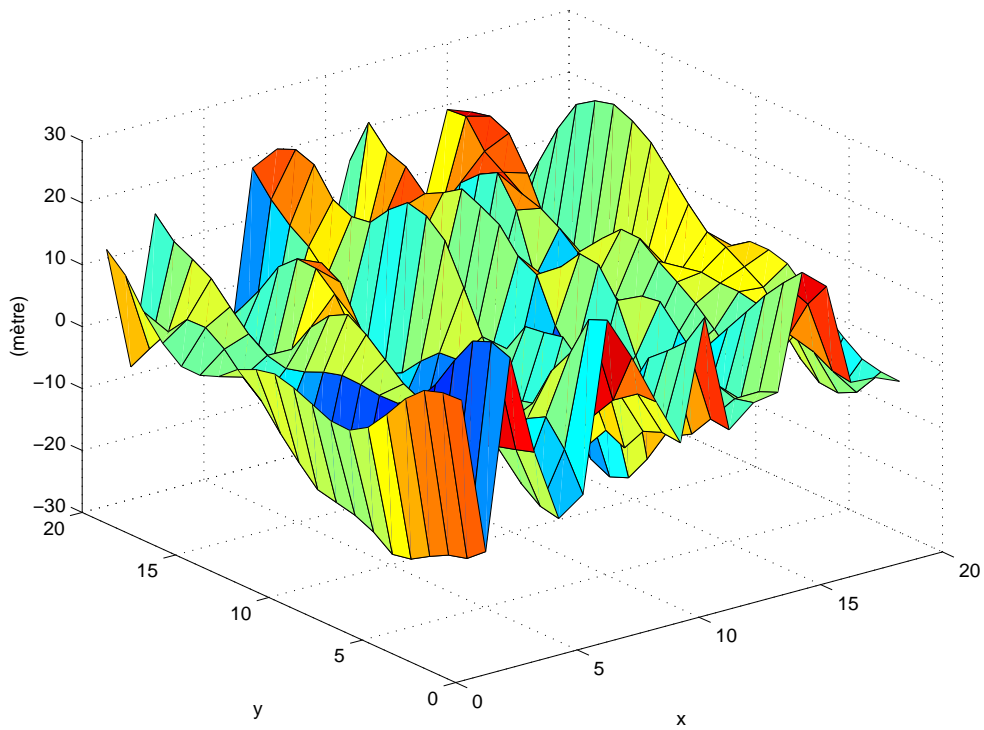


L'écart après contrôle par rapport à l'observation du champ du potentiel au temps final

FIG. V.24 – *Expérience avec le nombre d'espace de temps sur 2*

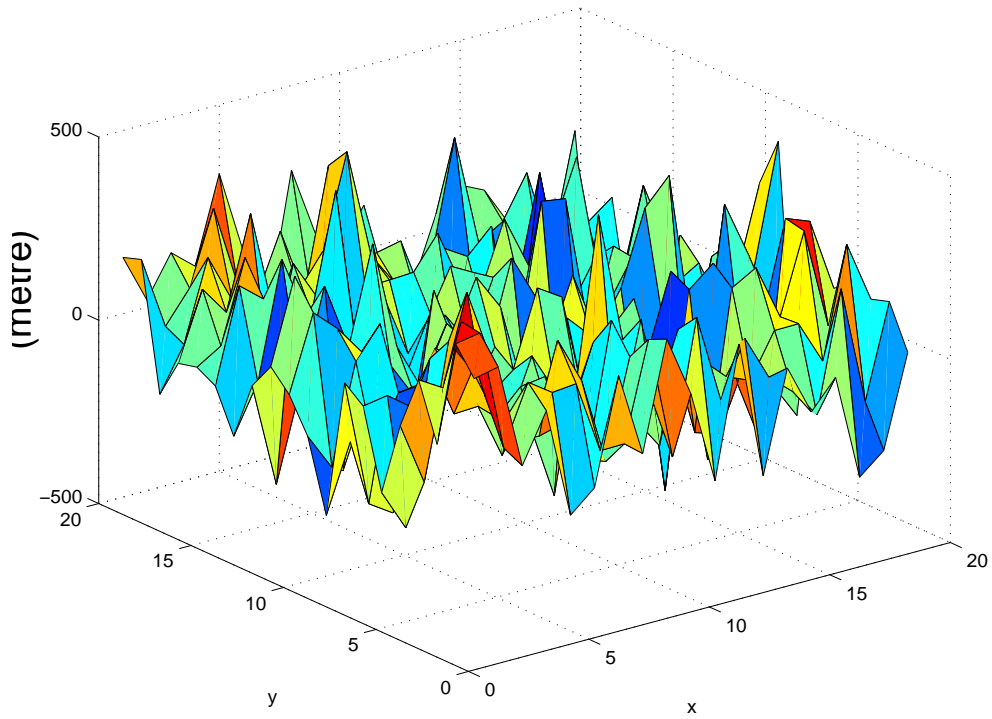


L'écart avant contrôle par rapport à l'observation du champ potentiel au temps final

