

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (INP Toulouse)

Discipline ou spécialité :

Surfaces Interfaces Continentales Hydrologie

Présentée et soutenue par :

M. NABIL EL MOÇAYD

le mercredi 1 mars 2017

Titre :

La décomposition en polynôme du chaos pour l'amélioration de l'assimilation de données ensembliste en hydraulique fluviale

---

Ecole doctorale :

Sciences de l'Univers de l'Environnement et de l'Espace (SDUEE)

Unité de recherche :

Centre Européen de Recherche et Formation Avancées en Calcul Scientifique (CERFACS)

Directeurs de Thèse :

M. OLIVIER THUAL

MME SOPHIE RICCI

Rapporteurs :

M. DRISS OUAZAR, ECOLE DES MINES DE RABAT

M. JACQUES SAINTE-MARIE, CEREMA

Mme CLEMENTINE PRIEUR, UNIVERSITE GRENOBLE ALPES

Membres du jury :

Mme CLEMENTINE PRIEUR, UNIVERSITE GRENOBLE ALPES, Présidente

M. CEDRIC GOEURY, EDF CHATOU, Membre

M. OLIVIER THUAL, INP TOULOUSE, Membre

M. SYLVAIN BIANCAMARIA, LEGOS, Membre

À ma femme Kaoutar , à mes parents, à mon fils Imrane.

Remerciements :

Un travail de thèse est sans aucun doute une aventure humaine avant d′être scientifique. Pour ma part, cette aventure m′a permis de croiser le chemin de beaucoup de gens qui minspiré dmanière ou dune autre. Je n′aurai pas pu arriver au bout de ce travail sans leur générosité et leur soutien inconditionné.

Je tiens tout d′abord à remercier mon directeur de thèse, monsieur Olivier THUAL, pour la confiance quil m′a accordée en acceptant d′encadrer ce travail doctoral, pour ses multiples conseils et pour toutes les heures qu′il a consacrées à diriger cette recherche. J′aimerais également lui dire à quel point j′ai apprécié sa grande disponibilité et son respect sans faille des délais serrés de relecture des documents que je lui ai adressés.

Je souhaiterai aussi remercier ma co-directrice de thèse, Madame Sophie RICCI, pour sa sympathie, sa confiance qu′elle m′a accordée, et sa patience avec moi. J′ai apprécié toutes les discussions scientifiques et la grande liberté qu′elle m′a accordée pour gèrer mon travail de thèse. Enfin, j′ai été extrêmement sensible à ses qualités humaines d′ écoute et de compréhension tout au long de ce travail doctoral.

Mes remerciements vont aussi à Clémentine PRIEUR, Jacques SAINTE-MARIE et Driss OUAZAR qui ont accepté, avec beaucoup de générosité, la fonction de rapporteurs de thèse. Leurs commentaires et leurs critiques m′ont été d′une grande aide afin d′envisager mon travail sous un autre angle.

Je souhaiterai aussi exprimer ma gratitude à Sylvain BIANCAMARIA et à Cédric GOEURY d′avoir accepter de faire partie du jury de thèse. Je les remercie pour les échanges scientifiques et le temps consacré à mon travail.

Je remercie le CNES d′avoir financé cette thèse. M ais aussi EDF R&D qui a non seulement participé au financement de cette thèse, mais mà aussi permis de travailler dans ses locaux pendant trois mois. Mon séjour à Chatou était très intéressant d?un point de vue scientifique mais aussi humain. Il m′a permis de croiser des gens que je tiens à remercier : Riad Ata et Kamal El Kadi Abderrezzak.

Durant ces trois ans de thèse, j′ai pu aussi profiter de l′encadrement des personnes qui méritent que je remercie : Merci à Nicole GOUTAL, Sébastien BOYAVAL et Didier LUCOR. Ils ont pu répondre à chacune de mes questions, aussi absurdes soient elles avec la plus grande générosité.

Un travail de thèse est aussi une opportunité de travailler chaque jour avec des gens qui font que le laboratoire soit un lieu agréable à vivre, même durant les périodes rudes. Je tiens donc à remercier toutes les personnes du laboratoire qui ont pu apaiser les temps difficiles : Milka, Laurent, Marie-pierre, Odile, Paul-Arthur, Xavier, Emilia, Mèlanie, Anthony, Youssef, Frans, Marie, Vita, Catherina, Mathias, Thierry, Christophe, Eric, Christian, Rym, Philippe,... Le laboratoire de recherche constitue aussi un endroit où on peut croiser beaucoup de doctorants qui font que la douteur dthèse soit partagée, et pour ce soulagement commun je les remercie : Sébastien, Elizabeth, Gildas, Margot, Thomas, Johan, Yohan, Said, Rémy, Marie, Antoine,.... J‘aimerai aussi dire Merci à celles qui nous supportent moralement mais aussi administrativement et qui font que notre travail de thèse soit allégé des taches

administratives : Isabelle, Nathalie et Chantal.
Durant ce travail de thèse j′ai été aussi supporté par beaucoup de gens qui ont cru en moi (sans forcnt connae le thème de mon travail). Je tiens donc à dire merci à Stéphane, Rachel, l′equipe JKA Toulouse, l′equipe Aqua Ionique, Abdellatif Sekairiba, Saad Sekairiba, Hassan Idrissi, Said Abidi, Najoua Hnach.
Finalement je ne pourrai pas oublier de remercier mon support inconditionnel, la raison qui fait que j′arrive au bout de cette thèse, et ma motivation de chaque jour : Ma famille. Je remercie donc mon épouse de m′avoir supporté, mon enfant qui vient d′arriver, mes oncles qui ont toujours cru en moi, mes frères pour leur soutien, ma mère pour sa bienveillance et mon père qui est ma source dinépuisable.
Je remercie aussi toutes personnes qui ne voient pas leur noms apparaitre dans cette liste mais qui le méritent.

"La connaissance est une navigation dans un océan d'incertitudes à travers des archipels de certitudes." Edgar Morin

Table des matières

Table des matières v

Table des figures ix

I Introduction 1

II Hydraulique fluviale 13

1 Équations de Saint-Venant 15
1.1 Équations de Navier-Stokes . . . 15
1.1.1 Formulation des équations . . . 15
1.1.2 Équations de Navier-Stokes incompressibles . . . 16
1.1.3 Conditions aux limites . . . 16
1.2 Dérivation des équations de Saint-Venant . . . 18
1.2.1 Equations de Saint-Venant . . . 18
1.2.2 Equations de Saint-Venant filaires . . . 19
1.2.3 Courbes caractéristiques . . . 20

2 Modélisation numérique des équations de Saint-Venant 23
2.1 Introduction . . . 23
2.2 Volumes finis . . . 23
2.3 Solveur de l'approximation de Riemann de Roe . . . 24
2.4 Termes sources . . . 25
2.5 Conditions aux limites . . . 26

3 Simplification des équations de Saint-Venant 27
3.1 Introduction . . . 27
3.2 Classification des régimes d'écoulement . . . 27
3.3 Régime stationnaire . . . 28
3.3.1 Nombre de Froude . . . 29
3.3.2 Régime uniforme . . . 29
3.3.3 Courbes de remous (régime stationnaire graduellement varié) . . . 30
3.3.4 Régime brusquement varié . . . 31

3.4 Régime instationnaire . . . 32
3.4.1 Équations de Saint-Venant uni-dimensionnelles . . . 32
3.4.2 Approximation ondes de crues . . . 33
3.4.3 Ondes de crues diffusantes . . . 33

III Quantification d'incertitudes 35

4 Généralités sur les méthodes stochastiques en quantification d'incertitudes et en analyse de sensibilité **41**
4.1 Variable aléatoire et espace de probabilité associé . . . 41
4.1.1 Définitions . . . 41
4.1.2 Propriétés . . . 42
4.1.3 Moments statistiques . . . 42
4.1.4 Formulation d'un problème stochastique . . . 43
4.1.5 Estimation de la densité de probabilité . . . 45
4.2 Dépendance et probabilité conditionnelle . . . 46
4.2.1 Notion de dépendance . . . 46
4.2.2 Probabilité conditionnelle . . . 46
4.2.3 Théorème de Bayes . . . 46
4.3 Vecteur aléatoire et espace de probabilité associé . . . 47
4.3.1 Définitions . . . 47
4.3.2 Distribution . . . 47
4.3.3 Moments statistiques . . . 48
4.4 Analyse de sensibilité . . . 49
4.4.1 Généralités . . . 49
4.4.2 Analyse de sensibilité globale (ANOVA) . . . 50
4.4.3 Estimation des indices de Sobol' . . . 52

5 Réduction de dimension de l'espace incertain **53**
5.1 Introduction . . . 53
5.2 Echantillonnage d'un champs aléatoire . . . 53
5.2.1 Réalisations d'un processus stochastique . . . 53
5.2.2 Échantillonnage d'un processus stochastique . . . 54
5.2.3 La décomposition de Karhunen-Loève . . . 54
5.2.4 Cas particulier du processus gaussien . . . 55
5.3 POD "*Proper Orthogonal Decomposition*" . . . 56
5.3.1 Généralités . . . 56
5.3.2 Méthode de construction de la POD . . . 56

6 Polynômes du chaos **59**
6.1 Généralités . . . 59
6.2 Cadre d'approximation polynomiale . . . 60
6.2.1 Contexte probabiliste . . . 60

6.2.2 Base d'approximation 61
6.2.3 Troncature 62
6.3 Calcul des coefficients polynomiaux 65
6.3.1 Méthodes intrusives 65
6.3.2 Méthodes non-intrusives 66
6.4 Quantification des incertitudes et analyse de sensibilité 73
6.5 Limitations de l'approximation polynomiale 74
6.5.1 Polynômes du chaos en grande dimension 74
6.5.2 Polynômes du chaos en présence de discontinuité 78
6.5.3 Polynômes du chaos pour les simulations longues 80

7 L'assimilation de données 81
7.1 Introduction 81
7.2 Le vocabulaire de l'assimilation de données 81
7.3 La modélisation des erreurs et les matrices de covariance d'erreur
82 7.3.1 Remarques préliminaires sur les erreurs, les matrices de covariance et de corrélation d'erreur 83
7.3.2 Les erreurs liées à l'état du modèle 83
7.3.3 Les erreurs liées aux observations 84
7.4 Le Best Linear Unbiased Estimator - BLUE 84
7.5 Le filtre de Kalman et ses dérivations 86
7.5.1 Le filtre de Kalman 86
7.5.2 Le filtre de Kalman Etendu ou Extended Kalman Filter (EKF) 87
7.5.3 Le filtre de Kalman d'ensemble - EnKF 87
7.6 Vers l'EnKF basé sur la décomposition en chaos polynomial . . . 88
7.6.1 Limitations de l'EnKF 88
7.6.2 Amélioration de l'EnKF à l'aide de la décomposition en polynômes du chaos 88
7.6.3 Limitations de l'approximation polynomiale dans le cadre de l'EnKF 89

IV Quantification d'incertitudes en hydrauliques 91

8 Besoin de quantification d'incertitudes dans le contexte hydraulique 93
8.1 Introduction 93
8.2 État de l'art 94
8.2.1 La condition à la limite en débit 94
8.2.2 Paramètres du modèle 94
8.3 Outils utilisés dans la thèse 95
8.3.1 Modélisation hydraulique 95
8.4 Quantification d'incertitudes 96
8.5 Couplage modélisation hydraulique et outils de quantification d'incertitudes 96

9 Décomposition en polynômes du chaos en régime stationnaire **97**
9.1 PCE dans le cas permanent . . . 97
9.1.1 Résumé . . . 97
9.1.2 Article soumis dans la revue Environmental Modeling and Assessment . . . 98
9.1.3 Discussion . . . 139
9.2 PC-EnKF pour l'assimilation de données SWOT . . . 143
9.2.1 Introduction . . . 143
9.2.2 Génération des observations de type SWOT . . . 143
9.2.3 Correction de paramètres à l'aide de PC-EnKF . . . 144
9.2.4 Algorithme de correction de paramètres . . . 144
9.2.5 Comparaison entre l'EnKF et le PC-EnKF pour la correction des paramètres de frottement et de débit . . . 145
9.2.6 Impact des caractéristiques des données SWOT sur la réanalyse . . . 149

10 PCE dans le cas instationnaire **153**
10.1 Paramètres incertains . . . 153
10.1.1 Introduction . . . 153
10.1.2 Description du cadre d'expérience . . . 153
10.1.3 Analyse de sensibilité . . . 154
10.1.4 Décomposition polynomiale . . . 155
10.1.5 Vers l'assimilation de données pour la correction du paramètre . . . 157
10.1.6 Prévisibilité du modèle réduit . . . 160
10.2 Champs incertains . . . 162
10.2.1 Cadre d'étude . . . 162
10.2.2 Représentation stochastique du débit non-permanent . . . 162
10.2.3 Approximation polynomiale . . . 164

V Conclusion et perspectives **169**

VI Annexes **175**
.1 Rapport Technique : Quantification des incertitudes liées aux pa-ramètres hydrauliques : Application aux coefficients de frottement avec le modèle Mascaret Marne Amont Global . . . 177
.2 Rapport Technique : Etude d'incertitudes autour de l'équation de Manning-Strickler . . . 194

Bibliographie **213**

Table des figures

1 Principe de la mesure SWOT par système de radar en double fauchée. Source : site internet de SWOT du CNES 5
2 Différence en l'altimétrie nadir et l'altimétrie à fauchée utilisée dans SWOT. Source : Site internet SWOT du CNES. 5
3 Une représentation schématique des différentes étape d'une étude de quantification d'incertitudes. 8

1.1 profil de l'écoulement longitudinal [a] et transversal [b] 17
1.2 Les régimes d'écoulement donnés par la théorie des caractéristiques : [a] régime fluvial, [b] régime torrentiel 21

2.1 Schéma représentant le calcul d'une cellule en volumes finis 24

3.1 Les différents régimes d'écoulement, source : [27] 28
3.2 Les courbes de remous de faible pente, source : [136] 30
3.3 Les courbes de remous de forte pente, source : [136] 31
3.4 Exemple de passage des écoulements graduellement variés et brusquement variés, source : [136] 32
3.5 Méthode de quantification des incertitudes selon [66]. 38

6.1 Approximation PC de degré 6 pour la fonction $g(x) = x\sin(x)$ avec une base du polynôme d'Hermite et du polynômes de Legendre pour une variable aléatoire gaussienne $x \sim \mathcal{N}(0,1)$ 63
6.2 Approximation PC de degré 8 et de degré 3 pour la fonction $g(x) = x\sin(x)$ avec une base du polynôme d'Hermite et pour une variable aléatoire gaussienne $x \sim \mathcal{N}(0,1)$ 64
6.3 Comparaison entre l'utilisation d'une grille pour un nombre de variables égal à 2 basée sur le produit tensoriel complet (à droite) et une grille basée sur la technique de Smolyak avec un niveau de quadrature $\ell = 5$ (à gauche). Source : [152] 70
6.4 Principe de la stratégie de troncature basée sur la q-norm. Source : [15] . . . 77

7.1 Comparaison entre l'erreur de convergence du gPC-EnSRF et de l'EnSRF standard. Source : [79] 89

9.1 Sensibilité de l'approximation polynomial par rapport à discrétisation numérique. 140
9.2 Indices de Sobol', et indices totaux pour le cas de l'équations de Manning. . . 141
9.3 Les modes spectraux correspondants aux deux représentations polynomiales . 142
9.4 L'estimation de la variance à l'aide des deux représentations polynomiales et la comparaison avec la variance estimée avec une représentation bivariée [a], ainsi que l'estimation de l'erreur entre les deux variances [b] 142
9.5 La correction des paramètres de frottements et de débit à l'aide de l'algorithme PCEnKF et l'EnKF 146
9.6 L'estimation de la covariance entre la hauteur d'eau et les coefficients de frottement, et la covariance entre la hauteur d'eau et le débit pour PC-EnKF (rouge) et EnKF(bleu) 147
9.7 L'estimation de la variance sur toute la Garonne [a]. Calcul de l'erreur entre l'analyse et l'observation pour le PC-EnKF et l'EnKF [b] 148
9.8 La correction des paramètres de frottements en utilisant des données de type SWOT 150
9.9 L'erreur relative (à l'observation) due à l'analyse réelle et l'analyse issue des observations moyennées. 151

10.1 L'évolution temporelle des indices de Sobol 154
10.2 Estimation de la moyenne et l'écart-type de la hauteur d'eau à Marmande et comparaison avec les données d'observation 156
10.3 Estimation de la variance à l'aide des coefficients polynomiaux dans le temps 157
10.4 Estimation de l'incertitude de la hauteur d'eau avant et après l'analyse et comparaison avec l'observation 159
10.5 Spectre des coefficients polynomiaux pour l'étape de propagation (en bleu) et l'étape d'analyse (en rouge) pour les jours : 4, 5, 6 et 7 160
10.6 Erreur sur l'estimation de la variance entre le modèle réduit crée après l'analyse et celui avant la propagation 161
10.7 a - Les premières valeurs propres de la fonction de corrélation exponentielle b - Les 7 premiers vecteurs propres 163
10.8 a - La matrice de corrélation estimée avec l'équation 10.9 b -Erreur sur l'estimation 164
10.9 a - La sensibilité par rapport au degré polynomial b -Indices de Sobol pour le premier pas de temps 167
10.10Evolution temporelle des indices de Sobol 167

I

Introduction

Contexte général

Selon le rapport [1] "*The Human cost of weather related disasters 1995-2015*" produit en commun par l'UNISDR (United Nations Office for Disaster Risk Reduction) et le CRED (Centre for Research on the Epidemology of Disasters), les risques d'inondation sont ceux qui ont la plus grande probabilité d'occurence dans les 20 ans entre la COP 1 et la COP21 parmi toutes les catastrophes naturelles en rapport avec le changement climatique. Entre 1995 et 2015, ils présentent 47% du nombre total des catastrophes naturelles et impactent la vie de plus de 2.3 milliard de personnes dans le monde. Le nombre de crues est estimé à une moyenne de 171 dans la décade entre 2005 et 2014 dans le monde. Le 23 octobre 2007, l'union européenne adopte la directive européenne Inondation [2], elle vise à limiter les dégâts humains, économiques, sociaux et environnementaux des inondations, en encadrant la gestion des risques d'inondation au niveau européen. Plusieurs acteurs scientifiques, politiques et même civils se sont organisés pour améliorer la gestion de l'eau au niveau européen (et même au niveau mondial). En France, Le Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations (SCHAPI) et les Services de Prévisions des Crues (SPC) ont été créés afin de prévoir et anticiper des inondations dévastatrices avec la collaboration de Météo-France. Le portail Vigicrues [3], destiné au grand public et aux services de la sécurité civile, héberge des cartes de vigilance, renseignant l'ampleur de la crue.
Suite à cette dynamique, plusieurs laboratoires de recherche se sont impliqués dans des études qui visent à améliorer la gestion du risque d'inondation dans le cadre opérationnel. Ces études s'organisent en trois axes principaux :

- comprendre les phénomènes liés aux inondations,
- développer des outils et des modèles pour prévoir,
- améliorer la gestion des ouvrages hydrauliques qui interagissent avec les mécanismes provoquant les inondations.

Afin de répondre à ces différentes problématiques, les laboratoires de recherche ont dû s'appuyer sur plusieurs outils dont :

- les bases de données historiques sur les inondations,
- les données des mesures in-situ ou satellitaire,
- la simulation expérimentale,
- la simulation numérique.

Aujourd'hui et en résultat de ces efforts de recherche, la compréhension de la physique de la dynamique des écoulements en rivière a permis de modéliser mathématiquement ces processus. Cette compréhension fondamentale physique a permis de construire des modèles numériques permettant de simuler des inondations. Ces modèles, bien qu'ils soient performants et fiables, sont entachés d'erreurs. Les observations offrent une solution complémentaire. De ce fait, des efforts considérables ont été faits pour jauger des rivières, afin d'obtenir des informations sur l'écoulement et ainsi mieux prévoir les inondations. Ces mesures sont faites

1. http://www.unisdr.org
2. http://eur-lex.europa.eu
3. http://www.vigicrues.gouv.fr/

de manière continue ou instantanée. Cependant, comme le montre ce même rapport [4], les inondations se produisent plus fréquemment dans les pays de l'Asie et l'Afrique, où les cours d'eau ne sont pas jaugés et par conséquent, l'hydrodynamique est plus difficilement prévisible. L'un des principaux objectifs de la future mission spatiale SWOT est d'estimer le niveau d'eau ainsi que les volumes d'eau dans de tels endroits.

La mission satellitaire SWOT (*Surface Water Ocean Topography*) pour l'observation des rivières

L'observation de l'état de l'eau dans les rivières s'appelle l'hydrométrie. Cela vise à mesurer les différents paramètres caractérisant les écoulements dans les cours d'eau. La limnimétrie vise, par exemple, à estimer la côte de la surface de l'eau. Elle peut être estimée soit de façon discontinue à l'aide d'une échelle limnimétrique, soit de façon continue à travers les stations d'observation limnimétrique, où la mesure de la hauteur d'eau se fait grâce à un dispositif de mesure et d'enregistrement des variations de la hauteur de l'eau en fonction du temps. Dans les deux cas, ces observations sont localisées spatialement et ne peuvent donner une information complète sur l'hydrodynamique de la rivière. De plus, lors d'une crue, la lecture des données sur une échelle limnimétrique est une tâche difficile voire impossible quand le cours d'eau détruit l'équipement de mesure. Il existe des portails qui gèrent ces données issues de mesures tels que le site internet SANDRE [5] pour la France. Un inconvénient majeur de la mesure *in-situ* réside dans le fait qu'elle ne présente qu'une information partielle de l'écoulement de la rivière vu que la mesure est ponctuelle et locale. Cette limitation est levée par la future mission SWOT qui vise à fournir une couverture globale à haute résolution spatiale de la côte d'eau sur les océans et les continents.

La future mission SWOT, prévue en Avril 2021 et organisée conjointement par les deux agences spatiales française (CNES) et américaine (NASA), vise (entre autres) à estimer le débit des rivières et à estimer les variations de volumes de stock d'eau dans les hydrosystèmes de surface. La mission consiste à embarquer dans l'espace un radar interférométrique à large fauchée dénommée (KaRIn) fonctionnant en bande Ka. Cet instrument fonctionne sur le principe de SAR (Radar à synthèse d'Ouverture [6]) ; il permet d'éclairer deux fauchées de 50 km de part et d'autre de la trace (Fig. 1). Les ondes sont émises (avec un angle inférieur à 4.5) et reçues grâce aux deux antennes qui fonctionnent de manière simultanée et qui sont situées aux deux extrémités d'un mât de 10 m d'envergure (Fig. 1). L'avantage principal de l'utilisation de l'altimétrie à fauchée par rapport à l'altimétrie nadir est de fournir une image bidimensionnelle avec une résolution spatiale au lieu de données ponctuelles le long de la trace du satellite au nadir (Fig. 2).

4. http://www.unisdr.org
5. `http://www.sandre.eaufrance.fr/`
6. Page Wikipédia sur le principe de SAR

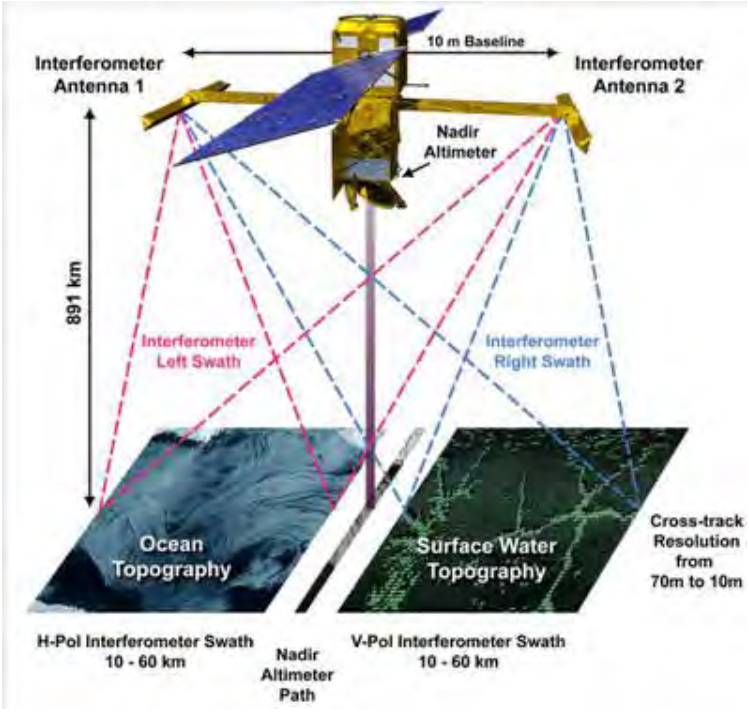


FIGURE 1 – *Principe de la mesure SWOT par système de radar en double fauchée. Source : site internet de SWOT du CNES*

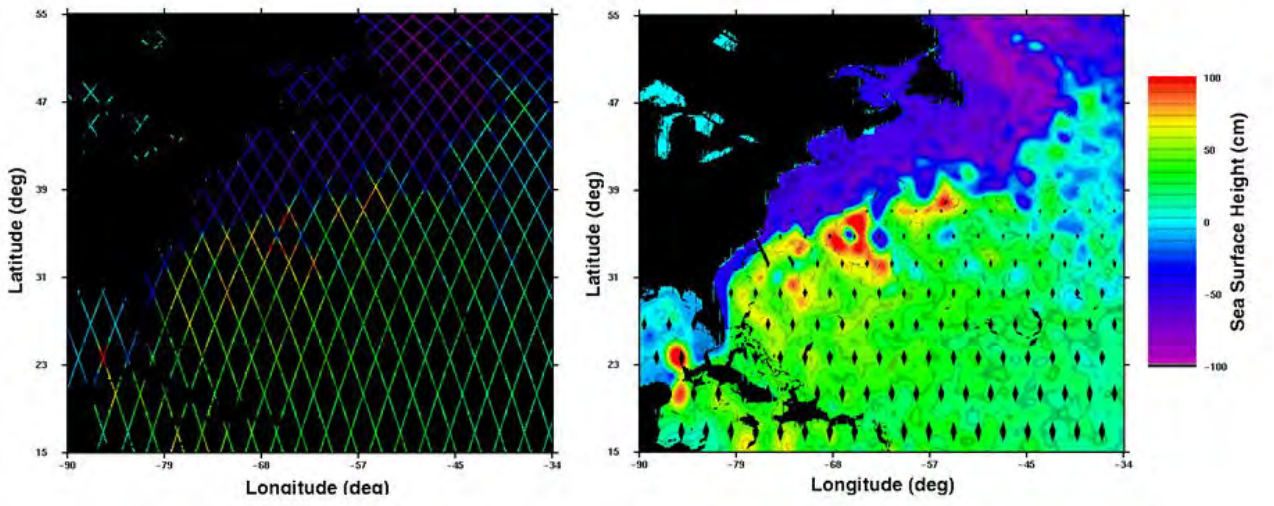


FIGURE 2 – *Différence en l'altimétrie nadir et l'altimétrie à fauchée utilisée dans SWOT. Source : Site internet SWOT du CNES.*

SWOT pourra alors délivrer des cartes de hauteurs d'eau avec une résolution spatiale de $10m$ pour le point le plus loin et de $60m$ pour le point le plus proche du nadir. On pourra avoir une résolution plus fine en moyennant sur la trace du satellite. Le satellite sera en mesure d'observer les hauteurs d'eau des lacs de superficie supérieure à $250m \times 250m$

Paramètre	SWOT	*in-situ*
Précision de la mesure	Dépendante de la définition des profils en travers	Plus exacte
Échantillonnage temporel	tout les 21 jours	Continue
Disponibilité de la mesure	45 jours après	Instantanée
Taille des rivières	Largeur > 100m	Petite taille
Nature de la mesure	Toute au long de la rivière	Ponctuelle
Description physique de l'hydrodynamique	2 D	0 D

TABLE 1 – *Tableau récapitulatif des différences entre la mesure in-situ et la mesure SWOT.*

et des fleuves ayant une largeur supérieure à $100m$. La mesure donnée par SWOT pourra alors compléter la mesure in-situ. Le tableau 1 rassemble les points de comparaison entre la mesure SWOT et les mesures *in-situ*.

Modélisation numérique

La modélisation mathématique des écoulements à surface libre a fait l'objet des travaux de Saint-Venant [118] au XIXe siècle. Cette avancée dans la compréhension des processus physiques des écoulements en rivières est un héritage de la révolution scientifique dans la compréhension de la mécanique des fluides entre le XVIIe siècle et le XVIIIe siècle. Cette avancée va révolutionner, par la suite, plusieurs domaines liés à la mécanique des fluides (tels que l'hydrodynamique, l'aéronautique, l'industrie du pétrole, la combustion, etc). Cette révolution scientifique a permis de construire des modèles mathématiques, d'une simple équation linéaire aux équations aux dérivées partielles non-linéaires, qui permettent de décrire, à des niveaux de simplifications différents, l'accélération d'une particule fluide soumise à des contraintes diverses, internes ou externes. Le XXe siècle a connu l'évolution vers la simulation numérique. Les équations devenaient maintenant estimables avec les machines de calcul. Les équations de Saint-Venant, comme les équations de Navier-Stokes, pouvaient maintenant être résolues à l'aide de schémas numériques. Ceci a permis de regrouper la communauté scientifique autour d'une nouvelle thématique : l'analyse numérique. On a réussi alors, à s'intéresser à la modélisation du climat, de l'atmosphère, de l'océan et des écoulements dans les surfaces continentales. L'expertise, reçue alors en modélisation numérique des équations de Navier-Stokes, a permis d'écrire des modèles qui s'intéressent à l'évolution du climat de façon générale [5, 88]. Il fallait attendre les années 1980, quand la puissance des ordinateurs devenait attractive, pour que les simulations numériques soient faisables dans des temps suffisamment courts. C'est ainsi qu'au cours des années 1980 et 1990, on est arrivé à construire des modèles numériques avec une base physique forte et simulable en des temps raisonnables. Comme décrit par [75], on avait ainsi construit une *planète numérique* proche de notre planète *réelle* et sur laquelle il était possible de conduire des expérimentations variées. La simulation numérique appliquée à l'hydrodynamique a profité aussi des ces différents champs d'expertise. Plusieurs modèles numériques virent le jour pour différents champs d'application en hydraulique (prévision des crues [57], contrôle des canaux d'irrigation[87],

dimensionnement de centrale hydroélectriques, etc) et adaptés pour les différentes physiques simulés (présence de méandres [40], confluences [8], passage de l'eau dans des ponts, etc).
Aux États-unis d'Amérique, l'US Army Corps of Enginneers développa le modèle HEC-RAS [7]. Ce logiciel permet de simuler les écoulements dans des rivières pour une physique 1D en régime permanent et une physique 1D et 2D pour le régime non permanent. Il est aussi utilisé pour simuler le transport de sédiments et pour modéliser la qualité de l'eau dans les rivières. En Angleterre, différents modèles simulant l'hydrodynnamique ont vu le jour tels que LISFLOOD-FP [8] de l'université de Bristol. En France, la compétence dans la modélisation hydrodynamique n'est pas négligée, car la compétence de l'IRSTEA dans le domaine de l'irrigation a permis de mettre en place le logiciel SIC [9]. Depuis 1987, EDF R&D , selon [60], s'attèle à produire des codes industriels (Telemac et Mascaret [10]) simulant l'hydrodynamique des rivières.
Malgré les évolutions considérables des modèles numériques en hydrodynamique, ces derniers sont entachés d'incertitude. Les facteurs qui contribuent majoritairement à l'apparition de l'incertitude dans les modèles numériques sont :

— Incertitudes dues à la modélisation : les équations de Saint-Venant traduisent de façon plus ou moins fidèle les phénomènes physiques présents dans les écoulements en rivière.

— Incertitudes numériques : ils proviennent de la discrétisation, la convergence et la méthode de résolution numérique ainsi que le maillage utilisé dans le modèle.

— Incertitudes épistémiques : La modélisation empirique des coefficients de frottement, la description de la plannimétrie, etc...

— Incertitudes stochastiques : La variabilité naturelle qui existe dans la description des conditions aux limites.

La précision du résultat d'une simulation numérique est contrainte par toutes ces sources d'incertitude. Par conséquent, il est impératif d'associer à chaque simulation une incertitude, afin de bien comprendre les résultats de la simulation et par la suite prévoir la physique réelle. C'est la raison pour laquelle il est important d'étudier la quantification d'incertitude dans les modèles hydrodynamiques.

La quantification des incertitudes

Les incertitudes, liées aux composantes d'un système dynamique (description approchée des données, simplification des équations), aux échelles spatiales (par exemple : les échelles de turbulence négligées) et à sa modélisation approchée (discrétisation numérique), se traduisent en erreurs sur les résultats de la modélisation numérique en géosciences. Il est donc essentiel de quantifier les incertitudes. Bien que l'analyse numérique ait connu des progrès importants dans la modélisation des incertitudes lors des dernières années, le concept de la quantification d'incertitude reste relativement peu utilisé en géosciences.

7. http://www.hec.usace.army.mil/software/hec-ras/
8. http://www.bristol.ac.uk/geography/research/hydrology/models/lisflood/
9. http://sic.g-eau.net/preambule?lang=fr
10. http://www.opentelemac.org/

Afin de mener une étude de quantification des incertitudes, il est de coutume, dans la communauté scientifique, de se placer dans un contexte probabiliste. Les paramètres du modèle qui sont supposés incertains sont considérés comme des variables aléatoires (entrées incertaines dans la figure Fig. 3). Dans le cadre présent des géosciences, quelques paramètres sont des champs physiques, que l'on définira en terme de vecteurs aléatoires. On supposera par ailleurs que tous ces paramètres sont proprement définis dans un espace de probabilité, et qu'ils sont décrits par une loi de probabilité a priori. Comme illustré sur la figure 3, l'étape initiale d'une étude de quantification d'incertitude passe par la définition d'un espace de probabilité pour des variables aléatoires ou par des vecteurs aléatoires. La représentation probabiliste des champs physiques est notamment une question largement abordée dans la communauté afin de prendre en compte les incertitudes associées à ces champs.
La définition de ce contexte probabiliste permet d'associer le notion d'incertitude au résultat d'une simulation numérique une incertitude. Ce résultat porte le nom de variable d'intérêt (ou variable pronostique, ou quantité d'intérêt) (Fig. 3). Cette variable est aussi considérée comme une variable aléatoire (ou vecteur aléatoire). La connaissance de l'espace de probabilité (ou ses paramètres caractéristiques tels que la moyenne ou la variance), associé à la quantité d'intérêt, permet d'étudier l'incertitude du modèle.
La deuxième étape de la quantification des incertitudes est l'étape de propagation. Cette

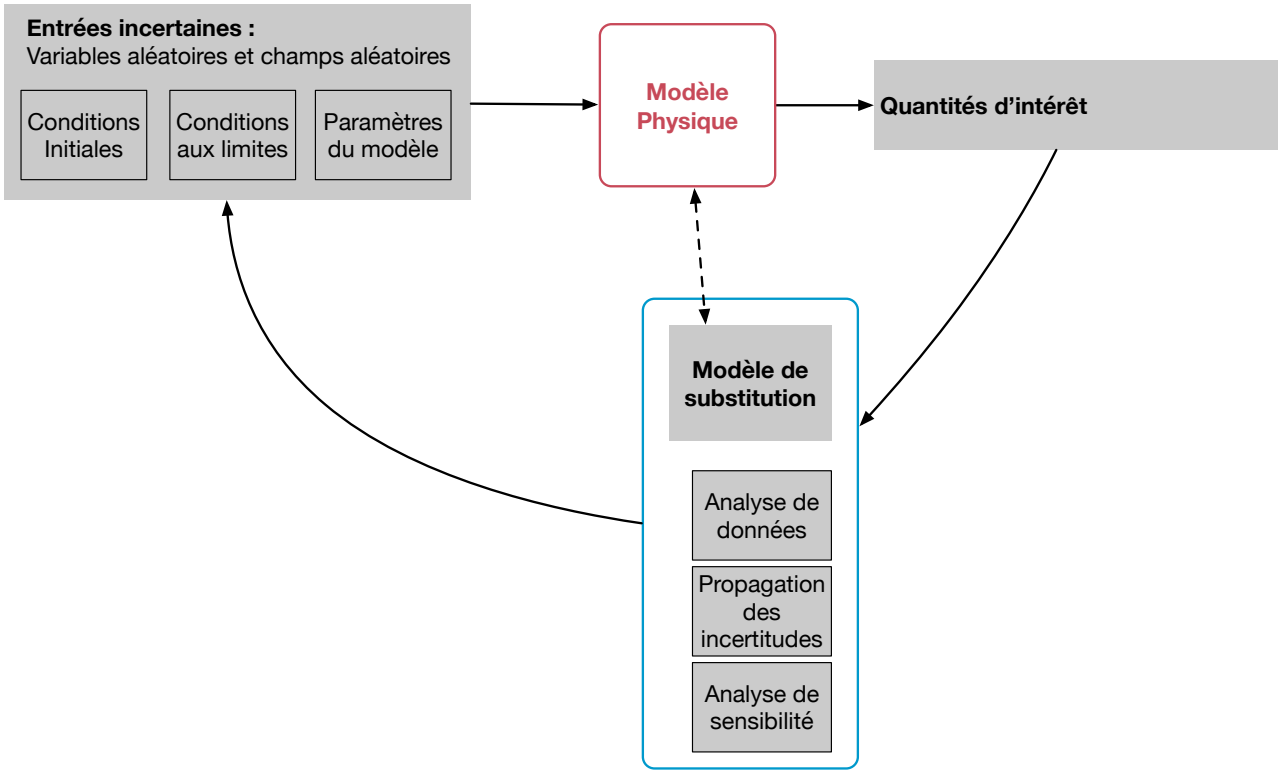


FIGURE 3 – *Une représentation schématique des différentes étape d'une étude de quantification d'incertitudes.*

étape est primordiale afin d'estimer l'incertitude sur la variable pronostique (aussi appelé

variable d'intérêt). L'approche la plus classique utilisée dans la communauté scientifique est la méthode des simulations Monte-Carlo. Elle consiste à échantillonner les paramètres incertains et à construire un ensemble de réalisations de ces paramètres, puis d'intégrer le modèle pour chaque réalisation. L'ensemble généré par les solutions du modèle est utilisé pour estimer de manière empirique les statistiques de la variable d'intérêt. L'avantage de la méthode Monte-Carlo est qu'elle est robuste, facile à appliquer et que la convergence des statistiques n'est conditionnée par aucune hypothèse sur la dimension des paramètres incertains. Cependant, l'inconvénient principal réside dans le fait qu'elle nécessite un large nombre de réalisations, ce qui peut rapidement augmenter le coût de calcul.
Une méthode assez courante dans la communauté scientifique afin d'accélérer une étude de quantification des incertitudes est l'utilisation d'un modèle de substitution [31]. Comme illustré sur la figure 3, on utilise un modèle substitution afin de propager les incertitudes, d'effectuer une analyse de sensibilité ou une analyse de données. Ce modèle est moins coûteux que le modèle direct, mais conserve les mêmes propriétés statistiques que celui-ci. Une méthode de construction des modèles réduits la plus utilisée en quantification d'incertitude est la décomposition en polynômes du chaos [131]. Basée sur une méthode spectrale, ce modèle réduit consiste à exprimer les solutions stochastiques du modèle physique comme une représentation polynomiale (en utilisant une base de polynômes orthogonaux) des paramètres incertains. Les différentes statistiques de la variable d'intérêt peuvent alors être estimées à l'aide des coefficients polynomiaux.
La quantification de l'incertitude des modèles hydrodynamiques est une clé essentielle pour estimer l'erreur sur la prévision du modèle. Cependant, le modèle n'en demeure pas moins inexact. L'incertitude tendant à s'amplifier à mesure que l'échéance de prévision croit, il est indispensable de rajouter une information issue de l'observation, qui via une méthode d'assimilation de données permet d'améliorer le modèle et ainsi améliorer les prévisions.

Assimilation de données

L'assimilation de données est une discipline qui s'est développée au cours de la seconde moitié du XXe siècle. Elle hérite de l'intérêt de la communauté scientifique à la prévision numérique du temps. Beaucoup d'études au début du XXe siècle ont démontré la difficulté de la prévision numérique du temps : [13] montre que la base d'une bonne prévision consiste à connaître de façon précise les lois d'évolution de l'atmosphère (du temps en général) mais aussi d'avoir une connaissance précise de l'état de l'atmosphère à un instant donné. On utilise effectivement les observations du temps pour initialiser les modèles numériques. Cependant, vu la nature des observations, entachées d'erreurs de mesure et la nature chaotique des équations de l'évolution du temps, les premiers essais de la prévision se révélèrent complètement irréalistes [113]. En 1963, Lorrenz montre que la dynamique chaotique des équations de la dynamique est à l'origine de la restriction pratique de la prévisibilité des modèles. En effet, deux états de l'atmosphère très proches mais distincts simulés par le même modèle divergent très rapidement l'un de l'autre : c'est "l'effet papillon". Il est ainsi nécessaire de recaler la trajectoire du modèle, faute de quoi la prévision finira par diverger de la réalité. Le recalage

est fait à partir des observations qui servent à donner une vision riche mais partielle de l'instant présent. Les techniques mathématiques, qui permettent de recaler le modèle de façon optimale (en un sens mathématique) à l'aide des observations afin de produire des prévisions proches de la réalité, sont connues sous le nom de méthodes d'assimilation de données.
Les méthodes d'assimilation de données sont donc des techniques mathématiques, qui visent à réaliser un compromis statistiquement optimal entre les observations et la simulation issue du modèle numérique, en tenant compte des erreurs respectives des mesures et du modèle (que l'on nomme généralement ébauche). Il existe deux grandes familles classiques de techniques en assimilation de données :

— **Méthodes variationnelles** : C'est une généralisation du concept de la minimisation moindres carrés d'une fonction objectif. La solution peut être obtenue à l'aide des méthodes de contrôle optimal (aussi connu sous le nom de méthodes adjointes) [30].

— **Méthodes de filtrage :** Basées principalement sur le filtre de Kalman, très populaire en sciences de l'ingénieur aussi dérivé sous la forme stochastique par le filtre de Kalman d'Ensemble qui utilise un ensemble de simulations, afin de représenter l'incertitude du système.

La communauté s'intéresse à des méthode hybrides qui combinent astucieusement les avantages offerts par les méthodes d'ensemble pour estimer la probabilité des prévisions ainsi que les avantages des méthodes variationnelles [33].

Vers la prévision des crues

L'utilisation des techniques d'assimilation de données en hydraulique fluviale est assez récente, par rapport à son application dans le domaine de la météorologie. Selon les sources d'incertitudes considérées dans le modèle hydraulique (paramètres, forçages ou état) les approches d'assimilation de données seront différentes. Généralement, l'assimilation de données a pour but de corriger l'état hydraulique [8]. Cependant, quand certains paramètres (ou forçages) du modèle qui sont influents, sont mal connus (comme c'est le cas pour cette thèse), on peut utiliser l'assimilation de données pour la mise à jour des valeurs des paramètres (ou forçages) [57]. Finalement l'assimilation de données peut s'intéresser à la fois à l'estimation des paramètres ainsi qu'à la correction de l'état simulé (c'est l'estimation duale) [99].
Dans leur article, [112] utilisent un filtre de Kalman étendu pour corriger les forçages à l'amont du modèle hydraulique (Mascaret) combiné avec une correction de l'état hydraulique utilisant un filtre de Kalman. Cette procédure permet d'améliorer la prévision des hauteurs d'eau à moyenne échéance. Pour la correction de l'état hydraulique, beaucoup d'études ont permis de démontrer l'applicabilité de l'assimilation de données pour la prévision de crues en opérationnel à l'aide d'un filtre de Kalman [86, 112], ou encore un filtre de Kalman d'ensemble [8, 100] ou de filtrage particulaire [93].
L'apport des observations satellitaires dont les données arrivent après une échéance de plusieurs jours, comme SWOT, ne permettent pas de corriger l'état hydraulique en temps réel,

car les observations ne sont pas instantanées. Elle sont par conséquent, utilisées principalement pour la correction des paramètres du modèle hydraulique. On cite à titre d'exemple les études de [104] qui utilise des données synthétiques de type SWOT pour corriger les coefficients de frottement de modèle hydrologiques sur le bassin du Niger. Les travaux de [34, 156] mettent en évidence l'apport des données altimétriques pour la correction de la bathymétrie sur les fleuves Amazone et Ohio.
La compétence en assimilation de données du Cerfacs, ainsi que les diverses collaborations avec le SCHAPI, EDF R&D et Météo France (avec sa composante recherche au CNRM) ont permis de démontrer que l'assimilation de données améliore la prévision des modèles hydrauliques à travers la correction de paramètres [112]. Cette dynamique au sein du cette collaboration a permis l'implémentation de l'assimilation de données au seins des SPC (Service de prévision des crues). Dans son travail de thèse, Johan Habert [57], prévisionniste au sein du SPC SAMA, s'est intéressé à l'application de l'assimilation de données pour l'amélioration de la prévision des crues en temps réel en utilisant une méthode de filtre de Kalman étendu. La thèse de Sébastien Barthélémy [8] s'inscrit dans un contexte collaboratif similaire. Ces travaux améliorent la prévision des crues au temps réel en dépassant les limites d'usage déterministe des codes hydrauliques en apportant une incertitude à la prévision d'ensemble en utilisant un filtre de Kalman d'Ensemble. Cette thèse s'intéresse de plus à l'amélioration de la modélisation hydraulique en étudiant le couplage 1D-2D.
Dans la complémentarité de ces travaux, la présente thèse s'intéresse à accélérer les méthodes du filtre de Kalman d'Ensemble en estimant les incertitudes avec un modèle réduit basée sur la décomposition en chaos polynomial. Cette méthode a été, notamment, testée dans la thèse de Mélanie Rochoux [115] dans le contexte de la modélisation et la propagation des feux de foret et a donné des résultats suffisamment prometteurs pour la poursuivre ici, dans le contexte hydraulique.