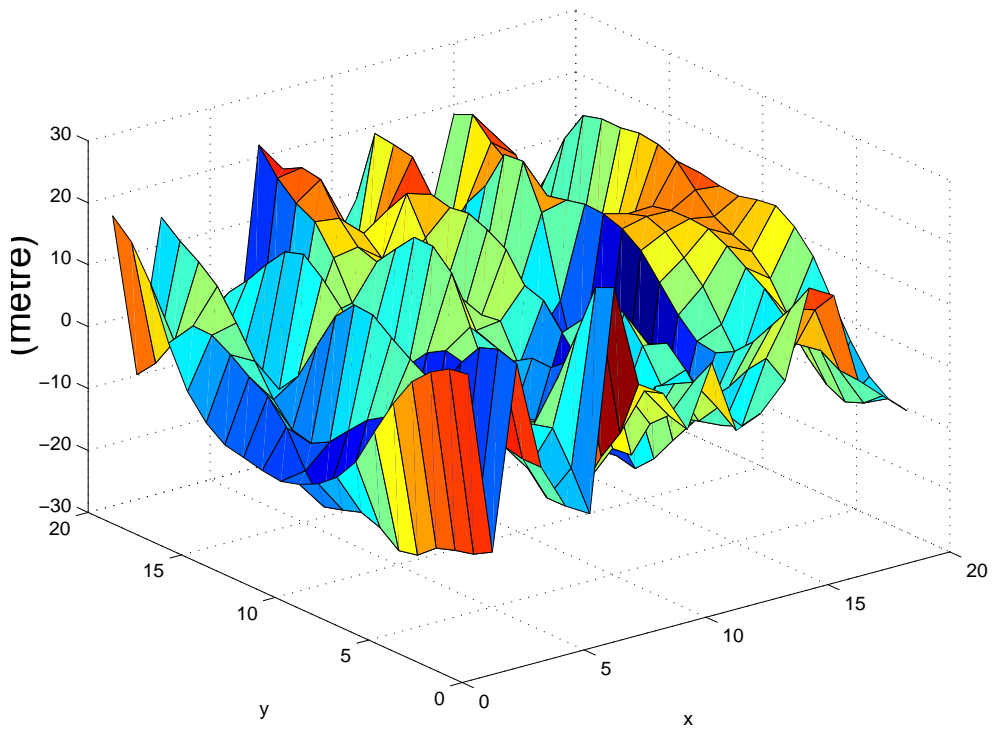


L'écart après contrôle par rapport à l'observation du champ du potentiel au temps final

FIG. V.25 – *Expérience avec le nombre d'espace de temps sur 3*

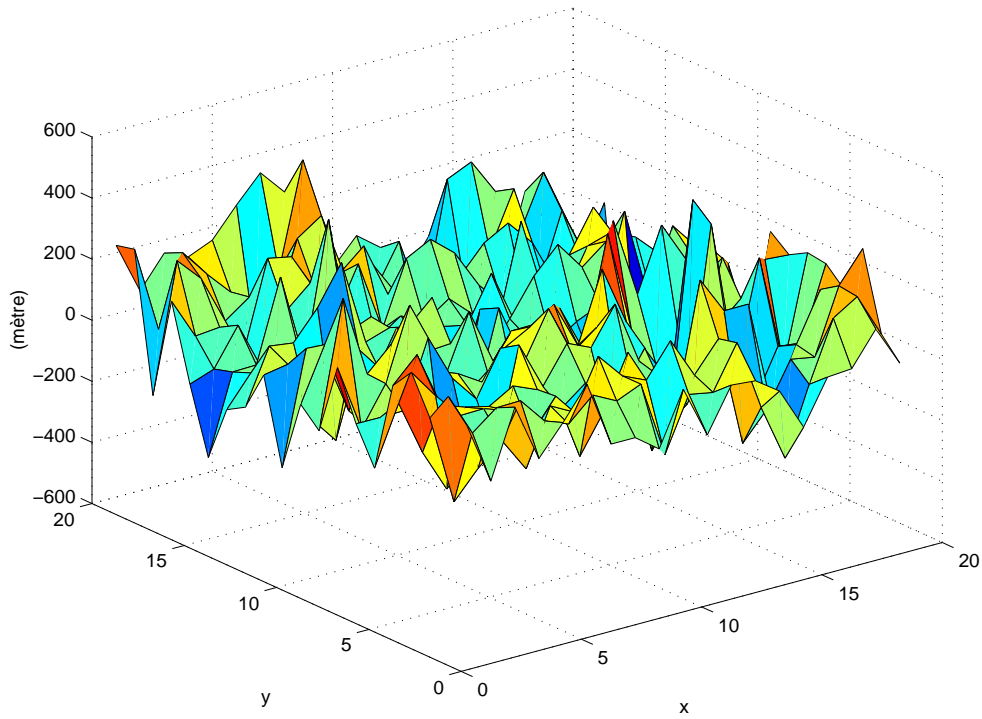


L'écart avant contrôle par rapport à l'observation du champ potentiel au temps final

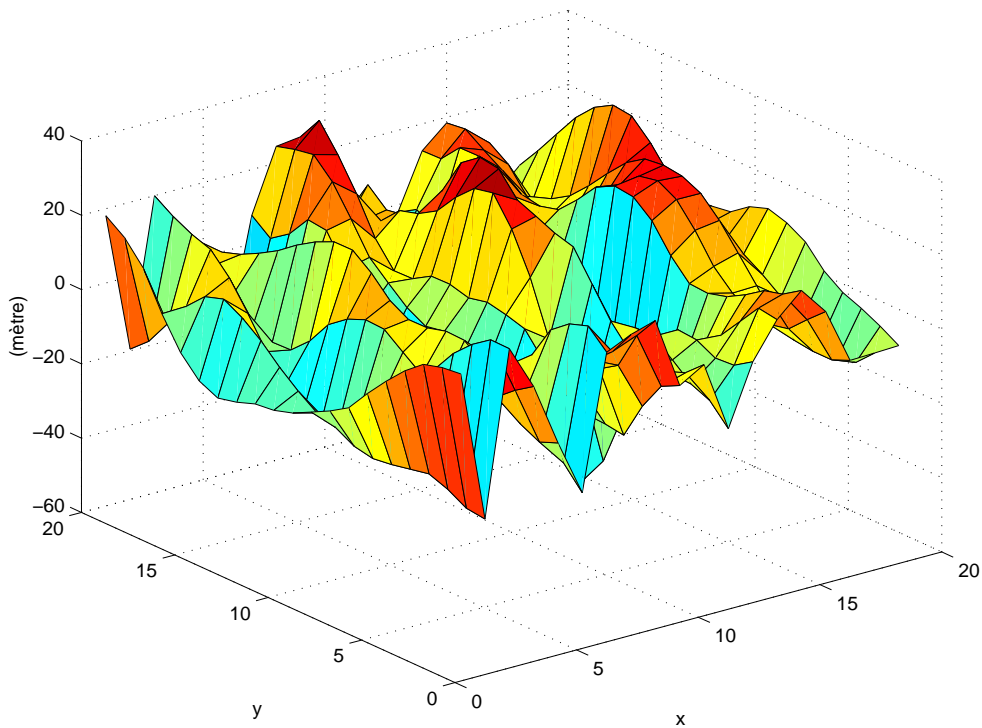


L'écart après contrôle par rapport à l'observation du champ du potentiel au temps final

FIG. V.26 – *Expérience avec le nombre d'espace de temps sur 6*



L'écart avant contrôle par rapport à l'observation du champ du potentiel au temps final



L'écart après contrôle par rapport à l'observation du champ du potentiel au temps final

FIG. V.27 – *Expérience avec le nombre d'espace de temps sur 8*

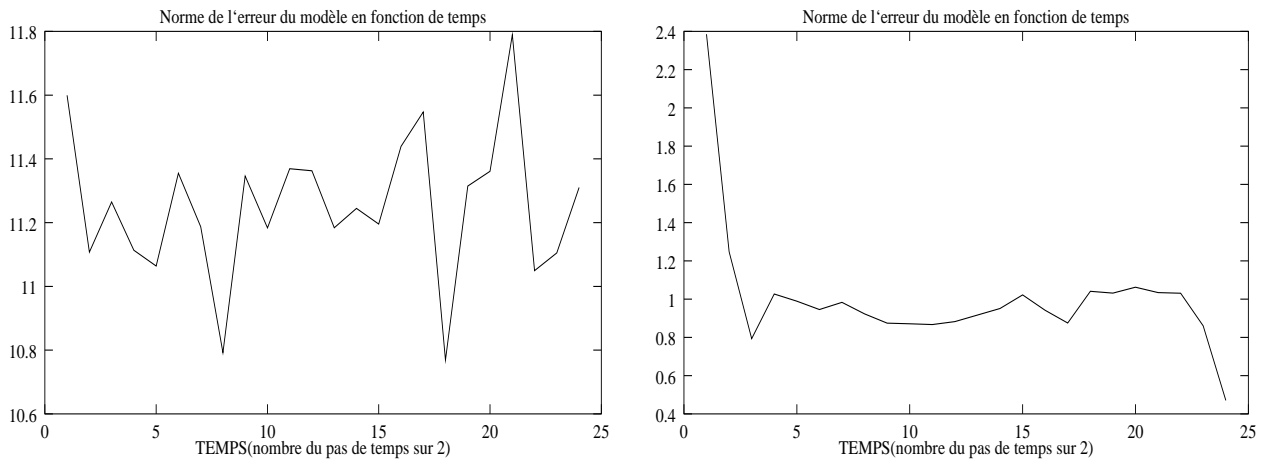


FIG. V.28 – *Expérience avec $M_s = 2$. La norme de l'erreur avant contrôle (à gauche) et de l'erreur optimale (à droite) en fonction du temps.*