Objectif et cadre collaboratif de la thèse

Ce travail de thèse s'inscrit dans un cadre collaboratif entre le Cerfacs, le CNES (co-financeur de la thèse) et EDF (co-financeur de la thèse). L'objectif de la thèse est d'étudier la quantification d'incertitudes et d'améliorer les algorithmes ensemblistes d'assimilation de donnée tels que le filtre de Kalman d'Ensemble, afin de corriger des paramètres incertains tels que la bathymétrie ou le coefficient de frottement, notamment grâce à des données de type SWOT. Une première approche consiste à utiliser la décomposition en chaos polynomial, d'une part pour mener une étude de quantification d'incertitude en hydraulique fluviale et d'autre part pour accélérer les algorithmes d'assimilation de données pour corriger les paramètres tels que les coefficients de frottements. Dans un second temps, on explore parmi les méthodes existantes, celles qui permettent de réduire la dimension d'un champ physique pour étudier l'incertitude du modèle qui en résulte. Celle-ci sont utiles pour quantifier et réduire les incertitudes liées aux champs de bathymétrie et/ou aux forçages hydrologiques .

PLAN DU MANUSCRIT

Ce manuscrit est composé de six parties, chacune divisée en chapitres.
L'introduction est donnée en première partie.
La deuxième partie présente les équations de Saint-Venant, qui décrivent les écoulements dans les rivières. On s'intéresse notamment à une représentation 1D des équations. On formule une discrétisation numérique de ces équations, telle qu'implémentée dans le logiciel Mascaret. Le dernier chapitre de cette partie propose quelques simplifications des équations de Saint-Venant.
La troisième partie de ce manuscrit présente les méthodes de quantification et de réduction des incertitudes. On présente notamment le contexte probabiliste qui permet de bien définir un problème de quantification d'incertitudes et d'analyse de sensibilité. On propose ensuite de réduire la dimension d'un problème stochastique quand on traite de champs aléatoires dans le contexte des modèles géophysiques. Les méthodes de décomposition en polynômes du chaos sont ensuite présentées ; on présente notamment les différentes stratégies pour le calcul des coefficients polynomiaux. Cette partie dédiée à la méthodologie s'achève par un chapitre consacré à l'assimilation de données ensemblistes (typiquement le filtre de Kalman d'Ensemble) et à l'utilisation des modèles réduits dans ce cadre.
La quatrième partie de ce manuscrit est dédiée aux résultats. On commence par identifier les sources d'incertitudes en hydraulique que l'on s'attache à quantifier et réduire par la suite. Un article en cours de révision détaille la méthode et la validation d'un modèle réduit polynomial pour les équations de Saint-Venant en régime stationnaire lorsque l'incertitude est majoritairement portée par les coefficients de frottement et le débit d'apport à l'amont. On met à profit l'utilisation du modèle réduit pour réduire le coût de calcul du filtre de Kalman d'Ensemble dans le cadre d'un exercice d'assimilation de données synthétiques de type SWOT. On cherche à reconstruire les coefficients spatialisés de frottement et le débit d'apport à l'amont. On s'intéresse précisément à la représentation spatiale de la données telle que vue par SWOT : couverture globale du réseau, moyennage spatial entre les pixels observés.
On s'intéresse enfin à la construction du modèle réduit en régime instationnaire. On suppose dans un premier temps que l'incertitude est liée aux coefficients de frottement. Il s'agit à présent de juger de la nécessité du recalcul des coefficients polynomiaux au fil du temps et des cycles d'assimilation de données.
On suppose dans un deuxième temps que l'incertitude est portée par le débit en amont du réseau, qui est un vecteur temporel. On procède à une décomposition de type Karhunen-Loève pour réduire la taille de l'espace incertain.
Pour finir, les conclusions et les perspectives de ce travail sont présentées en cinquième partie. La sixième partie comporte des annexes de ce travail, qui portent sur la quantification d'incertitudes liées à la description locale des coefficients de frottement.

II

Hydraulique fluviale

Équations de Saint-Venant 1

1.1 Équations de Navier-Stokes

1.1.1 Formulation des équations

Les équations de Navier-Stokes sont décrites à partir de 4 équations aux dérivées partielles, et servent à décrire le mouvement d'une particule fluide (on ne s'intéresse ici qu'à l'hydrodynamique et donc on n'inclut pas les équations qui décrivent la conservation de l'énergie). Dans un système de coordonées cartésiennes en dimension 3, l'équation de la conservation de mouvement couplée avec les équations de mouvement forment les équations de base en hydrodynamique. Ces équations servent à la description des écoulements quelque soit leur nature : stationnaire, instationnaire, compressible ou incompressible. Elles ont un large champ d'applications en sciences ainsi qu'en ingénierie. Développées initialement par Claude-Louis Navier et George Gabriel Stokes, ces équations découlent de l'application de la seconde loi de Newton sur une particule fluide en mouvement qui subit des forces liées aux vitesses de déformation (gradients de vitesse) et à un champ de pression, ce qui permet d'avoir une description assez complète de la mécanique des fluides. Leur formulation mathématique est décrite par les équations :

$$\begin{cases} \partial_t \rho + \overrightarrow{div}(\rho \overrightarrow{V}) = 0 & \text{(équation de continuité)}, \\ \rho(\partial_t \overrightarrow{V} + \overrightarrow{V}.grad\overrightarrow{V}) = div.\mathbf{\Sigma} + \overrightarrow{f} & \text{(équation de mouvement)}, \end{cases} \tag{1.1}$$

où ρ représente la densité du fluide, $\overrightarrow{V} = (u, v, w)$ le vecteur vitesse dans le système des coordonées cartésiennes, $\mathbf{\Sigma}$ le tenseur des contraintes et $\overrightarrow{f}$ les forces volumiques. On remarque que l'équation de conservation de la quantité de mouvement comporte essentiellement deux inconnues : la vitesse et le tenseur des contraintes. Or, il existe une loi de comportement (*semblable à la loi de Hooke en élasticité linéaire*) qui permet d'écrire le tenseur des contraintes en fonction des dérivées partielles des composantes de la vitesse pour les fluides newtoniens. L'équation suivante :

$$\mathbf{\Sigma} = -p\mathbf{I} + \lambda(div\overrightarrow{V})\mathbf{I} + 2\mu\mathbf{D}, \tag{1.2}$$

décrit le comportement rhéologique d'un fluide newtonien, qui dépend de manière linéaire homogène et isotrope du gradient de vitesse. Le lecteur peut se référer au livre [135] pour

plus de détails sur la construction de ces modèles. Dans l'équation (1.2), p est la valeur du champ de pression, (λ et μ) sont les coefficients de Lamé, la matrice $\mathbf{D}$ représente la matrice des taux de déformation :

$$\mathbf{D} = \frac{1}{2}\begin{pmatrix} 2\partial_x u & \partial_x v + \partial_y u & \partial_x w + \partial_z u \\ \partial_x v + \partial_y u & 2\partial_y v & \partial_y w + \partial_z v \\ \partial_x w + \partial_z u & \partial_y w + \partial_z v & 2\partial_z w \end{pmatrix}.$$

1.1.2 Équations de Navier-Stokes incompressibles

Le fluide considéré ici est l'eau. Dans cette thèse, on ne s'intéressera pas à la dynamique sédimentaire que l'on supposera faible. Par conséquent, on peut émettre l'hypothèse que la densité est constante le long d'une trajectoire. Dans ce cas, l'équation de continuité se réduit à :

$$div(\overrightarrow{V}) = 0.$$

Ainsi, les équations de Navier-Stokes pour les fluides newtoniens incompressibles s'écrivent de la manière suivante :

$$\begin{cases} div(\overrightarrow{V}) = 0 & \text{(équation de continuité)} \\ \rho(\partial_t \overrightarrow{V} + \overrightarrow{V}.\overrightarrow{grad}\overrightarrow{V}) = -\overrightarrow{grad}p + \rho\overrightarrow{g} + \mu\nabla^2\overrightarrow{V} & \text{(équation de mouvement),} \end{cases} \tag{1.3}$$

- $\partial_t \overrightarrow{V}$ exprime l'accélération due à l'instationnarité ;
- $\overrightarrow{V}.\overrightarrow{grad}\overrightarrow{V}$ exprime l'accélération due à la convection ;
- $-\overrightarrow{grad}p$ est le gradient de pression ;
- $\rho\overrightarrow{g}$ exprime les forces de volume compte tenu de l'hypothèse de pression hydrostatique ;
- $\mu\nabla^2\overrightarrow{V}$ sont les effets visqueux, μ représente le coefficient de viscosité dynamique.

On remarque que pour le système des équations présentées ici, le champ de pression ne constitue pas une variable thermodynamique. Le champ de pression s'adapatera seul avec la dynamique du fluide par la contrainte d'incompressibilité. De ce fait, la fermeture du système d'équation ne nécessite pas une équation de la thermodynamique du fluide, mais juste d'une condition initiale et des conditions aux limites.

1.1.3 Conditions aux limites

Afin de compléter les équations de Navier-Stokes, il est important de définir la géométrie sur laquelle on va les résoudre. Dans le cas des rivières, l'écoulement est à surface libre et est, par conséquent, limité par : le fond de la rivière, la surface de l'eau et des frontières latérales (mur, quai, digue, etc).

Surface libre et fond On définit la fonction $\Phi(x, y, t)$ par :

$$\Phi(x, y, t) = z - Z_s(x, y, t).$$

La condition aux limites cinématiques, qui spécifie que la vitesse normale du fluide à la surface est égale à la vitesse de cette surface s'écrit :

$$\frac{\partial \Phi}{\partial t} + \overrightarrow{V}.\overrightarrow{grad}\Phi = 0. \tag{1.4}$$

Notons U^s, V^s et W^s les composantes de vitesse d'un point de la surface libre. L'équation $\frac{d\Phi}{dx} = 0$ s'écrit alors :

$$W^s - \frac{\partial Z_s}{\partial t} - U^s \frac{\partial Z_s}{\partial x} - V^s \frac{\partial Z_s}{\partial y} = 0. \tag{1.5}$$

La normale à la surface libre, orientée vers les z croissants est le vecteur $\vec{n_s} = \vec{grad}\Phi$, qui a pour composantes $\vec{n_s}(-\frac{\partial Z_s}{\partial x}, -\frac{\partial Z_s}{\partial y}, 1)$. Finalement l'équation (1.5) s'écrit :

$$\frac{\partial Z_s}{\partial t} - \vec{U^s}.\vec{n_s} = 0. \tag{1.6}$$

De la même façon, on obtient la condition au fond, en tenant compte de l'orientation de la normale $\vec{n_f}$ vers le bas :

$$\frac{\partial Z_f}{\partial t} - \vec{U^f}.\vec{n_f} = 0. \tag{1.7}$$

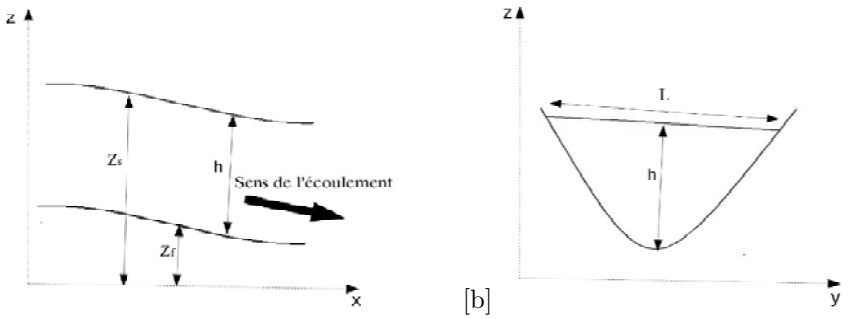
Pression hydrostatique Dans le cadre des équations de Saint-Venant, on utilisera l'hypothèse de pression hydrostatique, qui consiste à simplifier l'équation de la vitesse verticale, en négligeant la diffusion, les termes sources et l'accélération. On obtient alors :

$$\frac{\partial p}{\partial z} = -\rho g. \tag{1.8}$$

Ainsi en intégrant l'équation (1.8) sur la direction longitudinale, on obtient l'équation qui décrit la pression en un point

$$p = p_{atm} + \rho g(Z_s - z). \tag{1.9}$$

On ne tient pas compte dans cette formulation des traceurs actifs (matières en suspension, température, salinité, etc). On renvoie le lecteur à [60] pour une formulation plus complète qui prend en compte les différents traceurs qui peuvent exister lors de l'écoulement de l'eau en rivière.



[a] [b]

FIGURE 1.1 – *profil de l'écoulement longitudinal [a] et transversal [b]*

1.2 Dérivation des équations de Saint-Venant

1.2.1 Equations de Saint-Venant

Les équations portant le nom du célèbre scientifique français "*Jean-Claude Adhémar Barre, comte de Saint-Venant*", publiées en 1871 [118], décrivent les écoulements à surface libre pour les eaux peu profondes d'où leur appelation anglaise "*shallow water equations*". L'écoulement de l'eau à surface libre dans les canaux ou rivières est un processus varié dans l'espace (la vitesse, la hauteur d'eau et par conséquent le débit varient dans l'espace). Les équations de Saint-Venant permettent d'estimer ces dernières grandeurs pour chaque point de l'espace à l'aide de l'équation de continuité et l'équation de conservation de quantité de mouvement, qui comme pour les équations de Navier Stokes, sont basées sur des équations à dérivées partielles ayant pour variables le débit (où les variables de vitesses) et la hauteur d'eau en fonction du temps et de l'espace.
Dans un système tridimensionnel, les équations de Navier-Stokes écrites pour la description des écoulements dans des canaux ou des rivières s'avèrent être complexes, et demanderaient une grande quantité de données avec une description spatiale de la géométrie. Une simplification consiste à moyenner ces équations sur les sections d'une rivière en faisant l'hypothèse de phénomènes lentement variables dans la direction longitudinale. Nous présentons tout d'abord le cas d'une section infiniment large à fond plat avant d'aborder, au paragraphe suivant, le cas général. Dans ce cas, nous définissons la "vitesse débitante" par :

$$u = \frac{1}{h}\int_{Zf}^{Zs} V_x dz. \tag{1.10}$$

Cette grandeur moyennée sera appelée vitesse par abus de langage.

Équations de Saint-Venant Les équations de Navier-Stokes moyennées sur la verticale, sous certaines hypothèses simplificatrices, conduisent aux équations de Saint-Venant :

$$\begin{aligned} &\partial_t h + \partial_x(hu) = 0, \\ &\partial_t(hu) + \partial_x(huu) = -gh\tfrac{\partial Zs}{\partial x} + hf_x + h\tfrac{\tau}{\rho}, \end{aligned} \tag{1.11}$$

avec $\tau = \mu_e \partial_x u$.
On peut retrouver les détails du calcul qui mène à ces équations à partir des équations (1.3), (1.7), (1.6) et (1.9) dans l'ouvrage [60]. Ces équations sont les équations de Saint-Venant unidimensionnelles sous forme dite conservative.

Modélisation du frottement La contrainte du frottement exercée par une plaque plane horizontale sur un écoulement qui lui est parallèle en dimension 1 s'exprime suivant la formule :

$$\tau = -\frac{1}{2}\rho C_f u|u|, \tag{1.12}$$

avec ρ la masse volumique (kg/m^3), C_f le coefficient de frottement (sans dimension), u la vitesse de l'écoulement (m/s). En hydraulique le coefficient C_f est remplacé par le coefficient

C (coefficent de Chézy) avec $Cf = \sqrt{\frac{2g}{C}}$. De ce fait le terme de frottement dans l'équation (1.12) devient $\tau = -\frac{g}{hC^2}u|u|$. La loi de Manning-Strickler remplace le coefficient C par la formule $C = K_s R_h^{1/6} = K_s h^{1/6}$ où K_s est le coefficient de Strickler et R_h est le rayon hydraulique (le rayon hydraulique est défini comme le quotient de la surface mouillée figure(1.1 [b]) sur le périmètre mouillé, dans le cas d'un canal infiniment large $R_h = h$). Ainsi pour un canal de largeur infinie, le rayon hydraulique est égal à la hauteur d'eau. Finalement on écrit le terme de frottement :

$$\tau = -\frac{\rho g u|u|}{K_s^2 h^{1/3}}. \tag{1.13}$$

Forme conservative En injectant l'équation 1.13 dans les équations de Saint-Venant 1.11 et en négligeant les termes de forces de volumes externes f_x, on obtient les équations suivantes :

$$\begin{cases} \partial_t h + \partial_x(hu) = 0, \\ \partial_t(hu) + \partial_x(huu) = -gh\frac{\partial Zs}{\partial x} + \frac{ghu|u|}{{K_s}^2 R_h^{1/3}}. \end{cases} \tag{1.14}$$

1.2.2 Equations de Saint-Venant filaires

Les écoulements en rivière peuvent fréquemment être considérés comme des écoulements filaires, car l'écoulement ne se fait que dans le sens de l'axe de la rivière (figure 1.1 [a]). Afin de réécrire une forme des équations de Saint-Venant adaptée à ce raisonnement, on intègre les équations (1.14) par rapport à la largeur L (voir figure 1.1 [b], on relâche l'hypothèse de largeur infinie). Nous définissons donc la surface mouillée par $A_h = \int_{Z_f}^{Z_s} L dz$ et le périmètre mouillé $P_h = \int_{Z_f}^{Z_s} dz$. On notera que le rayon hydraulique est donné par la formule $R_h = \frac{A_h}{P_h}$. Afin d'obtenir les équations de Saint-Venant filaires, on moyenne les équations (1.14) par rapport à la largeur L [55]. Une deuxième méthode pour obtenir les équations filaires est d'appliquer le principe fondamental de la dynamique sur un tronçon de la rivière avec les hypothèses de pression hydrostatique, pente faible, effets visqueux négligés devant les frottements pariétaux [54]. On obtient :

$$\begin{aligned} &\partial_t A_h + \partial_x Q = 0, \\ &\partial_t(A_h Q) + \partial_x(\tfrac{Q^2}{A_h}) = -gA_h\partial_x z - gA_h J, \end{aligned} \tag{1.15}$$

où Q est le débit ;
-A_h est la surface mouillée ;
-g est l'accélération de la pesanteur ;
-J est le terme de frottement pariétal défini par l'équation selon les notations de [136] :

$$J = \frac{Q^2}{K_s^2 A_h^2 R_h^{4/3}}. \tag{1.16}$$

Afin de procéder à la modélisation numérique, des études ([52] et [126]) préconisent l'utilisation de la forme conservative des équations (1.15). En effet, la forme conservative permet de montrer le caractère hyperbolique des équations de Saint-Venant, et donc de traiter les solutions discontinues (présence de ressaut hydraulique). Afin d'écrire une équation sous forme

conservative il suffit de la mettre sous la forme suivante :

$$\partial_t W + div(F(W)) = B(W) \tag{1.17}$$

où W est le vecteur d'état ; $F(W)$ est un terme de flux ; et $B(W)$ est un terme source.
On remarque que, dans l'équation de conservation de quantité de mouvement (1.15), le terme $-gA_h\partial_x z$ empêche de garder le caractère conservatif. Afin de rendre les équations conservatives, on remplace le terme de pression par la formule $P = -g\int_0^h A_h dz$. On peut écrire donc que :

$$\partial_x P = gA_h\partial_x h + g\int_{Z_f}^{Z_s} \partial_x A_h(x,z,t)dz, \tag{1.18}$$

et en remplaçant h par $z - Z_f$ on obtient :

$$\partial_x P = gA_h\partial_x z - gA_h\partial_x Z_f + g\int_{Z_f}^{Z_s} \partial_x A_h dz. \tag{1.19}$$

Finalement on peut remplacer le terme $-gA_h\partial_x z$ dans l'équation de Saint-Venant filaire par l'équation (1.19) pour obtenir :

$$\partial_t(A_h Q) + \partial_x(\frac{Q^2}{A_h} + P) = g\int_{Z_f}^{Z_s} \partial_x A_h dz - gA_h\partial_x Z_f - gA_h J. \tag{1.20}$$

On pose dans la suite :

$$W = \begin{pmatrix} A_h \\ Q \end{pmatrix}, \tag{1.21}$$

$$F(W) = \begin{pmatrix} Q \\ \frac{Q^2}{A_h} + P \end{pmatrix}, \tag{1.22}$$

et

$$B(W) = \begin{pmatrix} 0 \\ g\int_0^h \partial_x A_h dz - gA_h\partial_x Z_f - gA_h J \end{pmatrix}. \tag{1.23}$$

On obtient alors les équations de Saint-Venant filaires sous la forme conservative donnée par l'équation 1.17.

$$\partial_t W + \partial_x F(W) = B(W). \tag{1.24}$$

1.2.3 Courbes caractéristiques

L'équation (1.17) est dite hyperbolique si la matrice jacobienne du terme $F(W)$ admet des valeurs propres réelles. Pour le cas des équations de Saint-Venant, la matrice jacobienne

du terme de flux s'écrit de la manière suivante :

$$\left(\frac{\partial F(W)}{\partial W}\right) = \begin{pmatrix} 0 & 1 \\ -\frac{Q^2}{A_h^2} + \frac{\partial P}{\partial A_h} & \frac{2Q}{A_h} \end{pmatrix}. \tag{1.25}$$

La matrice jacobienne (1.25) possède deux valeurs propres réelles si le terme $\frac{\partial P}{\partial A_h}$ est positif, i.e si la hauteur d'eau est positive. Les deux valeurs propres de la matrice sont :

$$\begin{cases} \lambda^+ = \frac{Q}{A_h} + c \\ \lambda^- = \frac{Q}{A_h} - c \end{cases} \tag{1.26}$$

avec

$$c = \sqrt{\frac{\partial P}{\partial A_h}}. \tag{1.27}$$

Pour le cas d'une géométrie rectangulaire, on peut montrer que $c = \sqrt{gh}$ (on renvoie le lecteur à [136] pour la construction de ces équations pour un cas mono-dimensionnel). Les deux valeurs propres définissent la pente des courbes caractéristiques :
-la courbe caractéristique $\mathcal{C}^+$ est définie par l'équation suivante : $\frac{dx}{dt} = \lambda^+$,
-la courbe caractéristique $\mathcal{C}^-$ est définie par l'équation suivante : $\frac{dx}{dt} = \lambda^-$.
Les courbes caractéristiques présentent les trajectoires possibles ainsi que la vitesse qu'une

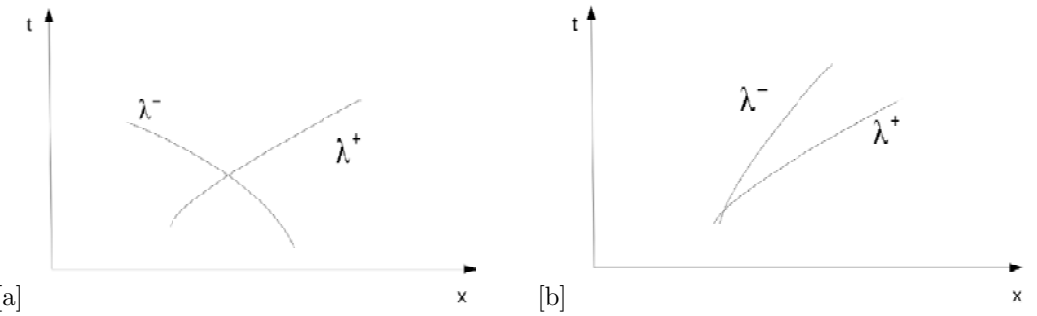


FIGURE 1.2 – *Les régimes d'écoulement donnés par la théorie des caractéristiques : [a] régime fluvial, [b] régime torrentiel*

information peut avoir lors d'un écoulement. Ces caractéristiques sont intéressantes pour déterminer les conditions aux limites. La courbe $\mathcal{C}^-$ définit le régime de l'écoulement. En effet, si la vitesse λ^- est positive le régime d'écoulement est torrentiel (figure 1.2 [b]), aucune information ne remonte la rivière ; toutes les informations se dirigent de l'amont vers l'aval avec deux vitesses différentes : λ^+ et λ^-. Pour la modélisation numérique ce régime impose de donner deux conditions aux limites à l'amont seulement. Réciproquement, si la vitesse λ^- est négative, le régime d'écoulement est fluvial (figure 1.2 [a]), une information remonte de l'aval vers l'amont à la vitesse λ^- et descend vers l'aval à la vitesse λ^+. En modélisation, cela veut dire que pour le régime fluvial, on a besoin de deux conditions aux limites différentes une à l'aval et une à l'amont.

Modélisation numérique des équations de Saint-Venant

2

2.1 Introduction

La modélisation numérique des équations de Saint-Venant a connu différents développements au cours des dernières années. Plusieurs méthodes de résolution numérique des équations de Saint-Venant s'appuient sur les développements déjà utilisés en méthodes numériques pour les équations de Navier-Stokes [80]. Dans les années 90, plusieurs études se sont intéressées à la résolution directe des équations de Saint-Venant par différentes méthodes [39], [44], [2], [127], etc.... Aujourd'hui il existe plusieurs codes utilisés en industrie comme en recherche qui résolvent les équations de Saint-Venant (Mascaret [53], Telemac [59], MIKE11 [108], MIKE21 [146], HEC-RAS [22], etc.) pour des applications pratiques : écoulements fluviaux, écoulements côtiers [124], rupture de barrages [51], transport sédimentaires [141], etc. Dans cette thèse, on a utilisé le code industriel Mascaret développé par EDF/CEREMA.
Les méthodes utilisées pour la résolution numérique des équations de Saint-Venant sont : les différences finies [150], les volumes finis [21, 1], les éléments finis [61], les méthodes Lagrangiennes [142, 63], les méthodes spectrales [74]. Il est cependant possible d'utiliser des méthodes hybrides qui consistent à utiliser plus d'une méthode de résolution dans le même code pour traiter différents termes des équations de Saint-Venant afin d'améliorer la robustesse d'un code. On s'intéressera, dans cette partie en particulier, à la méthode des volumes finis en utilisant l'approximation de Riemann de Roe, car c'est une des méthodes utilisée dans le cas de Mascaret.

2.2 Volumes finis

La formulation volumes finis est décrite par rapport à un bilan local des flux dans un petit volume de contrôle. Elle est caractérisée par sa formulation conservative (d'où la nécessité d'utiliser la forme conservative des équations de Saint-Venant), et a démontré son efficacité pour la résolution des problèmes hyperboliques.

On applique une intégrale de surface $\mathcal{S}$ sur l'équation 1.17 :

$$\int_{\mathcal{S}} \partial_t W d\mathcal{S} + \int_{\mathcal{S}} div(F(W)) d\mathcal{S} = \int_{\mathcal{S}} B(W) d\mathcal{S}. \tag{2.1}$$

Supposons que $\mathcal{V}_i$ est un petit volume de contrôle, on applique le théorème de Green sur l'équation 2.1 et on obtient :

$$\int_{\mathcal{S}_i} \frac{dW}{dt} d\mathcal{S}_i + \oint_{\partial \mathcal{V}_i} F(W) \vec{n} d\mathcal{L}_i = \int_{\mathcal{S}_i} B(W) d\mathcal{S}_i, \tag{2.2}$$

où $\vec{n}$ est le vecteur unitaire normal à la frontière, et $d\mathcal{L}_i$ la longueur de la frontière. En pratique on approche le deuxième terme de l'équation 2.2 par la formule suivante :

$$\oint_{\partial \mathcal{V}_i} F(W) \vec{n} d\mathcal{L}_i = F(W)_{i+1/2} - F(W)_{i-1/2}, \tag{2.3}$$

où $F(W)_{i+1/2}$ représente le flux normal traversant la face $i+1/2$. Afin de résoudre numériquement l'équation (2.1), il suffit de trouver une solution approchée du terme $F(W)_{i+1/2}$.

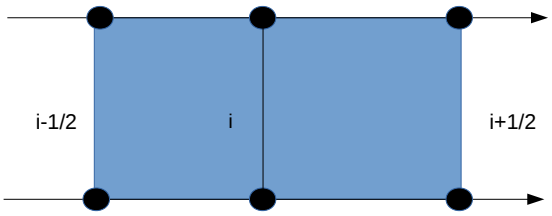


FIGURE 2.1 – *Schéma représentant le calcul d'une cellule en volumes finis*

2.3 Solveur de l'approximation de Riemann de Roe

Le solveur de l'approximation de Riemann proposé par Roe permet de calculer le flux $F(W)_{i+1/2}$ dans l'équation 2.3 :

$$F(W)_{i+1/2} = \frac{F(W)_{i+1} + F(W)_i - |\tilde{\mathcal{Q}}_{i+1/2}|(W_{i+1} - W_i)}{2} \tag{2.4}$$

où $\mathcal{Q}_{i+1/2}$ est l'approximation du jacobien du flux à la normale, défini par l'équation 1.25. Dans son papier Roe [117] approche le terme $\tilde{\mathcal{Q}}_{i+1/2}$ en moyennant les termes de l'équation, et la matrice de Roe s'écrit donc :

$$\tilde{\mathcal{Q}} = \begin{pmatrix} 0 & 1 \\ -\tilde{u}^2 + \tilde{c} & \tilde{u} \end{pmatrix} \tag{2.5}$$

avec $\tilde{u} = \frac{\sqrt{A_{h_{i+1}}}u_{i+1}+\sqrt{A_{h_i}}u_i}{\sqrt{A_{h_{i+1}}}+\sqrt{A_{h_i}}}$ et $\tilde{A_h} = \frac{\sqrt{A_{h_{i+1}}}A_{h_{i+1}}+\sqrt{A_{h_i}}A_{h_i}}{\sqrt{A_{h_{i+1}}}+\sqrt{A_{h_i}}}$ où $u = \frac{Q}{A_h}$ et c désigne la célérité.
Les valeurs propres de la matrice $\tilde{Q}$ sont de la forme :

$$\begin{array}{l} \lambda_1 = \frac{Q}{A_h} + c, \\ \lambda_2 = \frac{Q}{A_h} - c, \end{array} \tag{2.6}$$

et les vecteurs propres s'écrivent de la façon suivante :

$$r^1 = \begin{pmatrix} 1 \\ \tilde{u} + \tilde{c} \end{pmatrix} \quad r^2 = \begin{pmatrix} 1 \\ \tilde{u} - \tilde{c} \end{pmatrix} \tag{2.7}$$

ainsi on peut écrire les termes W_{i+1} et W_i dans une base générée par les vecteurs propres r^1 et r^2 :

$$\begin{array}{l} W_{i+1} = \alpha^1_{i+1} r^1 + \alpha^2_{i+1} r^2, \\ W_i = \alpha^1_i r^1 + \alpha^2_i r^2, \end{array} \tag{2.8}$$

ce qui permet d'obtenir la valeur de $W_{i+1/2}$ avec la résolution du problème de Riemann

$$W_{i+1/2} = \alpha^1_i r^1 + \alpha^2_{i+1} r^2. \tag{2.9}$$

Finalement on retrouve l'expression du flux numérique $F(W)_{i+1/2}$

$$F(W)_{i+1/2} = \frac{1}{2}[F(W)_{i+1} + F(W)_i - (\alpha^1_i \lambda_1 r^1 + \alpha^2_{i+1} \lambda_2 r^2)]. \tag{2.10}$$

Le solveur d'approximation de Riemann de Roe a été proposé initialement pour la résolution de l'équation d'Euler (qui ressemble aux équations de Saint Venant pour le cas d'un écoulement sur un canal prismatique). La méthode a prouvé sa robustesse pour simuler correctement les écoulements discontinus et les chocs. Cependant, des oscillations indésirables peuvent se produire si les termes sources sont importants (discontinuités provoquées au voisinage des points soniques). Afin de traiter ces cas particuliers, on peut changer de méthode de résolution des équations de Saint-Venant en remplaçant la méthode d'approximation de Riemann de Roe par une méthode de projection (travaux de thèse de Yu-e [126]). Une deuxième méthode consiste à modifier le calcul des flux aux voisinages de ces points en ajoutant une correction entropique à la valeur propre.

2.4 Termes sources

Comme cité précédemment, la résolution des équations de Saint-Venant est très sensible aux termes sources. En effet, ils sont responsables, avec les conditions aux limites, des discontinuités et des chocs qui peuvent apparaître lors d'un écoulement en rivière. C'est la raison pour laquelle il est important d'avoir un bon traitement numérique des termes sources. De façon générale, les équations de Saint-Venant présentent trois termes sources : les apports des débits, la géométrie et le frottement. En reprenant l'équation (2.1), on a vu que l'on n'a

pas traité le terme $B(W)$ qui comporte les termes source.
Il reste à déterminer le deuxième terme de l'équation (2.2). Une des méthodes utilisées afin de discrétiser le terme source dans un schéma volumes finis en utilisant le schéma de Roe est celle présentée dans [143]. Afin d'estimer l'intégrale on la divise en deux parties :

$$\int_{\mathcal{S}_i} B(W)d\mathcal{S}_i = \int_{\mathcal{S}_i G} B(W)d\mathcal{S}_i + \int_{\mathcal{S}_i D} B(W)d\mathcal{S}_i, \tag{2.11}$$

où S_iD et S_iG sont les deux moitiés des éléments $i+1/2$ et $i-1/2$ comme montré sur la figure 2.1. Ainsi la moyenne du terme source est estimée à partir d'un calcul barycentrique sur le terme source des éléments aux frontières.

$$B(W_i) = \frac{1}{A_{h_i}}[A_{h_{i-1/2}}\psi_D(W_i, W_{i-1}) + A_{h_{i+1/2}}\psi_G(W_{i+1}, W_i)]. \tag{2.12}$$

Finalement, il suffit d'estimer les deux grandeurs $\psi_D(W_i, W_{i-1})$ et $\psi_G(W_{i+1}, W_i)$ afin d'estimer le terme source discret. L'article [9] propose une écriture de ces deux termes de la façon suivante :

$$\begin{aligned} \psi_D(W_i, W_{i-1}) &= [I + |\tilde{\mathcal{Q}}_{i-1/2}|\tilde{\mathcal{Q}}^{-1}_{i-1/2}]B(W_{i-1/2}), \\ \psi_G(W_{i+1}, W_i) &= [I - |\tilde{\mathcal{Q}}_{i+1/2}|\tilde{\mathcal{Q}}^{-1}_{i+1/2}]B(W_{i+1/2}). \end{aligned} \tag{2.13}$$

2.5 Conditions aux limites

Comme cité dans le paragraphe 1.2.3, les conditions aux limites sont déterminées par la théorie des caractéristiques. Ainsi, selon le régime de l'écoulement, la nature des conditions aux limites que l'on doit imposer change. Afin d'évaluer les flux numériques, on a montré dans le paragraphe 2.3 que l'on a besoin d'avoir le flux aux deux frontières à l'amont et à l'aval. Le calcul d'un flux suppose la connaissance de deux données : la hauteur d'eau (où la côte d'eau) et le débit.

3 Simplification des équations de Saint-Venant

3.1 Introduction

Les sections précédentes ont montré la complexité de la résolution des équations de Saint-Venant. Leur résolution numérique est un progrès qui a pu se faire avec l'avancée dans l'ère du numérique. Avant cette avancée technologique des équations plus simples que les équations de Saint-Venant étaient utilisées sous quelques hypothèses faites sur la nature de l'écoulement. Ces équations représentent une simplification des équations de Saint-Venant. On s'intéressera dans cette partie à présenter quelques équations simplifiées des équations de Saint-Venant, que l'on utilisera dans cette thèse par la suite.

3.2 Classification des régimes d'écoulement

Afin de simplifier les équations de Saint-Venant, on s'intéressera à la physique de l'écoulement qui permet de négliger quelques termes selon le régime de l'écoulement. On retiendra la classification que propose [27] dans son ouvrage, car elle nous semble la plus complète :

- Régime stationnaire
  - Régime uniforme
  - Régime graduellement varié
  - Régime rapidement varié
- Régime instationnaire
  - Régime instationnaire uniforme
  - Régime instationnaire graduellement varié
  - Régime instationnaire rapidement varié

La figure 3.1 montre une visualisation des différents régimes d'écoulements selon la classification ci-dessus.

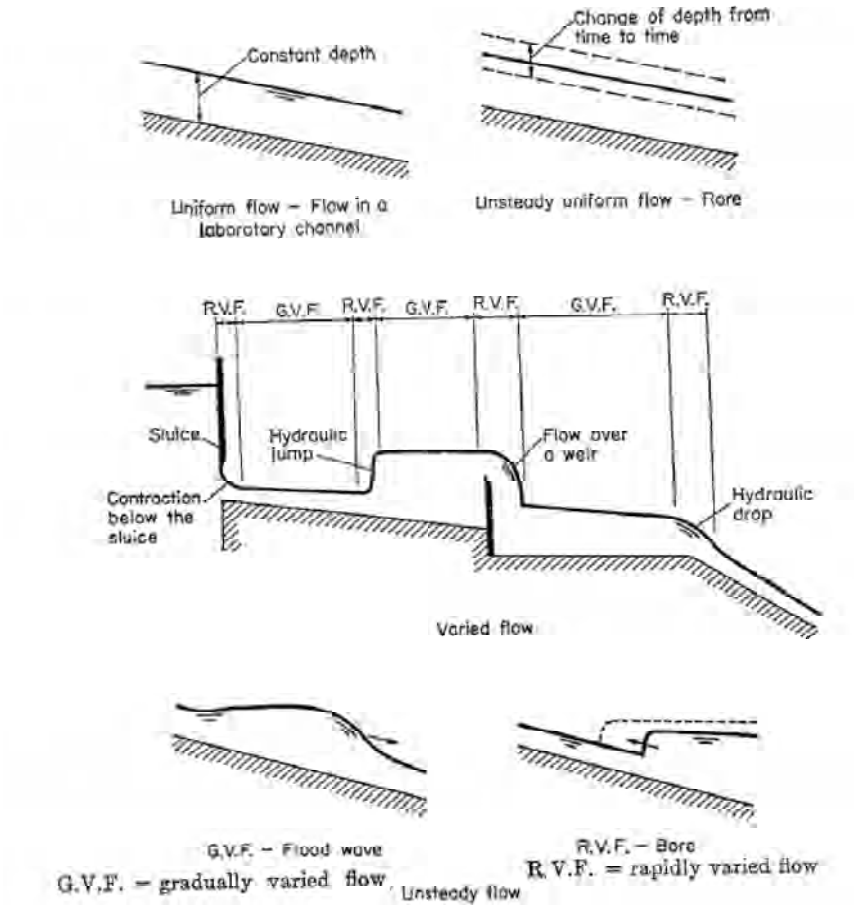


FIGURE 3.1 – *Les différents régimes d'écoulement, source : [27]*

3.3 Régime stationnaire

Un écoulement suit un régime stationnaire si la côte d'eau ne change pas durant le temps. En négligeant les termes qui ont une dépendance temporelle, l'équation 1.15 devient alors de la forme suivante :

$$\begin{cases} \partial_x Q = 0, \\ \partial_x(\frac{Q^2}{A_h}) = -gA_h\partial_x z - gA_h J. \end{cases} \tag{3.1}$$

L'équation de continuité exprime la continuité du débit à travers la rivière. Afin de simplifier la deuxième équation on introduit le facteur de planimétrie $b_h = \frac{dA_h}{dh}$. L'équation de conservation de quantité de mouvement devient alors :

$$-\frac{Q^2}{A_h^2}\partial_x A_h + \frac{gA_h}{b_h}\partial_x(A_h) = -gA_h\partial_x(Z_f) - gA_h J. \tag{3.2}$$

Afin d'alléger les notations on note $I = -\partial_x Z_f$ qui est la pente du fond. L'équation (3.2) devient :

$$\frac{g}{b_h}(1 - \frac{b_h Q^2}{g A_h^3})\partial_x(A_h) = g(I - J). \tag{3.3}$$

3.3.1 Nombre de Froude

L'écoulement de l'eau en rivière est majoritairement piloté par la viscosité (les frottements au fond) et la gravité. Le paramètre qui permet de quantifier les effets de viscosité en mécanique des fluides est le nombre de Reynolds. Ce nombre traduit l'importance des effets du frottement vis à vis des termes d'inertie, il est décrit par l'équation ci-dessous :

$$R_e = \frac{Q R_h}{\nu_e A_h}, \tag{3.4}$$

où ν_e est la viscosité cinématique. Selon le rapport entre les forces d'inertie et les forces de frottements, l'écoulement sera appelé laminaire si l'écoulement est piloté principalement par les frottements et turbulent si les forces d'inertie sont prépondérantes.
Comme les forces de gravité ont aussi un effet considérable sur l'écoulement, William Froude introduit un nombre qui permet de comparer l'effet de l'énergie potentielle à l'énergie cinétique et donc de comparer les forces de l'inertie à la gravité pour un écoulement. On peut écrire le nombre de Froude que l'on notera Fr par la suite de la manière suivante :

$$Fr = \frac{Q\sqrt{b_h}}{g\sqrt{A_h^3}}. \tag{3.5}$$

Le nombre de Froude traduit la relation d'équilibre entre les forces de gravité et les forces d'inertie, la hauteur d'eau qui résulte de cet équilibre est appelée la hauteur critique notée h_c. Dans la littérature, beaucoup d'études comparent le nombre de Froude au nombre de Mach en acoustique. D'où l'appellation du régime super-critique pour $Fr > 1$ et sous-critique pour $Fr < 1$. Ces deux régimes sont aussi connus sous les noms de torrentiel et fluvial respectivement (c'est la notation que l'on gardera pour la suite de ce manuscrit).
L'introduction du nombre de Froude permet de simplifier l'équation (3.3) :

$$(1 - Fr^2)\partial_x(A_h) = b_h(I - J). \tag{3.6}$$

3.3.2 Régime uniforme

On appelle le régime uniforme, un régime pour lequel l'écoulement est parallèle à la pente, i.e : il n'y a pas de variabilité spatiale, la hauteur d'eau reste la même le long de la rivière. Ce régime est établi quand $\partial_x A_h = 0$ et donc $I = J$. Pour ce régime il y a un équilibre entre les forces dues au frottement et les forces dues à la gravité. La hauteur d'eau correspondant à ce régime est appelée la hauteur normale notée h_n, et ne dépend que du débit du fluide, de la forme de la section ainsi que du coefficient de frottement. Pour un canal rectangulaire

de largeur L, elle est écrite sous la forme :

$$h_n = (\frac{Q^2}{IL^2{K_s}^2})^{3/10} \tag{3.7}$$

3.3.3 Courbes de remous (régime stationnaire graduellement varié)

Quand la hauteur d'eau est différente de la hauteur critique l'écoulement suit la forme donnée par les courbes de remous décrites par l'équation (3.6), avec l'hypothèse ($b_h \neq 0$:

$$\partial_x(h) = \frac{I - J}{1 - Fr^2}. \tag{3.8}$$

Une deuxième méthode pour obtenir cette équation est d'utiliser l'équation de Bernoulli en régime stationnaire et de considérer que la perte de charge spécifique n'est due qu'aux frottements au fond, on peut retrouver le détail de cette méthode dans [136]. Les courbes de remous permettent de décrire les solutions stationnaires de l'équation de Saint-Venant. Il existe six formules différentes de solutions (3 en régime fluvial et 3 en régime torrentiel) que l'on classifiera sous 2 classes : faible pente et forte pente. On distingue les deux régimes par les valeurs de h_n et h_c : le régime faible pente (respect. forte pente) est défini quand $h_n > h_c$ (respect. $h_n < h_c$). Afin de simplifier les explications on remplace l'équation 3.8 par l'équation démontrée par [136] :

$$\partial_x(h) = \mathcal{F}(h) = I\frac{1 - (h/h_n)^{-10/3}}{1 - (h/h_c)^{-3}}. \tag{3.9}$$

Faible pente

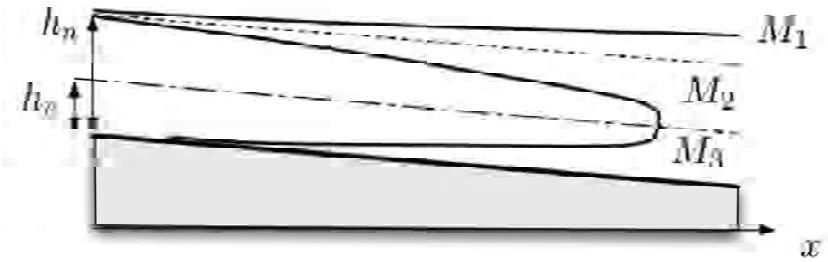


FIGURE 3.2 – *Les courbes de remous de faible pente, source : [136]*

Le régime faible pente est défini quand la hauteur normale h_n est supérieure à la hauteur critique h_c. Trois courbes de remous sont définies alors et sont de type M "*mild*" et sont présentées dans la figure 3.2 :

— Quand $h > h_n > h_c$, le terme $1-(h/h_n)^{-10/3}$ est négatif ainsi que le terme $1-(h/h_c)^{-3}$. Le terme définissant $\partial_x h$ est positif, donc la hauteur est une fonction croissante de x. Quand $h/h_n \to \infty$ la fonction $\mathcal{F}(h) \to 0$ (dans l'équation 3.9) donc à l'aval la hauteur d'eau possède une tangente horizontale. Cette courbe est appelée M_1.

— Quand $h_n > h > h_c$, la fonction $\mathcal{F}(h)$ est négative, donc la courbe de la hauteur d'eau est décroissante de x. Quand $h/h_n \to \infty$ la fonction $\mathcal{F}(h) \to \infty$, la hauteur d'eau prend une tangente verticale quand elle arrive à la valeur de la hauteur critique h_c. Cette courbe est appelée M_2.

— Finalement quand $h_n > h_c > h$, la fonction de la dérivée de la hauteur d'eau est positive, donc la hauteur d'eau est croissante, à l'aval il y aura formation d'un ressaut hydraulique car passage de régime torrentiel à fluvial. Cette courbe est appelée M_3.

Forte pente



FIGURE 3.3 – *Les courbes de remous de forte pente, source : [136]*

Le régime forte pente est défini quand la hauteur normale h_n est inférieure à la hauteur critique h_c 3.3. Trois courbes de remous sont définies alors et sont de type S "*steep*" et sont présentées dans la figure 3.3

— Quand $h > h_c > h_n$, le terme $1-(h/h_n)^{-10/3}$ est négatif ainsi que le terme $1-(h/h_c)^{-3}$. Donc le terme définissant $\partial_x h$ est positif, donc la hauteur est une fonction croissante. Quand $h/h_n \to \infty$ la fonction $\mathcal{F}(h) \to 0$ (dans l'équation 3.9) donc à l'aval la hauteur d'eau possède une tangente horizontale. Elle a aussi tendance à avoir une tangente verticale à l'amont. Cette courbe est appelée S_1.