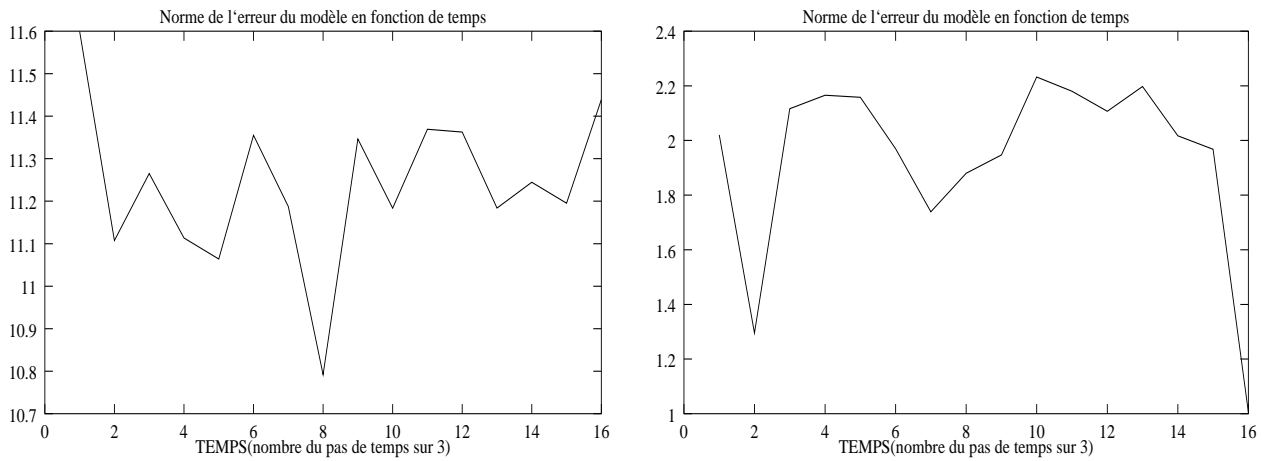


FIG. V.29 – *Expérience avec $M_s = 3$. La norme de l'erreur avant contrôle (à gauche) et de l'erreur optimale (à droite) en fonction du temps.*

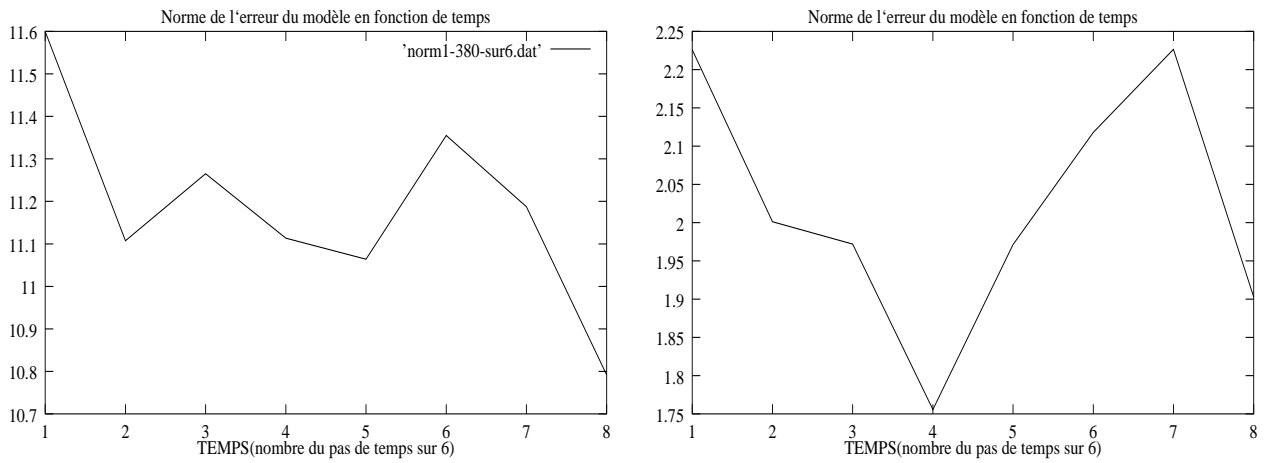


FIG. V.30 – *Expérience avec $M_s = 6$. La norme de l'erreur avant contrôle (à gauche) et de l'erreur optimale (à droite) en fonction du temps.*

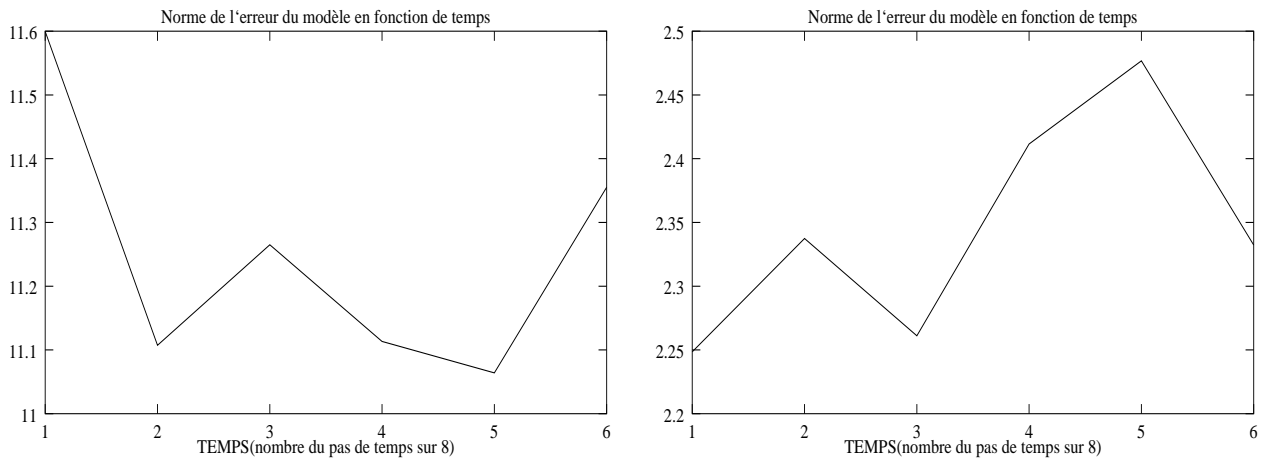


FIG. V.31 – *Expérience avec $M_s = 8$. La norme de l'erreur avant contrôle (à gauche) et de l'erreur optimale (à droite) en fonction du temps.*

Conclusion générale

Nous avons dans ce travail développé une procédure d'assimilation pour un modèle du transport des sédiments en rivière. La diminution notable de la fonction coût et de la norme de son gradient permet de conclure au succès de la méthode et à la validation du code du modèle adjoint. Avec ce code adjoint, on peut aussi étudier d'autres problèmes de contrôle optimal dans les travaux hydrauliques. Dans Zhu [76], le problème de contrôle optimal tenant compte de draguer des sédiments dans un canal de navigation et celui de la sensibilité de l'endroit de dragage ont été étudiés, en introduisant une mesure de la profondeur du canal de navigation.

Le grand inconvénient de la méthode variationnelle reste son coût de calcul important, due au stockage de la trajectoire du système dynamique que l'on a besoin pour la construction du code adjoint, à la grande taille de l'espace de minimisation.

L'expérience montre qu'il exige une analyse approfondie préalable de chaque problème particulier, dans un premier temps en vue d'éviter de stocker ou bien approcher tous les trajectoires de variables intermédiaires, et ensuite afin d'améliorer et d'adapter au cadre particulier des algorithmes d'optimisation pour assimiler afin d'assurer une bonne qualité de l'identification des variables de contrôle.

L'hypothèse sous jacente aux méthodes suppose que le modèle utilisé dans la méthode variationnelle est exact. On s'est intéressé enfin dans ce mémoire à l'estimation de l'erreur liée au modèle et on a pris en compte une étude dans la direction de recherche concernant la réduction de la taille de l'espace de contrôle dans le cas où une méthode est mise en oeuvre d'estimer des ingrédients incertains du contrôle de telle sorte que on les détermine avec le recours à des espaces fonctionnels engendrés par un groupe des vecteurs. Les vecteurs, trouvés par l'algorithme de Lanczos et l'application itérative du code linéaire tangent et adjoint, sont les directions le long desquelles les perturbations destinées au temps initial s'amplifient avec le taux de croissance le plus grand dans une période de temps limitée. Dans le cas d'une longue période de temps, au point de vue théorique, ces vecteurs tendent vers les vecteurs de Lyapunov locaux dominants, ce sera

intéressant d'identifier l'erreur du modèle dans l'espace engendré par ces derniers vecteurs.

Dans ce mémoire, on a travaillé avec un modèle simplifié qui s'adapte au problème à long terme. Les techniques sont espérées aux applications dans le cadre plus réel et pour des modèles plus complexes, par exemple, en granulométrie étendue, afin d'assurer une prévision plus précise de l'évolution du fond de la rivière dans le cas d'une structure compliquée.

Par exemple, un problème actuel très important à confronter est la prévision de l'écoulement des crues, où il se limite au transport solide en suspension, le problème peut être résolu par un couplage avec d'autres modèles en météologie ou en océanographie en introduisant d'autres sources des mesures comme la précipitations ou le mouvement de la marée.