— Quand $h_c > h > h_n$, la fonction $\mathcal{F}(h)$ est négative, donc la courbe de la hauteur d'eau est décroissante. La courbe est tangente à h_n à l'aval et elle a une tangente verticale à l'amont, la courbe est nommée S_2.

— Finalement quand $h_c > h_n > h$, la fonction de la dérivée de la hauteur d'eau est positive, donc la hauteur d'eau est croissante, à l'aval la courbe est tangente à h_n. Cette courbe est la courbe S_3.

3.3.4 Régime brusquement varié

Le régime brusquement varié (R.V.F dans la figure 3.1) est un phénomène local où la hauteur d'eau change brusquement dans une petite distance. Quelques exemples de régimes brusquement variés sont les ressauts hydrauliques et les chutes d'eau. Sur la figure 3.4 on remarque que le passage de la courbe S_2 à la courbe M_3 se fait sur une petite distance que l'on arrive à peine à la déterminer à l'œil, c'est le régime brusquement varié, on peut le

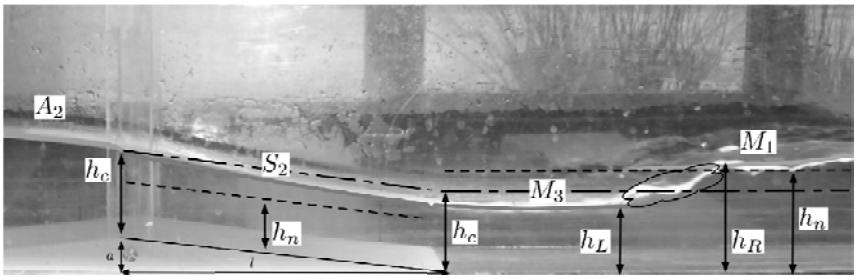


FIGURE 3.4 – *Exemple de passage des écoulements graduellement variés et brusquement variés, source : [136]*

remarquer aussi sur la zone du ressaut hydraulique et le passage de la courbe M_3 à la courbe M_1.

Ressaut hydraulique

Le ressaut hydraulique est défini mathématiquement comme une discontinuité, car la hauteur d'eau passe brusquement de la hauteur h_L à la hauteur h_R (figure 3.4). Physiquement, il s'agit d'une zone tourbillonnaire qui se caractérise par une variation brutale de la hauteur d'eau. Le ressaut hydraulique apparait lors du passage d'un écoulement torrentiel à un écoulement fluvial quand le nombre de froude $Fr = 1$. L'équation de la courbe de remous 3.8 n'est alors plus définie d'où la discontinuité mathématique.

3.4 Régime instationnaire

L'étude du régime instationnaire en hydraulique est nécessaire pour bon nombre d'applications telles que la prévision des crues. On appelle un régime instationnaire quand l'état hydraulique (débit et/ou la hauteur d'eau) change en fonction du temps. Dans ce cadre il existe plusieurs équations simplificatrices des équations de Saint-Venant afin d'étudier la propagation des ondes en rivière.

3.4.1 Équations de Saint-Venant uni-dimensionnelles

Lorsque la largeur de la rivière est très grande (par rapport à la longueur) nous pouvons supposer que l'écoulement est uni-dimensionnel et donc que le rayon hydraulique est égal à la hauteur d'eau. Ceci permet de simplifier l'équation 1.15. Dans [136], les équations de Saint-Venant uni-dimensionnelles sur un canal à fond plat sont présentées de la manière suivante :

$$\begin{cases} \partial_t h + \partial_x(hu) = 0, \\ \partial_t(hu) + u\partial_x(u) + g'\partial_x h = gI - \frac{gu|u|}{Ks^2h^{4/3}}. \end{cases} \tag{3.10}$$

où $g' = gcos(\alpha)$, α étant l'angle que fait le canal avec l'axe horizontal. La paramétrisation du frottement est faite à partir de l'équation de Manning-Strickler (1.13).

3.4.2 Approximation ondes de crues

La dynamique de rivière décrite par les équations de Saint-Venant peut être associée à une dynamique d'onde de crue lorsqu'on s'intéresse à des petites perturbations $(\tilde{h}, \tilde{u})$ autour d'un état d'équilibre $(\hat{h},\hat{u})$. On définit ainsi les variables d'état de la manière suivante :

$$\begin{array}{ll} h = \hat{h} + \tilde{h}, & \text{la hauteur d'eau,} \\ u = \hat{u} + \tilde{u}, & \text{la vitesse moyenne de l'écoulement.} \end{array} \tag{3.11}$$

On se place ici dans le cadre de "l'approximation d'ondes de crues". Celle ci est définie par [136] lorsqu'on néglige l'accélération horizontale et le terme de pression. Dans ce cadre d'approximation les équations 3.10 s'écrivent de la manière suivante :

$$\begin{cases} \partial_t h + \partial_x(hu) = 0 \\ 0 = gI - \frac{gu|u|}{Ks^2h^{4/3}}. \end{cases} \tag{3.12}$$

On s'intéresse aux régimes où $u \geq 0$, la résolution de l'équation de quantité de mouvement dans l'équation 3.12 donne alors la relation de Strickler définie par :

$$u(h) = KsI^{1/2}h^{2/3}. \tag{3.13}$$

En considérant un état d'équilibre $(\hat{h},\hat{u})$, où $\hat{h}$ et $\hat{u}$ sont deux constantes vérifiant la relation de Strickler 3.13 ($\hat{u} = KsI^{1/2}\hat{h}^{2/3}$) et en reportant dans l'équation de continuité puis en négligeant les termes d'ordre 2, on obtient :

$$\partial_t\tilde{h} + \frac{5\hat{u}}{3}\partial_x(\tilde{h}) = 0. \tag{3.14}$$

L'équation 3.14 est une équation d'advection. La perturbation $\tilde{h}$ est advectée à la vitesse $\frac{5\hat{u}}{3}$ par rapport à l'écoulement moyen.

3.4.3 Ondes de crues diffusantes

Lorsque l'on ajoute aux hypothèses de l'approximation des ondes de crues, le fait que la pente du canal (ou de la rivière) peut être négligée, on obtient les équations d'ondes de crues diffusantes. L'équation 3.10 devient alors :

$$\begin{cases} \partial_t h + \partial_x(hu) = 0, \\ g'\partial_x h = gI - \frac{gu|u|}{Ks^2h^{4/3}}. \end{cases} \tag{3.15}$$

Sachant que la relation de Strickler est toujours vérifiée pour l'état d'équilibre, l'équation 3.14 devient :

$$\partial_t\tilde{h} + \frac{5\hat{u}}{3}\partial_x(\tilde{h}) = \frac{\hat{u}\hat{h}}{2tan(\alpha)}\partial_x^2\tilde{h}. \tag{3.16}$$

On remarque l'ajout du terme supplémentaire qui représente une diffusion. La petite perturbation sera advectée ainsi que diffusée (son amplitude sera diminuée et sa longueur de portée sera augmentée).

III

Quantification d'incertitudes

Introduction

Le but de la quantification des incertitudes est d'identifier les sources d'incertitudes prépondérantes dans un modèle physique (e.g paramètres, forçage externes, conditions de surface, condition initiales) et d'analyser leur impact sur les quantités d'intérêt simulées par le modèle (variables pronostiques, probabilité de dépassement de seuil). Ceci permet d'associer à chaque prévision une incertitude. Ce champ a reçu une attention particulière durant ces dernières années car la précision du résultat de la simulation dépend significativement du degré de certitude (ou d'incertitude) que l'on porte aux données d'entrée. Par conséquent, afin de mieux comprendre les résultats des simulations numériques et par la suite améliorer la prévision de la physique, il est nécessaire de prendre en compte ces incertitudes dans les simulations. Un problème de quantification des incertitudes est un problème qui vise à estimer l'incertitude sur les sorties d'une simulation numérique en fonction des incertitudes sur la connaissance de ses paramètres en entrée.

Représentation stochastique de l'incertitude

À cause de la nature "aléatoire" de l'incertitude, l'approche probabiliste pour traiter un problème de quantification d'incertitudes est de considérer les données incertaines du modèle comme des variables aléatoires ou des processus aléatoires, et de reconsidérer le modèle numérique réel déterministe comme un modèle stochastique (Figure 3.5 + étape A). Il existe différentes manières de quantifier les incertitudes d'un modèle numérique et d'étudier la sensibilité à un paramètre. Une des techniques les plus populaires utilisée en quantification d'incertitudes est l'échantillonnage de Monte-Carlo. Cette méthode consiste à générer des réalisations (indépendantes) des paramètres aléatoires basées sur leurs densités de probabilités. Pour chaque réalisation, la valeur du paramètre est fixée et le problème devient déterministe. En résolvant chaque problème déterministe correspondant à chaque réalisation, on obtient un ensemble de solutions. À partir de cet ensemble, il est possible d'extraire différentes informations statistiques telles que la moyenne, la variance ou la distribution de probabilité de la variable d'intérêt. Bien que la méthode Monte-Carlo soit une approche directe, robuste et simple à mettre en place, elle souffre d'une convergence assez lente (on a besoin d'un grand nombre de simulations afin d'atteindre une estimation précise des statistiques) [41]. Afin de surmonter ce problème, d'autres techniques d'échantillonnage ont été développées. Celles-ci visent à accélérer la convergence de la méthode brute comme l'échantillonnage hypercube latin (LHS) [83], ou l'échantillonnage quasi Monte-Carlo [42]. Cependant, d'autres difficultés liées au design de la méthode rendent leur mise en pratique

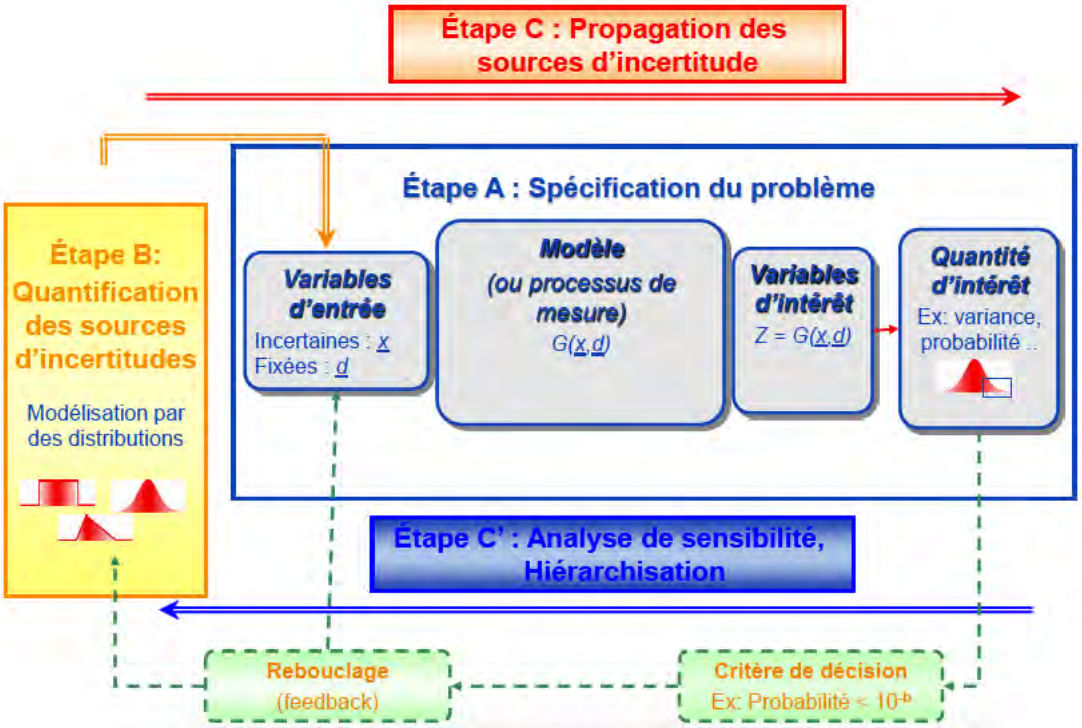


FIGURE 3.5 – *Méthode de quantification des incertitudes selon [66].*

difficile.

Il existe des alternatives aux méthodes d'échantillonnage qui se basent sur la réécriture du modèle numérique (méthodes intrusives), telles que les méthodes de perturbations, les équations de moments et les méthodes basées sur les opérateurs. On renvoie le lecteur vers [152] pour plus de détails sur ces méthodes qui ne sont pas traitées dans cette thèse. Une dernière approche plus récente est d'utiliser un méta-modèle. Celui-ci peut être construit en réecrivant les équations du modèle numérique (approche intrusive) ou en utilisant quelques évaluations du modèle (approche non-intrusive).

Un méta-modèle, aussi appelé modèle réduit, surface de réponse[134] ou modèle de substitution (*surrogate model*) [31] a pour objectif d'imiter le comportement général du modèle déterministe à un coût de calcul moindre ; ce qui est utile pour mener des études de propagation d'incertitudes et d'analyse de sensibilité. Il existe une multitude de méthodes pour construire un métamodèle dans la littérature. On s'intéresse dans cette thèse en particulier aux polynômes du chaos. Cette approche permet de représenter intrinsèquement la quantité d'intérêt comme une fonction polynomiale des variables en entrée.

L'étude d'incertitudes et l'exploration des codes numériques en utilisant les polynômes du chaos ont largement été détaillées dans la littérature [11, 132, 149]. La majorité de ces études traitent de codes numériques résolvant des équations de la mécanique des structures via différentes méthodes (différences finies, volumes finis, éléments finis, etc...). Ces équations sont pour la plupart des équations différentielles ordinaires ou des équations aux dérivées partielles de nature elliptique ou parabolique. Cependant, peu de travaux traitent de l'ex-

pansion en polynômes de chaos des équations de nature hyperbolique telles que les équations de Saint-Venant [105].

Réduction de l'incertitude via l'assimilation de données

Une fois l'incertitude quantifiée et le(s) paramètre(s) le(s) plus influent(s) déterminé(s), à l'aide des informations issues de la mesure expérimentale (que l'on nommera une observation), on réduit l'incertitude associée pour fournir des prévisions plus fiables, recentrées autour d'une trajectoire plus réaliste. Il existe principalement deux grandes classes de méthodes inverses, utilisées en géosciences, qui visent à utiliser une observation pour améliorer la performance d'un modèle numérique. La première classe de méthodes consiste à corriger la distribution de probabilité par l'inférence bayesienne. Cette méthode n'est pas considérée dans cette thèse, mais on renvoie le lecteur vers [90, 91] pour plus de détails. La deuxième classe de méthodes permet de corriger les variables estimées et de réduire leur incertitude, que ce soit un paramètre ou l'état du modèle. On parle d'assimilation de données. Cette méthode a été développée pour les besoins opérationnels de la prévision numérique du temps et a été largement adaptée à d'autres domaines des géosciences (e.g. océanographie [111], hydrologie[24], incendie de forêt [115], chimie atmosphérique [92], etc). Il existe une multitude d'algorithmes dont le filtre de Kalman et ses variantes telles que le filtre de Kalman d'ensemble (EnKF) [37]. Ce dernier se base sur l'échantillonnage Monte-Carlo pour formuler la matrice de gain de Kalman de manière stochastique. [79] montre dans son ouvrage que l'utilisation des polynômes du chaos pour estimer les statistiques permet d'accélérer l'algorithme EnKF (en réduisant le nombre d'appel au code numérique) et aussi d'améliorer la qualité de ces résultats (en réduisant l'erreur numérique due à l'échantillonnage).

Plan de la partie II

Cette partie sur la quantification d'incertitudes se présente comme suit. On rappelle, dans un premier chapitre (chap 4), les notions de probabilités dont on a besoin en quantification des incertitudes et d'analyse de sensibilité. On définit les notions de variables aléatoires, d'espace de probabilités et des moments stochastiques associés, ainsi que les méthodes stochastiques pour la quantification d'incertitudes et l'analyse de sensibilité. Le deuxième chapitre (chap 5) vise à décrire les méthodes pour la réduction de dimension des champs aléatoires. Le troisième chapitre (chap 6) introduit la réduction de modèle par l'approche en polynômes du chaos et leur usage pour traiter les problèmes de quantification d'incertitudes et d'analyse de sensibilité. Finalement, le dernier chapitre de cette partie (chap 7) est dédié à l'assimilation de données. On s'intéresse particulièrement au filtre de Kalman d'ensemble et on montre comment l'utilisation des polynômes du chaos permet de réduire le coût de calcul de l'algorithme de l'EnKF dans une optique d'assimilation de données.

4 Généralités sur les méthodes stochastiques en quantification d'incertitudes et en analyse de sensibilité

4.1 Variable aléatoire et espace de probabilité associé

4.1.1 Définitions

On appelle un phénomène aléatoire tout mécanisme aléatoire complexe et qui ne peut être prédit, par exemple la taille d'un étudiant de l'INP Toulouse, son âge ou son poids. Une observation de ce phénomène sera appelée *une épreuve* que l'on note ω. Le traitement mathématique d'une épreuve requiert que l'on lui attribue un chiffre, qui représente une *variable aléatoire*. Supposons à titre d'exemple que le phénomène aléatoire observé est le choix d'un étudiant de l'INP Toulouse, suivi de l'attribution du chiffre 1 si c'est une fille et de 2 si c'est un garçon. On a donc construit la variable aléatoire considérée $X = X(\omega) \in \{1, 2\}$ où $\omega \in \Omega$ est un étudiant de l'ensemble Ω de tous les étudiants.
Tous les résultats possibles du tirage forment un ensemble noté Ω. En théorie des probabilités, on affecte pour chaque ensemble d'épreuves un nombre qui définit sa *probabilité d'occurrence* $\mathcal{P}$. On appelle événement A un sous-ensemble de Ω. Dans l'exemple précédent, un événement possible est "l'étudiant de l'INP Toulouse est une fille". L'ensemble de tous les événements possibles ayant une probabilité ainsi définie est appelé une $\sigma-$algèbre associée à Ω, et est noté $\mathcal{F}$. Finalement on définit l'espace de probabilité $(\Omega, \mathcal{F}, \mathcal{P})$.
Mathématiquement, une variable aléatoire est une application $X : (\Omega, \mathcal{F}, \mathcal{P}) \longrightarrow \mathcal{D}_X \subset \mathbb{R}$. De plus, lorsque l'ensemble des valeurs que peut prendre une variable aléatoire est fini ou infini dénombrable, la variable aléatoire est dite discrète. Dans le cas contraire, elle est appelée continue. Le cas des variables discrètes n'est pas considéré dans cette thèse.
Généralement, une variable aléatoire est caractérisée par sa fonction de répartition (CDF

cumulative density function) notée $F_X(x)$:

$$F_X(x) = \mathcal{P}(X \leqslant x) \quad \forall x \in \mathcal{D}_X. \tag{4.1}$$

Ainsi, la densité de probabilité (PDF *probability density function*) associée est définie par :

$$f_X(x) = \frac{dF_X(x)}{dx}. \tag{4.2}$$

Cette PDF donne la distribution des probabilités de la valeur aléatoire pour tous les événements.

4.1.2 Propriétés

La mesure de probabilité possède les propriétés suivantes :

$$\mathcal{P}(\emptyset) = 0 \tag{4.3}$$
$$\mathcal{P}(A) > 0 \qquad \forall A \in \mathcal{F} \tag{4.4}$$
$$\mathcal{P}(\Omega) = 1 \tag{4.5}$$
$$\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B) - \mathcal{P}(A \cap B) \quad \forall A, B \in \mathcal{F} \tag{4.6}$$
$$\mathcal{P}(A) < \mathcal{P}(B) \quad si \quad A \subset B \tag{4.7}$$
$$\mathcal{P}(\bar{A}) = 1 - \mathcal{P}(A) \quad si \quad A \in \mathcal{F}. \tag{4.8}$$

4.1.3 Moments statistiques

Les moments statistiques d'une variable aléatoire X sont des quantités caractéristiques importantes de sa distribution (en particulier de sa forme). On présente notamment dans ce paragraphe les notions de moyenne (ou d'espérance), de variance, de coefficients d'aplatissement et d'asymétrie.

Moyenne et variance. L'espérance mathématique $\mathbb{E}(X)$ est le moment d'ordre 1 ou la valeur moyenne μ_X telle que :

$$\mu_X = \mathbb{E}[X] = \int_\Omega X dX. \tag{4.9}$$

La variance (le moment centré d'ordre 2) de X décrit la variabilité ou la dispersion de la variable aléatoire autour de sa moyenne :

$$Var(X) = \int_\Omega (X - \mathbb{E}[X])^2 dX. \tag{4.10}$$

Plus généralement, le moment centré d'ordre n de X est défini comme suit :

$$m_n = \mathbb{E}(X^n) = \int_\Omega (X - \mathbb{E}[X])^n dX. \tag{4.11}$$

On notera que tous les moments peuvent être estimés de façon empirique. C'est l'objet de la section 4.1.4.

Écart type et coefficient de variation. L'écart-type et le coefficient de variation sont deux autres paramètres connus en statistiques pour mesurer la dispersion (comme la variance) d'une variable aléatoire.

— L'écart type σ_X d'une variable aléatoire X est la racine carrée de sa variance :

$$\sigma_X = \sqrt{Var(X)}. \tag{4.12}$$

— Le coefficient de variation (ou écart-type relatif) est un paramètre statistique adimensionnel donnant le rapport entre la moyenne et l'écart type.

$$CV(X) = \frac{\sigma_X}{\mu_X}. \tag{4.13}$$

Le plus souvent exprimé en %, le coefficient de variation permet de comparer deux variables aléatoires qui n'ont pas la même moyenne ou qui n'ont pas la même dimension. Ce coefficient devient sensible à manipuler quand la moyenne est proche de 0.

Coefficients d'asymétrie et d'aplatissement. Les moments centrés d'ordre 3 et 4 définissent respectivement les coefficients d'asymétrie et d'aplatissement de la distribution de X.

— Le coefficient d'asymétrie (*skewness*) δ_X est un paramètre de forme et mesure l'asymétrie de la distribution de la variable aléatoire X :

$$\delta_X = \mathbb{E}\big[(\frac{X-\mu_X}{\sigma_X})^3\big] = \frac{m_3}{Var(X)^{3/2}}. \tag{4.14}$$

— Le coefficient d'aplatissement (*kurtosis*) κ_X est un paramètre de forme. Il mesure l'aplatissement de la distribution, c'est-à-dire la disposition des masses de probabilité autour de leur centre :

$$\kappa_X = \mathbb{E}\big[(\frac{X-\mu_X}{\sigma_X})^4\big] = \frac{m_4}{Var(X)^2}. \tag{4.15}$$

4.1.4 Formulation d'un problème stochastique

Généralement lors d'un problème de quantification d'incertitude, on cherche à estimer l'incertitude d'une quantité d'intérêt (qui est la variable en sortie du modèle) par rapport à une variable en entrée incertaine. L'approche commune dans la communauté scientifique est de considérer que la quantité d'intérêt et la variable en entrée sont des variables aléatoires. L'incertitude de la quantité d'intérêt est alors traduite en terme de moments statistiques ou/et sa densité de probabilité. Autrement dit, un problème de quantification d'incertitude se résume à estimer les moments statistiques (pas tous ceux exprimés dans le paragraphe précédent obligatoirement) et/ou la densité de probabilité de la quantité d'intérêt, connaissant

l'incertitude (généralement densité de probabilité) sur la variable en entrée. Ce qui résume tout problème de quantification d'incertitude en trois étapes :

1. Caractériser l'incertitude de la variable en entrée.
2. Formuler un modèle stochastique à partir du modèle déterministe physique.
3. Estimer l'incertitude (moments statistiques et/ou densité de probabilité) de la quantité d'intérêt.

On remarque que l'étape 2 est primordiale afin de bien définir un problème stochastique qui permet d'étudier la quantification d'incertitude. En effet, il existe différentes approches dans la littérature qui permettent de formuler un modèle stochastique à partir d'un modèle déterministe (des exemples de méthodes peuvent être trouvés dans [131, 152]). L'approche que l'on utilisera principalement dans cette thèse est l'approche qui se base sur l'échantillonnage Monte-Carlo. Cette approche consiste à échantillonner la variable en entrée, et pour chaque échantillon de cette variable une simulation correspondante sera réalisée. L'ensemble des réalisations permettra d'estimer l'incertitude sur la variable d'intérêt.

Estimation des moments statistiques

Dans la méthode Monte-Carlo, les propriétés de la distribution d'une variable aléatoire sont examinées en utilisant des nombres aléatoires simulés. Historiquement, cette méthode a été utilisée la première fois dans les années 40, lors de la conception de la bombe nucléaire [58]. Von Neumann, Ulam, Fermi et Metropolis sont confrontés à d'importantes difficultés mathématiques pour modéliser la réaction en chaîne. Metropolis [95] propose alors de tirer au dé un premier noyau se désintégrant, puis de tirer le nombre de neutrons émis, puis leurs directions, etc. Ils produisent ainsi de grands nombres de réalisations du modèle et cernent ses propriétés. La méthode est baptisée Monte-Carlo en référence aux jeux de hasard. L'idée de base est d'explorer de façon aléatoire l'espace de probabilité induit par le problème de quantification. Chaque étape de cette exploration correspond à un échantillon. Ainsi, grâce à un grand nombre d'échantillons collectés, on approxime la solution du problème.
Les étapes pour l'estimation stochastique des moments statistiques se résument de la manière suivante :

1. Générer de manière identique et indépendante N_{sample} échantillons $x^{(i)}, i = 1, ..., N_{sample}$ suivant la densité de probabilité de X en utilisant un plan d'expérience convenable.
2. Pour chaque $x^{(i)}$, on calcule la réponse du modèle correspondant $\mathcal{M}(x^{(i)})$.
3. Estimer le moment statistique suivant la formule

$$\hat{m}_n = \frac{1}{N_{sample}} \sum_{i=1}^{N_{sample}} (\mathcal{M}(x^{(i)}) - \hat{\mu}_{\mathcal{M}(x^{(i)})})^n \approx m_n. \tag{4.16}$$

L'estimation de la moyenne par exemple est donnée par la formule suivante :

$$\hat{\mu} = \frac{1}{N_{sample}} \sum_{i=1}^{N_{sample}} (\mathcal{M}(x^{(i)}) \approx \mu. \tag{4.17}$$

La seule étape qui fait intervenir l'utilisation du modèle déterministe est l'étape 2, et elle demande d'effectuer plusieurs simulations du problème déterministe. L'avantage de cette méthode réside dans le fait que la taille du plan d'expérience ne dépend pas du nombre de variables aléatoires à l'entrée du modèle. De plus, lorsqu'on utilise un plan d'expérience basé sur l'échantillonnage Monte-Carlo, la convergence de l'erreur de l'estimation est inversement proportionnelle à la racine carrée du nombre d'échantillons N_{sample}, la méthode stochastique converge alors nécessairement (c.f. théorème centrale limite). Cependant, malgré la robustesse et la flexibilité de la méthode, la convergence reste très lente, le théorème centrale limite montre que l'erreur sur l'estimation des moment par voie stochastiques est $\frac{1}{\sqrt{N_{sample}}}$. Si par exemple on souhaite améliorer l'estimation d'un moment statistique d'un chiffre significative, il faudrait utiliser 100 fois plus de simulations. Ce qui, pour des codes assez coûteux, n'est pas envisageable. D'où l'utilisation de plans d'expériences moins coûteux tel que l'échantillonnage Latin Hypercube [83] (qui sont hors cadre de cette thèse).

Théorème central limite On présente ici le théorème central limite car c'est un théorème essentiel pour les études en analyse de probabilité généralement et plus particulièrement en quantification des incertitudes. Ce théorème montre en particulier que l'utilisation de l'échantillonnage Monte-Carlo converge les statistiques estimées avec cette approche avec une vitesse en $\mathcal{O}(\frac{1}{N})$, où N est le nombre d'échantillons utilisés. Soient X_1, X_2, ...,X_n une suite de variables aléatoires réelles, indépendantes et qui ont la même loi de probabilité. On suppose que $\mathbb{E}(X_i)=\mu$ et $\mathrm{Var}(X_i)=\sigma^2 < \infty$. On écrit $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, et $U_n = \sqrt{n}(\frac{\bar{X}-\mu}{\sigma})$. La distribution de U_n converge alors vers une loi normale centrée réduite quand $n \longrightarrow \infty$. Le théorème central limite démontre que toute somme pondérée de variables aléatoires indépendantes et identiquement distribuées tend vers un loi gaussienne. De plus, le théorème montre que la convergence de cette loi à une vitesse en $\mathcal{O}(\frac{1}{n})$.

4.1.5 Estimation de la densité de probabilité

Afin d'obtenir une estimation de la densité de probabilité (PDF) d'une variable aléatoire X, on utilise les simulations Monte-Carlo $\{x^{1)}, ..., x^{N_{sample})}\}$ de la variable aléatoire. On construit un histogramme représentatif des différents échantillons de la variables aléatoire. L'histogramme se construit de la manière suivante : on échantillonne l'espace de variabilité de la variable aléatoire selon des classes définies à l'aide d'une longueur caractéristique h_{histo}. Puis la densité de probabilité est définie à partir du nombre d'échantillons qui existent dans cette classe par rapport aux nombre d'échantillons total. Cette méthode est donc une méthode étagée (par le choix des classes), et par conséquent est discontinue. L'histogramme

est un estimateur très élémentaire pour des analyse de statistiques rigoureuses.
Une deuxième méthode consiste à construire une représentation régulière de la fonction de densité de probabilité en utilisant l'estimateur du noyau (*kernel smoothing*) proposé par [145]. La fonction de densité de probabilité peut être approchée par la fonction suivante :

$$\hat{f}_X(x) = \frac{1}{N_K h_K} \sum_{i=1}^{N_K} K[\frac{x - x^{(i)}}{h_K}]. \tag{4.18}$$

$K(X)$ est une fonction positive adaptée que l'on nomme le noyau (*kernel*), h_K est une paramètre nommé fenêtre. Plusieurs fonctions noyaux sont connus comme le noyau gaussien (qui est la fonction de densité normale). D'autres noyaux existent dans la littérature, on renvoie le lecteur à [145] pour plus de détails. De même, il existe plusieurs représentations de la fenêtre que l'on peut retrouver dans [125]. L'idée de base de cette méthode est de créer autour de chaque échantillon *une bosse* (dont la forme est représentée mathématiquement par la fonction $K(X)$ et la largeur par le paramètre h_K). L'estimation de la densité de probabilité est donc la somme moyennée de ces bosses.
L'approximation de la fonction de densité de probabilité est très coûteuse en terme de nombre d'échantillons.

4.2 Dépendance et probabilité conditionnelle

4.2.1 Notion de dépendance

En théorie des probabilités, la notion de dépendance est très importante. Intuitivement, deux événements aléatoires sont indépendants si l'un n'a aucune influence sur l'autre. Supposons que A_1 et A_2 soient deux événements. Les deux événements sont indépendants si $\mathcal{P}(A_1 \cap A_2) = \mathcal{P}(A_1)\mathcal{P}(A_2)$.

4.2.2 Probabilité conditionnelle

La probabilité conditionnelle décrit la probabilité d'occurrence d'un événement sachant l'occurrence d'un autre événement. Soient A_1 et A_2 deux événements. On définit la probabilité $\mathcal{P}(A_1|A_2)$ de l'événement A_1 sachant que A_2 s'est réalisé par :

$$\mathcal{P}(A_1|A_2) = \frac{\mathcal{P}(A_1 \cap A_2)}{\mathcal{P}(A_2)}. \tag{4.19}$$

Quand les deux événements sont indépendants, on remarque que : $\mathcal{P}(A_1|A_2) = \mathcal{P}(A_1)$.

4.2.3 Théorème de Bayes

Le théorème de Bayes est un résultat fondamental en théorie des probabilités issu des travaux de Thomas Bayes et généralisé par Laplace. Il est une conséquence immédiate des probabilités conditionnelles et des probabilités totales. L'énoncé du théorème est le suivant :

Théorème 1 *Soient A_1 et A_2 deux événements non nuls. Connaissant les probabilités $\mathcal{P}(A_1)$, $\mathcal{P}(A_2)$ et $\mathcal{P}(A_2|A_1)$, la probabilité conditionnelle de A_1 sachant A_2 s'exprime comme suit :*

$$\mathcal{P}(A_1|A_2) = \frac{\mathcal{P}(A_2|A_1)\mathcal{P}(A_1)}{\mathcal{P}(A_2)}. \tag{4.20}$$

Le théorème de Bayes est utilisé en inférence statistique (aussi appelée inférence bayésienne) pour mettre à jour ou actualiser les estimations d'une probabilité ou d'un paramètre quelconque, à partir des observations et des lois de probabilité de ces observations. Il est notamment au cœur des algorithmes d'assimilation de données.

4.3 Vecteur aléatoire et espace de probabilité associé

Dans les applications liées à l'environnement généralement et à l'hydraulique spécifiquement, les paramètres considérés (la hauteur d'eau, le débit, la bathymétrie , etc) sont des champs physiques, souvent discrétisés sous forme de vecteurs qui représentent la variabilité spatiale ou temporelle de ces paramètres dans le domaine d'étude considéré. Afin d'étudier des problèmes comportant des champs physiques incertains, on introduit la notion de vecteur aléatoire.

4.3.1 Définitions

On définit $\mathbf{X} = (X_1, ..., X_n)$ comme un vecteur aléatoire de dimension n si toutes ses composantes $\{X_1, ..., X_n\}$ sont des variables aléatoires réelles. Autrement dit, un vecteur aléatoire est une collection d'un nombre fini de variables aléatoires.

4.3.2 Distribution

Chaque composante X_j du vecteur aléatoire est une variable aléatoire réelle et possède donc une loi de probabilité dite marginale F_{X_j}.

Densité conjointe. La loi de probabilité du vecteur aléatoire est la densité de probabilité conjointe de ses composantes.

$$F_{\mathbf{X}}(\mathbf{x}) = \mathcal{P}(X_1 \leq x_1, ..., X_n \leq x_n), \qquad \mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n. \tag{4.21}$$

La fonction de répartition est définie par la densité de probabilité conjointe des composantes de $\mathbf{X}$, elle s'écrit de la façon suivante :

$$F_{\mathbf{X}}(x_1, ..., x_n) = \int_{-\infty}^{x_1} ... \int_{-\infty}^{x_n} f_{\mathbf{X}}(y_1, ..., y_n) dy_1 ... dy_n. \tag{4.22}$$

Densités marginales. Si un vecteur aléatoire $\mathbf{X}$ possède une densité de probabilité $f_{\mathbf{X}}$, on peut définir la densité marginale pour chacune de ses composantes, ici pour X_j :

$$F_{X_j}(x_j) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f_{\mathbf{X}}(y_1, ..., y_n) dy_1 ... dy_{j-1} dy_{j+1} ... dy_n. \tag{4.23}$$

4.3.3 Moments statistiques

Pour les champs aléatoires, on peut aussi définir des quantités statistiques déjà, en partie, introduites dans le paragraphe 4.1.3.

Espérance et covariance. La moyenne d'un vecteur aléatoire est définie par :

$$\mu_{\mathbf{X}} = \mathbb{E}[\mathbf{X}] = (\mathbb{E}[X_1], \mathbb{E}[X_2], ..., \mathbb{E}[X_n]). \tag{4.24}$$

La matrice de covariance de $\mathbf{X}$, notée $\mathbf{C_X}$, est définie par :

$$\mathbf{C_X} = (cov(X_i, X_j))_{i,j=1,...,n}, \quad cov(X_i, X_j) = \mathbb{E}[(X_i - \mu_{X_i})(X_j - \mu_{X_j})]. \tag{4.25}$$

La matrice de covariance est une matrice carrée symétrique, semi-définie positive (donc diagonalisable). La diagonale de la matrice de covariance correspond à la variance de chaque élément du vecteur aléatoire, i.e. $cov(X_i, X_i) = Var(X_i)$. Les éléments extra-diagonaux de cette matrice représentent les covariances des couples de variables du vecteur aléatoire. La $i-$ème colonne de la matrice $\mathbf{C_X}$ représente la fonction de covariance entre la variable X_i et le vecteur $\mathbf{X}$. Dans le cas où le vecteur aléatoire représente la discrétisation d'un champ aléatoire dans l'espace (ou dans le temps), cette matrice représente la matrice de covariance spatiale (ou temporelle). Par exemple, pour un vecteur aléatoire correspondant à la hauteur d'eau d'une rivière, la $i-$ème colonne de la matrice de covariance correspond à la covariance entre la hauteur d'eau au point i et la hauteur d'eau dans tout le reste de la rivière $\mathbf{X}$. Dans le cas où le vecteur aléatoire comprend des paramètres physiques distincts, la covariance entre deux variables physiques de dimensions différentes montre le lien entre les statistiques des deux variables.

Corrélation. La matrice de corrélation est une forme normalisée de la matrice de covariance $\mathbf{C_X}$:

$$corr(X_i, X_j) = \frac{cov(X_i, X_j)}{\sigma_{X_i} \sigma_{X_j}}. \tag{4.26}$$

La matrice de corrélation permet de mesurer l'intensité de la dépendance qui peut exister entre les différentes variables du vecteur aléatoire. Deux variables X_i et X_j seront corrélées si leur corrélation $corr(X_i, X_j)$ s'approche de 1 et décorrélées si leur corrélation s'approche de 0.

Tout comme la matrice de covariance, la matrice de corrélation est une matrice carrée symétrique. Par définition, les éléments diagonaux de cette matrice sont égaux à 1, et les éléments extra-diagonaux représentent la corrélation entre chaque couple de variables du

vecteur aléatoire. La corrélation entre deux variables aléatoires telle que définie par l'équation (4.26) revient à décrire la relation entre ces deux variables aléatoires à l'aide d'une droite. On parle de corrélation linéaire aussi connue sous le nom de "corrélation de Pearson". Il existe d'autres manières d'écrire une corrélation telle que "la corrélation de Spearman" qui est largement utilisée en sciences du climat pour mesurer une dépendance statistique non paramétrique entre deux variables, mais qui ne sera pas traitée dans cette thèse.

Covariance croisée entre deux vecteurs aléatoires. Supposons que $\mathbf{X}$ et $\mathbf{Y}$ sont deux vecteurs aléatoires. La covariance croisée entre deux vecteurs aléatoires est la matrice qui s'écrit de la façon suivante.

$$C_{\mathbf{XY}} = \mathbb{E}[(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{Y} - \mu_{\mathbf{Y}})^T]. \tag{4.27}$$

Stochastiquement, cette matrice peut être estimée en utilisant l'équation 4.17. Soit $(\mathbf{X}_i)1 \leq i \leq N_{sample}$ (resp. $(\mathbf{Y}_i)1 \leq i \leq N_{sample}$) un ensemble de réalisations du vecteur $\mathbf{X}$ (resp. $\mathbf{Y}$). On écrit la matrice de covariance croisée entre le vecteur $\mathbf{X}$ et $\mathbf{Y}$ de la manière suivante :

$$C_{\mathbf{XY}} = \frac{1}{N_{sample}}(\mathbf{X_i} - \mu_{\mathbf{X}})(\mathbf{Y_i} - \mu_{\mathbf{Y}})^T]. \tag{4.28}$$

Cette matrice est particulièrement importante pour la définition du filtre de Kalman d'ensemble. Elle permet de représenter les erreurs de l'état du modèle dues à l'incertitude sur un paramètre. On notera aussi que la matrice de covariance $C_{\mathbf{X}}$ n'est qu'un cas particulier de la covariance croisée entre le vecteur $\mathbf{X}$ et lui même ($\mathbf{C_X} = C_{\mathbf{XX}}$).

Auto-corrélation. Une autre notion importante pour la description d'un champ stochastique est l'auto-corrélation. On définit la fonction d'auto-corrélation d'un processus stochastique $\{X_t\}_{t\in T}$ par la fonction $R_X : T \times T \to \mathbb{R}$ définit de la manière suivante :

$$R_X(t, s) = \mathbb{E}[X_s X_t], \qquad s, t \in T. \tag{4.29}$$

Une des propriétés intéressantes par rapport à la fonction d'auto-corrélation est le fait que la continuité de cette fonction assure que le processus soit carré intégrable. Une démonstration de cette propriété se trouve par exemple dans [3].

4.4 Analyse de sensibilité

4.4.1 Généralités

L'analyse de sensibilité est un outil important pour les domaines d'applications qui utilisent un modèle numérique. Dans son ouvrage Saltelli [121] définit quatre objectifs de l'analyse de sensibilité :

— *Factors prioritization* : Déterminer les variables à prendre en compte pour la réduction d'incertitudes afin de maximiser la réduction de l'incertitude de la quantité d'intérêt ;

— *Factors fixing* : Déterminer les variables qui n'ont pas suffisamment d'influence sur l'incertitude du modèle et que l'on peut fixer dans un exercice de réduction d'incertitudes ;

— *Variance cutting* : Déterminer les variables à fixer pour réduire l'incertitude à une valeur donnée ;

— *Factors mapping* : Cartographier l'influence des variables en entrée selon un domaine de valeurs de la sortie.

Cette analyse n'est pas une tâche facile en pratique car la relation entre les variables en entrée et les variables en sortie du modèle n'est ni linéaire ni monotone dans la majorité des modèles en géosciences. Il existe beaucoup de méthodes d'analyse de sensibilité dans la littérature. On peut en trouver un résumé dans l'ouvrage de Saltelli [121], ou plus récemment dans le papier de Xu et Gertner [155] ainsi que dans le livre issu du travail du réseau scientifique "MEXICO" [38]. Généralement, on peut classifier ces méthodes en deux groupes :

— Analyse de sensibilité locale qui consiste à étudier l'influence sur la quantité d'intérêt d'une (de) petite(s) perturbation(s) autour d'une (des) valeur(s) nominal(s) de la (des) variable(s) en entrée. Cette approche vise à calculer des indices de sensibilité basés sur les dérivées partielles du modèle en un point précis [56, 25, 140].

— Analyse de sensibilité globale : elle s'intéressent à la quantification de l'incertitude dans l'ensemble du domaine de variation possible des variables à l'entrée. Ces méthodes sont développées dans un cadre statistique, et sont composées de techniques issues de la théorie des plans d'expérience (pour l'exploration des codes de calcul à grand nombre d'entrées), de méthodes de type Monte-Carlo rendues possibles grâce aux nouvelles capacités informatiques (pour des analyses de sensibilité quantitatives et fines) et de la théorie de l'apprentissage statistiques (pour les codes coûteux et complexes). On remarquera que dans le cas d'un modèle linéaire les approches globale et locale sont identiques.

On ne s'intéressera dans la suite de la thèse qu'à la deuxième classe pour l'analyse de sensibilité globale.

4.4.2 Analyse de sensibilité globale (ANOVA)

Il existe une littérature foisonnante qui s'intéresse à l'analyse de sensibilité globale. Un état de l'art des méthodes est présenté dans [119] et rassemble les méthodes sous deux groupes :

— Méthodes basées sur la régression : l'analyse de sensibilité est donnée à partir de la régression linéaire entre les variables en entrée et les variables en sortie. Les principales limitations de ces méthodes, comme présentés dans [120], se résument en deux points : si le modèle n'est pas strictement monotone, la linéarisation n'est pas valable. Deuxièmement la régression à tendance à sous-estimer les effets totaux qui proviennent de l'interaction entre les variables en entrée.

— Méthodes basées sur la variance : L'analyse de sensibilité est décrite à partir de la décomposition de la variance de la réponse du modèle comme la somme de la contribution

de chaque variable en entrée, ainsi que toutes les combinaison possible d'interaction entre eux. En statistique, on les connait sous le nom de ***ANOVA***(ANalysis Of VAriance) [36]. Plusieurs indices de sensibilité ont été créés avec cette méthode : indices FAST (Fourier amplitude sensitivity test)([122]), indices de Sobol' [129]. Dans la suite, l'attention sera portée sur les indices de Sobol'.

On nomme $\mathcal{M}$ le modèle numérique et on suppose que le vecteur $\mathbf{X} = (X_1, ..., X_M)$ comporte les variables en entrée que l'on supposera indépendants. On écrit $Y = \mathcal{M}(\mathbf{X})$ la réponse du modèle et on la suppose de variance finie. Comme prouvé dans [36], le modèle $\mathcal{M}(\mathbf{X})$ peut être décomposé sous forme d'une combinaison linéaire d'effets principaux et d'interactions vérifiant (4.31) de façon unique [129] :

$$Y = \mathcal{M}(\mathbf{X}) = \mathcal{M}_0 + \sum_{i=1}^{M} \mathcal{M}_i(X_i) + \sum_{1 \leq i \leq j \leq M} \mathcal{M}_{i,j}(X_i, X_j) + ... + \mathcal{M}_{1,2,...M}(\mathbf{X}). \tag{4.30}$$

Ainsi tous les termes de (4.30) sont orthgonaux :

$$\begin{aligned} &\mathcal{M}_0 = \mathbb{E}[\mathcal{M}(\mathbf{X})] \\ &\mathcal{M}_i(X_i) = \mathbb{E}[\mathcal{M}(\mathbf{X})|X_i] - \mathcal{M}_0 \\ &\mathcal{M}_{i,j}(X_i, X_j) = \mathbb{E}[\mathcal{M}(\mathbf{X})|X_i, X_j] - \mathcal{M}_i(X_i) - \mathcal{M}_j(X_j) - \mathcal{M}_0. \end{aligned} \tag{4.31}$$

$$\begin{aligned} &\mathbb{E}[\mathcal{M}_{i_1,...,i_s}(X_{i_1}, ..., X_{i_s})\mathcal{M}_{j_1,...,j_t}(X_{j_1}, ..., X_{j_t})] = 0 \\ &s, t = \{1, ..., M\} \quad , \quad \{i_1, ..., i_s\} \neq \{j_1, ..., j_t\}. \end{aligned} \tag{4.32}$$

Cette orthogonalité permet de définir directement la variance du modèle :

$$Var(\mathcal{M}(\mathbf{X})) \equiv V = \sum_{i=1}^{M} V_i + \sum_{1 \leq i \leq j \leq M} V_{i,j} + ... + V_{1,...,M} = \sum_{\mathcal{I} \subset \{1,...,M\}} V_{\mathcal{I}}, \tag{4.33}$$

où $V_{\mathcal{I}} = Var(\mathcal{M}_{\mathcal{I}}(X_{\{\mathcal{I}\}})$ pour tout $\mathcal{I} \subset \{1, ..., M\}$, avec $X_{\{\mathcal{I}\}} = (X_i)_{i \in \mathcal{I}}$. Ceci permet de définir l'indice de Sobol' du premier ordre associé aux entrées $X_{\{\mathcal{I}\}} = (X_i)_{i \in \mathcal{I}}$:

$$S_{\mathcal{I}} = \frac{V_{\mathcal{I}}}{V}. \tag{4.34}$$

L'indice de Sobol' de l'ordre associé à X_i s'écrit alors :

$$S_i = \frac{V_i}{V} = \frac{Var(\mathcal{M}(\mathbf{X}|X_i))}{Var(\mathcal{M}(\mathbf{X}))}. \tag{4.35}$$

De plus, il est possible de définir un indice de sensibilité totale S_i^T :

$$S_i^T = \sum_{\mathcal{I} \subset (1,...,M), i \in \mathcal{I}} S_{\mathcal{I}}. \tag{4.36}$$

Il correspond à la somme des indices de Sobol' du premier ordre associés aux sous-ensembles d'entrées comprenant X_i.

4.4.3 Estimation des indices de Sobol'

En pratique, on estime les indices de Sobol' 4.34 et 4.36 à l'aide des simulations Monte-Carlo. Comme présenté précédemment à l'aide d'un $N-$échantillon des variables aléatoires $\mathbf{X}$ on peut estimer la moyenne et la variance en utilisant les équations 4.17 et 4.16.
Afin d'estimer la variance V_i, on utilise deux échantillons indépendants $\mathbf{X} = \{\mathbf{x}^{(i)}, i = 1, ..., N\}$ et $\tilde{\mathbf{X}} = \{\tilde{\mathbf{x}}^{(i)}, i = 1, ..., N\}$ et on applique les formulations suivantes :

$$\begin{array}{l} \mathbf{x^{(k)}} = \{x_1^{(k)}, ..., x_M^{(k)}\} \\ \tilde{\mathbf{x}^{(k)}} = \{\tilde{x}_1^{(k)}, ..., \tilde{x}_M^{(k)}\} \end{array} \tag{4.37}$$

L'estimation de V_i à l'aide des simulations Monte-Carlo est définie alors de la façon suivante :

$$V_i \equiv \frac{1}{N} \sum_{k=1}^{N} \mathcal{M}(x_1^{(k)}, ..., x_M^{(k)}) \mathcal{M}(\tilde{x}_1^{(k)}, ..., \tilde{x}_{i-1}^{(k)}, x_i^{(k)}, \tilde{x}_{i+1}^{(k)}, ..., \tilde{x}_M^{(k)}) - \hat{\mathcal{M}}_0^2 \tag{4.38}$$

Une formule identique est obtenue pour les variances partielles des ordres plus élevés, on renvoie le lecteur à [62] pour plus de détails.

Réduction de dimension de l'espace incertain

5

5.1 Introduction

Généralement, les champs physiques utilisés dans les modèles numériques (en entrée comme en sortie du modèle) en géosciences ont une description spatiale ou temporelle continue. La discrétisation est le processus qui permet d'approcher un champ aléatoire par un vecteur aléatoire. Pour garder une bonne approximation, la dimension du vecteur aléatoire, correspondant à la discrétisation, est souvent importante. Pour les besoins de quantification d'incertitude et d'analyse de sensibilité, il est important de réduire leur dimension. On présente dans ce qui suit deux méthodes de réduction de dimension de l'espace incertain : La transformation de Karhunen-Loève qui permet de réduire l'espace des champs aléatoires en entrée afin de les échantillonner, et la décomposition orthogonale propre qui permet de réduire l'espace des sorties du modèle.

5.2 Echantillonnage d'un champs aléatoire

5.2.1 Réalisations d'un processus stochastique

Afin de définir la notion de réalisation pour un processus stochastique, on utilise la définition donnée par [148]. Soit T un ensemble d'indexation (par exemple le temps ou l'espace), E un espace mesurable (par exemple le débit ou la bathymétrie) et $(\Omega, \mathcal{F}, \mathcal{P})$ une $\sigma-$algèbre qui représente l'espace de probabilité. La définition classique d'un processus stochastique X indexé par T, associé à l'espace d'état E et à l'espace de probabilité $(\Omega, \mathcal{F}, \mathcal{P})$, est une collection $\{X_t, t \in T\}$ dont les valeurs aléatoires vivent dans E et sont décrites statistiquement par $(\Omega, \mathcal{F}, \mathcal{P})$. Par conséquent, $t \in T$ on obtient la transformation mathématique suivante :

$$X_t : \Omega \to E. \tag{5.1}$$

Ceci permet de définir la notion de réalisation pour un champ aléatoire. Pour chaque réalisation $\omega \in \Omega$, le champ associé est la fonction :

$$X(\omega) : \begin{array}{ll} T \to & E \\ t \mapsto & X_t(\omega). \end{array} \tag{5.2}$$

On renvoie le lecteur pour plus de détails sur la notion d'un processus stochastique ou d'un champ aléatoire vers [148].

5.2.2 Échantillonnage d'un processus stochastique

Afin de procéder à un échantillonnage d'un processus stochastique, on propose dans la littérature de paramétrer les champs stochastiques. Le problème de paramétrisation peut être écrit de la façon suivante :
Soit $(X_t, t \in T)$ un processus stochastique, où t est un indice qui appartient à un ensemble d'indices T (On rappelle que t peut représenter le temps ou l'espace). Afin d'échantillonner X_t, une méthode intuitive serait de considérer la version discrétisée de X_t. On discrétise la variable dans le domaine T :

$$(X_{t_1}, ..., X_{t_n}) \qquad t_1, ..., t_n \in T. \tag{5.3}$$

On note que la discrétisation de X_t n'est qu'une approximation du processus. La précision numérique implique d'utiliser une discrétisation très fine, et donc d'avoir un nombre n grand. Cependant, ceci augmenterait le nombre d'échantillons du vecteur. Pour remédier ce problème, il existe des méthodes permettant de réduire la dimension du problème en maintenant une approximation satisfaisante. On se contente ici de présenter la méthode de Karhunen-Loève que l'on utilisera dans la suite de la thèse.

5.2.3 La décomposition de Karhunen-Loève

La décomposition de Karhunen-Loève ([82, 68]) est l'une des techniques les plus connues pour la réduction de dimension pour la représentation des champs stochastiques. Elle consiste à décomposer le champ $\mathbf{X}_t$ en une partie déterministe (la moyenne $\mu_{\mathbf{X}_t}$), et une partie aléatoire décomposée sur la base des valeurs propres λ_i et fonctions propres $v_i(t)$ de la fonction d'autocovariance $\mathcal{C}(t_1, t_2)$ sous la forme

$$\mathbf{X}_t(\omega) = \mu_{\mathbf{X}_t} + \sum_{i=1}^{\infty} \sqrt{\lambda_i} v_i(t) \zeta_i(\omega). \tag{5.4}$$

L'ensemble de fonctions sur lesquelles chaque réalisation du champ $\mathbf{X_t}$ est défini par le problème aux valeurs propres suivant :

$$\forall i \in \mathbb{N} \qquad \int_{t_2 \in T} \mathcal{C}(t_1, t_2) v_i(t_2) dt_2 = \lambda_i v_i(t_1) \qquad t_1 \in T. \tag{5.5}$$

L'existence des valeurs propres et des vecteurs propres de la fonction de covariance est due à son caractère borné, symétrique et défini positif (théorème de Mercer [3]). Ainsi l'ensemble des $\{v_i\}$ forme une base complète orthogonale de l'espace des fonctions de carré sommables. Par conséquent, chaque réalisation de $\mathbf{X}_t$ peut être développée sous cette base (Eq. 10.3). Les variables $\zeta_i(\omega)$ sont des variables aléatoires indépendantes qui ont les propriétés suivantes :

$$\mathbb{E}[\zeta_i] = 0 \qquad \mathbb{E}[\zeta_i \zeta_j] = \delta_{ij}, \tag{5.6}$$

et sont définies par

$$\zeta_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_T \big(\mathbf{X}_t(\omega) - \mu_{\mathbf{X}_t}(t)\big) v_i(t) dt, \qquad \forall i. \tag{5.7}$$

En pratique, la somme infinie dans l'équation 10.3 est tronquée sous la forme suivante.

$$\mathbf{X_t}(\omega) \approx \mu_{\mathbf{X}_t} + \sum_{i=1}^{d} \sqrt{\lambda_i} v_i(t) \zeta_i(\omega). \tag{5.8}$$

Il reste à déterminer l'ordre de troncature d qui permet d'obtenir une approximation avec une précision satisfaisante. La détermination de d est liée à la vitesse de décroissance des valeurs propres. L'approche la plus commune [76, 152] est d'examiner la décroissance des λ_i et de ne garder que les d premières valeurs propres de façon à ce que la contribution du reste des valeurs propres soit négligeable. En général, on considère que la contribution est négligeable à partir d'un pourcentage ϵ_{KL} élevé définit comme un critère de seuil sur $\mathbf{X}_t$ définit avec une norme associée. Ce critère est la part de variance expliquée, et s'écrit généralement.

$$\text{Crit}(d) = \frac{\int_T Var(\mathbf{X_t^{(d)}}) dt}{\int_T Var(\mathbf{X_t}) dt} = \frac{\sum_{i=1}^{d} \lambda_i}{\sum_{i=1}^{\infty} \lambda_i} \geq \epsilon_{KL}. \tag{5.9}$$

Par conséquent, pour un seuil ϵ_{KL} donné, si le processus est fortement corrélé (longueur de corrélation grande) la décomposition de Karhunen-Loève nécessite moins de termes qu'un processus dont la longueur de corrélation est petite.
L'utilité de la décomposition de Karhunen-Loève réside dans la résolution du problème aux valeurs propres intégral. Ce problème (équation 10.4) est une équation dite de Fredholm de seconde espèce homogène [114]. Dans des cas très particuliers, l'équation 10.4 admet des solutions analytiques [131]. Dans le cas général, on a recours à une résolution numérique [7]. Une fois les fonctions propres et valeurs propres estimées, l'échantillonnage d'un processus aléatoire est fait à partir de l'échantillonnage les variables aléatoires ζ_i. Généralement la taille du vecteur $\boldsymbol{\zeta} = (\zeta_1, ..., \zeta_d)$ est significativement plus petite que la taille du vecteur $\mathbf{X}_t$

5.2.4 Cas particulier du processus gaussien

La décomposition tronquée de Karhunen-Loève 5.8 approche un processus aléatoire par une somme finie de fonctions orthonormales pondérées par des variables aléatoires centrées réduites et décorrélées. Pour des variables aléatoires gaussiennes, la corrélation et

la dépendance sont des notions équivalentes. De plus, la combinaison linéaire de variables aléatoires gaussiennes est une variables aléatoire gaussienne. Par conséquent, si $\mathbf{X}_t(\omega)$ est un processus gaussien, alors les variables aléatoires ζ_i dans les équations 10.3 et 10.5 sont des variables aléatoires gaussiennes centrés, réduites et indépendantes. En conclusion, l'équation 10.3 montre qu'un processus gaussien peut être représenté par un nombre fini de variables aléatoires gaussiennes et indépendantes. Dans le cas d'un processus non gaussien et carré intégrable, cette décomposition est encore valable. Cependant, la caractérisation probabiliste des variables aléatoires devient délicate [152].

5.3 POD "*Proper Orthogonal Decomposition*"

5.3.1 Généralités

La décomposition orthogonale aux valeurs propres est une technique qui permet d'approximer un système de dimension élevée par un autre de dimension faible. Cette méthode consiste à déterminer une base de modes propres orthogonaux représentatifs de la physique simulée. Les modes propres sont obtenus par résolution de l'intégrale de Fredholm, le noyau de cet intégrale est construit à partir d'un ensemble de simulations construite à l'aide d'un plan d'expériences. La propriété intéressante de cette représentation réside dans le fait que les fonctions propres associées au problème sont optimales dans le sens de la représentation énergétique (que l'on définira plus loin) ce qui permet de les utiliser pour construire une représentation réduite de la physique.
Historiquement, la POD a été introduite pour traiter un problème de turbulence dans le but d'extraire les structure cohérentes d'un écoulement [45]. Elle peut être vue comme une décomposition de Fourrier lorsque les directions de l'écoulement ne peuvent pas être supposées périodiques et homogènes. La POD a été appliquée dans plusieurs domaines d'applications dont le traitement de signal, l'activité neurale, compression de données et contrôle optimal. Le point commun de toutes ces applications est l'extraction des caractères dominants d'un ensemble de données, permettant d'accéder ainsi à une réduction de données [158].
La POD est utilisée en quantification d'incertitudes afin de réduire la dimension d'un vecteur aléatoire à la sortie du modèle. L'incertitude est par conséquent, portée sur chaque direction propre. Généralement, on utilise un plan d'expérience qui couvre le plus possible l'espace incertain tel que les suites à faibles discrépance afin de construire une première surface de réponse, ensuite on interpole les points de cette surface afin de construire une surface de réponse plus lisse [20, 89, 116].

5.3.2 Méthode de construction de la POD

La technique POD se base sur la réduction Karhunen-Loève présentée dans la section 5.2.3. Elle consiste à projeter la solution du modèle $\mathcal{M}$ dans une base finie et orthonormée $\{\phi_i\}_{i\in\mathbb{I}}$, où $\mathbb{I}$ est un ensemble discret fini d'indices. Selon la caractérisation du problème

que l'on souhaite traiter, la POD peut être considérée comme l'analyse en composantes principales (SVD) ou comme un problème qui consiste à trouver la base POD en utilisant la matrice snapshots. Supposons que $Y(\mathbf{X},t) = \mathcal{M}(\mathbf{X},t)$ est la réponse du modèle, et qu'elle a une dépendance temporelle. Le processus $Y(\mathbf{X},t)$ peut être décomposé comme suivant :

$$Y(\mathbf{X},t) = \sum_{i\in\mathbb{I}} a_i(t)\phi(\mathcal{X}). \tag{5.10}$$

L'orthonormalité des ϕ_i implique :

$$a_i(t) =< Y(.,t), \phi_i >, \tag{5.11}$$

on considère l'ensemble

$$Y_{snap} = \{Y(.,t)\ t \in \mathbb{T}\}, \tag{5.12}$$

avec $\mathbb{T}$ l'ensemble des temps indexés par $\mathbb{I}$.Y_{snap} nommé *matrice des snapshots* contient un ensemble qui peut provenir de mesures, d'observations ou de résultats de simulation. Si l'on possède N_{snap} échantillons (snapshots) de $Y(\mathbf{X},t)$, la réduction du modèle se traduit par la formule suivante :

$$Y(\mathbf{X},t) \approx \hat{Y}(\mathbf{X},t) = \sum_{i=1}^{N_{snap}} a_i(t)\phi_i(\mathbf{X}). \tag{5.13}$$

Le calcul des vecteur $\{\phi_i\}$ se fait en décomposant la matrice de covariance C_{snap} de la matrice des snapshots : $C_{snap} = \frac{1}{N_{snap}} Y_{snap} Y_{snap}^T$ en vecteurs propres. L'inconvénient de cette méthode est qu'elle nécessite le calcul des vecteurs propres. Dans de nombreux cas, la matrice C_{snap} est très pleine et mal conditionnée. Le calcul des valeurs propres et vecteurs propres peut devenir coûteux en terme de calcul informatique. Dans la littérature on trouve des méthodes d'estimation de calcul des valeurs propres les plus influentes en premier lieux, l'utilisation d'algorithme tel que celui de Lanczos [71] est fortement recommandé dans ce cas.

Polynômes du chaos 6

6.1 Généralités

La décomposition spectrale est largement utilisée en analyse numérique. Dans le contexte probabiliste, elle consiste à projeter une variable aléatoire (réponse d'un modèle) dans une base convenable de dimension fini $(\psi_j(\mathbf{X})_{0\leq j\leq N_{base}})$ de la manière suivante :

$$\mathcal{M}(\mathbf{X}) \approx \sum_{j=0}^{N_{base}} \alpha_j \psi_j(\mathbf{X}). \tag{6.1}$$

La théorie des polynômes du chaos consiste à construire la base à partir de polynômes orthogonaux. Par conséquent, définir un modèle réduit à partir de cette définition consiste à calculer les coefficients polynomiaux α_j. Une méthode intuitive pour calculer ces coefficients, est de remplacer la réponse du modèle par son approximation dans les équations aux dérivées partielles, les coefficients sont ainsi calculés directement par la méthode Galerkin. Ces méthodes ont été présentées pour la première fois pour résoudre un problème de mécanique des structures avec des variables en entrée aléatoires et présentés dans [131]. Cette méthode est appelé *intrusive*, car elle demande de modifier le code existant. Alternativement, les méthodes *non intrusives* consistent à utiliser le code comme une boite noire. L'obtention de la surface de réponse ne peut être faite que par le biais des évaluations du modèle. Ces méthodes permettent d'obtenir une approximation polynomiale comme celle de l'équation (6.1) avec une méthode stochastique.

Historiquement, le terme "*polynômes du chaos* a été introduit par N.Wiener afin de modéliser des phénomènes physiques chaotiques dans le cadre de la turbulence homogène [147]. Il a été utilisé la première fois par Ghanem et Spanos [131] dans le contexte de la propagation d'incertitudes comme alternative aux méthodes Monte-Carlo. Son application repose sur le corollaire du théorème de Cameron-Martin [23].

Les travaux de [153, 154] ont permis de généraliser les travaux de [131] sur les polynômes du chaos pour d'autres classes de distributions de probabilités que la loi gaussienne, en se basant sur le schéma d'Askey [6] qu'ils ont nommé "gPC *generalized Polynomial Chaos*" (polynômes du chaos généralisés). Par simplicité, on gardera la désignation polynômes du chaos aussi pour les gPC dans toute la suite de cette thèse.

Par ailleurs, les polynômes du chaos ont prouvé leur efficacité dans la propagation et le trai-

tement d'incertitudes [76, 152, 151], l'étude de sensibilité [132] et aussi pour accélérer les algorithmes de résolution des problèmes inverses [79, 90, 91].

6.2 Cadre d'approximation polynomiale

6.2.1 Contexte probabiliste

On rappelle que le modèle physique est représenté par $Y = \mathcal{M}(\mathbf{X})$, où $\mathbf{X} = \{X_1, ..., X_M\}$ est le vecteur aléatoire en entrée de dimension M, et Y est la quantité d'intérêt, réponse du modèle. On définit par la suite un contexte probabiliste vu que les variables en entrée sont affectées par l'incertitude.
Soit $(\Omega, \mathcal{F}, \mathcal{P})$ un espace de probabilité, où Ω est l'espace des évènements muni d'une σ-algèbre $\mathcal{F}$ et d'une mesure de probabilité $\mathcal{P}$. Supposons la variable aléatoire $X_i(\omega) : \Omega \to \mathcal{D}_{X_i} \subset \mathbb{R}$, sa réalisation sera notée x_i.

On appelle $\mathcal{L}^2_{\mathbb{R}} \equiv \mathcal{L}^2(\Omega, \mathcal{F}, \mathcal{P}, \mathbb{R})$ un espace des variables aléatoires de variance finie si :

$$\mathbb{E}(X^2) = \int_\Omega X^2(\omega) d\mathcal{P}(\omega) = \int_{\mathcal{D}_X} x^2 f_X(x) dx < \infty \tag{6.2}$$

où f_X est la distribution de probabilité de X. Cet espace est un espace de Hilbert muni du produit scalaire :

$$< X_1, X_2 >_{\mathcal{L}^2_{\mathbb{R}}} \equiv \mathbb{E}(X_1 X_2) = \int_\Omega X_1(\omega) X_2(\omega) d\mathcal{P}(\omega) = \int_{\mathcal{D}_X} x_1 x_2 f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 \tag{6.3}$$

où f_{X_1,X_2} est la probabilité jointe du vecteur aléatoire $\{X_1, X_2\}^T$. Ce produit scalaire définit la norme $||X||_{\mathcal{L}^2_{\mathbb{R}}} = \sqrt{\mathbb{E}(X^2)}$.

On suppose que le modèle physique, est une fonction d'une variable aléatoire $X(\omega), \omega \in \Omega$ définie par une distribution de probabilité f_X. On supposera que la quantité d'intérêt Y est aussi une variable aléatoire telle que $Y \in \mathcal{L}^2_{\mathbb{R}}$.
On note $\mathcal{L}^2$ l'espace des fonctions d'une variable aléatoire (ou un vecteur aléatoire) carré intégrable muni de la distribution de probabilité des valeurs en entrée. On s'intéresse ici à décomposer Y dans une base adaptée et orthogonale de $\mathcal{L}^2$ muni du produit scalaire correspondant à la densité de probabilité $f_{\mathbf{X}}(\mathbf{x})$. La théorie du chaos homogène veut que cette base soit construite à partir de polynômes orthogonaux. Historiquement on utilisait les polynômes multivariés d'Hermite puisque les variables aléatoires étaient considérées gaussiennes [131].
Les travaux de [130] montrent que si l'on suppose que les composantes du vecteur aléatoire $\mathbf{X}$ sont indépendantes, la variable aléatoire Y peut s'écrire sous la forme de l'équation (6.1), où la série converge au sens de la norme de $\mathcal{L}^2$.

$$\lim_{P \to \infty} ||\mathcal{M}(\mathbf{X}) - \sum_{j=0}^{P} \alpha_j \psi_j(\mathbf{X})||^2_{\mathcal{L}^2} = 0 \tag{6.4}$$

l'équation (6.4) montre que la construction de l'approximation polynomiale nécessite la connaissance de la base polynomiale $\psi_j(\mathbf{X})$, le calcul des coefficients α_j polynomiaux ainsi que le degré du polynôme P (troncature).

6.2.2 Base d'approximation

Suite à l'hypothèse de l'indépendance entre les composantes du vecteurs aléatoires, la densité conjointe $f_{\mathbf{X}}(\mathbf{x})$ s'écrit :

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{M} f_{X_i}(x_i) \tag{6.5}$$

où les $f_{X_i}(x_i)$ représentent les densités marginales de la i-ème composante X_i du vecteur $\mathbf{X}$. On définit l'espace des polynômes orthogonaux suivant le mesure de probabilité f_{X_i}.

$$W^{i,d_i} = \{v : D_{X_i} \to \mathbb{R} : v \in span\{\phi_m x_i\}_{m=0}^{d_i}\}, \quad i = 1, ..., M, \tag{6.6}$$

où $\{\phi_m(x_i)\}$ est l'ensemble des polynômes orthogonaux qui respectent les critères d'orthogonalité :

$$\int_{D_{X_i}} \phi_m(x_i)\phi_n(x_i) f_{X_i}(x_i) dx_i = ||\phi_m(x_i)||^2 \delta_{mn} \tag{6.7}$$

avec δ_{mn} le symbole de Kronecker et

$$||\phi_m(x_i)||^2 = \int_{D_{X_i}} \phi_m^2(x_i) f_{X_i}(x_i) dx_i \tag{6.8}$$

un facteur de normalisation. Afin de simplifier, dans toute la suite on supposera aussi que la base est orthonormale.
Dans l'équation 6.7, la fonction f_{X_i} joue le rôle d'un poids d'intégration, ce qui permet de définir le type des polynômes orthogonaux $\{\phi_m\}$. Pour la plupart des densités de probabilité connues, il existe des correspondances avec une famille de polynômes orthogonaux connue. Les correspondances sont déterminées par [153]. Par exemple, les polynômes orthogonaux qui respectent la propriété 6.7 pour une loi normale sont les polynômes d'Hermite (c'est le polynômes du chaos homogène tel qu'il a été introduit par Wiener [147] et utilisé par Ghanem et Spanos [131] pour la propagation d'incertitudes). Quand la distribution de la variable aléatoire est uniforme, la famille des polynômes qui respecte 6.7 est la famille des polynômes de Legendre.

À titre d'exemple, la figure 6.1 représente l'approximation polynomiale de la fonction $g(x) = x\sin(x)$ en supposant que la variable aléatoire x suit une loi normale centrée réduite. La fonction tracée en rouge est celle de l'approximation en polynômes du chaos en utilisant les polynômes d'Hermite, qui selon [153] sont les plus appropriés pour représenter des variables aléatoires gaussiennes. Alors que la fonction verte représente l'approximation polynomiale en utilisant une base des polynôme de Legendre. On remarque que la courbe rouge converge plus

rapidement vers la solution exacte (la courbe en bleu) que la courbe verte à degré polynomial égale ($P = 6$). La solution avec les polynômes d'Hermite a convergé alors que celle avec les polynômes de Legendre est encore loin de la solution.

Une fois la famille de polynômes orthogonaux définie pour chaque composante du vecteur aléatoire. On peut construire une famille de polynômes multi-variés en utilisant le produit tensoriel. La famille des polynômes orthogonaux pour le vecteur aléatoire $\mathbf{X}$ s'écrit de la façon suivante :

$$W_M^P = \bigotimes_{|\mathbf{d}| \leq P} W^{i,d_i}, \tag{6.9}$$

où le produit tensoriel est effectué pour toutes les combinaisons possibles de l'ensemble des indices $\mathbf{d} = (d_1, ..., d_M) \in \mathbb{N}^M$ qui correspondent à $|\mathbf{d}| = \sum_{i=1}^{N} d_i \leq P$. De cette façon, W_M^P est l'espace des polynômes orthogonaux M-variés d'un degré total égal à P au plus. Soit $\{\psi_m(\mathbf{x})\}$ un élément de W_M^P. Chaque élément est construit à partir du produit des polynômes univariés dans chaque direction x_i, $i = 1, ...M$

$$\psi_m(\mathbf{x}) = \phi_{m_1}(x_1)...\phi_{m_M}(x_M), \quad m_1 + ... + m_M \leq P, \tag{6.10}$$

où m_i est le degré du polynôme univarié $\phi(x_i)$ dans la direction de x_i pour $1 \leq i \leq M$. On peut facilement montrer que les polynômes multivariés $\psi_m(\mathbf{x})$ respectent aussi la propriété d'orthogonalité (équation 6.7). Le cardinal de l'ensemble W_M^P qui détermine le nombre de fonctions de la base est donné par la formule :

$$dim(W_M^P) = \binom{M+P}{M}. \tag{6.11}$$

6.2.3 Troncature

L'approximation en polynôme du chaos de degré P du modèle $\mathcal{M}(\mathbf{X})$ dans l'équation (6.1) est obtenu en projetant $\mathcal{M}$ dans l'espace W_M^P :

$$\mathcal{M}(\mathbf{X}) \approx \hat{\mathcal{M}}(\mathbf{X}) = \sum_{j=0}^{M_t} \alpha_j \psi_j(\mathbf{X}), \quad M_t = dim(W_M^P) \tag{6.12}$$

où α_j sont les coefficients de Fourrier (aussi appelés modes) qui sont définis par la formule suivante :

$$\alpha_j = \int_{D_\mathbf{X}} \mathcal{M}(\mathbf{x})\psi_j(\mathbf{x}) f_\mathbf{X}(\mathbf{x}) d\mathbf{x} = \mathbb{E}[\mathcal{M}(\mathbf{x})\psi_j(\mathbf{x})], \quad 1 \leq j \leq M_t. \tag{6.13}$$

L'erreur due à cette projection que l'on nommera "erreur de troncature" est définie par la relation suivante :

$$\begin{aligned} \epsilon &= \| \mathcal{M} - \hat{\mathcal{M}} \|_{\mathcal{L}^2} \\ &= (\mathbb{E}[(\mathcal{M}(\mathbf{x}) - \hat{\mathcal{M}}(\mathbf{x}))^2])^{1/2}, \forall \mathbf{x} \in D_\mathbf{X}. \end{aligned} \tag{6.14}$$

La théorie de l'approximation polynomiale garantit l'existence d'une telle décomposition ainsi que la convergence de l'erreur ϵ vers 0 quand P augmente [152]. La figure 6.2 montre l'effet de l'erreur de troncature. En reprenant l'exemple mentionné plus haut, on utilise une

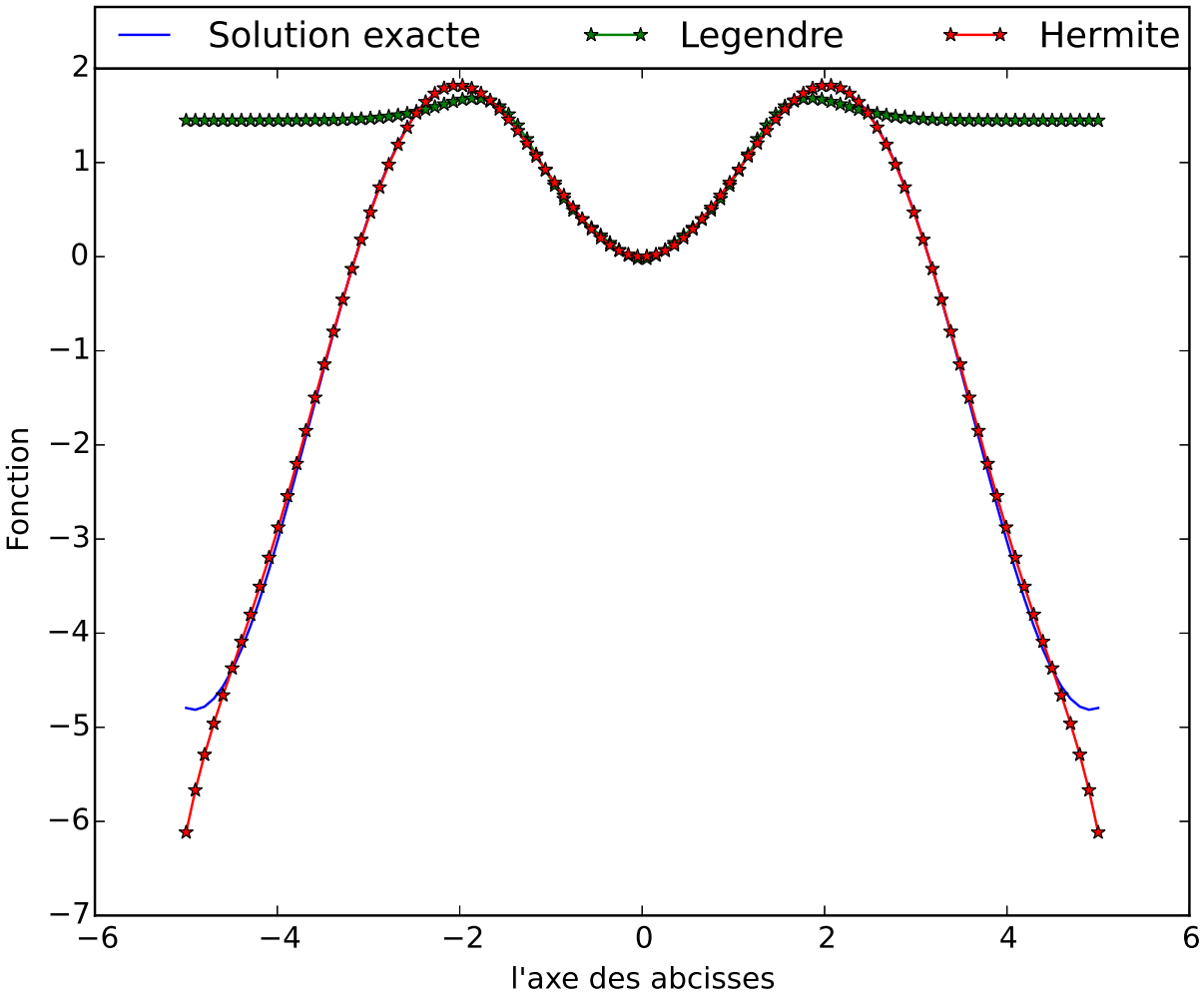


FIGURE 6.1 – *Approximation PC de degré 6 pour la fonction* $g(x) = x\sin(x)$ *avec une base du polynôme d'Hermite et du polynômes de Legendre pour une variable aléatoire gaussienne* $x \sim \mathcal{N}(0,1)$

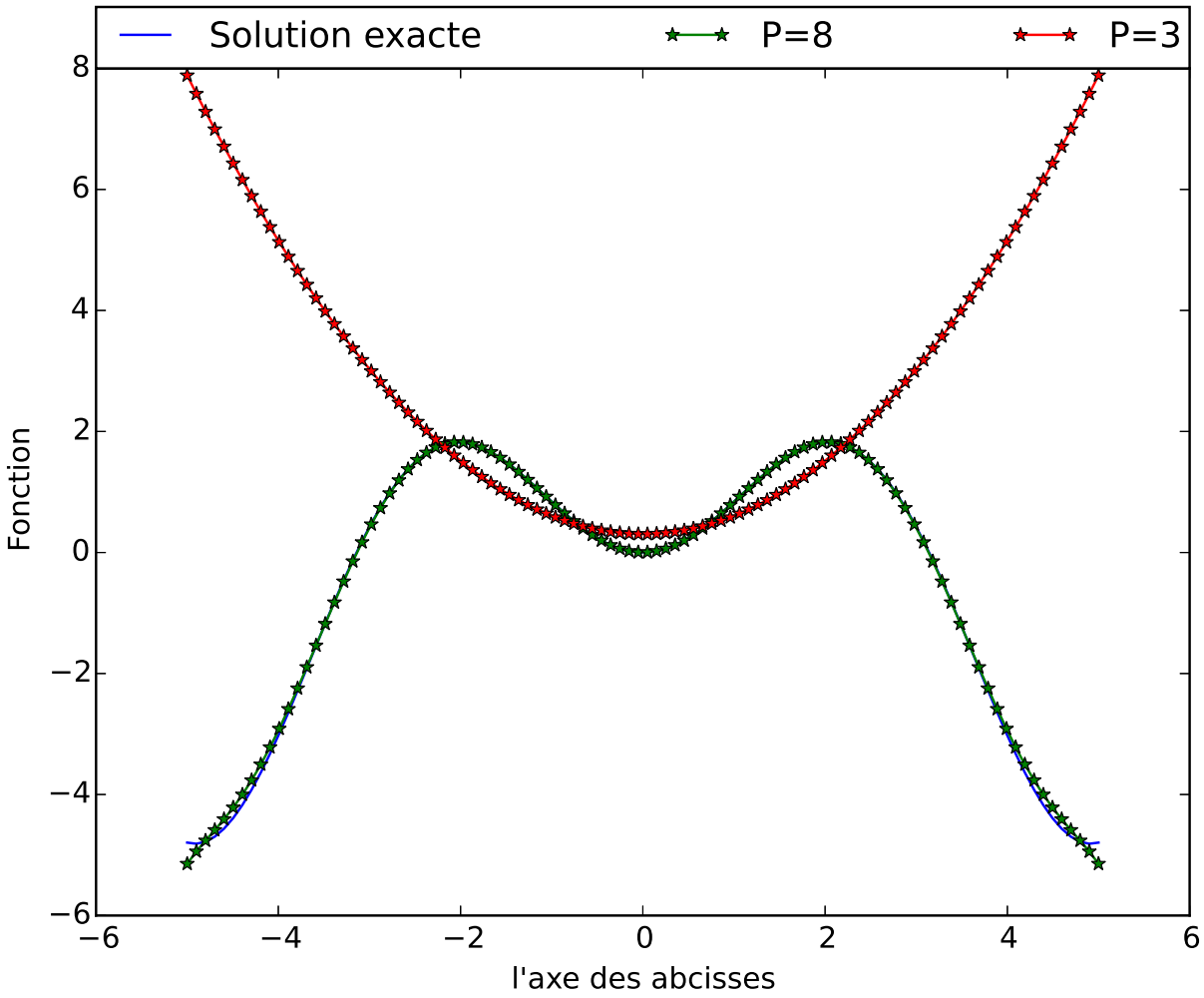


FIGURE 6.2 – *Approximation PC de degré 8 et de degré 3 pour la fonction* $g(x) = x\sin(x)$ *avec une base du polynôme d'Hermite et pour une variable aléatoire gaussienne* $x \sim \mathcal{N}(0,1)$

base de polynômes d'Hermite pour approcher la fonction $g(x) = x\sin(x)$, en supposant que x suit une loi normale centrée réduite. On remarque que l'approximation PC pour $P = 3$ (courbe en rouge) est encore assez loin de la solution exacte, et en augmentant le degré polynomiale, à $P = 8$ (courbe en vert) l'erreur se réduit.

6.3 Calcul des coefficients polynomiaux

On explicite ici les différentes méthodes utilisées dans la littérature pour le calcul des coefficients polynomiaux. Toutefois, une attention particulière sera portée aux méthodes non intrusives.

6.3.1 Méthodes intrusives

Dans cette partie, on présente les méthodes intrusives pour la construction de l'approximation polynomiale. Dans la littérature on retrouve plusieurs nominations de ces méthodes, telles que les méthodes stochastiques Galerkin [151] ou les méthodes éléments finies stochastiques spectrales (*Spectral Stochastic Finite Element Method*) [131]. Ces méthodes ont été les premières à être utilisées dans le cadre du chaos polynomial pour la propagation d'incertitudes [49, 85, 105, 131]. L'idée principale est de reformuler les équations aux dérivées partielles (ou les équations différentielles) de façon à ce que le calcul des coefficients polynomiaux soit effectué implicitement pendant la résolution de ces équations.
Supposons que le modèle $Y = \mathcal{M}(s, t, \mathbf{X})$ s'écrit de la façon suivante (sans perte de généralité) dans un domaine physique $\mathcal{D} \subset \mathbb{R}^l, l = 1, 2, 3$ et $T > 0$:

$$\begin{cases} \partial_t Y_t(s,t,\mathbf{X}) = \mathcal{L}(Y), & \mathcal{D} \times [0,T] \times \mathbb{R}^M, \\ \mathcal{B}(Y) = 0, & \partial\mathcal{D} \times [0,T] \times \mathbb{R}^M, \\ Y = Y_0, & \mathcal{D} \times \{t = 0\} \times \mathbb{R}^M, \end{cases} \tag{6.15}$$

où $\mathcal{L}$ est un opérateur différentiel, $\mathcal{B}$ représente les conditions aux limites, Y_0 est la condition initiale et $\mathbf{X} = (X_1, ..., X_M) \in \mathbb{R}^M, M \geq 1$ est l'ensemble des variables aléatoires indépendantes qui caractérisent les paramètres incertains de l'équation. Par ailleurs, on suppose que $\{\psi_k(\mathbf{X})\}$ est la base polynomiale orthonormale adéquate qui vérifie les conditions définies dans la section 6.2.2. La projection de Y s'écrit alors de la façon suivante :

$$\hat{Y} = \sum_{k=0}^{M_t} \alpha_k(s,t)\psi_k(\mathbf{X}), \qquad \alpha_k(s,t) = \mathbb{E}[Y(s,t,\mathbf{X})\psi_k(\mathbf{X})]. \tag{6.16}$$

Le calcul des coefficients suppose donc la connaissance de la solution Y. La méthode intrusive basée sur l'approche Galerkin s'appuie sur le nouveau problème suivant. On cherche une

solution $\hat{Y}$ qui résout le système suivant :

$$\begin{cases} \mathbb{E}[\partial_t \hat{Y}(s,t,\mathbf{X})\psi_k(\mathbf{X})] = \mathbb{E}[\mathcal{L}(\hat{Y})\psi_k(\mathbf{X})], & \mathcal{D}\times[0,T]\times\mathbb{R}^M, \\ \mathbb{E}[\mathcal{B}(\hat{Y})\psi_k(\mathbf{X})] = 0, & \partial\mathcal{D}\times[0,T]\times\mathbb{R}^M, \\ \alpha_k = \mathbb{E}[Y_0\psi_k(\mathbf{X})], & \mathcal{D}\times\{t=0\}\times\mathbb{R}^M, \end{cases} \tag{6.17}$$

L'équation ci-dessus, montre que l'approximation polynomiale suppose la résolution d'un système d'équations déterministes couplées, la taille du système est $\begin{pmatrix} M_t + M \\ M_t \end{pmatrix}$ correspondant au nombre de coefficients à estimer.
Les méthodes intrusives pour l'approximation polynomiale, étaient un développement naturel pour les méthodes "classiques" éléments finies. Ce qui permettait de développer plus facilement des théories sur la convergence et l'estimation de l'erreur [43]. Malgré les efforts pour trouver des solutions adaptatives pour cette classe de méthodes [73, 105], elle présente un nombre considérable d'inconvénients qui sont cités dans [133] et que l'on résume ici :

— Les problèmes relatifs aux ressources informatiques ne sont pas négligeables à cause du nombre d'équations important à résoudre.

— Il n'est pas possible de dissocier un élément de la réponse d'un autre dans la résolution des équations, vu que le système est couplé. Pour les équations de Saint-Venant, cela revient à résoudre forcément les variables du débit et de la hauteur d'eau même si on ne traite que l'incertitude par rapport à la hauteur d'eau.

— Pour des équations qui présentent des non-linéarités, la projection polynomiale peut s'avérer assez difficile (problème pour trouver des schémas numériques adaptés).

— Les méthodes intrusives nécessitent la modification des codes numériques. Leur utilisation est par conséquent impossible pour la décomposition en polynômes du chaos des simulations numériques faites à l'aide de codes industriel, car pour ces derniers les schémas utilisés ne sont pas souvent disponible pour l'utilisateur.

Il est clair que pour le suite de cette thèse, il est n'est pas souhaitable d'utiliser ce genre de méthode, vu que l'on utilise un code industriel pour la résolution des équations de Saint-Venant. C'est la raison pour laquelle que l'on présente les méthodes non-intrusives.

6.3.2 Méthodes non-intrusives

Introduction

Les méthodes non-intrusives se basent sur un ensemble d'évaluations (réalisations) déterministes du modèle afin d'estimer les coefficients polynomiaux [76]. Autrement dit, on utilise le modèle comme une *boite noire* afin d'associer à chaque réalisation de $\mathbf{X}$ sa valeur correspondante du modèle $\mathcal{M}(\mathbf{X})$. L'avantage de ces méthodes réside dans le fait qu'elle utilise le modèle comme boite noire. Aucune modification n'est nécessaire pour générer les quantités d'intérêt. Sur le plan de la gestion de la ressource informatique, les méthodes non-intrusives ont l'avantage d'être facilement parallélisables comparées aux méthodes Galerkin. L'utilisation d'un modèle/ code comme une boite noire permet de rendre la méthode applicable

à n'importe quel modèle indépendamment de la complexité qu'il présente (multiphysiques, couplés, fortement non-linéaire, etc...) [76].
L'inconvénient des méthodes non-intrusives réside principalement sur le nombre de réalisations au modèle dont on a besoin pour construire l'approximation. Elles souffrent en particulier de ce que l'on appelle le fléau de dimensionalité "*curse of dimensionnality*" où le nombre de simulations nécessaires augmente exponentiellement avec la dimension du problème, i.e le nombre de composants du vecteur $\mathbf{X}$ [26]. Par conséquent, beaucoup de travaux visent à réduire le coût de ces méthodes [15, 17, 72, 109, 149].
[133] classe les méthodes non-intrusives qui existent dans la littératures selon deux approches pour le calcul des coefficients polynomiaux : la régression et la projection.

Méthodes de projection

La projection spectrale repose sur l'orthogonalité de la base polynomiale. Le calcul du coefficient α_i est fait de la manière suivante

$$\begin{aligned} \alpha_i &= < \mathcal{M}(\mathbf{X}), \psi_i(\mathbf{X}) > \\ &= \int_\Omega \mathcal{M}(\mathbf{x}) \psi_i(\mathbf{x}) f_X(\mathbf{x}) d\mathbf{x} \end{aligned} \tag{6.18}$$

en utilisant le produit scalaire défini en (6.13) et en utilisant l'orthonormalité de la base polynomiale. L'estimation numérique de l'intégrale est fait classiquement à l'aide des règles de quadrature.

Règles de quadrature univariée

$$I(f) = \int_\Theta f(x)\rho(x) \tag{6.19}$$

Soit $Q_N(f)$ une règle de quadrature, l'approximation d'une intégrale d'une fonction numérique (Eq. 6.19) d'une variable aléatoire x, définie dans un espace de probabilité Θ par une densité de probabilité $\rho(x)$, par la règle de quadrature Q_N s'écrit de la manière suivante :

$$Q_N(f) = \sum_{i=1}^{N} \omega^{(i)} f(x^{(i)}), \tag{6.20}$$

où $\omega^{(i)}$ sont les poids de la formule de la quadrature, et les $x^{(i)}$ sont les nœuds ou les points d'intégration et N est le nombre de points de quadrature nécessaires pour l'approximation de l'intégration. On définit ainsi l'erreur liée à l'estimation de l'intégrale numérique par la formule de quadrature de la manière suivante :

$$R_N^{quad}(f) = I(f) - Q_N(f) \tag{6.21}$$

On dira que la règle de quadrature est exacte si $R_N^{quad}(f) = 0$. Ceci permet de définir le degré de précision (ou *la précision polynomiale*) d'une règle de quadrature comme le nombre

N pour lequel la règle de quadrature est exacte, et ne l'est plus à partir de $N+1$.
Afin de définir une règle de quadrature il suffit de calculer les points et les poids de quadrature. La littérature permet de définir plusieurs règles qui permettent d'estimer les nœuds d'intégration. Globalement, on peut classer les méthodes de détermination sous deux grandes classes : des méthodes semi-déterministes où les points de quadratures sont estimés de façon aléatoires et les poids sont calculés à partir des positions des nœuds d'intégration, et des méthodes déterministes où les poids et les points de quadratures sont déterminés complètement par le nombre de points N. On s'intéressera dans cette thèse à la deuxième classe et plus particulièrement aux règles de quadrature de Gauss.
Les formules de Gauss permettent de choisir les nœuds de façon à avoir la précision polynomiale la plus élevée. Le problème s'écrit de la manière suivante :

— On cherche quel est le nombre de nœuds nécessaires afin de calculer la valeur suivante :

$$\int_{\Theta} f(x)\rho(x) \approx \sum_{i=0}^{N} \omega^{(i)} f(x^{(i)}), \tag{6.22}$$

— le problème revient donc à chercher $2N+2$ inconnues (N points de quadrature $x^{(i)}$ + N poids $\omega^{(i)}$).