Bibliographie

- [1] D. A.LYN, P. GOODWIN. Stability of a general PREISSMANN scheme. *Journal of Hydraulic Engineering*, 113(1):16–28, 1987.
- [2] A. BENNETT. *Inverse Methods in Physical Oceanography*. 346 pp. Cambridge University Press, 1992.
- [3] A. BENNETT, P. MCINTOSH . Open ocean modeling as an inverse problem-tidal theory. *J. Phys. Oceanogr.*, 12 : 1004–1018, 1982.
- [4] R. BUIZZA. Localization of optimal perturbation using a projection operator. *Q. J. R. Meteorol. Soc.*, 120:1647–1681, 1994.
- [5] R. BUIZZA, J. TRIBBIA, F. MOLTENI, T. PALMER. Computation of optimal unstable structures for a numerical weather prediction model. *Tellus*, 45A:388–407, 1993.
- [6] R. BUIZZA, T. PALMER. The singular vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, 52:1434–1456, 1995.
- [7] W. CHAO, L. CHANG . Development of a four-dimensional variational system using the adjoint method at GLA. part I: Dynamics. *Mon. Wea.Rev.*, 120(1661-1673), 1992.
- [8] J. P. CHOLLET. *Ecoulement non permanent sur fond mobile de rugosité instationnaire–Modèle mathématique*. PhD thesis, Université Science et Médecine de Grenoble (France), 1977.
- [9] R. L. COOLEY, S. A. MOIN. Finite element solution of Saint-Venant equations. *Journal of the hydraulics division, ASCE*, 102(HY6):759–775, 1976.
- [10] P. COURIER, O. TALAGRAND. Variational assimilation of meteorological observations with the direct and adjoint Shallow-Water equations. *Tellus*, 42A:531–549, 1990.
- [11] P. COURTIER. *Application du contrôle optimal à la prévision numérique en météorologie*. PhD thesis, l'Université Paris VI, 1987.

- [12] J. A. CUNGE, F. M. HOLLY. *Practical aspects of computational river hydraulics*. Pitman, London, 1980.
- [13] J. A. CUNGE, N. PERDREAU. Modèle mathématiques fluviaux à fond mobile. méthodes et formulation numérique pour les programmes char2 et char3. Technical report, SGREAH NT 1676, 1971.
- [14] J. A. CUNGE, N. PERDREAU. Mobile bed fluvial mathematical models. *La Houille Blanche*, 7, 1973.
- [15] R. DALEY. Atmospheric data analysis. *Cambridge Atmospheric and Space Science Series*, eds. Cambridge University Press, 1991.
- [16] S. DAS , R.W. LARDNER . On the estimation of parameters of hydraulic models by assimilation of periodic tidal data. *J. geophys. Res.*, 96 : 15187–15196, 1991.
- [17] J.C. DERBER. *The variational 4-D assimilation of analysis using filtered models as constrains*. PhD thesis, University of Wisconsin-madison, 1985.
- [18] M. DE VRIES. Consideration about non-stead. In *papier 3. 8. 11e Congrès de l'AI RH*, Leningrad, 1965.
- [19] Wuhan Institute of Hydraulic Division of River Sediment Engineering and Electric Engineering. *River Sediment Engineering* (in Chinese). Beijing Hydraulic Press, 1982.
- [20] J. DOUGLAS . The application of stability analysis in the numerical solution of quasi-linear parabolic differential equations. *Trans. Am. Soc.*, 89:484–518, 1958.
- [21] J. DOUGLAS , H. H. RACHFORD. On the numerical solution of heat conduction in two and three space variables. *Trans. Amer. Math. Soc.*, 82:421–439, 1956.
- [22] J. DOUGLAS, J. GUNN. A general formulation of alternating direction methods. *Num. Math.*, 6:428–453, 1964.
- [23] H. A. EINSTEIN. *The bed load function for sediment transportation in open channels*. Technical Bulletin 1026, US dept. of agriculture, Soil Conservation Service Washington, 1950.
- [24] F. ENGELUND, E. HANSEN. *A Monograph on sediment transport in alluvial streams*. Teknisk Forlag Copenhagen, 1967.
- [25] A. F. FILIPPOV. On stability of difference equations. *Doklady AN SSSR*, 100(6), 1955.

- [26] M. GHIL, P. MALANOTTE-RIZZOLI . Data assimilation in meteorology and oceanography. *Adv. Geophys.*, 33:141–266, 1990.
- [27] J. C. GILBERT, G. LE VEY, J. MASSE . La différentiation automatique de fonctions représentées par des programmes. Technical Report 1557, recherche INRIA, 1991.
- [28] S. K. GODUNNOV, K. A. SEMENDJAEV. On the numerical solution of heat conduction in two and three space variables. *Trans. Amer. Math. Soc.*, 82:421–439, 1956.
- [29] G. H. GOLUB, C. F. VAN LOAN. *Matrix Computations*. The Johns Hopkins University Press, 1983.
- [30] A. GRAMMELTVEDT. A survey of finite-differences schemes for the primitive equations for a barotropic fluid. *Mon. Wea. Rev.*, 97:387–404, 1969.
- [31] A. GRIEWANK, G. F. CORLISS. Workshop on automatic differentiation of algorithms. In SIAM proceeding series, editor, *Automatic differentiation of algorithms: Theory implementation and application*, 1991.
- [32] JR. HARLAND , J.J. O'BRIEN. Assimilation of scatterometer winds into surface pressure field using a variational method. *J. Geophys. Res.*, 91 : 7816–7836, 1986.
- [33] P. L. HOUTEKAMER. Variation of the predictability in a low order spectral model of the atmospheric circulation. *Tellus*, pages 177–190, 43A.
- [34] M. F. KARIM, J. F. KENNEDY. IALLUVIAL: A computer based flow and sediment routing model for alluvial streams and its application to the Missouri river. Technical Report IIHR report 250, University of Iowa, USA, 1982.
- [35] J. F. LACARRA, O. TALAGRAND. Short range evolution of small perturbation in a barotropic model. *Tellus*, 40A:81–95, 1988.
- [36] R. W. LARDNER, S. K. DAS. Optimal estimation of eddy viscosity for a quasi-three-dimensional numerical tidal and storm surge model. *Inter. J. for Numer. Methodes in Fluids*, 18:295–312, 1994.
- [37] P. LASCAUX, R. THEODOR. *Analyse numérique matricielle appliquée à l'art de l'ingénieur*. Tome II. MASSON, 1993.
- [38] F. X. LEDIMET, O. TALAGRAND. Variational algorithms for analysis and assimilation of meteorological observation:theoretical aspects. *Tellus*, 38A:97–110, 1986.
- [39] F.X. LEDIMET, O. TALAGRAND. Variational and optimization methods in meteorology - A view. Technical report, Supercomputer Computations Research Institute, Florida State University, Tallahassee, 1989.