L'idée de la quadrature de gauss est de chercher une formule qui soit exacte pour l'ensemble des polynômes de degré au plus $2N+1$ ce qui revient à résoudre l'ensemble des équations suivantes :

$$\int_{\Theta} x^k \rho(x) \approx \sum_{i=0}^{N} \omega^{(i)} (x^k)^{(i)}, \qquad \forall k \in [0, 2N+1] \tag{6.23}$$

La résolution des $2N+2$ équations précédentes permet de calculer les points et les poids de quadratures. La précision polynomiale de cette méthode est $2N+1$.
Pour des lois de probabilité connues, les points de quadratures suivant la règle de Gauss [70, 50, 138] sont connues, par exemple pour des variables aléatoires normales les points de la règle de quadrature de degré N, sont les racines du polynômes d'Hermite de degré $N+1$ et ont une précision polynomiale $2N+1$ (quadrature de Gauss-Hermite). De même, pour une loi uniforme les points de quadrature sont les racines du polynôme de Legendre (quadrature de Gauss-Legendre). Pour les deux dernières règles, les poids d'intégration sont les valeurs de la loi de probabilité aux points de quadrature.

Tensorisation des règles de quadrature Afin de traiter complètement le problème de la détermination des coefficients polynomiaux dans le cadre du chaos polynomial, il faut déterminer des règles de quadrature pour des vecteurs de variables aléatoires. Supposons que $\mathbf{x} = \{x_1, ..., x_{dim}\}$ est le vecteur des variables aléatoires, on suppose aussi que $Q^{(i)}_{N_i}(f), \forall i \in [1, ..., dim]$ sont les règles de quadrature correspondantes à chaque variable aléatoire de $\mathbf{x}$. Il est possible de construire une règle de quadrature pour le vecteur $\mathbf{x}$ à partir de la tensorisation de toutes les règles de quadrature grâce à de l'hypothèse d'indépendance entre les variables

aléatoires. On suppose que $\mathbf{Q^{FT}}$ est la règle correspondante pour le vecteur $\mathbf{x}$:

$$\mathbf{Q^{FT}} = \bigotimes_{i=1}^{dim} Q_{N_i}^{(i)}. \tag{6.24}$$

En général, on suppose que la précision polynomiale est la même pour toutes les règles de quadrature, mais ceci n'est pas tout le temps valide ne serait ce que du fait que toutes les variables aléatoires ne sont pas décrites par la même densité de probabilité. Le deuxième inconvénient de l'utilisation du produit tensoriel est le fléau de la dimensionnalité.
Finalement, la complexité de la méthode de quadrature basée sur le produit tensoriel complet est $N(\mathbf{l}) = \prod_{i=1}^{d} N(l_i)$, qui a une croissance exponentielle par rapport à la dimension du problème.

Grille de Smolyak Afin de réduire le coût de cette méthode et en gardant la qualité de l'approximation, on a introduit les quadratures de Smolyak [69, 128]. Elles ont pour but de supprimer quelques évaluations du modèle. On définit d'abord, l'opérateur $\Delta_l^Q = Q_l - Q_{l-1}$, et $\Delta_1^Q = Q_1$. L'équation 6.24 devient alors :

$$\begin{aligned}
\mathbf{Q_l^{FT}} R &= (Q_{l_1} \otimes ... \otimes Q_{l_d})R \\
&= ((\textstyle\sum_{i_1=1}^{l_1} \Delta_{i_1}^Q) \otimes ... \otimes (\sum_{i_d=1}^{l_d} \Delta_{i_d}^Q))R \\
&= \textstyle\sum_{i_1=1}^{l_1} ... \sum_{i_d=1}^{l_d} (\Delta_{i_1}^Q \otimes ... \otimes \Delta_{i_d}^Q)R \\
&= \textstyle\sum_{\mathbf{i} \in \mathfrak{L}_\mathbf{l}^{FT}} (\Delta_{i_1}^Q \otimes ... \otimes \Delta_{i_d}^Q)R \\
&= \textstyle\sum_{\mathbf{i} \in \mathfrak{L}_\mathbf{l}^{FT}} \mathbf{\Delta_i}^Q R
\end{aligned} \tag{6.25}$$

où

$$\mathfrak{L}_\mathbf{l}^{FT} = \{\mathbf{i} \in \mathbb{N}_+^d, i_j \leq l_j, j = 1, ..., d\}$$

et R est une fonction carré intégrable. Les quadratures de Smolyak sont définies en prenant un ensemble d'indices $\mathfrak{L}_\mathbf{l}$ qui appartient à $\mathfrak{L}_\mathbf{l}^{FT}$. En général, le sous-ensemble $\mathfrak{L}_\mathbf{l}$ est défini à partir d'un niveau total de quadrature ℓ [128].

$$\mathfrak{L}_\ell = \{\mathbf{i} \in \mathbb{N}_+^d, \sum_{j=1}^{d} i_j \leq \ell + d - 1\}$$

La figure 6.3 montre la réduction du nombre d'appels aux modèle grâce à la technique de Smolyak. Sur la figure on considère que le modèle à deux variables en entrée $d = 2$ et que le niveau de quadrature $\ell = 5$. Dans cette exemple on réduit le nombre de simulations de 1089 points de quadrature en utilisant le produit tensoriel complet à 145 simulations en utilisant la technique de Smolyak. On renvoie le lecteur à [152, 149] pour plus de discussion sur l'utilisation des grilles creuses pour l'estimation des intégrales.
Finalement, afin d'estimer les coefficients polynomiaux α_j dans l'équation (6.18) il suffit d'appliquer la technique de Smolyak, en choisissant les bons points de quadratures donnés par [138] et en remplaçant R par $\mathcal{M}\psi_j$ dans l'équation (6.25).

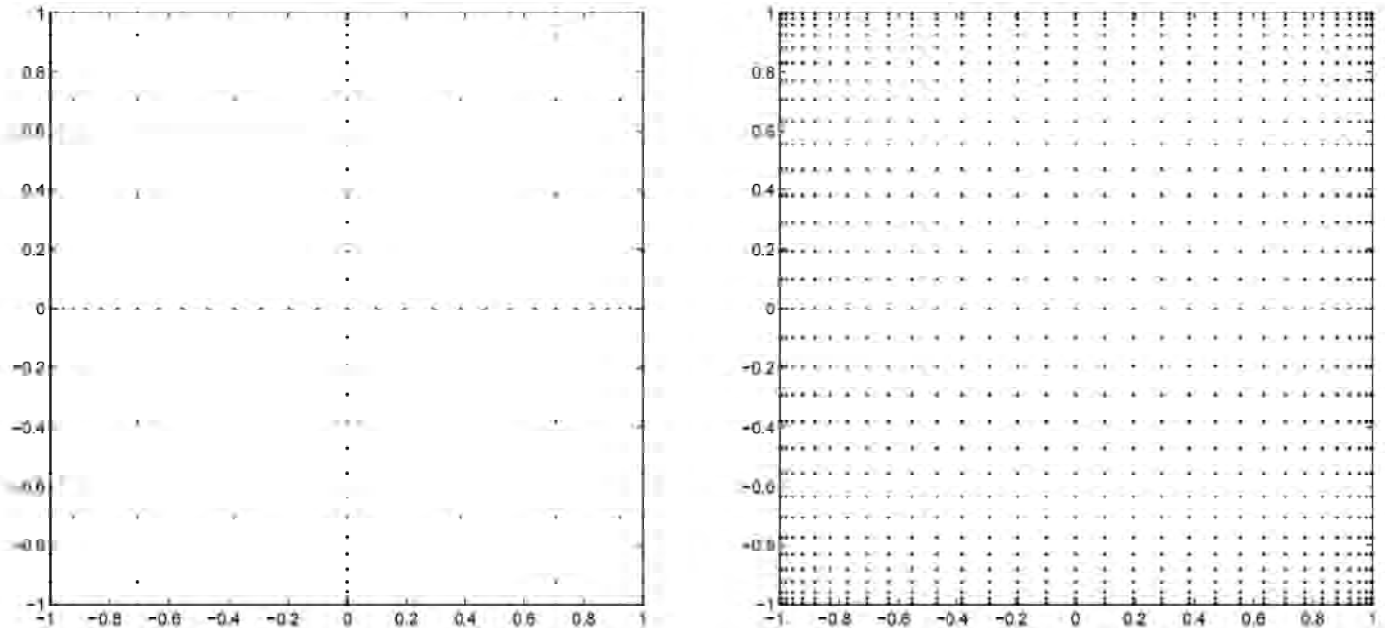


FIGURE 6.3 – *Comparaison entre l'utilisation d'une grille pour un nombre de variables égal à 2 basée sur le produit tensoriel complet (à droite) et une grille basée sur la technique de Smolyak avec un niveau de quadrature $\ell = 5$ (à gauche). Source : [152]*

Algorithme d'estimation des coefficients par projection. L'estimation des coefficients polynomiaux par la méthode de projection se résume finalement à estimer l'intégrale Eq. (6.18) à l'aide d'une règle de quadrature Eq. (6.20) adaptée selon la dimension du problème (utilisation de grille creuse par exemple). Les points de quadrature étant définis dans un espace stochastique normalisé, il faut définir des échantillons dans l'espace physique, correspondant à la physique simulée, à l'aide d'une transformation isoprobabiliste. L'équation (6.18) devient alors :

$$\begin{aligned} \alpha_i &= \mathrm{Q}_{N_{quad}}(\mathcal{M}(\mathbf{X})\psi_i(\mathbf{X})) \\ &= \sum_{i=1}^{N_{quad}} \boldsymbol{\omega}^{(i)} \mathcal{M}(\mathbf{X}^{(i)}(\boldsymbol{\xi}^{(i)}))\psi_i(\boldsymbol{\xi}^{(i)}) \end{aligned} \tag{6.26}$$

où $\boldsymbol{\xi^{(i)}} = \{\xi_1^{(i)}, ..., \xi_M^{(i)}\}$ et $\boldsymbol{\omega^{(i)}} = \{\omega_1^{(i)}, ..., \omega_M^{(i)}\}$, avec $\xi_j^{(i)}$ les points de quadrature de la $j-$ème dimension en entrée et les $\omega_j^{(i)}$ les poids correspondants.

Algorithm 1 Méthode de projection pour l'estimation des coefficients

1: Définir un ordre polynomial P.
2: Définir une règle de quadrature adaptée (selon la dimension du problème), et identifier les points et les poids de quadrature.
3: Définir dans l'espace des entrée les réalisations correspondantes à chaque points de quadrature (Faire une transformation isoprobabiliste entre l'espace physique et l'espace stochastique normalisé).
4: Intégrer le modèle pour toutes les réalisations de l'espace incertain.
5: Calculer les coefficients à l'aide de la formule Eq. (6.26).

Méthode de régression

La méthode de régression a pour but de minimiser l'erreur quadratique moyenne de l'approximation de la réponse du modèle par le polynôme du chaos [11]. En utilisant la notation suivante :

$$\boldsymbol{\alpha} = \{\alpha_0, ..., \alpha_P\}^T \tag{6.27}$$

$$\boldsymbol{\Psi}(\mathbf{X}) = \{\psi_0(\mathbf{X}), ..., \psi_P(\mathbf{X})\}^T. \tag{6.28}$$

Le problème de régression se limite à trouver la valeur de $\hat{\boldsymbol{\alpha}}$ qui minimise la fonction suivante :

$$\mathcal{J}(\boldsymbol{\alpha}) \equiv \mathbb{E}[(\boldsymbol{\alpha^T \Psi(X)} - \mathcal{M}(\mathbf{X}))^2] \tag{6.29}$$

La condition de minimalité $\frac{d\mathcal{J}}{d\boldsymbol{\alpha}}(\hat{\alpha}) = 0$ donne l'équation suivante :

$$\mathbb{E}[\boldsymbol{\Psi}(\mathbf{X})\boldsymbol{\Psi}^{\mathbf{T}}(\mathbf{X})]\hat{\alpha} = \mathbb{E}[\boldsymbol{\Psi}(\mathbf{X})\mathcal{M}(\mathbf{X})]. \tag{6.30}$$

Cette équation donne la similitude entre les deux méthodes pour l'estimation de coefficients polynomiaux. Par conséquent on retrouve que le modèle polynomial dont les coefficients sont estimés par la méthode de projection minimise bien l'erreur quadratique moyenne.
Afin d'estimer les coefficients avec une méthode de régression, on considère un plan d'expérience $\boldsymbol{\mathcal{X}} = \{\mathbf{x^{(1)}}, ..., \mathbf{x^{(N_r)}}\}^T$ du vecteur aléatoire en entrée $\mathbf{X}$. L'estimation des coefficients polynomiaux suivant la méthode de régression donne la formule suivante :

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{\Psi}^{\mathbf{T}}(\boldsymbol{\mathcal{X}})\boldsymbol{\Psi}(\boldsymbol{\mathcal{X}}))^{-1}\boldsymbol{\Psi}^{\mathbf{T}}(\boldsymbol{\mathcal{X}})\mathbf{Y}. \tag{6.31}$$

Avec $\boldsymbol{\Psi(\mathcal{X})} = \{\boldsymbol{\Psi}(\mathbf{x^{(1)}}), ..., \boldsymbol{\Psi}(\mathbf{x^{(N_r)}})\}$.
La matrice $(\boldsymbol{\Psi^T(\mathcal{X})\Psi(\mathcal{X})})^{-1}$ est appelée matrice d'information. La projection spectrale est aussi nommée une méthode de *quasi-régression* [101], car la matrice d'information est ignorée dans ce cas. En effet, Le terme de l'espérance mathématique à gauche de l'équation 6.30 devrait tendre vers la matrice unité, car il représente la matrice de corrélation du vecteur aléatoire $\psi(\mathbf{X})$ dont les composantes sont décorrélés par définition. Ce qui permet de retrouver l'équation qui définit la méthode de projection :

$$\hat{\boldsymbol{\alpha}} = \mathbb{E}[\boldsymbol{\Psi}(\mathbf{X})\mathcal{M}(\mathbf{X})]. \tag{6.32}$$

Afin de tenir compte de toute l'information obtenue par le plan d'expérience, il est important que la matrice d'information soit de rang plein. C'est la raison pour laquelle, il important que la taille du plan d'expérience N_r soit supérieur au nombre d'éléments de la base polynomiale P. L'inversion de la matrice d'information peut s'avérer difficile si elle est mal-conditionnée. En général, on résout le problème décrit par l'équation (6.31) en utilisant une technique numérique plus robuste comme la décomposition en valeurs singulières [14]. Les techniques d'échantillonnage de la variable en entrée peuvent s'appuyer sur des plans d'expérience aléatoires [15, 133], ou sur des plans d'expérience déterministes (tels que les grilles creuses de Smolyak) [11].

Algorithme d'estimation des coefficients polynomiaux par la méthode de régression On présente ci-dessous l'algorithme correspondant à l'estimation des coefficients par la méthode de régression.

Algorithm 2 Méthode de régression pour l'estimation des coefficients

1: Définir un ordre polynomial P.
2: Définir un plan d'expérience permettant de définir des échantillons.
3: Intégrer le modèle pour chaque élément du plan d'expérience, et estimer la matrice $\mathbf{Y}$ de l'équation 6.31.
4: Calculer la matrice d'information en intégrant les éléments de la base polynomial sur le plan d'expérience.
5: Calculer les coefficients à l'aide de la formule Eq. (6.31).

6.4 Quantification des incertitudes et analyse de sensibilité

La projection de la réponse d'un modèle $Y = \mathcal{M}(\mathbf{X})$ sur une base polynomiale orthonormale permet de calculer les statistiques de la réponse du modèle à partir des coefficients polynomiaux. Les travaux de [132] montrent que les statistiques de la variables Y s'écrivent de la façon suivante :

$$\begin{aligned} m_Y &= \alpha_0 \\ VarY &= \sum_{i=1}^{N_{PC}} \alpha_i^2 \end{aligned} \tag{6.33}$$

De même tous les moments statistiques peuvent être estimés à l'aide des coefficients polynomiaux [11]. En particulier, les coefficients d'asymétrie et d'aplatissement :

$$\begin{aligned} \delta_Y &= \frac{1}{VarY^{3/2}} \sum_{i=1}^{N_{PC}} \sum_{j=1}^{N_{PC}} \sum_{k=1}^{N_{PC}} \mathbb{E}[\psi_i \psi_j \psi k] \alpha_i \alpha_j \alpha_k \\ \kappa_Y &= \frac{1}{VarY^2} \sum_{i=1}^{N_{PC}} \sum_{j=1}^{N_{PC}} \sum_{k=1}^{N_{PC}} \sum_{l=1}^{N_{PC}} \mathbb{E}[\psi_i \psi_j \psi_k \psi_l] \alpha_i \alpha_j \alpha_k \alpha_l \end{aligned} \tag{6.34}$$

L'analyse de sensibilité peut être aussi faite l'aide des coefficients polynomiaux. [132] montre que les indices de Sobol' peuvent être estimés à l'aide des coefficients polynomiaux. Afin de définir les indices de Sobol', on réécrit la décomposition polynomiale de façon cohérente avec ce qui est présenté dans [132] :

$$Y \equiv \mathcal{M}(\mathbf{X}) \approx \sum_{i=1}^{N_{PC}} \alpha_i \psi_i(\mathbf{X}), \tag{6.35}$$

on définit aussi l'ensemble $\mathfrak{J}_{i_1,...,i_s}$ comme l'ensemble des coefficients pour lesquels seulement les indices $(i_1, ..., i_s)$ sont non nuls. On notera, à titre d'exemple que l'ensemble $\mathfrak{J}_i$ correspond aux termes polynomiaux qui ne dépendent que d'un seul paramètre. On peut réécrire maintenant l'équation 6.35 en fonction des paramètres dont elle dépend :

$$\begin{aligned} &\mathcal{M}(\mathbf{X}) \approx \alpha_j + \sum_{i=1}^{M} \sum_{j \in \mathfrak{J}_i} \alpha_j \psi_j(X_i) \\ &+ \textstyle\sum_{1 \le i_1 < i_2 \le M} \sum_{j \in \mathfrak{J}_{i_1,i_2}} \alpha_j \psi_j(X_{i_1}, X_{i_2}) \\ &+ ... + \textstyle\sum_{1 \le i_1 < ... < i_s \le M} \sum_{j \in \mathfrak{J}_{i_1,...,i_s}} \alpha_j \psi_j(X_{i_1}, ..., X_{i_s}) \\ &+ ... + \textstyle\sum_{j \in \mathfrak{J}_{i_1,...,i_d}} \alpha_j \psi_j(X_{i_1}, ..., X_{i_d}). \end{aligned} \tag{6.36}$$

Dans l'équation précédente, il est clair que chaque terme de la forme $\displaystyle\sum_{j \in \mathfrak{J}_{i_1,...,i_s}} \alpha_j \psi_j(X_{i_1}, ..., X_{i_s})$ est le terme polynomial qui ne dépend que de tout l'ensemble des paramètres $i_1, ..., i_s$ et seulement de ceux ci. Finalement, l'indice de Sobol' peut s'écrire de la façon suivante :

$$S_{i_1,\ldots,i_s} = \frac{\sum_{j \in \mathfrak{J}_{i_1,\ldots,i_s}} \alpha_j^2}{\sum_{j=1}^{N_{PC}} \alpha_j^2} \tag{6.37}$$

6.5 Limitations de l'approximation polynomiale

6.5.1 Polynômes du chaos en grande dimension

La section précédente a présenté différentes méthodes pour la construction d'un modèle réduit à l'aide d'une base de polynôme du chaos. Que ce soit les méthode intrusives ou non-intrusives, le nombre d'éléments de la base devient très large quand la dimension du problème augmente. Quand on tronque la décomposition polynomiale de façon à ne retenir que les polynômes multivariés correspondants à un degré total inférieur à P, le nombre de termes augmente exponentiellement avec la dimension du problème :

$$N_{PC} = \frac{(M+P)!}{M!\ P!}. \tag{6.38}$$

En utilisant la formule de Stirling $M! \approx \sqrt{2\pi M}(\frac{M}{e})^M$ il est assez clair que le problème augmente exponentiellement avec la dimension M. Ce problème rend l'application d'un méta-modèle basé sur le polynôme du chaos assez compliqué dans des domaines tels que les géosciences. C'est la raison pour laquelle plusieurs études se sont intéressées à réduire le coût de la méthode en utilisant plusieurs voies de recherche. La première voie est d'utiliser des plans d'expériences moins coûteux tels que la grille de Smolyak [96]. Malheureusement, l'introduction de cette grille fait apparaitre l'erreur d'aliasing. Les travaux de la thèse de J. Winokur [149] ont introduit une méthode permettant de réduire cette erreur, on présentera l'idée générale de cette méthode dans la première sous-section. Une deuxième méthode consiste négliger certains coefficients polynomiaux et notamment ceux qui représentent des effets croisés [15].

Les méthodes adaptatives pseudo spectrales

Le risque de l'utilisation de la grille de Smolyak pour estimer les coefficients polynomiaux réside dans l'erreur d'aliasing : l'orthogonalité définie par l'équation 6.7 n'est pas garantie pour tous les éléments de la base et pour tous les points de quadratures. En effet, si la dimension du problème est importante et que le problème requiert l'augmentation du degré de troncature, la taille de la grille de Smolyak qui donne une approximation des coefficients polynomiaux sera elle aussi de grande taille [4]. En effet, l'approximation pseudo spectrale définie dans la partie 6.3.2 permet d'approcher numériquement les coefficients de l'approximation polynomiale α_i avec la formule suivante :

$$\alpha_i = < \mathcal{M}(\mathbf{X}), \psi_j(\mathbf{X}) > = \sum_{i=1}^{N_{quad}} \omega^{(i)} \mathcal{M}(\mathbf{X}^{(i)}) \psi_j(\mathbf{X}^{(i)}) \tag{6.39}$$

Le calcul du coefficient polynomial est exact quand la règle de quadrature considérée est exacte, par contre l'utilisation d'une tensorisation des règles de quadrature implique l'utilisation d'un plan d'expérience très large. En conséquence, on a introduit l'algorithme de Smolyak (équation 6.25) qui permet d'alléger le cout de calcul des coefficients polynomiaux. Cependant, l'utilisation des grilles basées sur l'algorithme de Smolayk induit une erreur sur le calcul des coefficients polynomiaux que l'on nomme *erreur d'aliasing*. Supposons que $\{\alpha_i^{exact}, i = \{1, ..., N_{PC}\}\}$ est l'ensemble des coefficients polynomiaux en prenant leurs valeurs exactes et $\{\alpha_i^{Smolyak}, i = \{1, ..., N_{PC}\}\}$ est l'ensemble des coefficients polynomiaux calculé avec la méthode de Smolyak. L'erreur d'aliasing $\varepsilon_{aliasing}$ est définie de la manière suivante :

$$\varepsilon_{aliasing} = \sum_{i=1}^{N_P C} (\alpha_i^{exact} - \alpha_i^{Smolyak})^2 \tag{6.40}$$

Il est évident que l'erreur d'aliasing n'est pas nulle, car le calcul du coefficient n'est basé que sur une approximation numérique de l'intégrale. Cette approximation va aussi avoir un influence sur l'hypothèse d'orthogonalité de la base polynomiale. En reprenant l'équation qui définit la méthode de projection, on définit le produit scalaire calculé avec la méthode de Smolyak de la façon suivante $<,>_{Smolyak}$:

$$\begin{aligned}
\alpha_i^{Smolyak} &= < \mathcal{M}(\mathbf{X}), \psi_i(\mathbf{X}) >_{Smolyak} \\
&= < \textstyle\sum_{j=0}^{\infty} \alpha_j, \psi_j(\mathbf{X}), \psi_i(\mathbf{X}) >_{Smolyak} \\
&= \textstyle\sum_{j=0}^{\infty} \alpha_j < \psi_j(\mathbf{X}), \psi_i(\mathbf{X}) >_{Smolyak} \\
&= \textstyle\sum_{j=0}^{N_{PC}} \alpha_j < \psi_j(\mathbf{X}), \psi_i(\mathbf{X}) >_{Smolyak} + \sum_{j=N_{PC}+1}^{\infty} \alpha_j < \psi_j(\mathbf{X}), \psi_i(\mathbf{X}) >_{Smolyak}
\end{aligned} \tag{6.41}$$

Dans cette équation on approche le produit scalaire numériquement à l'aide d'une méthode de Smolyak. Ceci revient à estimer toutes les grandeurs $< \psi_i(\mathbf{X}), \psi_j(\mathbf{X}) >$ avec la méthode des quadratures. Théoriquement cette grandeur vaut δ_{ij} mais son estimation numérique ne garantit en aucun cas la validité théorique. On remplace maintenant la valeur de $\alpha_j^{Smolyak}$ dans l'équation 6.40.

$$\varepsilon_{aliasing} = \sum_{i=1}^{N_P C} (\alpha_i^{exact} - \underbrace{\sum_{j=0}^{N_{PC}} \alpha_j < \psi_j(\mathbf{X}), \psi_i(\mathbf{X}) >_{Smolyak}}_{\text{Aliasing interne}} - \underbrace{\sum_{j=N_{PC}+1}^{\infty} \alpha_j < \psi_j(\mathbf{X}), \psi_i(\mathbf{X}) >_{Smolyak}}_{\text{Aliasing externe}})^2 \tag{6.42}$$

Cette équation permet de découper l'erreur d'aliasing en deux parties : une erreur d'aliasing interne qui est due à l'erreur de l'approximation du produit scalaire entre deux éléments de le base tronquée, alors que l'alisasing externe représente l'erreur commise sur l'estimation de l'orthogonalité pour chaque élément de la base tronquée avec les autres éléments qui ne sont pas considérés lors de la troncature. Pour plusieurs règles de quadrature utilisées dans la pratique, [138] montre que l'erreur due à l'aliasing interne est non nulle, et que son ordre de grandeur est égale à celui de la valeur du coefficient correspondant . L'article de [29] prouve que les règles de quadratures usuelles, qui ont une précision polynomiale, permettent d'annuler l'erreur d'aliasing interne en prenant le nombre de points de quadrature qui cor-

respond bien. Cependant, la seule méthode qui permet de réduire l'erreur d'aliasing externe est de tronquer l'approximation polynomiale jusqu'au moment où l'ordre de grandeur des coefficients est très petit.
Le but des méthodes adaptatives est de chercher la grille de points de quadrature basée sur un algorithme de Smolyak afin de réduire les erreurs d'aliasing sans augmenter le cout de calcul significativement. Une méthode adaptative est présenté dans [48] afin de déterminer l'ensemble $\mathfrak{L}_{\mathbf{l}}$ de manière itérative de façon à sélectionner les points de grilles qui peuvent réduire l'erreur d'aliasing. On renvoie le lecteur à [149, 29] pour plus de détails sur cette méthode.

Stratégie de troncature de la décomposition polynomiale

Classiquement, on calcule tous les coefficients. Ici, on propose de n'en calculer que certains. On considère la décomposition polynomiale suivante :

$$Y \equiv \mathcal{M}(\mathbf{X}) = \sum_{\mathbf{i}\in\mathcal{I}_M, \mathcal{I}_M\subset\mathbb{N}^M} \alpha_{\mathbf{i}}\psi_{\mathbf{i}}(\mathbf{X}) \tag{6.43}$$

On utilises ici la formulation multivariée de la décomposition polynomiale. Cette écriture permet de comprendre la stratégie de troncature. En effet, le choix d'un ensemble non vide fini d'indices multivariés $\mathbf{i} = \{i_1, ..., i_M\} \in \mathbb{N}^M$ permet de définir complètement une représentation polynomiale. Classiquement lorsqu'on a définit une stratégie de troncature, on suppose que le degré de l'approximation polynomiale est P si $||\mathbf{i}||_1 = \sum_{k=0}^{M} i_k \leq P$. De cette façon nous pouvons définir un ensemble $\mathcal{A}^{M,P}$qui contient les indices de l'approximation polynomiale que l'on défini selon la dimension du problème et le degré de troncature de la manière suivante :

$$\mathcal{A}^{M,P} = \{\mathbf{i} \in \mathbb{N}^M, ||\mathbf{i}||_1 \leq P\}. \tag{6.44}$$

Il est maintenant évident de montrer que le cardinal de cet ensemble est le nombre de termes de la décomposition polynomiale.

$$card(\mathcal{A}^{M,P}) = \frac{(M+P)!}{M!\ P!} = N_{PC}. \tag{6.45}$$

Cette relation permet de montrer que la stratégie de troncature choisie a un impact sur le nombre de termes de la décomposition polynomiale. Dans la littérature il existe différentes études qui s'intéressent à la recherche de la stratégie de troncature optimale qui permet d'avoir une bonne approximation polynomiale en utilisant le moins de termes possibles. Les stratégies de troncature existantes dans la littérature s'appuient sur le principe de *sparsity-of-effects* introduit par [98]. Ce principe dit que pour la plupart des modèles, l'incertitude est principalement expliquée par les effets principaux et les effets d'interaction de faible ordre.
On se limitera ici à présenter les travaux de [16, 15] sur la stratégie des ensembles d'indices

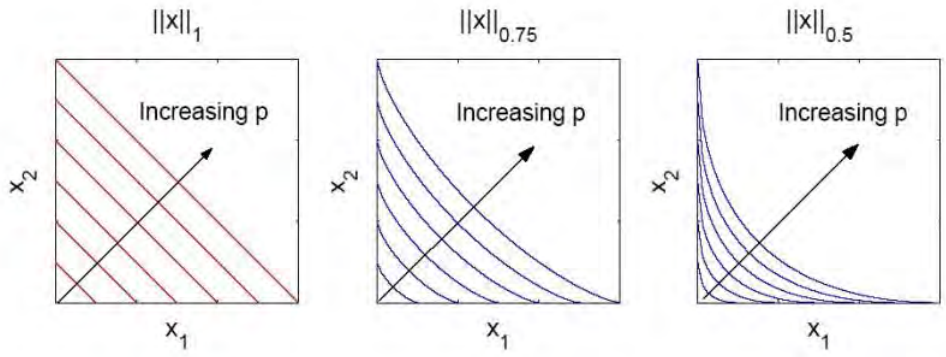


FIGURE 6.4 – *Principe de la stratégie de troncature basée sur la q-norm. Source : [15]*

hyperboliques. Pour cela, on définit l'ensemble suivant :

$$\mathcal{A}_q^{M,P} = \{\mathbf{i} \in \mathbb{N}^M, ||\mathbf{i}||_q \leq P\}, \tag{6.46}$$

où

$$||\mathbf{i}||_q = (\sum_{k=1}^{M} i_k^q)^{1/q} \tag{6.47}$$

[16] propose de remplacer l'ensemble $\mathcal{A}^{M,P}$ dans l'approximation polynomiale par l'ensemble $\mathcal{A}_q^{M,P}$ qui est basé sur une $q-$norme, $0 < q < 1$. L'utilisation de cette quasi-norme a pour effet de négliger les termes correspondant aux interactions d'ordre élevé. Quand $q = 1$, on retrouve la stratégie classique de troncature. Plus q s'approche de 0, moins de termes sont retenus dans la base. La figure 6.4 montre l'ensemble des éléments retenus pour une espace de dimension 2 en fonction de la valeur de q pour des valeurs de degré polynomiaux (p dans cette figure) qui augmentent. La réduction de la taille de la base polynomiale permet de réduire le cout du calcul.

Polynomes du chaos creux basé sur la méthode "LAR"

Les résultats avantageux de la stratégie des ensembles d'indices hyperboliques présentés ci-dessus ont permis de pousser la réflexion un peu plus et de chercher à déterminer les éléments de la base qui ont le plus d'interaction avec le modèle. Dans sa thèse [15], Blatman définit une méthode itérative afin de chercher une représentation polynomiale creuse que l'on présentera ici en utilisant une méthode LAR.
Least Angle Regression (LAR) [35] est une méthode qui permet de sélectionner à partir d'une base polynomiale, le terme qui a le plus d'impact sur le modèle. L'idée est donc de minimiser la distance entre le modèle et cet élément jusqu'à ce qu'au moment ou un autre élément de la base devient aussi important que le premier. On intègre donc le nouvel élément dans la base à retenir et on continue jusqu'au ce qu'on atteint un seuil d'erreur défini au préalable.
Cette méthode permet de construire la décomposition polynomiale en utilisant seulement les éléments de la base les plus influents. Quand $N_{PC} < N_{Re}$ (dans **Algorithm** 3) l'algorithme

Algorithm 3 LAR

Définir un ordre polynomial P et définir le nombre d'éléments maximum de la base N_{PC}.

Définir un budget maximum de réalisation N_{Re}.
Calculer les réalisations avec le modèle direct $\{Y^{(1)}, ..., Y^{(N_{Re})}\}$.
Initialiser le vecteur contenant les coefficients au vecteur nul, $\alpha_1, ..., \alpha_{N_{PC}} = 0$.
Initialiser le vecteur résiduel au vecteur des réalisations $\mathcal{R} = \{Y^{(1)}, ..., Y^{(N_{Re})}\}$.
$iter = 0$.
while $iter \leq min(N_{PC}, N_{Re})$ **do**
    Trouver le(s) élément(s) de la base $\Psi_{actual} = \{\psi_{i1}, ..., \psi_{iN_{actual}}\}$ le(s) plus corrélé(s) avec $\mathcal{R}$.
    Minimiser la distance entre $\mathcal{R}$ et Ψ_{actual} jusqu'au moment où un autre élément de la base ψ_j à la même corrélation avec $\mathcal{R}$ que Ψ_{actual}.
    Actualiser le vecteur des coefficients avec les valeurs de $\alpha_{\mathbf{actual}} = \alpha_{i1}, ..., \alpha_{i_{actual}}$
    Actualiser le vecteur de résidus $\mathcal{R} = ||Y - {\alpha_{\mathbf{actual}}}^T \Psi_{actual}||_2$.
    $iter \leftarrow iter + 1$
end while

donne les mêmes résultats que la méthode de régression. Cette méthode peut être améliorée en négligeant les éléments de la base dont les coefficients correspondants sont nuls (négligeables). Cette méthode est appelé LASSO (*Least absolute shrinkage and selection operator*). L'idée de cette méthode est de ne calculer que les coefficients non nuls de l'approximation polynomiale avec une méthode de régression. Cela revient à réécrire l'équation coût que l'on minimise avec la méthode de régression pour le calcul des coefficients polynomiaux de la manière suivante.

$$\mathcal{J}(\boldsymbol{\alpha}) \equiv \mathbb{E}[(\boldsymbol{\alpha}^T \Psi(X) - \mathcal{M}(\mathbf{X}))^2] + \underbrace{\mathcal{C}_{\mathcal{LASSO}} ||\boldsymbol{\alpha}||_1^2}_{\text{terme de pénalité}}, \tag{6.48}$$

où $||\boldsymbol{\alpha}||_1 = \sum_{i=1}^{N_{PC}} |\alpha_i|$, et $\mathcal{C}_{\mathcal{LASSO}}$ est une constante positive. Cette écriture permet d'obtenir une représentation polynomiale creuse, car le terme de pénalité va avoir un effet d'annuler les coefficients négligeables. Dans le cadre de l'algorithme du LAR, ce coefficient va être augmenté au fil des itération de la valeur 0 à une valeur maximale. L'algorithme modifié donne alors :

Par abus de langage on notera LAR pour référer à l'algorithme LAR-LASSO dans la suite de cette thèse.

6.5.2 Polynômes du chaos en présence de discontinuité

Les équations hyperboliques par exemple (comme c'est le cas des équations de Saint-Venant) présentent des discontinuités dans le domaine spatial (ressauts hydraulique c.f 3.3.4) dues au développement des ondes de chocs et de discontinuité de contact. Par conséquent, durant la propagation des incertitudes dues aux paramètres incertains, des discontinuités peuvent apparaitre. L'utilisation d'une base polynomiale pour la représentation des phénomènes de Gibbs peut devenir coûteuse [73].

Algorithm 4 LAR-LASSO

1: Définir un ordre polynomial P et définir le nombre d'éléments maximum de la base N_{PC}.
2: Définir un budget maximum de réalisation N_{Re}.
3: Calculer les réalisations avec le modèle direct $\{Y^{(1)}, ..., Y^{(N_{Re})}\}$.
4: Initialiser le vecteur contenant les coefficients au vecteur nul, $\alpha_1, ..., \alpha_{N_{PC}} = 0$.
5: Initialiser le vecteur résiduel au vecteur des réalisations $\mathcal{R} = \{Y^{(1)}, ..., Y^{(N_{Re})}\}$.
6: $iter = 0$.
7: **while** $iter \leq min(N_{PC}, N_{Re})$ **do**
8: Trouver les éléments de la base $\Psi_{actual} = \{\psi_{i1}, ..., \psi_{iN_{actual}}$ les plus corrélés avec $\mathcal{R}$ (la corrélation entre les vecteur doit être la même).
9: Minimiser la distance entre $\mathcal{R}$ et Ψ_{actual} jusqu'au moment où un autre élément de la base ψ_j à la même corrélation avec $\mathcal{R}$ que Ψ_{actual}.
10: Actualiser le vecteur des coefficients avec les valeurs de $\alpha_{\mathbf{actual}} = \alpha_{i1}, ..., \alpha_{i_{actual}}$
11: Si un coefficient α_j appartenant à $\alpha_{\mathbf{actual}}$ est nul, on l'enlève de l'ensemble.
12: Actualiser le vecteur de résidus $\mathcal{R} = ||Y - {\alpha_{\mathbf{actual}}}^T \Psi_{actual}||_2$.
13: $iter \leftarrow iter + 1$
14: **end while**

On propose de faire un état de l'art non exhaustif des méthodes utilisées afin de faire face à ce genre de problème. [73] montrent que dans le cas d'une instabilité de Rayleigh-Bénard (instabilités qui apparaissent dans les équations thermiques) le remplacement des bases polynomiales suivant le schémas de Askey par des bases d'ondelettes de Haar permet de mieux considérer le problème quand l'instabilité apparait. Cependant pour des solutions plus continues (absences de discontinuités) la base du chaos polynomial donne une meilleure approximation que la base d'ondelettes. Dans une publication de [32], l'utilisation de paramètres incertains basée sur des fonctions dérivatives, en utilisant une méthode de projection pseudo-spectrale pour l'estimation des coefficients polynomiaux, permet de mieux appréhender les solutions discontinues que les polynômes du chaos. Dans la thèse de [139], le problème est géré avec une méthode intrusive qui consiste à changer le schéma numérique pour la résolution de l'équation d'Euler, ainsi qu'elle discrétise les paramètres stochastiquement à l'aide d'une méthode de Galerkin en utilisant un espace de polynômes par morceaux. Une méthode originale proposée par Poette [105, 106] étudiée sur l'équation de Burger et l'équation d'Euler, montre que pour détourner le problème lié au phénomène de Gibbs, il faut porter l'approximation polynomiale sur des variables entropiques et non sur les variables conservatives avec une méthode intrusive. Finalement, ce dernier montre dans une deuxième étude [107], qu'une approche non-intrusive pour utiliser le chaos polynomial est possible même dans le cas d'apparition de discontinuité. Cette méthode consiste à calculer le chaos polynomial sur la solution, puis d'estimer la densité de probabilité à l'aide de ce modèle et ensuite construire une base qui correspond à cette densité de probabilité, finalement il faut reprendre le processus de manière itérative jusqu'au point où on atteint un seuil critique. Cette méthode présente l'avantage de n'utiliser qu'une seule fois les réalisations du modèle et d'être non-intrusive.

En conclusion, le phénomène de Gibbs est assez contraignant pour la représentation en polynôme de chaos. Les études présentés dans cette partie proposent tous de changer la projection de la base polynomiale par une autre base, et/ou de changer la méthode numérique par voie intrusive, de façon à prendre en compte les discontinuité mais cela ne va pas sans détruire la précision de l'approximation polynomiale. La dernière méthode présentée permet de construire une nouvelle base polynomiale avec une méthode non-intrusive, en gardant une bonne précision de la représentation. Dans la suite de cette thèse, afin d'éviter ces problème lié au discontinuités, on se placera exclusivement dans le régime fluvial.

6.5.3 Polynômes du chaos pour les simulations longues

La propagation de l'incertitude via un modèle représentant la physique, tend à déformer la densité de probabilité de l'incertitude des paramètres en entrée en une nouvelle densité de probabilité. L'incertitude tend à augmenter au fil du temps de la simulation. Plus l'incertitude est grande, et plus la décomposition en chaos polynomial sur la base des polynômes orthogonaux définis par le schémas d'Askey sur la distribution des paramètre en entrés devient coûteuse (Le degré de troncature nécessaire augmente) [65]. Afin d'éviter ce problème, [144] propose de décomposer l'espace des variables en entrée en une intersection de sous espaces, et effectuer une décomposition polynomiale qui soit valable pour chaque sous-espace. L'étude montre que ce raffinement permet de d'améliorer la qualité de la représentation polynomiale. Dans une autre étude, [47] propose de trouver une base polynomiale adaptée à la distorsion de la densité de probabilité. De la même façon que ce que présente [107], avec la seule différence que la base est calculée par rapport à la densité de probabilité du modèle et non du méta-modèle et donc le processus n'est pas itératif.

L'assimilation de données 7

7.1 Introduction

L'assimilation de données est un compromis entre un modèle et des données d'observation. Les observations représentent les données observées du système. Elles peuvent être de natures très différentes : données *in situ*, comme c'est par exemple le cas d'observations de concentrations de certains éléments chimiques dans l'air, données radars ou encore données satellitaires. Néanmoins celles-ci, de manière analogue aux modèles, peuvent être entachées d'erreurs. L'objet de l'assimilation est alors de combiner ces deux sources d'information en prenant en compte, si cela est possible, les incertitudes qui leurs sont liées pour donner la meilleure estimation possible de l'état du système (en un sens que l'on précisera ultérieurement). Cela a notamment deux applications directes. Tout d'abord la prévision, en effet donner la meilleure estimation possible de l'état du système permet d'améliorer l'état prévu du système à plus ou moins longue échéance. D'autre part cela permet la réanalyse des données passées de manière à donner la meilleure estimation de la trajectoire d'un système au cours du temps. Dans le cadre de cette thèse, on s'intéressera à la réanalyse des données.

Parmi les grandes familles de méthodes d'assimilation de données on peut citer d'une part les méthodes basées sur l'estimation stochastique comme le filtre de Kalman et les méthodes qui en dérivent : filtre de Kalman étendu, filtre de Kalman d'ensemble... D'autre part on peut citer les méthodes variationnelles dans lesquelles la meilleure estimation de l'état du système est obtenue en minimisant une certaine fonctionnelle. Ces deux types de méthodes ont néanmoins ceci de commun qu'elles peuvent être dérivées du filtrage bayésien selon que l'on cherche à déterminer l'espérance conditionnelle d'une variable aléatoire (cas des filtres de Kalman) ou son mode (assimilation variationnelle). Bien que ces familles soient différentes il existe des ponts entre elles et elles ont notamment un vocabulaire commun. Dans le cadre de cette thèse on ne s'intéressera qu'à la première famille.

7.2 Le vocabulaire de l'assimilation de données

Supposons que x (le paramètre à corriger) est une variable aléatoire. On peut décrire statistiquement cette variable par une densité de probabilité a priori $f_X(x)$, avec 'objectif de l'assimilation de données est de déterminer la valeur optimale du paramètre x qui permet

au résultat de la simulation de s'approcher de l'état réel, en se basant sur la connaissance à priori du système (que l'on nomme ébauche et notée x_b) et de l'observation $\mathbf{y^o}$. On nomme la nouvelle valeur du paramètre : l'analyse x_a.
On définit le vecteur de contrôle x comme étant l'ensemble des variables que l'on cherche à corriger avec l'assimilation de données. On pourrait par exemple chercher à corriger le coefficient de Strickler avec l'assimilation de données. Plus généralement le vecteur de contrôle peut être constitué des variables d'état, ou des paramètres du modèle, des conditions limites... De plus cette variable n'étant pas de la même dimension physique de la variable observée $\mathbf{y^o}$, on peut écrire la relation qui permet de lier $\mathbf{y^o}$ à x :

$$\mathbf{y^o} = \mathcal{G}(x) \tag{7.1}$$

où l'opérateur $\mathcal{G}$ est un opérateur qui permet de passer de l'espace de définition du vecteur de contrôle à l'espace de l'état observé. On peut maintenant considérer les différents statuts du vecteur de contrôle :

- après la correction du vecteur de contrôle avec l'assimilation d'une donnée d'observation on l'appellera analyse et on le notera x^a ;
- L'estimation *a priori* du vecteur de contrôle avant assimilation est appelé ébauche et on le notera x^b (b pour background) ;

7.3 La modélisation des erreurs et les matrices de covariance d'erreur

Comme on l'a précisé dans l'introduction l'objet de l'assimilation est de combiner les informations fournies par un modèle et celles fournies par des observations en prenant en compte les erreurs qui sont liées. Celles-ci occupent donc une place importante en assimilation de données. On donne dans cette section la définition des différents types d'erreur que nous pouvons rencontrer en assimilation de données ainsi que la définition des matrices de covariance d'erreur qui leurs sont associées et qui apparaissent dans les différentes méthodes d'assimilation.

On définit tout d'abord les erreurs comme étant la différence entre la grandeur considérée et l'état vrai. Lui-même étant inconnu et accessible de manière indirecte à travers les données d'observation, elles-mêmes entachées d'erreur.

En général on distingue trois types d'erreur liées aux modèles numériques [28] : l'erreur liée à l'état du modèle, l'erreur liée à la modélisation du système physique (discrétisation, choix des paramètres...) et l'erreur de représentativité du modèle. Mais avant de présenter les différents types d'erreur nous présentons quelques remarques qui nous seront utiles par la suite.

7.3.1 Remarques préliminaires sur les erreurs, les matrices de covariance et de corrélation d'erreur

Dans la suite de ce travail dans un souci de simplicité et sauf mention contraire on considérera que les différentes erreurs rencontrées suivent une loi normale de moyenne nulle. C'est la raison pour laquelle dans les sections suivantes lorsque nous présenterons les différents types d'erreurs nous ne nous intéresserons qu'aux moments d'ordre 2 de ces variables i-e les matrices de covariance.