- [40] J. M. LEWIS, J.C. DERBER. The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, 37A : 309–322, 1985.
- [41] J. L. LIONS. *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles*. Dunod, 1968.
- [42] G. I. MARCHUK. On the theory of the splitting-up method. In *Numerical Solutions Partial Differential Equations II, SYNSPADE 1970*. Allerton Press Inc., New York-London : Academic Press, 1971.
- [43] G. I. MARCHUK. *Methods of Numerical mathematics*. Springer-Verlag, 1982.
- [44] M. G. MARCHUK. *Problème de prévision météorologique*. Leningrad (traduction française, Armand Colin, 1968), 1967.
- [45] E. MEYER-PETER, R. MULLER. Formula for bed load transport. In *Proceeding of 2nd Congress, Intern. Assoc. Hyd. Res.*, Stockholm, 1948.
- [46] A. M. MOORE. Data assimilation in a quasi-geostrophic open ocean model of the gulf stream region using the adjoint method. *J. of Physical Oceanography*, 21:398–427, 1991.
- [47] I. M. NAVON. A review of variational and optimization methods in meteorology. In Y.K. Sasaki, editor, *Festival Volume of the international Symposium on Variational Methods in Geosciences*. Elsevier Science Pub. Co. Developments in Geomathematics, 1986.
- [48] I. M. NAVON, X. ZOU, J. DERBER, J. SELA. Incomplete observation and control of gravity waves in variational data assimilation. *Tellus*, 44A:273–296, 1992.
- [49] V. PANCHANG, J.E. RICHARDON . Inverse adjoint estimation of eddy viscosity for coastal flow models. *J. Hydr. Eng.*, 119 : 506–524, 1993.
- [50] V. PANCHANG, J.J. O'BRIEN . On the determination of hydraulic model parameters using adjoint state formulation. *Modeling Marine Syst.*, I : 5–18, 1988.
- [51] N. PERDREAU, J. A. CUNGE. Modèle mathématique et modèle physique fluvial à fond mobile. Technical report, SOGREAH NT 1657, 1971.
- [52] J. L. RAHUEL. *Modélisation de l'évolution du lit des rivières alluvionnaires à granulométrie étendue*. PhD thesis, INPG, 1988.
- [53] R. D. RICHTMYER, K.W. MORTON. *Difference Methods for Initial Value Problems*. New York : Wiley, 1967.

- [54] N. ROSTAING, E. HASSOLD . Basic functional representation of programs for automatic differentiation in the odyssee system. In *Proceeding of the workshop on High Performance Computing in the Geosciences*. NATO ASIE series, 1994.
- [55] N. ROSTAING, S.DALMAS, A. GALLIGO . Automatic differentiation in odyssee. *Tellus*, 45A : 558–568, 1993.
- [56] A. A. SAMARSKII. On convergence of the fractional steps methods for the heat equation. *Zh. Vych. Matem. Matem. Fiz.*, 2(6), 1962.
- [57] I. SZUNYOGH, E. KALNAY, Z. TOTH . A comparison of Lyapunov and optimal vectors in a low-resolution GCM. *Tellus*, 49A:200–227, 1997.
- [58] O. TALAGRAND, P. COURTIER. Variational assimilation of meteorological observation with the adjoint vorticity equation - part I. theory. *Q. J. R. Meteorol. Soc.*, 113:1311–1328, 1987.
- [59] R. TEMAM. Sur la stabilité et la convergence de la méthode des pas fractionnaires. *Thèse, Paris VI, 1967, Annali di Mat.*, 1965.
- [60] W. C. THACKER. Fitting models to inadequate data by enforcing spatial and temporal smoothness. *J. Geophys. Res.*, 93:10655–10665, 1988.
- [61] W.C. THACKER, R.B. LONG. Fitting dynamics to data. *J. Geophys. Res.*, 93:1227–1240, 1988.
- [62] W. A. THOMAS, A. L. PRASUHN. Mathematical modeling of scour and deposition. *Journal of Hydraulic Engineering, ASCE*, 109(4), 1977.
- [63] Z. TOTH, E. KALNAY. Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Met. Soc.*, 74:2317–2330, 1993.
- [64] Z. TOTH, E. KALNAY. Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, 1996. submit.
- [65] A. A. TSONIS. *Chaos : from theory to applications*. Plenum Press, New York, 1966.
- [66] J. P. VILA. *Sur la théorie et l'approximation numérique de problèmes hyperboliques non linéaires, applications aux équations de Saint-Venant et à la modélisation des avalanches de neige dense*. PhD thesis, Paris VI, 1986.
- [67] T. VUKICEVIC. Possibility of skill forecasts based on the finite-time dominant linear solutions for a primitive equation regional model. *J. Atmos. Sci.*, 50, 1993.

- [68] Z. WANG, I. M. NAVON, F. X. LEDIMET, X. ZOU . The second order adjoint analysis : Theory and applications. *Meteorol. Atmos. Phys.*, 50:3–20, 1992.
- [69] N. N. YANENKO. On convergence of the splitting-up method for the heat equations with variable coefficients. *Zh. Vych. Matem. Matem. Fiz.*, 1962.
- [70] N. N. YANENKO. *Méthode des pas fractionnaires*. Novosibirsk(traduction française, Armand Colin,1968), 1966.
- [71] J. Q. YANG, F. X. LE DIMET. Variational data assimilation in the transport of sediment in rivers. *Science in China (Series D)*, 41(5):pp 473–485, 1998.
- [72] H. C. YEE. Computational fluid dynamics. A class of high-resolution explicit and implicit shock-capturing methods. Technical report, NASA Ames Research Center, USA, 1989.
- [73] Q. C. ZENG, D.J. GUO, R. F. LI. Simulation of sediment and delta evaluation. *Natural scientific development*, 5(3), 1995.
- [74] Q. C. ZENG, D.J. GUO, Z. LIU, J. ZHU. Numerical modeling of silt sedimentation and relevant engineering problems. *Int. J. Sediment Res.*, 11(1):22–37, 1996.
- [75] Q. C. ZENG, Z.Z. JI, R.L. LI. The construction of difference schemes for evolution equations and the numerical simulation of offshore currents. *Atmospheric sciences*, pages 166–175, 1988.
- [76] J. ZHU, Q. C. ZENG, D. J. GUO. Adjoint sensitivity analysis of a mathematical model of sedimentation. *Chinese Jour. Sediment Res.*, 1998.
- [77] X. ZOU, I. M. NAVON, J. SELA. Control of gravitational oscillations in variational data assimilation. *Mon. Wea. Rev.*, 121 : 272–289, 1993.

Résumé

La prévision de la sédimentation d'une rivière requiert l'utilisation d'un modèle mathématique régissant l'écoulement et de données d'observation. Le but de ce travail est de proposer une méthode d'assimilation de données qui permet de reconstituer les champs en tenant en compte du modèle et des données d'observation. La méthode qui est proposée est fondée sur les techniques de contrôle optimal.

On présente les problèmes de sédimentation et leurs approximations numériques, un algorithme de décomposition est introduit et sa convergence est étudiée. En préalable à l'exploitation à des problèmes réels, on a vérifié la faisabilité de la méthode variationnelle d'assimilation de données pour trois types de problèmes de transport des sédiments : 1) la détermination du champ initial, 2) l'identification des paramètres, 3) l'estimation de l'erreur de la modélisation.

Les études de sédimentation sur le terrain conduisent à des problèmes numériques de très grande dimension, dans une dernière partie on s'intéresse à des techniques permettant la réduction de l'espace de contrôle pour obtenir des problèmes d'une taille raisonnable.

Mots clés : assimilation de données variationnelle, équations de Saint-Venant, sédimentation en rivière, optimisation, méthode adjointe, identification des paramètres, méthode de time-slitting, estimation d'erreur du modèle, vecteurs singuliers.

Abstract

The prediction of the sedimentation in rivers requires the utilization of mathematical models governing the flow and observational data. The objective of this work is to propose a data assimilation method to retrieve physical fields by combining the model, and the observation. This method is based on optimal control techniques.

We present sedimentation problems and their numerical approximations, as well as a splitting algorithm, and a study on convergence is carried out. Before working on real problems, variational data assimilation methods are developed and tested for their feasibility for three types of sediment transport problems : 1) determination of initial field, 2) identification of parameters, 3) estimation of model errors.

As the studies on real sedimentation fields will lead to numerical problems of great dimensionality, we are interested, in the last part, in the techniques of reducing the size of the control space in order to obtain problems with reasonable size.

Keywords : variational data assimilation, shallow-water equations, sedimentation in rivers, optimization, adjoint method, parameter identification, time-splitting method, model error estimation, singular vectors.