Dans le cas d'un **problème univarié** où les composantes de l'erreur $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)$ sont des variables aléatoires indépendantes alors l'erreur admet comme matrice de covariance une matrice diagonale avec sur celle-ci les variances relatives à chaque composante de l'erreur. Dans le cas où les composantes de l'erreur ne sont pas des variables aléatoires indépendantes celle-ci admet comme matrice de covariance d'erreur $\mathbf{P}$:

$$\mathbf{P} = \mathbb{E}\left[\varepsilon\varepsilon^T\right] = \begin{pmatrix} \mathbb{E}\left[\varepsilon_1\varepsilon_1\right] & \ldots & \mathbb{E}\left[\varepsilon_1\varepsilon_N\right] \\ \vdots & \ddots & \vdots \\ \mathbb{E}\left[\varepsilon_N\varepsilon_1\right] & \ldots & \mathbb{E}\left[\varepsilon_N\varepsilon_N\right] \end{pmatrix}$$

Dans le cas d'un **problème multivarié** où les variables sont indépendantes alors la matrice des covariances d'erreur se réduit à une matrice diagonale par blocs où chaque bloc décrit les covariances univariées propres à chaque variable. Si les erreurs relatives à chacune des différentes variables ne sont pas indépendantes alors la matrice des covariances d'erreur n'est pas diagonale par blocs. Dans ce cas les blocs diagonaux indiquent toujours les covariances univariées tandis que les blocs extra-diagonaux représentent les covariances entre les erreurs relatives à chaque variable. Supposons par exemple que l'on a 3 variables a, b et c de vecteurs d'erreur respectifs ε^a, ε^b et ε^c, alors la matrice des covariances d'erreur s'écrit :

$$\mathbf{P} = \mathrm{E}\left[\varepsilon\varepsilon^T\right] = \begin{pmatrix} \mathbf{P}^a & \mathrm{E}\left[\varepsilon^a\varepsilon^b\right] & \mathrm{E}\left[\varepsilon^a\varepsilon^c\right] \\ \mathrm{E}\left[\varepsilon^b\varepsilon^a\right] & \mathbf{P}^b & \mathrm{E}\left[\varepsilon^b\varepsilon^c\right] \\ \mathrm{E}\left[\varepsilon^c\varepsilon^a\right] & \mathrm{E}\left[\varepsilon^c\varepsilon^b\right] & \mathbf{P}^c \end{pmatrix}.$$

7.3.2 Les erreurs liées à l'état du modèle

L'erreur liée à l'état du modèle représente la différence entre l'état estimé du modèle x et l'état vrai. On définit ainsi, selon que l'on considère l'ébauche, l'état analysé ou prévu, trois erreurs liées à l'état du modèle.

Si on considère l'erreur d'ébauche ε^b à un temps donné on notera :

$$\varepsilon^b = x^b - x^t,$$

où x^t est la valeur du vecteur de contrôle qui correspond à l'état réel. Dans la foulée on définit la matrice des covariance d'erreur d'ébauche :

$$\mathbf{B} = \mathrm{E}\left[(\varepsilon^b - \mathrm{E}[\varepsilon^b])(\varepsilon^b - \mathrm{E}[\varepsilon^b])^T\right].$$

De manière analogue on note ε^a l'erreur d'analyse à un temps donné :

$$\varepsilon^a = x^a - x^t.$$

Et on définit la matrice des covariances d'erreur d'analyse :

$$\mathbf{A} = \mathrm{E}\left[(\varepsilon^a - \mathrm{E}[\varepsilon^a])(\varepsilon^a - \mathrm{E}[\varepsilon^a])^T\right].$$

7.3.3 Les erreurs liées aux observations

Les erreurs liées aux observations sont de deux ordres. On distingue tout d'abord les erreurs liées aux instruments de mesure. Ceux-ci sont par nature imparfaits et de fait toute mesure (ou observation) est entachée d'erreur. On distingue d'autre part les erreurs de représentativité. Celles-ci sont liées à la modélisation du système physique que l'on considère (en ce sens elles sont à rapprocher des erreurs de modélisation).On définit alors l'erreur d'observation :

$$\varepsilon^o = y^o - \mathcal{G}(x^t)$$

Et on définit la matrice des covariances d'erreur d'observation :

$$\mathbf{R} = \mathrm{E}\left[(\varepsilon^o - \mathrm{E}[\varepsilon^o])(\varepsilon^o - \mathrm{E}[\varepsilon^o])^T\right]$$

7.4 Le Best Linear Unbiased Estimator - BLUE

On souhaite construire un estimateur linéaire non biaisé qui serait le meilleur estimateur au sens où il minimise la variance de l'état analysé et dans lequel celui-ci serait décrit comme la combinaison linéaire d'une estimation *a priori* de notre système, une ébauche notée x^b et d'une observation y^o [46].

On suppose qu'un tel estimateur existe et que les erreurs d'ébauche et d'observation sont non biaisées (i-e $\mathbb{E}[\varepsilon^b] = 0$ et $\mathbb{E}[\varepsilon^o] = 0$), on l'écrit :

$$x^a = \mathbf{L}x^b + \mathbf{K}y^o \tag{7.2}$$

On suppose de plus que l'opérateur d'observation $\mathcal{G}$ est linéaire et on écrit $\mathbf{G}$. D'après les définitions des erreurs que nous avons données précédemment nous pouvons écrire :

$$x^a = x^t + \varepsilon^a \tag{7.3}$$
$$x^b = x^t + \varepsilon^b \tag{7.4}$$
$$y^t = \mathbf{G}x^t \tag{7.5}$$
$$y^o = \mathbf{G}x^t + \varepsilon^o \tag{7.6}$$

Alors en remplaçant les termes de l'équation (7.2) par les expressions données en (7.3)-(7.6) on peut écrire :

$$x^t + \varepsilon^a = \mathbf{L}x^t + \mathbf{L}\varepsilon^b + \mathbf{KG}x^t + \mathbf{K}\varepsilon^o \tag{7.7}$$

Ainsi en considérant l'espérance mathématique de chacun des termes de l'équation (7.7) on peut montrer que :

$$\mathbf{L} = \mathbf{I} - \mathbf{KG} \tag{7.8}$$

On peut alors réécrire (7.2) sous la forme suivante :

$$x^a = (\mathbf{I} - \mathbf{KG})\, x^b + \mathbf{K}y^o \tag{7.9}$$

$$x^a = x^b + \mathbf{K}\left(y^o - \mathbf{G}x^b\right) \tag{7.10}$$

Cette écriture est valable quelle que soit la matrice $\mathbf{K}$ or on souhaite que cet estimateur minimise la variance de l'état analysé. Minimiser la variance de l'état analysé est équivalent à minimiser la trace de la matrice des covariances d'erreur d'analyse, i-e $\mathrm{Tr}(\mathbf{A})$. On cherche donc la matrice $\mathbf{K}$ (également appelée matrice de gain) qui minimise $\mathrm{Tr}(\mathbf{A})$.

D'après (7.9) l'erreur d'analyse peut s'exprimer ainsi :

$$\varepsilon^a = (\mathbf{I} - \mathbf{KG})\, \varepsilon^b + \mathbf{K}\varepsilon^o \tag{7.11}$$

Ecriture de laquelle on peut déduire celle de $\mathbf{A}$:

$$\mathbf{A} = (\mathbf{I} - \mathbf{KG})\, \mathbf{B}\, (\mathbf{I} - \mathbf{KG})^T + \mathbf{KRK}^T \tag{7.12}$$

puis celle de $\mathrm{Tr}(\mathbf{A})$:

$$\mathrm{Tr}(\mathbf{A}) = \mathrm{Tr}(\mathbf{B}) + \mathrm{Tr}(\mathbf{KGBG}^T\mathbf{K}^T) - 2\mathrm{Tr}(\mathbf{BG}^T\mathbf{K}^T) + \mathrm{Tr}(\mathbf{KRK}^T) \tag{7.13}$$

On s'intéresse alors à la différence $d_{\mathbf{K}}\left[\mathrm{Tr}(\mathbf{A})\right]\mathbf{M} = \mathrm{Tr}(\mathbf{A})(\mathbf{K}+\mathbf{M}) - \mathrm{Tr}(\mathbf{A})(\mathbf{K})$ où $\mathbf{M}$ est une matrice quelconque. On peut montrer que (voir [18]) :

$$d_{\mathbf{K}}\left[\mathrm{Tr}(\mathbf{A})\right]\mathbf{M} = 2\mathrm{Tr}\left(\left\{\mathbf{K}\left(\mathbf{GBG}^T + \mathbf{R}\right) - \mathbf{BG}^T\right\}\mathbf{M}^T\right) \qquad \forall \mathbf{M} \tag{7.14}$$

On déduit de (7.14) que la matrice $\mathbf{K}$ qui minimise la variance de l'état analysé est donnée par l'expression :

$$\mathbf{K} = \mathbf{BG}^T\left(\mathbf{GBG}^T + \mathbf{R}\right)^{-1} \tag{7.15}$$

En injectant (7.15) dans l'égalité (7.10) on obtient l'expression de l'état analysé pour lequel la variance est minimale :

$$x^a = x^b + \mathbf{BG}^T\left(\mathbf{GBG}^T + \mathbf{R}\right)^{-1}\left(y^o - \mathbf{G}x^b\right)$$

Dans ce cas la matrice des covariances d'erreur de l'état analysé s'écrit :

$$\mathbf{A} = (\mathbf{I} - \mathbf{KG})\,\mathbf{B}. \tag{7.16}$$

7.5 Le filtre de Kalman et ses dérivations

Le filtre de Kalman est une méthode d'interpolation statistique dans laquelle les observations sont assimilées de manière séquentielle. On présente tout d'abord le filtre de Kalman proprement dit puis on présente quelques méthodes qui en sont dérivées dont le filtre de Kalman d'Ensemble.

7.5.1 Le filtre de Kalman

Le filtre de Kalman [67, 28] est une extension sous certaines hypothèses de la méthode BLUE aux problèmes d'évolution temporelle. Ces hypothèses s'énoncent en quatre points :

- les erreurs d'observation suivent une loi gaussienne de moyenne nulle et de matrice de covariance $\mathbf{R}$. On suppose de plus qu'elles sont décorrélées en temps i-e : $\mathrm{E}\left[\varepsilon_p^o \varepsilon_q^{oT}\right] = 0$, $p \neq q$;
- les erreurs d'ébauche suivent une loi gaussienne de matrice de covariance $\mathbf{B}$;
- les erreurs sont mutuellement décorrélées.

  Tout comme dans le BLUE l'opérateur d'observation est supposé linéaire et on suppose également que le modèle dynamique aussi est linéaire.

  On peut alors décrire le filtre de Kalman. Celui-ci se décompose en deux étapes. Une première étape dite étape d'analyse dans laquelle on met à jour l'état du système. On suppose que l'on dispose d'une ébauche au temps t_k notée x_k^b pour laquelle on connait sa matrice de covariance d'erreur notée $\mathbf{B}_k$. D'après les équations du BLUE la matrice de gain est donnée par :

$$\mathbf{K}_k = \mathbf{B}_k \mathbf{G}_k^T \left(\mathbf{G}_k \mathbf{B}_k \mathbf{G}_k^T + \mathbf{R}_k\right)^{-1}$$

  Alors l'analyse au temps t_k noté x_k^a est donné par l'expression :

$$x_k^a = x_k^b + \mathbf{B}_k \mathbf{G}_k^T \left(\mathbf{G}_k \mathbf{B}_k \mathbf{G}_k^T + \mathbf{R}_k\right)^{-1} \left(y_k^o - \mathbf{G}_k x_k^b\right) \tag{7.17}$$

  où $\mathbf{G}_k$ l'opérateur d'observation au temps t_k. Suivant les équations du BLUE la matrice des covariances d'erreur d'analyse $\mathbf{A}_k$ est donnée par l'expression :

$$\mathbf{A}_k = (\mathbf{I}_k - \mathbf{K}_k \mathbf{G}_k)\,\mathbf{B}_k \tag{7.18}$$

  L'étape de propagation consiste à propager avec le modèle l'analyse du temps t_k au temps t_{k+1} pour servir d'ébauche au temps t_{k+1}, soit :

Et de manière corollaire durant l'étape de propagation on propage également la matrice des covariances d'erreur d'analyse qui servira de matrice des covariances d'erreur d'ébauche au temps t_{k+1}.

7.5.2 Le filtre de Kalman Etendu ou Extended Kalman Filter (EKF)

Le filtre de Kalman étendu consiste en l'extension du filtre de Kalman aux modèles et/ou opérateurs d'observation faiblement non-linéaires. Dans ce cas seules les équations du filtre de Kalman (7.17) et (7.18) sont modifiées. L'équation (7.17) s'écrit alors :

$$x_k^a = x_k^b + \mathbf{B}_k \mathbf{G}_k^T \left(\mathbf{G}_k \mathbf{B}_k \mathbf{G}_k^T + \mathbf{R}_k\right)^{-1} \left(y_k^o - \mathcal{G}_k x_k^b\right)$$

où $\mathcal{G}_k$ et $\mathbf{G}_k$ désignent respectivement l'opérateur d'observation et son linéaire tangent au temps t_k.

Lorsqu'au moins l'opérateur $\mathcal{H}$ n'est pas linéaire l'analyse n'est plus optimal au sens où l'EKF ne minimise pas la variance de l'analyse.

Dans le cas où la taille du vecteur de contrôle est très grande comme cela peut être le cas en météorologie ou en océanographie les approches directes de type filtre de Kalman et filtre de Kalman étendu ne sont pas applicables du fait de la grande taille des matrices à manipuler. Pour contourner ce problème et approcher le filtre de Kalman des méthodes alternatives ont été développées. On présentera notamment le filtre de Kalman d'Ensemble.

7.5.3 Le filtre de Kalman d'ensemble - EnKF

Le filtre de Kalman d'ensemble ou EnKF pour Ensemble Kalman Filter a été introduit en 1994 par Evensen, [37]. Il est basé sur l'estimation statistique de l'état du modèle et des matrices de covariances d'erreur. Celui-ci ne fait pas d'hypothèse sur la linéarité du modèle contrairement au filtre de Kalman. Néanmoins pour donner une estimation satisfaisante des matrices de covariance d'erreur l'EnKF peut nécessiter la réalisation d'un grand nombre de réalisations perturbées du modèle. C'est en effet un inconvénient reconnu de l'EnKF qu'il converge lentement vers le filtre de Kalman. De plus il ne converge vers la distribution du filtrage bayésien que dans le cas où le modèle est linéaire et la loi de probabilité de l'état considéré est gaussienne. Cependant sa relative facilité de mise en place en fait un outil très répandu dans le domaine de l'assimilation de données.

On considère N réalisations (ou membres) du modèle réalisés à l'aide d'un plan d'expériences sur le vecteur du contrôle . Si on suppose que $y_b = \mathcal{M}(x_b)$ est la sortie du modèle.On peut estimer estimer stochastiquement les matrice de covariances croisées, notamment les matrices $C_{x_b y_b}$ et C_{x_b} (Eq. (4.28)) à l'aide des définitions utilisées dans la section 4.3. En remarquant que l'opérateur $\mathcal{G} = \mathcal{S} \circ \mathcal{M}$, avec $\mathcal{S}$ est un opérateur de sélection, qui permet de passer de l'espace des sorties du modèle discrétisées à l'espace

des observations, on pourra noter alors que $y'_b = \mathcal{G}(x_b)$. L'équation (7.15) pourra, alors, s'écrire :

$$\mathbf{K_{EnKF}} = C^T_{x_b y'_b}(C^T_{y'_b} + \mathbf{R})^{-1}, \tag{7.19}$$

7.6 Vers l'EnKF basé sur la décomposition en chaos polynomial

7.6.1 Limitations de l'EnKF

La source majeure des erreurs de l'EnKF réside dans l'erreur d'échantillonnage due à l'utilisation de la méthode Monte-Carlo. L'erreur définie ici est donc purement numérique et n'est pas une erreur de modélisation (ϵ). Notons cette erreur ϵ_{EnKF}. [78, 152] prouvent que ϵ_{EnKF} dépend non seulement du nombre d'échantillons N_{EnKF} utilisé $\epsilon_{EnKF} = \mathcal{O}(N_{EnKF})$, mais aussi du nombre p_{window} et de la taille ΔT des fenêtres d'assimilation :

$$\epsilon_{EnKF} = \mathcal{O}(N^{1/2}_{EnKF}, \Delta T^{p_{window}}). \tag{7.20}$$

L'équation (7.20) montre d'une part que l'erreur d'échantillonnage dépend de taille du plan d'expérience, résultat exprimé par le théorème central limite. D'autre part, elle dépend de la taille de la fenêtre d'assimilation comme une fonction puissance du nombre de fenêtre d'assimilation. Autrement dit, plus il y a de fenêtres d'assimilation et plus l'erreur d'échantillonnage augmente. Cependant, afin d'obtenir une meilleure précision sur l'analyse il est préférable de mettre à jour le plus fréquemment la valeur du paramètre (l'analyse). Il est, par conséquent, nécessaire d'augmenter de façon significative le nombre d'échantillons N_{EnKF}. Malheureusement, cette raison rend l'application de l'EnKF dans le cadre des modèles géophysiques assez coûteuse.

7.6.2 Amélioration de l'EnKF à l'aide de la décomposition en polynômes du chaos

[79] montrent que l'utilisation du modèle réduit basé sur la décomposition en chaos polynomiale à la place du modèle direct permet de réduire significativement l'erreur d'échantillonnage de l'EnKF. La figure 7.1, qui issue de l'article [79] montre ce dernier résultat en utilisant une variante de l'EnKF, l'EnSRF (Ensemble Square Root Filter) [137]. Le même article montre que l'augmentation du nombre d'échantillons utilisé avec la décomposition polynomiale pour construire la matrice de gain Eq. (7.19) n'augmente pas le coût, et permet d'atteindre une précision élevée sur l'estimation des statistiques. On reprend les avantages de l'utilisation de la décomposition polynomiale dans l'EnKF qui sont résumés dans [152] :

— La solution de l'étape de propagation, à l'aide de la décomposition polynomiale

(soit à l'aide d'une méthode intrusive soit à l'aide de la méthode de projection), est plus efficace que la méthode Monte-Carlo utilisée dans l'EnKF standard.

— L'utilisation d'un grand nombre d'échantillons sur un modèle réduit ne rajoute pas de coût de calcul significatif et réduit l'erreur d'échantillonnage.

— Pour des système linéaire dans un cadre gaussien, l'EnKF basé sur la décomposition polynomiale est équivalent au filtre de Kalman.

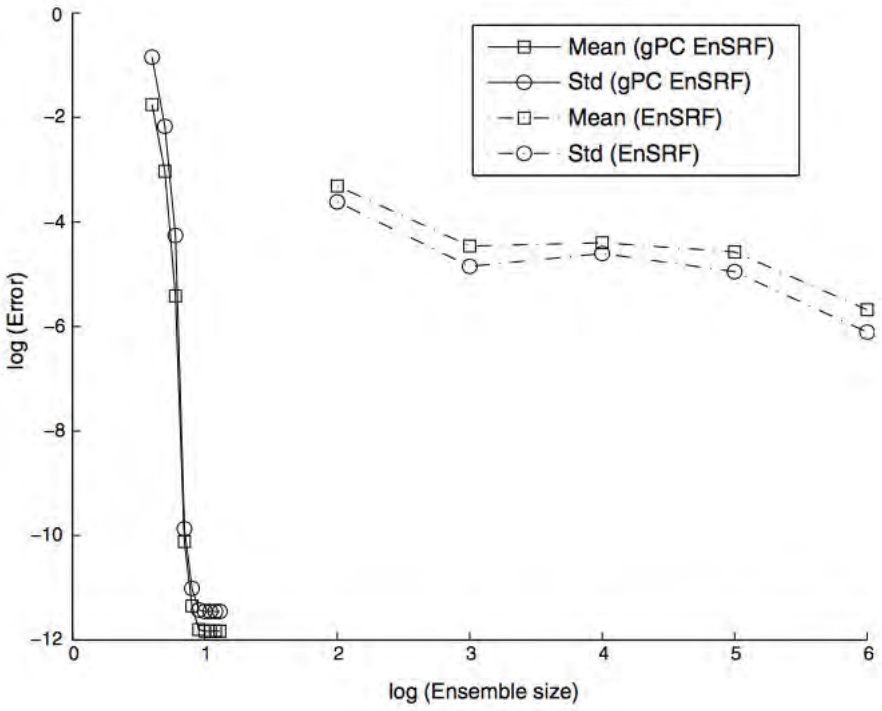


FIGURE 7.1 – *Comparaison entre l'erreur de convergence du gPC-EnSRF et de l'EnSRF standard. Source : [79]*

7.6.3 Limitations de l'approximation polynomiale dans le cadre de l'EnKF

On présente dans ce paragraphe les limitations dues à l'utilisation de la décomposition en polynômes du chaos qu'il faut bien garder à l'esprit dans le cadre de l'assimilation de données. [84] traitent des limitations du chaos polynomial dans le cadre de l'inférence Bayésienne ensembliste des problème inverses. L'assimilation de données n'étant qu'un cas particulier de l'inférence bayésienne, ces limitations restent valable pour l'EnKF.

La limitation principale de la décomposition en chaos polynomial provient de la troncature. En effet, la troncature réduit l'espace de validité de la décomposition polynomiale. Théoriquement, la décomposition polynomiale est valable pour tout l'espace incertain.

Cependant après troncature, celle-ci n'est valable que pour un sous-espace de l'espace incertain. Dans le cadre de l'assimilation de données, si l'espace incertain correspondant à l'observation n'est pas inclue dans le sous-espace modélisé par la décomposition polynomiale, on perd l'efficacité du chaos polynomial. Afin de prendre en compte un espace incertain plus large, il suffit d'augmenter l'ordre de troncature. Ce qui a pour conséquence d'augmenter le coût de l'approximation polynomiale.
Avec les séquences d'assimilation de données, l'incertitude sur l'analyse est réduite (c.f. Eq. (7.18)). La réduction de l'incertitude aura aussi pour effet de réduire l'espace incertain modélisé à l'aide de l'approximation polynomiale.
Vu ces limitations, il est important de reconstruire une décomposition polynomiale pour chaque fenêtre d'assimilation, de façon à prendre en compte l'incertitude de l'ébauche et celle de l'observation. Ces limitations ne sont pas valable seulement pour la décomposition en chaos polynomial mais aussi pour tout les modèles réduits.

IV

Quantification d'incertitudes en hydrauliques

8 Besoin de quantification d'incertitudes dans le contexte hydraulique

8.1 Introduction

En hydraulique, la simulation numérique consiste à estimer les composantes hydrauliques à l'aide d'un code de calcul basé sur les équations de Saint-Venant (ou Navier-Stokes pour le cas Telemac3D par exemple). Le résultat de cette simulation est entaché d'incertitudes, qui proviennent de plusieurs sources :

- Incertitudes dues à la modélisation : La modélisation mathématiques des équations de Saint-Venant se basent sur un nombre important de simplifications et d'hypothèses (c.f 1.2.1). Ces simplifications, nécessaires pour la compréhension de la physique, introduisent des erreurs intrinsèques à la modélisation.
- Incertitudes numériques : La simulation numérique repose sur l'approximation des équations de Saint-Venant continues en espace et dans le temps en un problème discret (c.f.2). Cette approximation induit des erreurs numériques.
- Incertitudes épistémiques : La modélisation empirique des coefficients de frottement (c.f 1.13 et la description de la planimétrie introduisent des erreurs pour la description des frottements au fond des rivières. Généralement, ces paramètres sont estimés à l'aide d'un calage. Le calage consiste à estimer la valeur de ces paramètres connaissant le résultat attendu de la simulation. On utilise souvent les données d'une crue dans le calage. Naturellement, cette estimation du paramètres n'est pas valables pour toutes les plages de variation du débit, ce qui introduit des erreurs de modélisation.
- Incertitudes stochastiques : La variabilité naturelle qui existe dans la description des conditions aux limites.

Pour toutes ces raisons, il est important d'associer une incertitude au résultat d'une simulation numérique. L'incertitude permet de mieux appréhender la gestion de la

ressource en eau [77]. D'autre part, dans le contexte de la prévision des crues, la quantification d'incertitudes permet d'associer un indice de confiance à la prévision. Dans le contexte de l'assimilation de données, il est important d'identifier les sources d'incertitudes les plus influents. Ces paramètres sont intégrés dans le vecteur de contrôle pour être corrigés via l'assimilation de données. Ce qui permet d'avoir une prévision moins incertaine.

8.2 État de l'art

Bien que la quantification des incertitudes des modèles numériques ait reçu une large attention durant les dernières décennies, la communauté en hydraulique est resté longtemps réticente par rapport à l'introduction de la quantification d'incertitude pour la modélisation hydrodynamique et hydrologique. Dans son article, [103] explique les raisons de cette réticence et montre que ces raisons n'ont pas lieu d'être. La quantification d'incertitudes est très dépendante de l'objet de la modélisation. Comme il a été évoqué dans la partie précédente, identifier les variables incertaines et leur affecter un loi de probabilité est une étape importante dans l'exercice de la quantification d'incertitudes. Cette section présente quelques études qui ont été faites dans le contexte de la quantification d'incertitudes en fonction des différentes sources d'incertitudes.

8.2.1 La condition à la limite en débit

Les conditions aux limites sont souvent sujette à l'incertitude. Pour la prévision des inondations, c'est la variable qui est la plus intuitivement pointée comme variable incertaine. Plusieurs études montrent que l'incertitude portée par la description de la condition en amont du débit explique de manière significative (surtout dans le contexte du prévision de crues) l'incertitude sur la hauteur de l'eau. Dans [110], on montre que l'amplitude du signal donnée à la condition au limite du débit a une grande influence sur l'incertitude du débit lors des écoulement qui simulent une crue. Les auteurs de [94] différencient les erreurs épistémiques des erreurs stochastiques qui ne sont pas réductibles, ces erreurs sont fondamentalement différentes. La condition à la limite, souffre souvent de ces deux natures d'erreurs. Par conséquent, elle est la plus influente dans un modèle hydraulique [19].

8.2.2 Paramètres du modèle

La bathymétrie

La bathymétrie est source majeure d'incertitudes dans un modèle hydraulique. Le plus souvent, on construit la bathymétrie à l'aide de la connaissance topographique de quelques points de rivière. La construction complète du modèle numérique qui représente la bathymétrie est faite à partir d'une interpolation des points connues. Comme mentionnée dans la partie 3.2, une incertitude sur la connaissance de la bathymétrie

peut induire un changement dans la nature de l'écoulement. Dans les modèles hydraulique tel que Mascaret, la bathymétrie est une donnée décrite de façon imparfaite et incomplète. Dans son article [64], utilise une méthode intrusive de perturbation de second ordre pour regarder les effets de l'incertitude sur la bathymétrie. Les résultats de la simulation sont écrits sous la forme d'une décomposition de Taylor et sont comparés à une méthode Monte-Carlo. Les résultats de cette étude montre que les effets de la non-linéarité du modèle par rapport à une perturbation de la bathymétrie sont totalement expliqués par les moments de second-ordre.

Coefficient de frottement

Comme présenté dans la partie 1.2.1, il existe différentes manière de modéliser le frottement dans le lit de la rivière. Les formules empiriques existantes dans la littérature font intervenir un coefficient (empirique par conséquent) nommé le coefficient de frottement. Dans la formule de Manning (utilisée dans cette thèse), il est représente le frottement physique de l'eau sur les parois de la rivière. En pratique, ce coefficient est calibré à partir de mesure in-situ. Du fait de la calibration et de cette modélisation empirique du coefficient de Manning, l'incertitude exprimée autour de sa valeur est grande, et influe beaucoup sur le résultat de la simulation hydraulique [19, 81, 10]. De part la nature de modélisation, le coefficient de frottement peut aussi palier à un manque d'information sur la bathymétrie localement. C'est l'une des contributions de la thèse [57]. Dans ce même contexte, on s'est intéressé à quantifier les incertitudes de la hauteur d'eau par rapport à une incertitude sur un coefficient de frottement défini localement à l'aide d'une méthode Monte-Carlo, cette étude a fait l'objet d'un rapport technique que l'on présente en annexe de ce Manuscrit (Annexe .1). Une étude qui explique les résultats obtenues à l'aide de l'équation de Manning Strickler sur ce cas est présentée aussi en annexe (Annexe .2).
Dans une étude un peu différente, [102] étudient la description spatiale des coefficients de frottement. Les auteurs se heurtent aux problèmes d'équifinalités, deux descriptions possibles du coefficient de frottement peuvent donner le même résultat. Cette étude montre la nécessité de bien distinguer dans les modèles hydrauliques entre les paramètres.

8.3 Outils utilisés dans la thèse

Dans cette partie, on décrit les moyens mis à disposition pour bien mener les différentes études qui seront décrit dans la partie résultats.

8.3.1 Modélisation hydraulique

Au cours de cette thèse, on utilise le modèle hydraulique Mascaret. Ce modèle est largement utilisé dans plusieurs domaines d'application en hydraulique. Les cas hydrauliques d'applications utilisés dans la thèse sont :

1. Cas de Canal : Ce cas est utilisé pour mener des études simplifiées d'hydraulique. Il permet la compréhension des méthodes que l'on souhaite utilisées sur un cas hydraulique simplifié.
2. Cas de la Garonne : Ce modèle est validé pour simuler les écoulements fluviaux entre Tonneins et La Réole. Il simule 50 km de l'aval de la Garonne. La modélisation est faite à partie de la description de 83 sections en travers de la bathymétrie. Le frottement est décrit sur 3 tronçons. Ce morceau de la Garonne est complètement décrit par une modélisation 1-D [12].

8.4 Quantification d'incertitudes

Il existent une multitudes de librairies qui sont à la disponibilité de la communauté scientifique pour traiter les études de quantification d'incertitudes. OpenTURNS [1] est une librairie scientifique *open-source* qui est dédiée au traitement des incertitudes. Cette bibliothèque est le fruit de la collaboration de EDF, Airbus, IMACS et Phimeca, qui y travaillent depuis 2005. OpenTURNS offre un large panel de méthodes utilisées dans la quantification d'incertitudes (dont la décomposition en chaos polynomial avec les différentes approches). Elle peut être utilisée comme une bibliothèque python, et bénéficie d'une communauté d'utilisateur et de développeurs (forum, formations, guide d'utilisateurs).

8.5 Couplage modélisation hydraulique et outils de quantification d'incertitudes

Afin de mener les études de quantifications d'incertitudes sur le modèles Mascaret, une plateforme qui permet de gérer les différentes tâche de travail selon la ressource informatique à disposition, est indispensable. Cette plateforme permet aussi de communiquer les différentes informations entre les routines. Typiquement, dans cette thèse entre le code de calcul et la routine python qui fait appel à la librairie OpenTURNS. Dans le cadre de cette thèse, on profite de la plateforme SALOME [2] qui permet de coupler une étude hydrodynamique avec une étude d'incertitudes à l'aide de la bibliothèque Openturns. Sauf que pour les différentes partie de la thèse ou on a eu à traiter la sensibilité de l'hydrographe, la plateforme ne permettait pas de perturber un hydrographe. On a utiliser donc la plateforme OpenPALM [3]. Dans le cadre de la collaboration entre le SCHAPI et le Cerfacs, ainsi que dans le cadre de la thèse de Sébastien Barthélémy, une maquette OpenPALM qui permet de perturber les hydrographes a été mise en place. On a profiter de cette dernière maquette pour l'utiliser lors d'une études de quantification d'incertitudes.

1. www.openturns.org
2. http://www.salome-platform.org/
3. http ://www.cerfacs.fr/globc/PALM$_W EB$/

9 Décomposition en polynômes du chaos en régime stationnaire

Introduction

Ce chapitre présente les résultats de la représentation en polynômes de chaos pour des simulations numériques en hydraulique fluvial en régime stationnaire, en utilisant le code Mascaret. Dans l'ensemble de ce chapitre, le débit et le coefficient de frottement sont considérés comme les sources d'incertitudes du modèle hydraulique. La quantité d'intérêt pour toute la suite du manuscrit est la hauteur d'eau à la sortie du modèle. Les résultats sont présentés pour le cas réel de la rivière de la Garonne, et plus précisément le tronçon entre Tonneins et la Réole. Ce chapitre se compose de deux sections. La première traite de la décomposition en chaos polynomiale pour le régime permanent. Cette étude à fait l'objet d'une ces travaux ont fait l'objet d'un article soumis au journal scientifique "*Environmental Modeling and Assessment*". Le deuxième chapitre utilise la décomposition polynomiale en régime permanent ainsi que les données SWOT pour mettre à jour la valeur des coefficients de frottement et du débit permanent.

9.1 PCE dans le cas permanent

9.1.1 Résumé

La quantification d'incertitudes accélérée par la représentation polynôme du chaos a donné lieu à la rédaction de l'article ci-dessous soumis à *Environmental Modeling and Assessment*. Comme présenté dans la partie 8, l'incertitude sur le paramètre de frottement et le débit se traduit par une incertitude sur l'estimation de la hauteur de l'eau (ou la côte d'eau) à travers les équations de Saint-Venant. La simulation numérique des équations de Saint-Venant est évaluée à l'aide du code hydraulique 1D MASCARET. Dans cette étude, on évalue la capacité de la représentation en polynômes du

chaos à estimer ces incertitudes, à coût informatique réduit. On s'intéresse notamment à représenter la distribution de la hauteur d'eau, ainsi que la corrélation spatiale. On compare notamment les résultats de la quantification d'incertitudes à l'aide du modèle réduit avec une méthode basée sur les simulations Monte-Carlo.
Dans cette étude, on a démontré de la faisabilité de la décomposition polynomiale dans un cas hydraulique fluvial. On a tout d'abord validé le modèle polynomial sur un canal rectangulaire à pente constante, et dont le coefficient de frottement est constant sur toute la longueur. Ensuite, on a mené l'étude sur un cas avec une géométrie réelle de la Garonne entre Tonneins et La Réole. La méthode de projection a démontré sa puissance pour l'estimation des coefficients polynomiaux face aux méthodes de régression, quand la dimension du problème est petite. Les méthodes LAR (c.f. 6.5.1) sont plus efficaces pour estimer les coefficients polynomiaux quand la dimension du problème est plus grande. Finalement, l'estimation de la variance, la matrice de corrélation et la loi de probabilité des hauteurs d'eau sont estimés à l'aide de la représentation polynomiale dans les deux cas traités (Canal et géométrie de la Garonne). Un polynôme de degré égal à 6 est suffisant pour estimer toutes ces grandeurs, sauf dans le cas où la géométrie locale est complexe. L'écoulement devient alors difficile à représenter avec un polynôme de degré 6, par conséquent la distribution des hauteurs d'eau en ces points n'est pas correctement représentée.

9.1.2 Article soumis dans la revue Environmental Modeling and Assessment

Manuscript
Click here to download Manuscript
Article_PCMascaret_V13.pdf

Environmental Modeling Assessment manuscript No.
(will be inserted by the editor)

Polynomial surrogate model for open-channel flows in steady state

Nabil El Moçayd · Sophie Ricci · Nicole Goutal · Mélanie C. Rochoux · Sébastien Boyaval · Cédric Goeury · Didier Lucor · Olivier Thual

Received: date / Accepted: date

Abstract Assessing epistemic uncertainties is considered as a milestone for improving numerical predictions of a dynamical system. In hydrodynamics, uncertainties in input parameters translate into uncertainties in simulated water levels through the hyperbolic shallow water equations. We investigate the ability of Polynomial Chaos surrogate to evaluate the probabilistic features of water level simulated by a 1-D hydraulic model (MASCARET) with the same accuracy as a classical Monte Carlo method but at a reduced computational cost. This study highlights that the water level probability density function and covariance matrix are better estimated with the Polynomial surrogate model than with a Monte Carlo approach on the direct model given a limited budget of direct model evaluations. The PC-surrogate performance is first assessed on an idealized channel with uniform geometry and then applied on the more realistic case of the Garonne River (France) for which a global sensitivity analysis using sparse Least-Angle Regression was performed to reduce the size of the stochastic problem. For both cases, the quadrature method

N. El Moçayd · S. Ricci · M.C. Rochoux
CECI, CERFACS – CNRS, 42 Avenue Gaspard Coriolis, 31057 Toulouse cedex 1, France
E-mail: elmocayd@cerfacs.fr

N. Goutal
Laboratory for Hydraulics Saint-Venant (LHSV), EDF R&D, 6 Quai Watier, 78401 Chatou, France

S. Boyaval
Laboratory for Hydraulics Saint-Venant (LHSV), 6 Quai Watier, 78401 Chatou, France

C. Goeury
EDF R&D, 6 Quai Watier, 78401 Chatou, France

D. Lucor
LIMSI, CNRS, Rue John Von Neumann, 91400 Orsay, France

O. Thual
CECI, CERFACS – CNRS, 42 Avenue Gaspard Coriolis, 31057 Toulouse cedex 1, France
INPT, IMFT – CNRS, Allée Camille Soula, 31400 Toulouse, France

that involves a limited number of direct model evaluations is compared with the least-square method for computing the coefficients when the surrogate is parameterized with respect to the local friction coefficient and the upstream discharge. The results showed that a PC-surrogate with total degree equal to 6 requiring 49 direct model evaluations is sufficient to represent the water level distribution (in the sense of the $\mathbb{L}_2$ norm). In locations where the flow dynamics is more complex due to bathymetry, a higher polynomial order is needed to retrieve the water level distribution. The use of a surrogate is thus a promising strategy for uncertainty quantification studies in open-channel flows and should be extended to unsteady flows. It also paves the way towards cost-effective ensemble-based data assimilation for flood forecasting and water resource management.

Keywords Uncertainty quantification · Hydraulic modeling · Surrogate model · Polynomial Chaos expansion · Sensitivity analysis · Covariance Matrix

1 Introduction

Water resource management and flood forecasting require a solid capacity of anticipating systems changes and thus rely on predictive skills of hydraulic models. These skills have greatly improved thanks to the advances in numerical modeling of free surface flow and the increase of computational resources. The shallow water equations (SWE) derived from the free surface Navier-Stokes equations are widely used for complex flow simulation. These equations have extensively proved that they are powerful tools to tackle most of open-channel hydraulics problems such as flood hazards [1] or dam breaking [41]. Nevertheless, the prediction of free surface flow characteristics requires the description of initial conditions, river geometry, boundary conditions and hydraulic physical parameters such as friction coefficients. These input data and parameters are affected by uncertainties that are either due to a lack of knowledge (*epistemic* uncertainties, for instance due to modeling assumptions or numerical approximations) or by the intrinsic randomness of the natural phenomena (*random* uncertainties, possibly induced by environmental and meteorological hazards), or both. The main idea of Uncertainty Quantification (UQ) is to quantify how errors in input parameters for the SWE translate into errors in the water level and discharge output variables, thus limiting the predictive capacity of hydraulic models and deterministic forecast abilities. Over the recent years, there was an increasing interest for UQ in hydraulics studies [7,8, 6] to identify, classify and eventually reduce the main sources of uncertainty in numerical models. Beyond these objectives, there is a need for propagating uncertainty through dynamical models based on partial derivative equations to better understand the hydraulic model structure and response.

For this purpose, the most classical methods are sample-based methods, which are non-intrusive to the forward model and derive from Monte Carlo techniques. Although Monte Carlo techniques are generic and robust for the

simulation of stochastic models, they are also computationally expensive due to their slow convergence rate as $1/\sqrt{N_{mc}}$ with N_{mc} the size of the sample. Despite the introduction of many-core hardware and parallel simulation tools, they become quickly intractable.

The cost of sample-based methods is significantly reduced when the forward model is replaced by a surrogate model that can be evaluated multiple times at almost no cost. There are various ways of formulating a deterministic or probabilistic surrogate such as Proper Orthogonal Decomposition [42], Gaussian processes or Kriging [23], Karhunen-Loève expansion [29], Polynomial Chaos expansion [22]. Polynomial Chaos (PC) approach has received much attention lately. Based on homogeneous chaos theory [35], it was first used by [29] to describe stochastic Gaussian variables in finite elements. It was finally extended to other distributions [39] using Askey's scheme and called *generalized Polynomial Chaos (gPC)* [38,19] (for simplicity purpose, the gPC-expansion approach is denoted by PC in the following). The PC-expansion method basically replaces the forward model by a polynomial expansion; it is also powerful to perform global sensitivity analysis as its coefficients relate to the statistics of the target variable [30]. The merits of PC-expansion were demonstrated in various fields, e.g. hydrology [9], structural mechanics [2], computational fluid dynamics [40,18], hydraulics [22], wildfire propagation [27,28]. Once identified and quantified, the uncertainty should be reduced, for instance using a data assimilation (DA) approach.

The Ensemble Kalman Filter (EnKF) [13] is a common ensemble-based DA algorithm that provides a correction to the model state and/or parameters; the correction can be regarded as a weight between observations and model predictions calculated via the Kalman gain matrix [25]. The Kalman gain matrix is computed from a stochastic estimate of the background error covariance matrix over an ensemble of input and output variables. A sample of the input uncertain variables is used to achieve an ensemble of forward model evaluations that provide a sample of the output variables. The propagation of the uncertainty is classically achieved with the forward model or, alternatively with a surrogate model. [20] showed that with a crude Monte Carlo technique for the ensemble generation, the convergence of the EnKF is slow and thus often not compatible with operational constraints or high-dimensional problems. A complementary approach between UQ and ensemble-based DA was presented in [21], where a PC-surrogate was used in place of the forward model to reduce the cost of the EnKF. Such method was implemented in [28] for regional-scale wildfire spread modeling and referred to as PC-EnKF; in the context of grassland controlled burns, the PC-EnKF algorithm replaced the wildfire spread model by a PC surrogate, whose coefficients were computed via a quadrature method, and achieved convergence in spite of the model nonlinearities with respect to the controlled input parameters (i.e. near-surface winds, biomass moisture content, biomass fuel properties).

The objective of the present work is to validate the PC-expansion strategy for hydraulic modeling of subcritical steady flows. We aim at demonstrating that the water level probability density function and covariance matrix are

better estimated with the PC-surrogate model than with a Monte Carlo approach on the direct model given a limited budget of direct model evaluations. The PC-surrogate is here expanded with respect to the upstream discharge and the spatially distributed friction coefficients. The friction coefficients are supposed to be independent random variables that are adjusted to account for various modeling errors. The formulation of the PC-surrogate model relies first on the choice of the polynomial basis for the input random variables, then on the identification of the maximum order of the polynomial decomposition that is sufficient to mimic the forward model behavior and that decides on the number of coefficients in the expansion, and finally on the computation of these coefficients by regression or spectral projection. The performance of the least-square (regression) and quadrature (spectral projection) methods is compared in this study. The merits of the quadrature approach are shown on two test cases: an idealized channel test case with uniform geometry and a real case on a 50-km reach of the Garonne River with complex bathymetry. The surrogate model for water level is validated in the sense of the $\mathbb{L}_2$ norm in the model output space and in the polynomial coefficients space with the appropriate probabilistic measure. The surrogate model is formulated at different locations of the Garonne hydraulic network for uncertainty quantification and sensitivity analysis studies with a particular interest for the Marmande observing station. The water level statistical moments, the probability density function (PDF) and the covariance matrix are estimated with the PC-surrogate at a lower computational cost than with a classical Monte-Carlo approach. In the long run, this work paves the way towards reduced-cost DA in hydraulics for flood forecasting [16] or water resource management provided the Kalman gain is stochastically estimated with the surogate model in place of the forward model. For model parameter estimation with DA, the Kalman gain matrix represents the statistics between the errors in the parameters and the errors in the simulated hydraulic state at the grid points where the observations are assimilated. In this case, the surrogate model is only necessary at the observing stations. For model state estimation, the Kalman gain matrix represents the statistics between the errors in the hydraulic state over the entire simulation domain. Thus, the surrogate model is necessary for every point of the domain.

The structure of the paper is as follows: Sect. 2 introduces the basic description of the hydraulic modeling. We recall the SWE for steady flows and present both idealized and Garonne River test cases chosen for this study. The theory and implementation of the PC expansion is presented in Sect. 3 along with the a posteriori error metrics for evaluating the surrogate performance and its statistics. In Sect. 4, results for the PC-expansion strategy are presented. Conclusions and perspectives are given in Sect. 5.

2 Hydraulic modeling

2.1 Shallow water equations (SWE)

Free surface flow dynamics in rivers are driven by the Navier-Stokes Equations (NSE). Nevertheless, the Shallow Water Equations (SWE) can be derived when the vertical dimension is supposed to be significantly smaller than the horizontal dimensions and when the bathymetry variations are supposed to be small. The SWE result from the vertical integration of the NSE under the hydrostatic pressure assumption. One-dimensional models are widely used in hydraulic engineering, for instance for flood forecasting [17]. The SWE form a hyperbolic system of partial differential equations, which represents subcritical and supercritical flows with hydraulic jumps. In the present work, only subcritical flows of rivers in plains are considered, consistently with the hydraulic model used at operational level for the Garonne River (see the application in Sect. 4.2).

2.1.1 One-dimensional flow

One-dimensional SWE are derived from two prime principles: mass conservation and momentum conservation. For most applications in rivers, the equations are written in terms of discharge (or flow rate) Q ($m^3\,s^{-1}$) and hydraulic section A (m^2) that relates to water level (or water height) h (m) such that $A \equiv A(h)$. The curvilinear abscissa in the simulation domain is denoted by s ranging from s_{in} upstream of the river to s_{out} downstream. Thus, the non-conservative form of the one-dimensional SWE reads [1,14,31]:

$$\begin{cases} \partial_t A(h) + \partial_s Q = 0 \\ \partial_t Q + \partial_s (Q(h)^2/A(h)) + gA(h)\partial_s h - gA(h)(S_0 - S_f) = 0 \end{cases} \tag{1}$$

with g the gravity, S_0 the channel slope and S_f the friction slope. These equations are usually combined with an equation for the friction slope S_f, here described with the Manning-Strickler formula:

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

where $R(h)$ is the hydraulic radius ($R(h) = A(h)/P(h)$) written as a function of the wet perimeter $P(h)$ and K_s is the Strickler friction coefficient ($m^{1/3}\,s^{-1}$).

In the present work, the MASCARET software is used to simulate the one-dimensional SWE, i.e. the water level h and discharge Q over the discrete hydraulic network for $s \in [s_{in}, s_{out}]$ (the pair (h, Q) is referred to as the hydraulic state in the following). MASCARET is part of the TELEMAC-MASCARET open-source modeling package developed by EDF (Electricité de France R&D) in collaboration with the CEREMA (Centre d'Etude et d'expertise sur les Risques, l'environnement, la Mobilité et l'Aménagement), and is commonly

used for dam-break wave simulation, reservoir flushing and flooding. The SWE are solved here with the steady kernel of MASCARET based on a finite difference scheme developed by [15]. The hydraulic model requires the following input data: bathymetry, upstream and downstream boundary conditions, lateral inflows, roughness coefficients and initial condition for the hydraulic state. The imperfect description of these data translates into errors in the simulated water level and discharge. It is thus of prime importance to understand the structure of these errors through uncertainty propagation as well as the contribution of each input variable on the hydraulic state variability through sensitivity analysis. This is one of the objectives of the present work.

2.1.2 Steady flow

For steady flow, Eq. 1 simplifies to:

$$\begin{cases} \partial_s Q = 0 \\ \\ \partial_s h = \dfrac{(S_0 - S_f)}{1 - Fr^2} \end{cases} \tag{3}$$

where Fr is the dimensionless Froude number

$$Fr^2 = \frac{Q}{gA^3}\frac{\partial A}{\partial h}. \tag{4}$$

The smooth solutions for Eq. 3 over $s \in [s_{in}, s_{out}]$ are called *backwater curves* when the downstream boundary condition is prescribed in a deterministic way. Assuming the channel geometry is rectangular ($A = h\,W$ with W the channel width) and the slope S_0 is constant implies for steady flow that h no longer depends on s. The balance between gravity and friction forces, known as the Manning equation, is then recast as the following equation for the water level h:

$$h = \left(\frac{Q}{K_s\, W\, \sqrt{S_0}}\right)^{-5/3}. \tag{5}$$

If the boundary conditions from the hydraulic domain are specified in coherence with Eq. 5, the entire flow is governed by the Manning equation.

2.2 Case study description

In the present work, two different case studies are considered: an idealized channel with uniform geometry on the one hand, and a reach of the Garonne River (South-West France) on the other hand.

2.2.1 Idealized channel with uniform geometry

The first case study is a 40-km open channel with uniform friction coefficient K_s and uniform slope $S_0 = 2.5 \times 10^{-4}\ m\,km^{-1}$. Upstream and downstream steady boundary conditions are defined by $Q(s_{in}) = Q$ and $h(s_{out}) = h_{out}$. The Manning equilibrium is preserved when h_{out} is computed with Eq. 5 for each pair (Q, K_s). For gradually-varying flow, h_{out} is set to $h_{out} = 10\ m$. In the following, Q and K_s are supposed to be independent physical random variables respectively described by a Gaussian PDF $\mathcal{N}(4031{,}400)$ and a uniform PDF $\mathcal{U}_{[15,60]}$. It should be noted that the mean values for Q and K_s correspond to the Manning equilibrium associated to $h_{out} = 10\ m$. The PDF for the Froude number was estimated over the sampling, it has a Gaussian-like shape and the probability for Fr to exceed 1 is about 10^{-6}. Thus it is legitimate to assume that all members in the sampling represent a subcritical flow. The water level thus describes a backwater curve of type M_1 or M_2 [31,34].

Figure 1 displays the water level h along the 40-km channel for different values of Q and K_s. The bathymetry (with constant slope S_0) is represented with a solid black line. The simulated water levels for $K_s = 15\ m^{1/3}\,s^{-1}$ and $K_s = 60\ m^{1/3}\,s^{-1}$ are represented in blue and cyan, respectively. Solid lines are associated with $Q = 4000\ m^3\,s^{-1}$; dashed lines are associated with $Q = 2000\ m^3\,s^{-1}$. In addition, the simulated water level for $Q = 8000\ m^3\,s^{-1}$ and $Ks = 15\ m^{1/3}\,s^{-1}$ is represented in blue dashed-dotted line. The flow is found to be uniform over the upstream part of the channel; the water level follows then a backwater curve to meet the downstream boundary condition. As expected, both friction and discharge have a significant impact on water level; the water level increases when the friction increases (i.e. when K_s decreases) and also when the discharge increases. When the water level prescribed by the downstream water level condition ($h_{out} = 10\ m$) is larger than the normal depth given by the Manning equilibrium (Eq. 5), the free surface describes a backwater curve of type M_1 (cyan/blue dashed and solid lines). On the opposite, when the water level is smaller than the normal depth, the water level describes a backwater curve of type M_2 (blue dashed-dotted line).

2.2.2 Garonne River test case

The second case study is a real hydraulic network over the Garonne River in France. The Garonne River flows from the Pyrenees to the Atlantic Ocean in the area of Bordeaux. It is approximately 647-km long and drains an area of $55,000\ km^2$. The present study focuses on a 50-km reach from Tonneins (curvilinear abscissa $s = 13\ km$) to La Réole (curvilinear abscissa $s = 62\ km$) with an observing station at Marmande as illustrated in Fig. 2a. The mean slope over the reach is $S_0 = 3.3\ m\,km^{-1}$ and the mean width of the river is $W = 250\ m$. The bank-full discharge is approximately equal to the mean annual discharge ($\overline{Q} = 1000\ m^3\,s^{-1}$). Despite the existence of active floodplains, this reach can be modeled accurately by a 1-D hydraulic model.

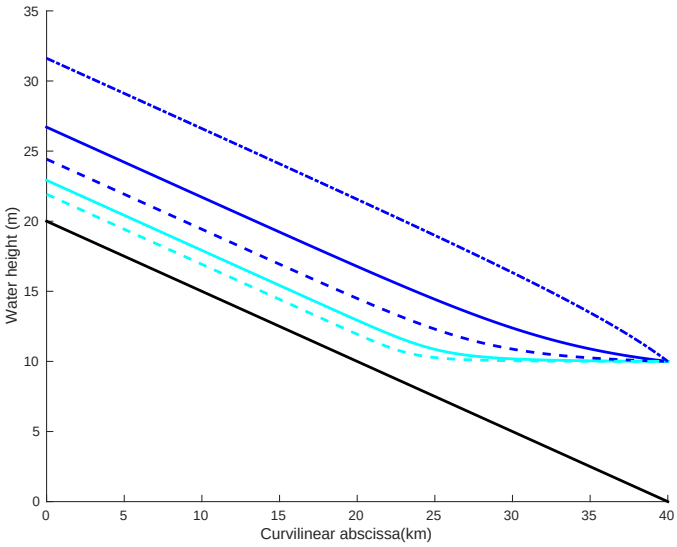


Fig. 1 Water level (or water height) along the 40-km idealized channel with uniform geometry and the downstream boundary condition $h_{out} = 10\ m$. The bathymetry is represented by the solid black line. Simulated water levels for $K_s = 15\ m^{1/3}\,s^{-1}$ (blue lines) and $K_s = 60\ m^{1/3}\,s^{-1}$ (cyan lines), $Q = 4,000\ m^3\,s^{-1}$ (solid lines) and $Q = 2,000\ m^3\,s^{-1}$ (dashed lines) represent M_1 backwater curves. Their counterpart for $K_s = 15\ m^{1/3}\,s^{-1}$ and $Q = 8000\ m^3\,s^{-1}$ (blue dashed-dotted line) represents a M_2 backwater curve.

The hydraulic model for the the Garonne River is built from 83 on-site bathymetry cross sections from which the full 1D bathymetry is interpolated. Friction is prescribed over 3 portions separated by dashed lines in Fig. 2b for the river channel and the floodplain; it is represented by the Strickler coefficients K_{s1}, K_{s2} and K_{s3}. It should be noted that Marmande is located at the beginning of the K_{s3} portion ($s = 36\ km$). The upstream boundary condition is prescribed with a steady flow $Q(s_{in}) = Q$; the downstream boundary condition is prescribed with a local rating curve RC established at La Réole that sets $h(s_{out}) = RC(Q_{out}) = h_{out}$. The hydraulic model has been calibrated using channel and floodplain roughness coefficients as free parameters [3]. The main sources of uncertainty considered in the present work are (Q, K_{s1}, K_{s2}, K_{s3}); in the following, they are considered as physical random independent variables defined by their PDFs $\mathcal{N}(4031,400)$ for discharge and $\mathcal{U}_{[15,60]}$ for the three friction coefficients.

3 Polynomial chaos surrogate

The formulation of the Polynomial Chaos (PC) surrogate [29,19,38] is presented for n random variables included in the random vector $\mathbf{x}$. The Strickler coefficient K_s and the constant river discharge Q (permanent flow) are known

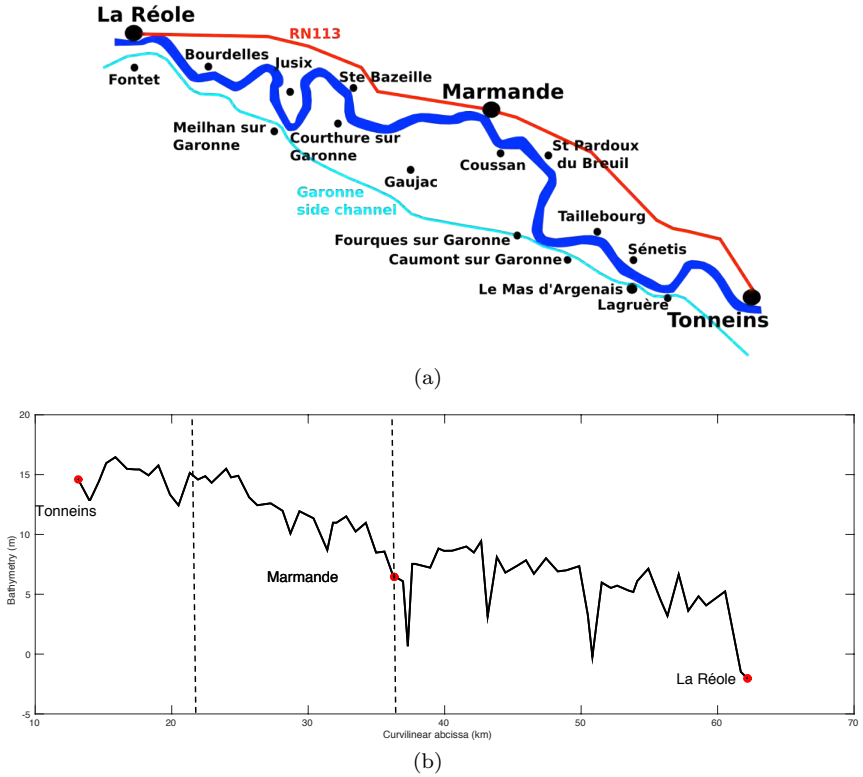


Fig. 2 Garonne River test case. a- Garonne reach between Tonneins (upstream) and La Réole (downstream). Marmande is located at the curvilinear abscissa $s = 36\ km$. b- Bathymetry profile along the curvilinear abscissa (km) between Tonneins and La Réole. The three portions of K_s are separated by vertical dashed lines.

to be the most relevant sources of uncertainty in water level predictions and are thus included in the random vector $\mathbf{x}$. If the Strickler coefficient is uniform over space, the stochastic dimension is $n = 2$. For the Garonne hydraulic model, there are three different values for K_s, implying that for that specific case the stochastic dimension is $n = 4$. For simplicity, we will denote the n-variate random vector $\mathbf{x}$ with $n \in \{2, 4\}$. In the following, details are given on the methods used to compute the PC coefficients (truncation strategy, projection strategy, choice of polynomial basis), derive the statistics of the water level h and evaluate the accuracy of the PC surrogate.

3.1 Polynomial chaos expansion for hydraulic modeling

The input random variables are supposed to be independent with finite variance and defined in a proper probabilistic space $(\Omega, \mathcal{A}, \mathcal{P})$ where Ω is the event space, $\mathcal{A}$ is the σ-algebra and $\mathcal{P}$ is the probability measure. The random

variables are cast in the random vector $\mathbf{x} \equiv \mathbf{x}(\omega) = [x_1, ..., x_n]^T$ with $\omega \in \Omega$ a random event, which evolves in a subset of the Euclidean space $\Gamma \subseteq \mathbb{R}^n$. The marginal PDF associated with each random variable x_i for $i = 1, \cdots, n$ is denoted by $\rho_i : \Gamma_i \longrightarrow \mathbb{R}$. Since random variables are assumed independent, the joint probability density of the random vector $\mathbf{x}$ can be written as follows:

$$\rho(\mathbf{x}) = \prod_{i=1}^{n} \rho_i(x_i). \tag{6}$$

In the following, the random variables in $\mathbf{x}$ defined in the input physical space are rescaled in the standard probabilistic space and noted $\boldsymbol{\zeta} \equiv \boldsymbol{\zeta}(\omega) = [\zeta_1, \ldots, \zeta_n]^T$ (0 mean and unit variance), to which the PC framework applies. In the following $\mathbf{x} = [Q, K_s]^T$ for $n = 2$ and $\mathbf{x} = [Q, K_{s1}, K_{s2}, K_{s3}]^T$ for $n = 4$. The physical model $\mathcal{M}$ (MASCARET in the present case) maps $\mathbf{x}$ onto the output space to formulate the spatially distributed water level h along the channel. This spatialized random vector is assumed to be of finite variance. Thus, we know from PC theory that there exists a convergent polynomial expansion, implying that h can be projected onto a stochastic space spanned by orthogonal polynomial functions.

The PC expansion is achieved for water level h at a given curvilinear abscissa x_s, thus h is a scalar. h is represented in terms of the same probabilistic space as the input random vector $\mathbf{x}$, implying the following formulation for h:

$$h = \mathcal{M}\left(\mathbf{x}(\boldsymbol{\zeta})\right) = \sum_{i=0}^{\infty} \hat{h}_i \, \Psi_i\left(\boldsymbol{\zeta}\right), \tag{7}$$

here $(\Psi_i)_{i=0,\cdots,\infty}$ designate the multi-variate polynomial functions that are defined as orthogonal with respect to the joint density $\rho(\mathbf{x})$ (Eq. 6), where i is the multi-index identifying the polynomial components in the multi-variate space, and where $\hat{h}_i$ are the corresponding deterministic coefficients (or modes).

The coefficients $\hat{h}_i$ do not change in space for the uniform steady case on which the Manning equation applies. However, they become spatially varying along the river network when considering the idealized channel case with backwater curves and when considering the Garonne River case that features a non-uniform bathymetry. When the water level varies over space, the PC expansion is achieved at each grid point of curvilinear abscissa s using the same set of forward model evaluations, thus providing spatially dependent expansion coefficients $\hat{h}_i(s)$ and water level representation:

$$h(s) = \sum_{i=0}^{\infty} \hat{h}_i(s) \, \Psi_i\left(\boldsymbol{\zeta}\right). \tag{8}$$

3.2 Truncation strategy

In practice, the sum in Eq. 7 needs to be truncated to a finite number of terms (noted N_{pc}) as follows:

$$h \cong \mathcal{M}_{pc}(\mathbf{x}(\boldsymbol{\zeta})) = \sum_{i=0}^{(N_{pc}-1)} \hat{h}_i \, \Psi_i \left(\boldsymbol{\zeta}\right), \tag{9}$$

where the coefficients $\hat{h}_i$ are the unknowns to be determined to build the PC surrogate (or metamodel) $\mathcal{M}_{pc}$ associated with a given maximum polynomial order P (also referred to as the total degree).

There are several ways of constructing the polynomial approximation space. The most common choice is to constrain the number of terms N_{pc} in the PC expansion by the number of random variables n and by the total degree P. Considering hierarchical polynomials, the number of terms N_{pc} to reach the user-defined polynomial order P is given by the following algebraic formula:

$$N_{pc} = \frac{(n+P)!}{(n!\ P!)} \tag{10}$$

meaning that all polynomials involving the n random variables of total degree less or equal to P are considered in the PC expansion.

Following this standard truncation strategy, building the approximation space may be too costly for high dimensional problems. New truncation strategies are reported in the literature (beyond the scope of this study) to reduce the number of terms to evaluate to build the PC expansion (thereby to decrease the computational cost of the PC method) as well as conditioning issues when n increases. These strategies explore alternative ways of selecting the N_{pc} polynomials by limiting high-order interaction terms (e.g. hyperbolic truncation scheme) or through the construction of sparse bases [24,4]. In any case, analyzing how the truncation error evolves with respect to P and for which value of P the convergence of the PC expansion is achieved is required. For this purpose, a sensitivity analysis of the PC expansion accuracy is carried out with respect to the total degree P in the following, for both the idealized channel case and the Garonne River application, see Sect. 4. The metrics used to evaluate the truncation error are presented in Sect. 3.7.

3.3 Choice of polynomial basis

The knowledge of the coefficients $\hat{h}_i$ fully characterizes the process h. Still, to obtain a relevant and accurate PC expansion, the projection should be performed on a suitable polynomial basis forming an orthogonal basis with respect to the joint PDF $\rho(\boldsymbol{\zeta})$, i.e. with regard to the following inner product $< \cdot, \cdot >$:

$$< \Psi_i(\boldsymbol{\zeta}), \Psi_j(\boldsymbol{\zeta}) >= \int_{\Gamma} \Psi_i(\boldsymbol{\zeta}) \, \Psi_j(\boldsymbol{\zeta}) \, \rho(\boldsymbol{\zeta}) \, \mathrm{d}\boldsymbol{\zeta} = \delta_{ij} \parallel \Psi_i \parallel^2, \tag{11}$$

with δ_{ij} the Kronecker delta-function extended to the n-variate case and with $\| \Psi_i \| = 1$ if the basis is constructed orthonormal. The choice for the basis functions is thus an essential step in the PC method and depends on the probability measure of the random variables. According to Askey's scheme, the Hermite polynomials form the optimal basis for random variables following Gaussian distribution, and the Legendre polynomials are the counterpart for uniform distribution [39].

In practice, the orthonormal basis is built using the tensor product of one-dimensional polynomial functions such that the ith two-dimensional basis function Ψ_i can be expressed as follows:

$$\Psi_i \equiv \Psi_i(\boldsymbol{\zeta}) = \otimes_{l=1}^{n} \Phi_{i_l}^{(l)} \zeta_l, \tag{12}$$

with $\Phi_{i_l}^{(l)}$ the one-dimensional polynomial basis associated with the lth random variable and i_l the corresponding one-dimensional index varying between 0 and P.

There is an intrinsic correspondence between the multi-index i and the one-dimensional index i_l to properly identify the polynomials involving the lth random variable. For instance, if $n = 2$ with $\mathbf{x} = [Q, K_s]^T$, Eq. 12 is recast as the following form:

$$\Psi_i \equiv \Psi_i(\boldsymbol{\zeta}) = \Phi_{i_1}^{(1)}(\zeta_1) \otimes \Phi_{i_2}^{(2)}(\zeta_2), \tag{13}$$

with $\Phi_{i_1}^{(1)} = P_{i_1}$ corresponding to the one-dimensional basis functions for K_s (i.e. Legendre polynomials) and $\Phi_{i_2}^{(2)} = H_{i_2}$ corresponding to the one-dimensional basis functions for Q (i.e. Hermite polynomials). The multi-index i is defined with the one-dimensional indices i_1 and i_2 specific to each one-dimensional basis function. There are several ways of defining the indexation reported in the literature [4,5]. In the following, for $n = 2$, the following notation will be preferred:

$$h \cong \mathcal{M}_{pc}(\mathbf{x}(\boldsymbol{\zeta})) = \sum_{i_1, i_2 \geq 0} \hat{h}_{i_1 i_2} P_{i_1}(\zeta_1) \otimes H_{i_2}(\zeta_2). \tag{14}$$

3.4 Computation of Polynomial Chaos modes

For a given polynomial basis (Eq. 9), there are several projection strategies to numerically compute the coefficients $\hat{h}_i$ $(i = 0, \cdots, N_{pc} - 1)$. The present work focuses on non-intrusive approaches, implying that no modification of the numerical model was required and that the coefficients were computed from ensemble of model evaluations $h^{(k)} = \mathcal{M}(\mathbf{x}^{(k)})$, with the index $k = 1, \cdots, N_e$ referring to the kth realization of the random vector $\mathbf{x}$. In the literature, two main families of non-intrusive approaches are reported: regression on the one hand, and spectral projection on the other hand. We present briefly below the ordinary least-square approximation that folds into the regression category and the Gaussian quadrature rule that folds into the spectral projection category.

The pseudo-spectral variant in the form of the Least-Angle Regression (LAR) approach is also introduced with the objective to select the most sensitive stochastic features for input parameters on the Garonne River application for which $n = 4$ and thus to reduce the dimension of the stochastic problem.

Least-square approximation strategy A first non-intrusive approach to estimate the coefficients $\hat{h}_i$ consists in setting up a least-square minimization problem that minimizes the truncation error (or residual) ϵ_{pc} in the $\mathbb{L}_2$ norm. ϵ_{pc} is defined as the distance between the (exact) model evaluations and the PC surrogate estimations for a finite set of randomly sampled input random variables of size $N_e = N_{ls}$ (where the subscript ls stands for least-square) such that:

$$\epsilon_{pc} = \mathcal{M}(\mathcal{X}) - \mathcal{M}_{pc}(\mathcal{X}) = \mathcal{H} - \sum_{i=0}^{(N_{pc}-1)} \hat{h}_i \Psi_i(\boldsymbol{\zeta}) \equiv \mathcal{H} - \hat{\mathbf{h}}^T \boldsymbol{\Psi}, \tag{15}$$

where $\mathcal{X} = [\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N_{ls})}]^T$ is the matrix of input samples of the random vector $\mathbf{x}$ and where $\mathcal{H} = [h^{(1)}, \cdots, h^{(N_{ls})}]^T$ is the vector of associated water level outputs. $\hat{\mathbf{h}} = [\hat{h}_0, \cdots, \hat{h}_{N_{pc}-1}]^T$ is the vector of the N_{pc} unknown coefficients and $\boldsymbol{\Psi}$ is the matrix of size $N_{ls} \times N_{pc}$ assembling the values of all the orthonormal polynomials at the experimental design nodes $\boldsymbol{\Psi}_{ki} = \Psi_i(\boldsymbol{\zeta}^{(k)})$, with $i = 0, \cdots, (N_{pc} - 1)$ and $k = 1, \cdots, N_{ls}$.

Determining $\hat{\mathbf{h}}$ as the ordinary least-square solution of the minimization problem formulated in Eq. 15 is a classical problem in linear algebra. The ordinary least-square solution is obtained by minimizing $\mathrm{LS}(\hat{\mathbf{h}}) = (\epsilon_{pc}^T \epsilon_{pc})$ such that:

$$\frac{\partial \mathrm{LS}}{\partial \hat{\mathbf{h}}} = -2\,\boldsymbol{\Psi}^T \mathcal{H} + 2\,\boldsymbol{\Psi}^T \boldsymbol{\Psi}\,\hat{\mathbf{h}} = 0, \tag{16}$$

leading to the following solution:

$$\hat{\mathbf{h}} = (\boldsymbol{\Psi}^T\,\boldsymbol{\Psi})^{-1}\,\boldsymbol{\Psi}^T\,\mathcal{H}. \tag{17}$$

In order to compute the PC coefficients, one should ensure that the matrix $(\boldsymbol{\Psi}^T \boldsymbol{\Psi})$ is invertible; the solution depends on the pre-conditioning of the matrix $\boldsymbol{\Psi}$ and thereby on the number and the choice of the realizations $\mathbf{x}^{(k)}$ ($k = 1, \cdots, N_{ls}$) of the random vector $\mathbf{x}$. The input space sampling is achieved in the present work with a classical Monte Carlo approach; the converged estimation of the PC coefficients requires a large size sample. When the direct model is too expensive, more efficient sampling methods – e.g. Quasi Monte Carlo, Latin Hypercube Sampling – are used. An iterative Monte Carlo approach is applied here for efficiency purpose: the coefficients are re-computed as the size of the sample E_{ls} increases from 10 to N_{ls}.

The least-square algorithm for an evaluation sample E_{ls} of dimension N_{ls} proceeds in the present study as follows:

- Choose the maximum polynomial order P
- Initialize the sample: $E_{ls} = \varnothing$

- For $l = 10, N_{ls}$ with increment $= 10$, do the following steps:
  - Add 10 members in the evaluation ensemble E_{ls} corresponding to realizations of the random vector $\mathbf{x}$ and formulate the input vector $\mathcal{X}$
  - Integrate the direct model for each new member in E_{ls} and formulate the output vector $\mathcal{H}$
  - Formulate the information matrix $\mathbf{\Psi}$ by evaluating the polynomial basis at the normalized nodes corresponding to the members $\mathcal{X}$ in the physical space
  - Compute the coefficient vector $\hat{\mathbf{h}}^l$ of length N_{pc} using Eq. 17
- Average the different sets of coefficients $\hat{\mathbf{h}}^l$ (l varying between 10 and N_{ls} in increments of 10) and formulate $\mathcal{M}_{pc}$

Least-angle regression When the stochastic dimension n of the problem is high, the number of model simulations N_e required to build the PC expansion becomes prohibitive when using a standard truncation scheme. In fact, it is very likely that the effective dimensionality of the problem is much lower than the size of the approximation PC space. The LASSO algorithm [32] is in this case an attractive modification of the ordinary least squares formulation that constrains the sum of the absolute regression coefficients. Weighted versions exist that make the approximation even more robust [33]. Another related model selection algorithm, the LAR is also very efficient in our framework. Here, the idea underlying the use of LAR is to build in an iterative manner, an optimal sparse basis and to compute a limited number of coefficients using a standard regression method. For this purpose, an active set of coefficients should be selected.

The N_{pc} polynomial functions in the standard PC expansion (i.e. when the basis is full) form the candidate set. At each iteration of the LAR algorithm, a polynomial is selected among these N_{pc} terms, based on its correlation with the current residual ϵ_{pc}, and is then added to the active set of polynomials (also referred to as regressors). The corresponding coefficients are computed so that every active polynomial is equicorrelated with the current residual. Thus, LAR builds a collection of surrogates that are less and less sparse along the iterations. Iterations stop either when the full basis has been selected, or when the maximum size of the experimental design E_{lar} has been reached. The final step is to choose the best surrogate according to an error criterion (see Sect. 3.7). More details about LAR can be found in [12, 5].

The LAR algorithm proceeds as follows:

- Choose the maximum polynomial order P
- Build the experimental design E_{lar}
  - Build the LAR input vector $\mathcal{X}$ of size $N_{lar} \approx 10^2 - 10^3$
  - Integrate the direct model for each member of $\mathcal{X}$ and form the output vector $\mathcal{H}$
  - Evaluate all components of the polynomial basis over the sample $\mathcal{X}$ to form the matrix of elements $\mathbf{\Psi}_{ki} = \Psi_i(\boldsymbol{\zeta}^{(k)})$, with $i = 0, \cdots, (N_{pc} - 1)$ and $k = 1, \cdots, N_{lar}$.

- Initialize the algorithm
  - Set the coefficient vector $\hat{\mathbf{h}}$ to 0
  - Set the residual vector ϵ_{pc} (of size N_{lar}) to the forward model output vector $\mathcal{H}$ ($\epsilon_{pc} = \mathcal{H} - \mathcal{M}_{pc}(\mathcal{X}) = \mathcal{H}$ since $\hat{\mathbf{h}} = 0$)
  - Set the active set (or regressors) $\mathbf{\Psi}_{lar} = \varnothing$
- Iterate: For $l = 1, \min(N_{pc}, N_{lar} - 1)$, do the following steps:
  - For all polynomial candidates $i = 0, (N_{pc} - 1)$, compute the correlation between the current residual ϵ_{pc} and the ith column of $\mathbf{\Psi}$ ($\mathbf{\Psi}_{ki}$ for $k = 1, N_{lar}$) evaluated on E_{lar}
  - Add to the active set $\mathbf{\Psi}_{lar}$, the vector Ψ_j (the jth column of $\mathbf{\Psi}$) that is the most correlated with the current residual ϵ_{pc}
  - Compute the vector of partial coefficients $\hat{\mathbf{h}}^l$ (or descent direction) associated to $\mathbf{\Psi}_{lar}$ by minimizing $\mathrm{LS}(\hat{\mathbf{h}}) = ((\mathcal{H} - (\hat{\mathbf{h}})^T \mathbf{\Psi}_{lar})^T (\mathcal{H} - (\hat{\mathbf{h}})^T \mathbf{\Psi}_{lar}))$. This consist in taking the largest possible step in the direction of $\mathbf{\Psi}_{lar}$ until the correlation between ϵ_{pc} and $\mathbf{\Psi}_{lar}$ becomes smaller than the correlation between some other vector Ψ_k in the candidate set and $\mathbf{\Psi}_{lar}$. This provides the vector of partial coefficients

$$\hat{\mathbf{h}}^l = \left(\mathbf{\Psi}_{lar}^T \mathbf{\Psi}_{lar}\right)^{-1} \mathbf{\Psi}_{lar}^T \epsilon_{pc}$$

  - Update the residual $\epsilon_{pc} = \epsilon_{pc} - (\hat{\mathbf{h}}^l)^T \mathbf{\Psi}_{lar}$ and store it for the current iteration
  - Update the coefficient vector $\hat{\mathbf{h}}$ with the contribution $\hat{\mathbf{h}}^l$: $\hat{\mathbf{h}} = \hat{\mathbf{h}} + \gamma^l \hat{\mathbf{h}}^l$ with γ^l the descent step defined such that every active polynomial is equicorrelated with the current residual [12,5]
  - Formulate the surrogate associated with the lth iteration
- Select the best surrogate $\mathcal{M}_{pc}$

It is worth noting that for a given total degree P, the LAR algorithm is valuable if the effective number of terms in the retained PC expansion is significantly less than the N_{pc} terms of the standard truncation scheme (Eq. 10). It is also worth mentioning that if $N_{lar} > N_{pc}$, then the last iteration of the LAR algorithm computes a PC surrogate on the full basis corresponding to the ordinary least-square solution. In general when using LAR, $N_{pc} > N_{lar}$, implying that N_{lar} surrogate candidates are formulated when iterations stop.

Spectral projection strategy The spectral projection strategy is an alternative to the least-square approach and relies on the orthogonality property of the polynomial basis. In this framework, the ith PC coefficient $\hat{h}_i$ is computed as follows:

$$\hat{h}_i = < h, \Psi_i(\boldsymbol{\zeta}) >, \tag{18}$$

using the inner product definition given in Eq. 11 and using the basis orthonormality property. The term $< h, \Psi_i(\boldsymbol{\zeta}) >$ involves an integral numerically

computed through a standard Gaussian quadrature rule:

$$< h, \Psi_i(\boldsymbol{\zeta}) > \cong \sum_{k=1}^{N_e} \mathcal{M}(\mathbf{x}^{(k)})\, \Psi_i(\boldsymbol{\zeta}^{(k)})\, w^{(k)}, \tag{19}$$

where $h^{(k)} = \mathcal{M}(\mathbf{x}^{(k)})$ corresponds to the direct model integration evaluated at the kth quadrature root $\mathbf{x}^{(k)}$ (in the physical space) with its associated weight $w^{(k)}$. Here k varies between 1 and N_e, with $N_e = (N_{quad})^n$ when N_{quad} quadrature points are selected for each input random variable. The spectral projection strategy is thus achieved over an evaluation sample E_{quad} of size $(N_{quad})^n$. In order to guarantee that a polynomial function can properly be represented by the PC expansion, N_{quad} should be constrained by the total polynomial degree P such that:

$$N_{quad} \geq (P + 1). \tag{20}$$

There are several ways to compute the term $< h, \Psi_i(\boldsymbol{\zeta}) >$ in Eq. 18 [19]. In this work, a standard quadrature rule is used since only a few quadrature points are enough to obtain a highly accurate result as shown in the following [2,4,37] and this method will only be applied for the case $n = 2$ (the kth quadrature root is defined as follows $\mathbf{x}^{(k)} = [Q^{(k)}; K_s^{(k)}]^T$). Still, the one-dimensional tensorisation method suffers from the curse of dimensionality; its applications may be limited by the size of the parameter space n since the number of terms evolving as $(P + 1)^n$ rapidly increases with respect to n. Alternatives such as the sparse Smolyak tensorisation [36,26] (beyond the scope of this work) could be applied to overcome this issue.

The spectral projection strategy proceeds as follows:

- Choose the maximum polynomial order P
- Choose the number of quadrature roots N_{quad} (according to Eq. 20) per stochastic dimension and identify the roots in the normalized space $\boldsymbol{\zeta}$
- Map the normalized roots onto the physical space to formulate the quadrature roots vector $\mathcal{X} = [\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N_{quad}^n)}]^T$, thus defining the input evaluation sample set E_{quad}
- Integrate the direct model for the quadrature roots $\mathcal{X}$ and formulate the output vector $\mathcal{H} = [h^{(1)}, \cdots, \mathbf{x}^{(N_{quad}^n)}]^T$
- Compute the PC coefficient $\hat{h}_i$ $(i = 0, \cdots, N_{pc} - 1)$ using Eqs. 18–19, and formulate $\mathcal{M}_{pc}$.

3.5 Notations for surrogate models

In the following, the PC expansion is denoted by $\mathcal{M}_{pc}$ with a superscript that gives the maximum polynomial order as well as the method for estimating the coefficients $\hat{h}_i$ $(i = 0, \cdots, N_{pc} - 1)$ along with the size of the sample used for this computation. $\mathcal{M}_{pc}^{ls,P}$ denotes a surrogate model of order P formulated with an ordinary least-square method, while $\mathcal{M}_{pc}^{quad,P}$ denotes a surrogate model

of order P formulated with a Gaussian quadrature method. The size of the evaluation sample is given in logarithmic form for the least-square method for simplicity purpose. For instance, $\mathcal{M}_{pc}^{ls3,P6}$ denotes a PC expansion of maximum order $P = 6$ with coefficients computed from a least-square method over a sample E_{ls} of size $N_{ls} = 10^3$. In the same line of thought, $\mathcal{M}_{pc}^{quad7,P6}$ denotes a PC expansion of maximum order $P = 6$ with coefficients computed from a quadrature method over a sample E_{quad} of size $N_{quad} = 7$ for each input variable following Eq. 20.

3.6 Statistical moments

Once the PC surrogate is built, the water level PDF and its main statistical moments, are estimated statistically or using the properties of the PC expansion.

Stochastic estimation The statistical moments of the water level PDF can be stochastically estimated using a validation sample E_{mc} of size N_{mc} (in the following $N_{mc} = 10^5$). The validation sample is generated through a standard Monte Carlo approach based on random parameters K_s and Q. The water level h can be computed either integrating the forward model $\mathcal{M}$ or the PC surrogate $\mathcal{M}_{pc}$ model (computed from Eq. 9) for the validation sample E_{mc} such that $h^{(k)} = \mathcal{M}(\mathbf{x}^{(k)})$ or $h_{pc}^{(k)} = \mathcal{M}_{pc}(\mathbf{x}^{(k)})$ for the kth member of E_{mc}.

We can stochastically estimate the first four moments of the water level PDF, i.e. the mean value, the standard deviation (STD), the skewness and the kurtosis for the forward model $(\mu_h, \sigma_h, \mathrm{skew}_h, \mathrm{kurt}_h)$ or for the PC surrogate model $(\mu_{h_{pc}}, \sigma_{h_{pc}}, \mathrm{skew}_{h_{pc}}, \mathrm{kurt}_{h_{pc}})$ as formulated here:

$$\mu_{h_{pc}} = \frac{1}{N_{mc}} \sum_{k=1}^{N_{mc}} h_{pc}^{(k)}, \tag{21}$$

$$\sigma_{h_{pc}} = \sqrt{\frac{1}{N_{mc}-1} \sum_{k=1}^{N_{mc}} \left(h_{pc}^{(k)} - \mu_h\right)^2}, \tag{22}$$

$$\mathrm{skew}_{h_{pc}} = \frac{1}{N_{mc}\,(\sigma_h)^3} \sum_{k=1}^{N_{mc}} \left(h_{pc}^{(k)} - \mu_h\right)^3, \tag{23}$$

$$\mathrm{kurt}_{h_{pc}} = \frac{1}{N_{mc}\,(\sigma_h)^4} \sum_{k=1}^{N_{mc}} \left(h_{pc}^{(k)} - \mu_h\right)^4. \tag{24}$$

Analytic derivation To avoid re-sampling the surrogate model, the statistical moments of the water level $\mathbf{h}$ can be directly derived from the coefficients:

$$\mu_{h_{pc}} = \hat{h}_0, \tag{25}$$

$$\sigma_{h_{pc}} = \sqrt{\sum_{0<|i|<P} \left(\hat{h}_i\right)^2}, \tag{26}$$

$$\text{skew}_{h_{pc}} = \frac{1}{\sigma^3_{h_{pc}}} \sum_{0<|i|,|j|,|k|<P} \hat{h}_i\, \hat{h}_j\, \hat{h}_k\, \mathbb{E}\left[\Psi_i(\boldsymbol{\zeta})\, \Psi_j(\boldsymbol{\zeta})\, \Psi_k(\boldsymbol{\zeta})\right], \tag{27}$$

$$\text{kurt}_{h_{pc}} = \frac{1}{\sigma^4_{h_{pc}}} \sum_{0<|i|,|j|,|k|,|l|<P} \hat{h}_i\, \hat{h}_j\, \hat{h}_k\, \hat{h}_l\, \mathbb{E}\left[\Psi_i(\boldsymbol{\zeta})\, \Psi_j(\boldsymbol{\zeta})\, \Psi_k(\boldsymbol{\zeta})\, \Psi_l(\boldsymbol{\zeta})\right], \tag{28}$$

with $|i|$ representing the polynomial degree of the corresponding term in the PC expansion (P being the user-defined maximum polynomial degree).

To measure the extent of variability in a given sample, one may be interested in computing the coefficient of variation (CV), also known as relative STD and defined as the ratio of the STD $\sigma_{h_{pc}}$ to the mean $\mu_{h_{pc}}$.

When the water level is varying in the space, the covariance matrix for water level along the curvilinear abscissa can be estimated from the PC coefficients. [38] provides an analytical estimation of the covariance between the water level at two different grid points m and n:

$$Cov(h(s_m), h(s_n)) = \sum_{0<|i|<P} \left(\hat{h}_i(s_m)\right) \left(\hat{h}_i(s_n)\right), \tag{29}$$

To evaluate the contribution of each random variable to the water level variance $(\sigma_h)^2$ and thus carry out a global sensitivity analysis, we use Sobol' indices directly derived from PC coefficients [30]. For any given random variable x_{i_s} in $\mathbf{x}$, Sobol' index reads:

$$S_{x_{i_s}} = \frac{1}{\sigma^2_{h_{pc}}} \sum_{i \in \mathcal{I}_{i_s}} \left(\hat{h}_i\right)^2, \tag{30}$$

with $\mathcal{I}_{i_s}$ the set of indices such that $0 < |i| < P$ and only the index i_s is nonzero (corresponding to the polynomials $\Psi_i(\boldsymbol{\zeta})$ only depending on the input random variable x_{i_s}). For instance, in the present case $n = 2$, Sobol' indices S_Q and S_{K_s} for Q and K_s (associated with index i_1 and i_2, respectively) are defined as follows:

$$S_Q = \frac{1}{\sigma^2_{h_{pc}}} \sum_{i_1 \geq 1, i_2 = 0} \left(\hat{h}_{i_1 0}\right)^2; \quad S_{K_s} = \frac{1}{\sigma^2_{h_{pc}}} \sum_{i_2 \geq 1, i_1 = 0} \left(\hat{h}_{0 i_2}\right)^2. \tag{31}$$

S_{Ks} (or S_Q) corresponds to a first-order sensitivity index by evaluating the standalone effect of K_s (or Q) on the water level variance $(\sigma_h)^2$.

3.7 Error metrics

PC theory guarantees the convergence of the expansion (assuming that the maximum polynomial order P is sufficient) with respect to the standard $\mathbb{L}_2$ error norm (mean square convergence) [38]. As a consequence, the accuracy of the PC surrogate $\mathcal{M}_{pc}$ over an independent sample E_{mc} is evaluated in $\mathbb{L}_2$. The criteria (LC) is computed on the PC coefficients space of $\hat{h}_{i_1 i_2}$ The criteria (LH) is computed on the water level output variable space. The leave-one-out error norm (LOO) is also introduced when applying the LAR method.

Water level metrics (LH) Assuming the forward model and the surrogate model are integrated over the validation sample E_{mc} of size N_{mc} to represent the water level h and h_{pc} respectively, the standard $\mathbb{L}_2$ error norm is defined as follows:

$$LH = \frac{1}{N_{mc}} \sum_{k=1}^{N_{mc}} \left(h^{(k)} - h_{pc}^{(k)} \right)^2. \tag{32}$$

In order to avoid integrating the forward model over the validation sample E_{mc}, the LOO error is introduced. This solution is of particular interest when one can not afford, due to computational cost limitations, to re-sample the forward model apart from the simulations dedicated to the PC formulation using sample E_{ls} or E_{quad}. When the dimension of the input space increases (for instance from $n = 2$ to $n = 4$ for the Garonne River application here), the number of coefficients rapidly grows, thus increasing the cost of both least square and quadrature methods. As a consequence, the LAR method is preferred; the decomposition on a sparse basis relying on a limited number of PC coefficients. Hence, the LOO metric is used to evaluate the LAR surrogate.

The LOO technique derived from statistical learning theory was used in [10, 5] to assess the PC convergence, and requires the formulation of several surrogates. Each surrogate is built excluding one point out of the evaluation sample E_{lar}; the accuracy of the surrogate is then calculated at this particular point. Following this theory, LOO is defined as follows:

$$LOO = \frac{\sum_{k=1}^{N_{lar}} \left(h^{(k)} - h_{pc}^{(-k)} \right)^2}{\sum_{k=1}^{N_{lar}} (h^{(k)} - \overline{h})^2} \tag{33}$$

with $\overline{h}$ denoting the sample-average water level and $h_{pc}^{(-k)}$ standing for the evaluation of the PC surrogate water level at $\mathbf{x}^{(k)}$, when the surrogate has been built using an experimental design in which $\mathbf{x}^{(k)}$ was excluded.

When applying the LAR strategy, N_{lar} surrogates should be formulated. In practice, [5] proposes a formulation of the LOO metric based on a single

formulation of the LAR model using the whole evaluation sample N_{lar} at once:

$$LOO \equiv \sum_{k=1}^{N_{lar}} \left(\frac{h^{(k)} - h_{pc}^{(k)}}{1 - D_k} \right)^2 \tag{34}$$

where D_k is the kth diagonal term of the matrix $\mathbf{\Psi}(\mathbf{\Psi}^T \mathbf{\Psi})^{-1} \mathbf{\Psi}^T$.

Coefficient metrics (LC) The truncation and projection errors produced by the PC expansion can also be assessed through the comparison of the PC-coefficients. For this purpose, a reference surrogate model is built using a least-square method that involves a sufficient number of forward model simulations ($N_{ls,ref} = 10^5$) to guarantee the convergence of Eq. 17. The vector of coefficients for this reference model is further denoted by $\hat{\mathbf{h}}^{ref}$. In the following, the coefficients $\hat{h}_i^{ref}$ are compared to those of surrogate models (noted $\hat{h}_i$) based on least-square or quadrature methods involving a smaller number of members (N_{ls} or N_{quad}^n) than $N_{ls,ref}$. The LC metric is defined in the sense of the $\mathbb{L}_2$ norm as follows:

$$LC = \sum_{i=0}^{N_{pc}-1} \left(\hat{h}_i^{ref} - \hat{h}_i \right)^2. \tag{35}$$

3.8 Numerical implementation

Due to the increasing interest in UQ over the last decade, a significant number of UQ-dedicated tools/libraries are now available for the scientific community. OpenTURNS (www.openturns.org) is an open-source (LGPL) scientific library usable as a Python module dedicated to the treatment of uncertainties in a structured industrial approach. EDF, Airbus and Phimeca have been working together since 2005 to build OpenTURNS designed to perform uncertainty treatment and reliability analyses. OpenTURNS offers a wide catalog of features for UQ (among them PC strategies [11] such as least-square, gaussian quadrature and LAR used in the present work) and benefits from a well-organized developers and users community (forum, training, user guides). It can either be used as a python module when the numerical model is implemented in python, or used as a component with a coupling platform. In the present study, MASCARET and OpenTURNS are both integrated components of the SALOME platform developed at EDF (www.salome-platform.org). This integrated framework allows for an efficient use of MASCARET as a python function driven by the flowchart of the PC algorithms implemented in OpenTURNS. For further information on the OpenTURNS library, the reader is referred to the online reference and use case guides.

4 Results

This section aims at studying the convergence of the PC expansion with respect to the choice of the truncation strategy, the projection strategy and the maximum polynomial order P.

4.1 Idealized channel

The PC-expansion approach presented in Sect. 3 is applied for the idealized channel test case with respect to the friction coefficient K_s and the discharge Q in steady flow ($\mathbf{x} = [Q, K_s]^T$ with $n = 2$). For uniform steady flow, the water level at the downstream boundary condition is set so that the flow is governed by the Manning equation (Eq. 5). This strategy is then applied for gradually-varied steady flow on the channel test case with a deterministic description of the downstream condition so that backwater curve appear.

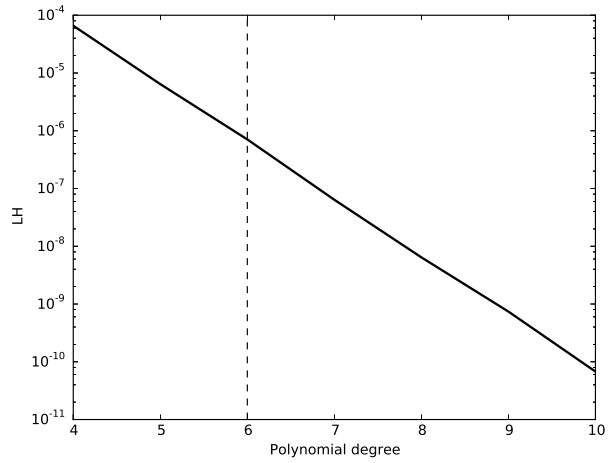


Fig. 3 Truncation error illustrated by the water level error LH (Eq. 32) in logarithmic scale with respect to the total degree P; idealized channel case with uniform steady flow (Manning equation). LH is computed over the validation sample $N_{mc} = 10^5$ for the least-square surrogate $\mathcal{M}_{pc}^{ls5,P}$ formulated over the evaluation sample $N_{ls,ref} = 10^5$ for P varying between 4 and 10.

4.1.1 Uniform steady flow

Truncation error In the first place, the truncation error is investigated in order to identify the appropriate maximum polynomial order P and thus the total number of coefficients N_{pc} to estimate in the PC expansion. In this perspective, a least-square surrogate is built here with a large enough number

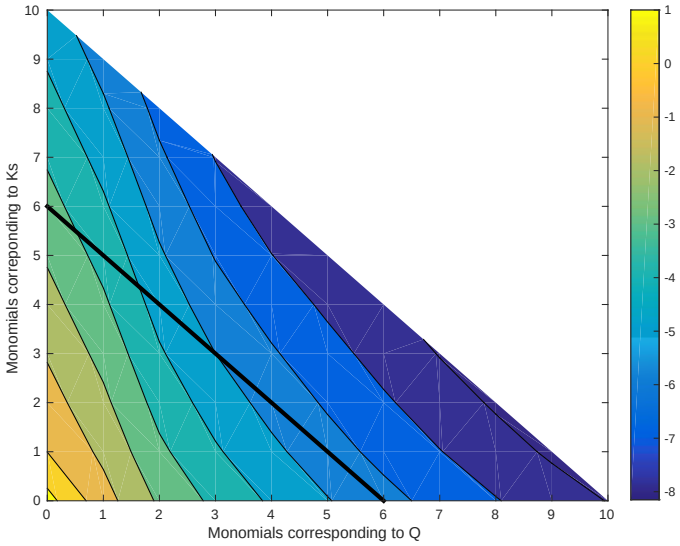


Fig. 4 Spectrum of coefficients for the least-square surrogate model; idealized channel case with uniform steady flow (Manning equation). The surrogate is formulated over the evaluation sample $N_{ls,ref} = 10^5$ and a maximum polynomial order $P = 10$. The colorbar describes the amplitude of the coefficients in logarithmic scale. The Hermite monomials (corresponding to Q) are displayed on the x-axis; the Legendre monomials (corresponding to K_s) are displayed on the y-axis. The bivariate coefficients $\hat{\mathbf{h}}_{i1,i2}$ are displayed in the lower triangular matrix along x- and y-axis. The oblique solid black line represents $P = 6$.

of the direct model simulations ($N_{ls} = 10^5$) to distinguish the truncation error from the error in the coefficient estimation. The least-square surrogate $\mathcal{M}_{pc}^{ls5,P}$ is formulated over the evaluation sample $N_{ls,ref}$ for varying P, then used to predict the water level h_{pc} over the validation sample E_{mc} of size N_{mc}. These water levels are validated against those obtained with a Monte Carlo direct model sampling to compute LH as described in Sect. 3.7. Figure 3 displays LH (Eq. 32) in logarithmic scale when P increases from 4 to 10. For instance, for $P = 6$, LH is approximately equal to 10^{-6}. The error $log(LH)$ linearly decreases by one order of magnitude as the polynomial order is incremented by 1 up to $P = 10$. The PC expansion is built here on a hybrid polynomial basis made of Hermite polynomials for Q and Legendre polynomials for K_s. The present convergence attests that the hybrid polynomial basis is adapted to properly describe the statistics of both input random variables and output variables.

The spectrum of coefficients $\hat{h}_{i1,i2}$ is then analyzed to check which terms in the PC expansion are contributing the most to the water level h. For this purpose, Fig. 4 shows the spectrum of coefficients associated with the surrogate $\mathcal{M}_{pc}^{ls5,P10}$ ($P = 10$ consistently with Fig. 3). The colorbar represents the amplitude of $\hat{h}_{i1,i2}$ in logarithmic scale. The first coefficient $\hat{h}_{00}$ (the mean water level) is displayed at the graph origin and features the highest magnitude. The monomials that only depend on Q (i.e. $\hat{h}_{i_1 0}$) are displayed on the x-axis; those

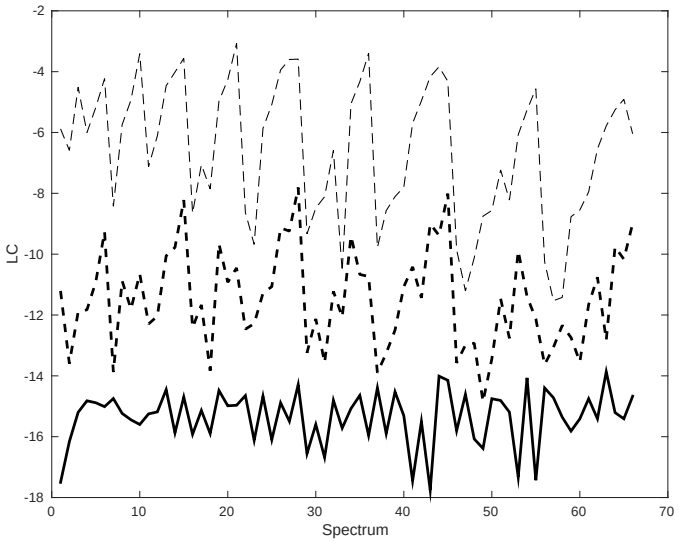


Fig. 5 Comparison of LC in logarithmic scale between the least-square surrogate $\mathcal{M}_{pc}^{ls2,P10}$ (thin dashed line), $\mathcal{M}_{pc}^{ls3,P10}$ (thick dashed line) and the quadrature surrogate $\mathcal{M}_{pc}^{quad11,P10}$ (thick solid line); idealized channel case with uniform steady flow (Manning equation). The x-axis is the coefficient index within the expansion spectrum (66 coefficients for $P = 10$).

that only depend on K_s (i.e. $\hat{h}_{0i_2}$) are displayed on the y-axis. The bivariate coefficients $\hat{h}_{i1,i2}$ are displayed in the lower triangular matrix along x-axis and y-axis. For instance, $x = 4$ and $y = 2$ represent the polynomial function $Q^4\,K_s^2$ of amplitude 10^{-6}. The K_s-monomials are found to be larger than the Q-monomials: given the PDFs for both random variables, the sensitivity to K_s is therefore larger than that to Q for uniform steady flow. This is confirmed by Sobol' indices: $S_Q = 0.06$ and $S_{K_s} = 0.94$ for variation coefficients equal to $CV_Q = 10$ % and $CV_{K_s} = 35$ %. These results depend on the hypothesis on the input random variables and especially on the choice of their distributions; for instance, if CV_Q was set larger than CV_{K_s}, the Q-monomials would have been larger than the K_s-monomials. The amplitude of the monomials is larger than that of the bivariate coefficients, meaning that the univariate coefficients are more significant than the bivariate coefficients. The value of the coefficients is also found to decrease as the polynomial degree increases. For instance, when the polynomial degree is beyond $P = 6$ (oblique solid black line in Fig. 4), all coefficients are smaller than 10^{-3}. Given the results on LH for the estimation of the water level in Fig. 3 and the results on the coefficient values in Fig. 4, it seems sufficient to truncate the PC expansion to $P = 6$, thus implying the computation of 28 coefficients for $n = 2$ (when $P = 10$, 66 coefficients should be computed).

Sampling error In the following, we investigate the size of the evaluation sample that is necessary to properly estimate the coefficients for $P = 10$ for both

least-square and quadrature strategies; the objective is now to focus on the projection error. In order to limit the computational cost induced by the direct model simulations, the size of the evaluation sample for the least-square surrogate is now reduced (previously set to $N_{ls,ref} = 10^5$); the surrogate models $\mathcal{M}_{pc}^{ls2,P10}$ and $\mathcal{M}_{pc}^{ls3,P10}$ are formulated with respectively $N_{ls} = 10^2$ and $N_{ls} = 10^3$. The quadrature surrogate $\mathcal{M}_{pc}^{quad11,P10}$ is formulated with the number of quadrature roots set to $N_{quad} = 11$ for each input random variable consistently with $P = 10$. Figure 5 displays the error LC as defined in Sect. 3.7 in logarithmic scale for each coefficient within the expansion spectrum (66 coefficients) and compares $\mathcal{M}_{pc}^{ls2,P10}$ (thin dashed line), $\mathcal{M}_{pc}^{ls3,P10}$ (thick black line) and $\mathcal{M}_{pc}^{quad11,P10}$ (thick solid line). LC decreases from 10^{-6} to 10^{-10} as the size of the evaluation sample for the least-square surrogate increases from $N_{ls} = 10^2$ to 10^3. $\mathcal{M}_{pc}^{quad11,P10}$ leads to a LC error value that is smaller to that of $\mathcal{M}_{pc}^{ls3,P10}$ for a significantly reduced computational cost (only 121 direct model evaluations) for all coefficients of the expansion spectrum; thus providing a reliable surrogate model for Eq. 3. The projection strategy can thus be trusted to provide a reliable surrogate model at a lower cost than the least-square strategy.

4.1.2 Gradually-varied steady flow

A similar analysis is carried out for gradually-varied steady flow to check the impact of more complex flow features (backwater curves) on the behavior and accuracy of the PC expansion. When the downstream boundary condition is prescribed to $h_{out} = 10m$ as Q and K_s are sampled within their respective PDF given in Sect. 2.2.1, the Manning equilibrium is no longer preserved over the entire channel. The upstream flow is still under steady flow conditions, while the flow is governed by the backwater equations (Eq. 3) in the downstream part of the river. It should be noted that within the sampling of the physical random variables, the flow remains subcritical and the water level describes a backwater curve of type M_1 or M_2. The surrogate is here formulated following the same strategy as for uniform steady flow: first, setting the maximum polynomial order P; then, setting the number of direct model evaluations for computing the coefficients. Since the flow is now spatially distributed, we compute a surrogate at each simulation grid point along the curvilinear abscissa using the same set of direct model evaluations, thus providing space-varying coefficients as described in Sect. 3.1.

Truncation error Figure 6 shows the water level error LH (Eq. 32) in logarithmic scale along the 40-km channel curvilinear abscissa for the least-square surrogate $\mathcal{M}_{pc}^{ls5,P}$ built with an evaluation sample $N_{ls,ref} = 10^5$. LH is computed over the validation sample E_{mc} with $N_{mc} = 10^5$. The colored lines correspond to LH when P varies from 4 (top curve in blue) to 10 (bottom curve in black). As highlighted in Fig. 3 for uniform flow, LH is small and in the present case remains below 10^{-6} when P is equal or larger than 6. LH decreases by one order of magnitude when P increases by 1 for $P = 4, \ldots, 7$ over

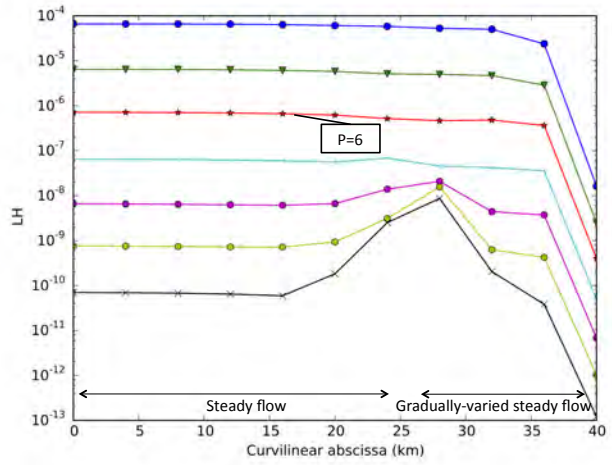


Fig. 6 Truncation error along the 40-km channel illustrated by the water level error LH (Eq. 32) in logarithmic scale for varying total degree P; idealized channel case with gradually-varied steady flow. LH is computed every 4 km over the validation sample $N_{mc} = 10^5$ for the least-square surrogate $\mathcal{M}_{pc}^{ls5,P}$ formulated over the evaluation sample $N_{ls,ref} = 10^5$ for $P = 4, \cdots, 10$ (top curve in blue for $P = 4$ to bottom in black for $P = 10$).

the entire computational domain. Increasing P beyond this threshold value does no longer guarantees a decrease of LH, especially where the flow transitions from uniform flow to backwater curve. It was shown (not illustrated here) that refining the mesh resolution by a factor of 2 (from 100-m to 50-m discretization) allows to further reduce LH when P increases, even where the flow dynamics gets more complex.

Sampling error In the following, $P = 6$ is set ($LH = 10^{-6}$ in coherence with Fig. 3 for uniform flow, here in the upstream part of the channel). We now investigate the size of the evaluation sample that is necessary to properly estimate the coefficients with the quadrature method. The performance of the surrogate is investigated when the number of quadrature roots N_{quad} varies from 4 to 10. Figure 7 displays the water level error LH (Eq. 32) in logarithmic scale along the curvilinear abscissa computed for the surrogate $\mathcal{M}_{pc}^{quad,P6}$ (top curve in blue for $N_{quad} = 4$ to bottom in black for $N_{quad} = 10$). Again LH is computed over the validation sample E_{mc} (with $N_{mc} = 10^5$). As expected from theory (Eq. 20), LH decreases as N_{quad} increases up to $N_{quad} = 7$ beyond which the error remains constant to 10^{-6}. For $P \geqslant 7$, the error sums up to the truncation error also represented in Fig. 6 for $\mathcal{M}_{pc}^{ls5,P6}$ with the red curve. It should be noted that in both Figs. 6–7, the LH error is smaller at the downstream boundary condition specified in a deterministic way. The transition between the uniform regime and the backwater curve (M_1 or M_2) occurs between 25 km and 30 km; the dynamics of this particular point is complex to represent with the surrogate, thus the error LH tends here to have

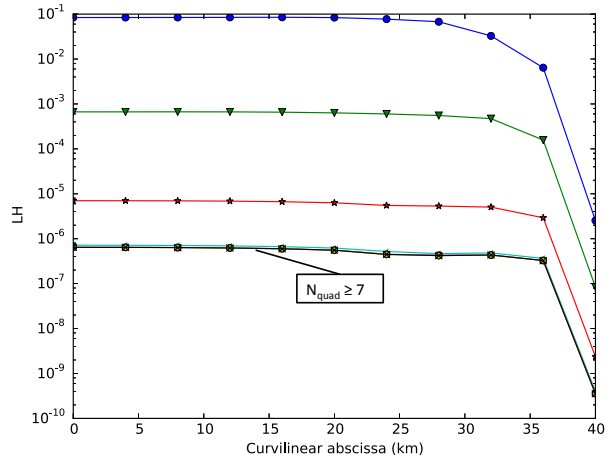


Fig. 7 Counterpart of Fig. 6 for the quadrature surrogate $\mathcal{M}_{pc}^{quad,P6}$ for $N_{quad} = 4, \ldots, 10$ (top curve in blue for $N_{quad} = 4$ to bottom curve in black for $N_{quad} = 10$) for $P = 6$. LH is computed every 4 km along the channel.

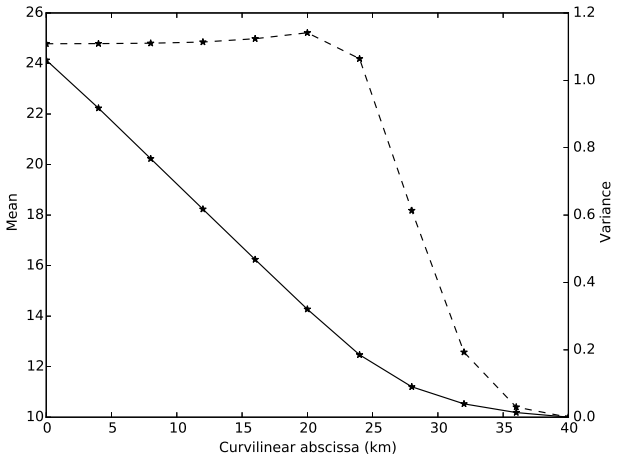


Fig. 8 Water level mean (m) in solid line (left y-axis) and variance (m^2) in dashed line (right y-axis) along the curvilinear abscissa; idealized channel case with gradually-varied steady flow. Comparison between the Monte Carlo stochastic estimates (lines) and quadraturre surrogate $\mathcal{M}_{pc}^{quad7,P6}$ ($P = 6$ and $N_{quad} = 7$) estimates (symbols).

larger values than at other locations over the reach, yet it remains extremely small (10^{-8}).

Statistical moments The next step consists in validating the statistical moments as defined in Sect. 3.6 for the retained quadrature surrogate $\mathcal{M}_{pc}^{quad7,P6}$;

the validation is carried out with respect to the classical Monte Carlo sampling over the validation sample $N_{mc} = 10^5$ (all backwater curves within the sampling describe M_1 backwater curves). The stochastically estimated mean and variance are displayed in Fig. 8 respectively on the left y-axis in solid line and on the right y-axis in dashed line along the 40-km channel. The mean and variance computed with the coefficients for $\mathcal{M}_{pc}^{quad7,P6}$ are represented with black stars on each line. The surrogate $\mathcal{M}_{pc}^{quad7,P6}$ thus offers a satisfying approximation of direct model water level over the whole computational domain, in terms of first and second order moments, for the present idealized channel case study.

4.2 Application on the Garonne River

The PC-expansion approach was validated for the idealized channel test case and is now applied to a real case of the Garonne River presented in Sect. 2.2.2, still in steady flow. The flow dynamics is governed by the SWE (Eq. 1) solved by MASCARET; the water level is non uniform as the river width, bathymetry and friction coefficients are spatially distributed over the Tonneins-La Réole reach. While the PC-expansion is formulated for the entire domain, a particular attention is paid to the water level at Marmande, where in-situ observations are available and where DA is applied in flood forecasting operational context.

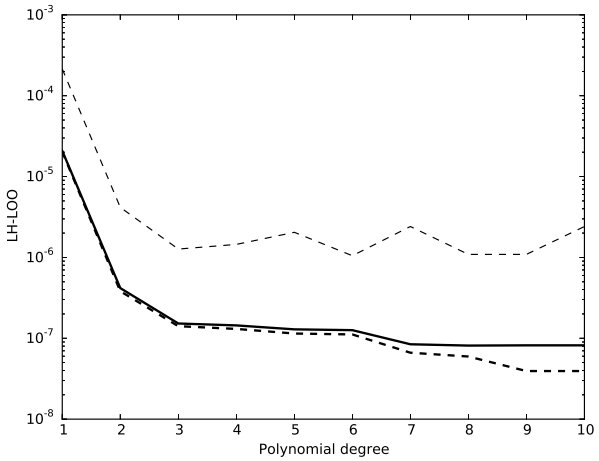


Fig. 9 Comparison of error metrics at Marmande observing station (36 km on the Garonne reach) for varying total degree P ($P = 1, \cdots, 10$): LOO in dashed lines ($\mathcal{M}_{pc}^{lar2,P}$ in thin dashed line; $\mathcal{M}_{pc}^{lar3,P}$ in thick dashed line); normalized LH in solid line computed over a validation sample E_{mc} of size $N_{mc} = 10^5$.

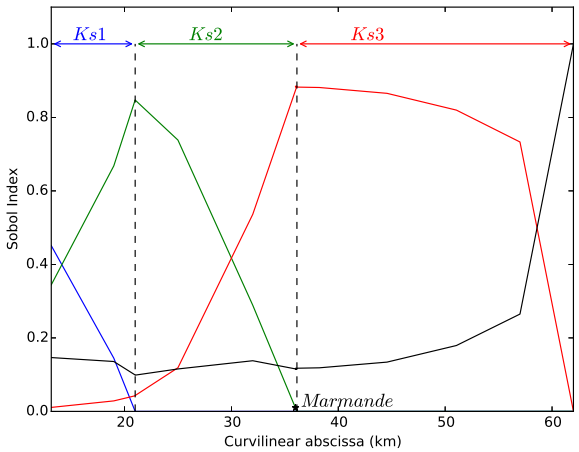


Fig. 10 Sobol' indices (first-order) computed with the LAR surrogate $\mathcal{M}_{pc}^{lar2,P4}$ formulated over an evaluation sample E_{lar} of size $N_{lar} = 10^2$ for a maximum polynomial order $P = 4$ with respect to K_{s_1} (blue line), K_{s_2} (green line), K_{s_3} (red line) and Q (black line). Vertical dashed lines indicate the limits between the different sections of K_s. The curvilinear abscissa at Marmande is $s = 36\ km$.

Sensitivity analysis with LAR strategy The cost of the PC-expansion is here potentially higher than for the idealized channel case study as there are four main sources of uncertainty (Q, K_{s_1}, K_{s_2}, K_{s_3}) in the Garonne hydraulic model. To ensure that this cost remains limited, the number of uncertain variables n should be reduced; the PC-expansion will then be formulated with respect to the most significant variables only. For this purpose, a sensitivity analysis is carried out and the sources of uncertainty are classified using the Sobol' indices defined in Sect. 3.6. The Sobol' indices are computed by means of a LAR surrogate as defined in Sect. 3.4. The accuracy of the LAR surrogate is tested by comparing $\mathcal{M}_{pc}^{lar2,P}$ and $\mathcal{M}_{pc}^{lar3,P}$ formulated over an evaluation sample E_{lar} of size $N_{lar} = 10^2$ and $N_{lar} = 10^3$.

Figure 9 displays the water level error at Marmande for $\mathcal{M}_{pc}^{lar3,P}$ measured with both the LH metric (Eq. 32) in solid line and the LOO metric (Eq. 34) in thick dashed line for the maximum polynomial order P varying between 1 and 10. The LOO error for $\mathcal{M}_{pc}^{lar2,P}$ is also plotted in thin dashed line to check the sensitivity of the LAR result to the size of the evaluation sample E_{lar}. LH is computed over a validation sample E_{mc} of size $N_{mc} = 10^5$ that samples the four random variables. For this comparison, LH is normalized by the water level variance and multiplied by N_{lar} to be consistent with the $L00$ formulation. Figure 9 shows that for $P = 4$, the LOO converges to 10^{-6} for $N_{lar} = 100$ and to 10^{-7} for $N_{lar} = 1000$, similarly to LH for $N_{mc} = 10^5$. This highlights the merits of the LAR method to reduce the computational cost, while providing an accurate representation of the water level.

The LAR method is thus used for sensitivity analysis with respect to (Q, K_{s_1}, K_{s_2}, K_{s_3}) to compute the Sobol' indices displayed in Fig. 10. Globally speaking, the impact of Q is homogeneous over the 60 km Garonne River reach. As the downstream boundary condition is prescribed by the local rating curve at Tonneins, the uncertainty in water level is totally explained by the uncertainty in discharge, thus S_Q is equal to 1 at the exit of the hydraulic network. The water level is also sensitive to the local K_s of the section as well as to the K_s of the section located downstream. In the upstream section ($0 - 20$ km), the water level is mostly sensitive to K_{s_1} and K_{s_2}, while the impact of K_{s_3} is smaller. In the second section ($20-36$ km), the water level is mostly sensitive to K_{s_2} and K_{s_3}. In the third section ($36-60$ km) that includes Marmande, the water level is only sensitive to K_{s_3} and Q. At Marmande, the sensitivity to K_{s_1} and K_{s_2} is zero, the Sobol' index for K_{s_3} is equal to 0.88 and that for Q is equal to 0.12. These values are strongly related to the prescribed PDFs on K_{s_3} and Q; in the present case, the variation coefficient for Q is $CV_Q = 10$ %, while the variation coefficient for K_{s_3} is $CV_{Ks_3} = 40$ %. Given the sensitivity analysis results, it is legitimate to only consider K_{s_3} and Q as significant input variables in the following in order to build a surrogate at Marmande in an efficient way so as to be coherent with operational framework. In the following, only these two variables are considered for the PC-expansion so that the random vector is now $\mathbf{x} = [Q, K_{s_3}]^T$ with $n = 2$; still, the surrogate is build at every 4 km of the computational domain.

Uncertainty quantification with least-square and quadrature strategies Figure 11 displays the water level error LH (Eq. 32) in logarithmic scale along the 60-km hydraulic network for the least-square surrogate $\mathcal{M}_{pc}^{ls5,P}$ built with an evaluation sample E_{ls} of size $N_{ls,ref} = 10^5$. The colored lines correspond to LH when $P = 1$ (blue solid line), $P = 6$ (purple solid line), $P = 11$ (green solid line) and $P = 15$ (black solid line). The error is found to decrease as the maximum polynomial order P increases and is of the order of 10^{-4} for $P = 6$. This error can be compared to that of the Manning equation displayed in Fig. 3 $LH = 10^{-6}$. Figure 11 also displays LH for a surrogate computed with the quadrature method $\mathcal{M}_{pc}^{quad,P6}$ setting $P = 6$ and varing $N_{quad} = 6, 7, 8$ (purple dashed, dashed-dotted and dotted lines). The error LH associated to $\mathcal{M}_{pc}^{quad7,P6}$ is similar to that of $\mathcal{M}_{pc}^{ls5,P6}$, while it only involves 49 evaluations of the direct model instead of 10^5. In conclusion, a PC expansion $\mathcal{M}_{pc}^{quad7,P6}$ with $P = 6$ and $N_{quad} = 7$ provides a satisfying description of the water level over the Garonne reach with respect to Q and K_{s_3} in the sense of the LH metric.

Figures 12–13 present the statistical moments along the 60-km Garonne reach, which are defined in Sect. 3.6 and computed here for the quadrature surrogates $\mathcal{M}_{pc}^{quad16,P15}$ and $\mathcal{M}_{pc}^{quad7,P6}$. The mean (solid line, left y-axis) and the variance (dashed line, right y-axis) are shown in Fig. 12a; the skewness (solid line, left y-axis) and the kurtosis (dashed line, right y-axis) are shown in Fig. 13a (the curves associated with both surrogates $\mathcal{M}_{pc}^{quad16,P15}$ and $\mathcal{M}_{pc}^{quad7,P6}$ overlap). In complement, Figs. 12b–13b present the errors in

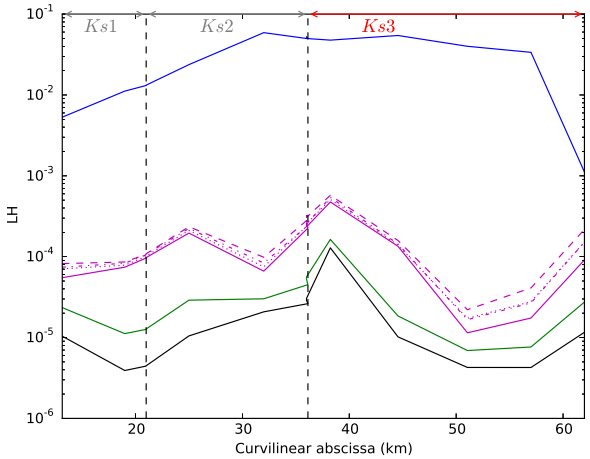


Fig. 11 Truncation and quadrature errors illustrated by the water level error LH (Eq. 32) in logarithmic scale along the curvilinear abscissa; Garonne River case in steady flow. LH is computed over the validation sample E_{mc} with $N_{mc} = 10^5$ for the following surrogates: least-square surrogates $\mathcal{M}_{pc}^{ls,P}$ $P = 1, 6, 11, 15$ (solid blue, purple, green and black line); and quadrature surrogates $\mathcal{M}_{pc}^{quad,P6}$ with $P = 6$ and $N_{quad} = 6, 7, 8$ (purple dashed, dashed-dotted and dotted lines).

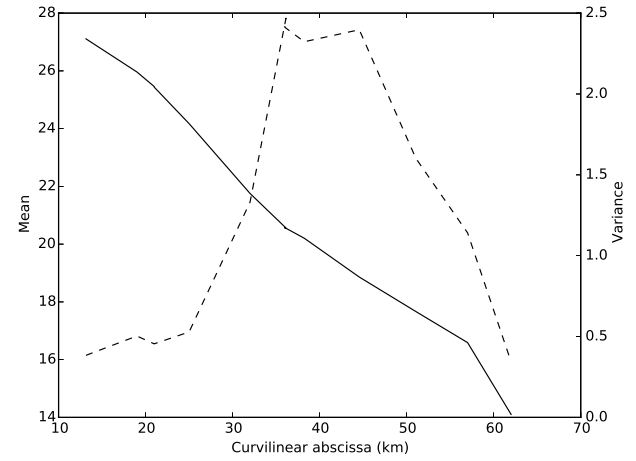
the statistical moments computed with respect to their stochastic estimation with a Monte Carlo sampling E_{mc}: red lines correspond to $\mathcal{M}_{pc}^{quad16,P15}$; blue lines correspond to $\mathcal{M}_{pc}^{quad7,P6}$. Results show that the estimation of the four statistical moments is satisfying with $P = 6$. The mean is estimated with an error of about 10^{-6} for $P = 6$ (blue solid line) that decreases by one order of magnitude for $P = 15$ (red solid line). The variance is estimated with an error of 10^{-4} for $P = 6$ (blue dashed line) that decreases by three orders of magnitude for $P = 15$ (red dashed line). The skewness and the kurtosis are respectively estimated with an error of 10^{-4} and 10^{-3} for $P = 6$ (solid blue and red lines) that decreases by one order of magnitude for $P = 15$ (dashed blue and red lines). The mean water level varies from 27 m upstream to 14 m downstream. The water level variance is about 0.5 m^2 in the first part of the reach and increases up to 2.5 m^2 in the area of Marmande, where the bathymetry strongly varies and the flow dynamics gets more complex with local changes of backwater curves. The skewness varies from -0.1 to 0.4 with a steep change in the Marmande area, suggesting that the water level PDF at Marmande is non-symmetric (non-gaussian) with a small shift on the left for smaller water level than the mean upstream the Garonne reach as well as a significant shift on the right for larger water levels than the mean over most of the Garonne reach, especially near Marmande. In this area, the water level PDF has a positive skew; large or extreme water levels values are more likely to occur. On the contrary, the kurtosis has negative values upstream and negative values

near Marmande, meaning that the PDF tends to be sharper than a gaussian PDF upstream, and smoother than a gaussian PDF downstream.

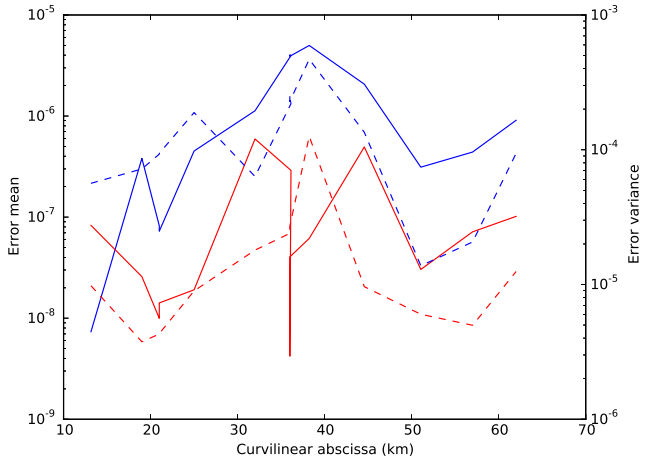
The observed skewness and kurtosis are probably due to the presence of multiple peaks in the PDF for this area of the Garonne reach. This is confirmed in Fig. 14. Figure 14a compares the PDF at the upstream boundary condition of the Garonne reach obtained with a Monte Carlo approach over E_{mc} (blue line) as well as quadrature surrogates $\mathcal{M}_{pc}^{quad16,P15}$ (red line) and $\mathcal{M}_{pc}^{quad7,P6}$ (green line). Figure 14b presents the counterpart of Figure 14a at Marmande. Upstream, the bathymetry is quite smooth, the flow dynamic represented by the surrogate is driven by Q (K_{s_3} has a small influence as shown in Fig. 15), and the relation between water level and discharge is fairly linear so that the water level PDF is almost gaussian. A PC expansion of total degree $P = 6$ thus provides a satisfying description of the PDF upstream the Garonne reach (the PDF estimated with $\mathcal{M}_{pc}^{quad7,P6}$ is close to that computed with a Monte Carlo approach – similar results are obtained up to Marmande). At Marmande, the flow dynamics is mostly driven by K_{s_3} (Q has a smaller impact as shown in Fig. 15). The flow dynamics is more complex since the bathymetry and the river width vary, the nonlinearities are stronger and the estimation of the PDF requires a higher total degree P for the PC expansion as shown in Fig. 14b.

These results demonstrate that the PC-expansion strategy can be used to formulate a surrogate for the SWE solved by the direct model over the Garonne reach between Tonneins and La Réole. The surrogate is built with respect to the most significant sources of uncertainty using the quadrature method to compute the coefficients. The required maximum polynomial order is higher when the flow becomes more complex.

Covariance estimation for DA In the context of DA, the formulation of the Kalman gain matrix relies on the estimation of the water level error covariance matrix that describes the spatial correlation between the errors along the river. The use of the PC surrogate model in place of the direct model allows for a significant cost reduction for stochastic estimate of these statistics, which are here formulated using the PC coefficients directly and which thus avoid resampling errors using formulas 29 and 26. Figure 16a displays the water level correlation matrix along the 60-km Garonne reach that is estimated using the retained quadrature surrogate $\mathcal{M}_{pc}^{quad7,P6}$ and parameterized with respect to Q and K_{s_3} only. These correlations are estimated with an error on the order of 10^{-7} with respect to those estimated with a classical Monte Carlo approach based on $N_{mc} = 10^5$ evaluations of the direct model as displayed in Fig. 16b. Two locations on the river are marked by vertical dashed lines at $s = 15$ km



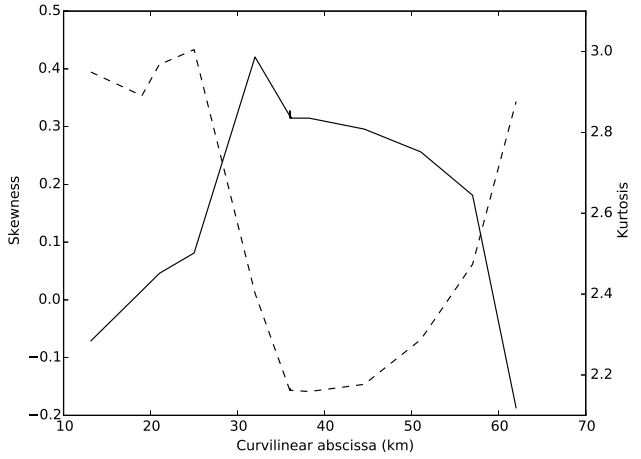
(a)



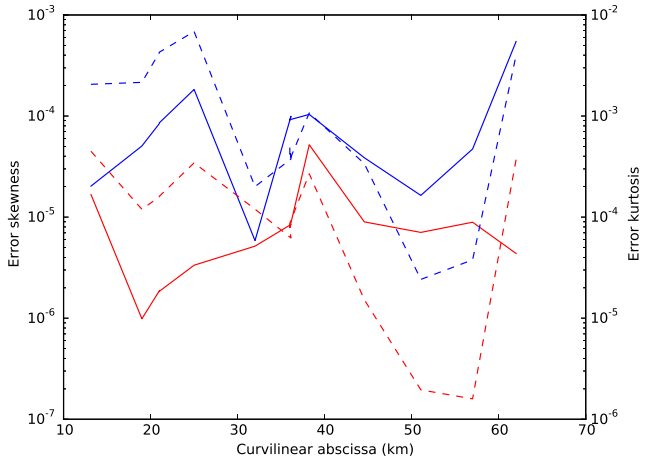
(b)

Fig. 12 a- Comparison of mean (left y-axis) and variance (right y-axis) statistical moments between quadrature surrogates $\mathcal{M}_{pc}^{quad16,P15}$ and $\mathcal{M}_{pc}^{quad7,P6}$ (results overlap). b- Error with respect to Monte Carlo stochastic estimation over $N_{mc} = 10^5$ members in the mean (left y-axis) and variance (right y-axis) estimates for quadrature surrogates $\mathcal{M}_{pc}^{quad16,P15}$ (red) and $\mathcal{M}_{pc}^{quad7,P6}$ (blue).

and $s = 36\ km$ (Marmande) that describe the water level error correlation functions for these respective locations. By definition, the correlation is equal to 1 at the location points where the correlations are expressed and decreases for when moving away to the other grid points along the channel. We first analyze the correlation function for the upstream location $s = 15\ km$ where



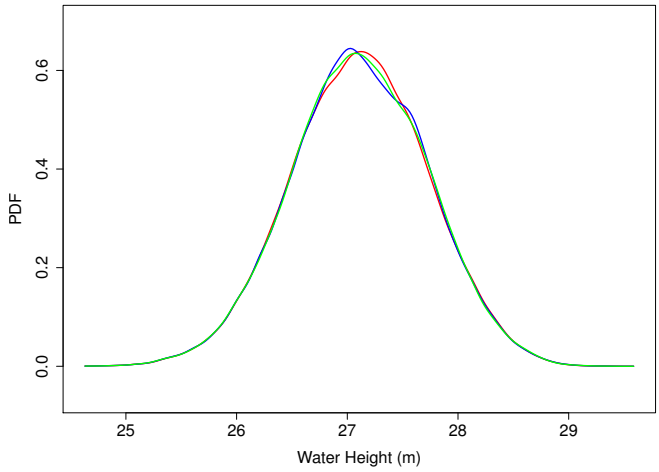
(a)



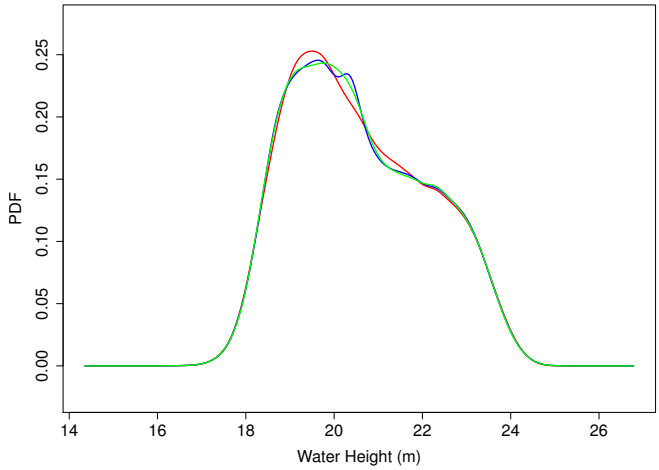
(b)

Fig. 13 Same caption as Fig. 12 but for the skewness and kurtosis statistical moments.

the error in water level is dominated by the error in Q as shown in Fig. 15 with $S_Q = 0.9$ and $S_{K_{s_3}} = 0.1$. The surrogate is formulated only with respect to Q and K_{s_3}. As a consequence, the downstream boundary condition prescribed with a rating curve that relates Q to h has a strong influence on the correlation function so that the correlation increases with the curvilinear abscissa close to the downstream boundary. On the contrary, at Marmande ($s = 36\ km$) and its neighboring, the error in water level is dominated by the error in K_{s_3} with $S_Q = 0.85$ and $S_{K_{s_3}} = 0.15$. As a consequence, the downstream boundary



(a)



(b)

Fig. 14 Water level PDF estimated with Monte Carlo sampling E_{mc} over $N_{mc} = 10^5$ members, either using the direct model (blue line) or quadrature surrogates $\mathcal{M}_{pc}^{quad7,P6}$ (red line) and $\mathcal{M}_{pc}^{quad7,P15}$ (green line) at $s = 15\ km$ (a) and $s = 36\ km$ at Marmande (b).

condition has a lesser impact on the water level error correlation function in the downstream part of the river (Fig. 15 shows that the Sobol' index $S_{K_{s_3}}$ drops in the end of the reach to be zero at the downstream boundary condition). The correlation remains close to 1 over a large section (approximately 20 km around Marmande) as the influence of K_{s_3} is homogeneous over the

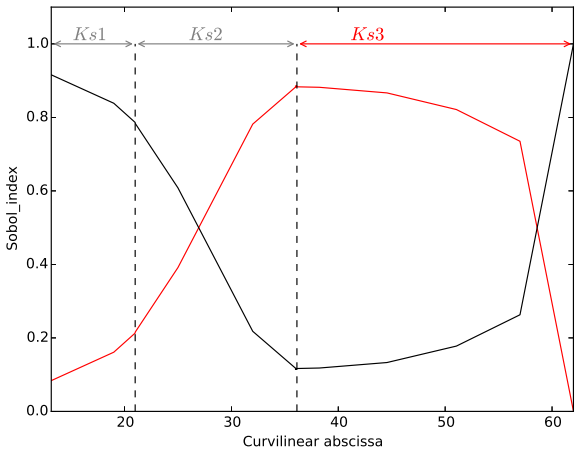


Fig. 15 Sobol' indices (first-order) computed with the quadrature surrogate $\mathcal{M}_{pc}^{quad7,P6}$ parameterized with respect to Q (black solid line) and K_{s_3} (red solid line) over the Garonne reach. Vertical dashed lines indicate the limits between the different sections of K_s. The curvilinear abscissa at Marmande is $s = 36\ km$.

section (K_{s_3} remains around 0.8). In the EnKF, the correlation matrix evolves over time to represent the spatially-distributed and temporally-dependent flow patterns. Estimating this correlation matrix through the use of a PC surrogate (using only 49 direct model evaluations of the water level) is thus a promising approach to reduce the cost of the EnKF, while representing the correlations with accuracy .

5 Conclusions and Perspectives

We investigated the use of a gPC-expansion strategy for hydraulic modeling of subcritical steady flows, both on an idealized channel test case and on a reach of the Garonne river. In this study, the surrogate model is used for uncertainty propagation and sensitivity analysis in place of the direct model solving the SWE. A major conclusion of this work is that the model output covariance matrix estimated with the reduced model build from a limited number of direct model evaluations is more reliable than that estimated with a classical MC approach with the same direct model evaluations budget.

The gPC-expansion was formulated with respect to two random variables, the friction coefficient of the river bed (K_s) and the upstream discharge (Q), on a hybrid basis combining Legendre and Hermite polynomials. The polynomial coefficients were computed with a quadrature method based on the evaluation of the direct model at the roots of the basis polynomials. The validation of the surrogate model is carried out thanks to $\mathbb{L}_2$ error metrics defined in the space of the simulated water level as well as in the space of the polynomial

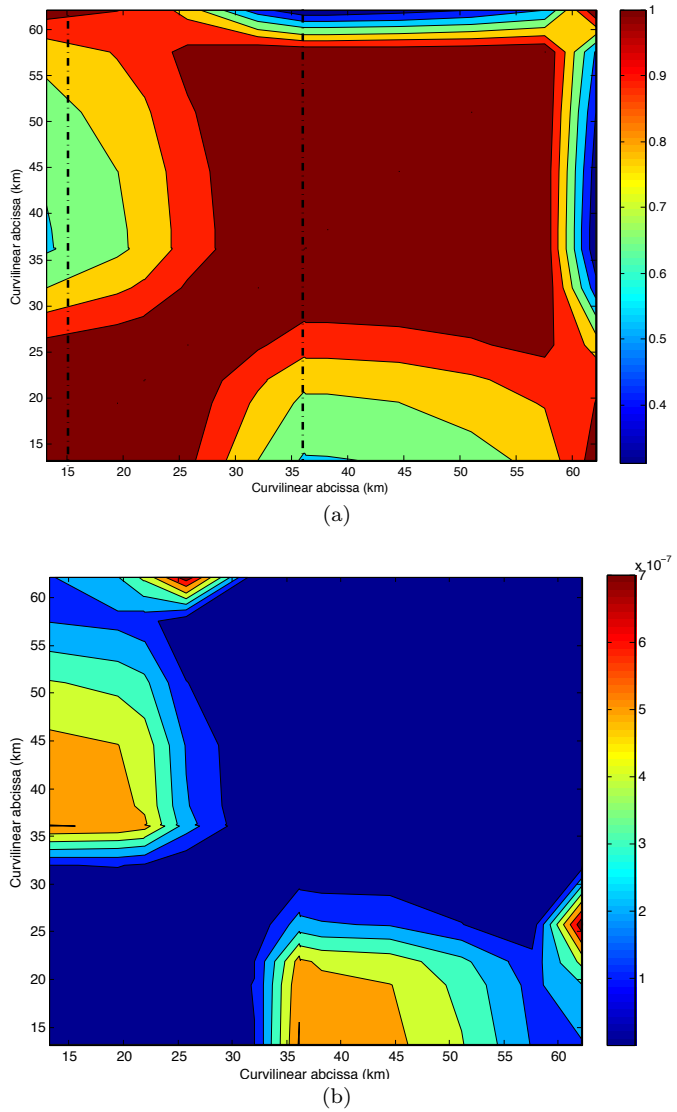


Fig. 16 a-Water level correlation matrix over the 60-km Garonne reach estimated with the retained surrogate $\mathcal{M}_{pc}^{quad7,P6}$ ($P = 6$, $n = 2$ with $\mathbf{x} = [Q, K_{s_3}]^T$). The colorbar represents the correlation in the water level error for each pair of grid points along the reach. Vertical dashed lines represent two grid points along the reach ($s = 15\ km$ and $s = 36\ km$ at Marmande). b- Error on the water level correlation matrix estimated $\mathcal{M}_{pc}^{quad7,P6}$ with respect to the correlation computed with a Monte Carlo sampling E_{mc} over $N_{mc} = 10^5$.

coefficients. These metrics are computed either over an independent sample, thus requiring additional direct model evaluations than those necessary to

the construction of the surrogate, or within the evaluations used to build the surrogate via the leave-one-out error metric.

On the rectangular channel with uniform slope (backwater curves), it was shown that the statistical moments for the water level are efficiently estimated with a reduced model built from a small number of the direct model evaluations. On the Garonne River case, where the (steady) flow is governed by the SWE that represent a system of hyperbolic partial differential equations. The dimension of the stochastic space is reduced with a LAR method consisting on a sparse evaluation of the coefficients. The Sobol' indices indicate that, at Marmande, the water level statistics are driven by the upstream discharge and the friction coefficient of the section where Marmande is located (K_{s_3}). A surrogate model of maximum polynomial order equal to 6 and parameterized with respect to Q and K_{s_3} only is thus formulated using the quadrature method based on 49 direct model evaluations. While a gPC-expansion of maximum polynomial order 6 suffices to properly estimate the first statistical moments, especially the mean, the variance and the covariance matrix that are mandatory in the context of DA, the estimation of the PDF requires a higher polynomial order in areas where the flow dynamics is complex.

In conclusion, the PC-surrogate with a quadrature approach provides a satisfying representation of the hydraulic flow and its statistics with a reduced computational cost related to a limited number of the direct model evaluations (about 50). The surrogate model can thus be used in place of the forward model for uncertainty propagation. In the context of UQ, the PDF, its main statistical moments and standard indices for sensitivity analysis can be retrieved. Turning to DA, the PC-surrogate allows for a cost-effective computation of background error covariances. This approach offers the possibility to provide crucial information for risk assessment in hydraulics and for data driven forecasting system, while meeting with computational constraints in operational framework.

The perspective for this work is to extend the PC-strategy to unsteady flows and identify how often the surrogate model should be re-computed according to the time scales of the flow dynamics. This is a crucial step for the complementary use of surrogate model within DA algorithms as the main objective is real-time flood forecasting. To begin with, the classification between the uncertainty sources (provided by sensitivity analysis for instance) can change in time and the surrogate model should eventually be formulated with respect to different variables over time. When the uncertain input variables include a time-varying variable (for instance the upstream discharge), its representations as a random variable may lead to high-dimensional problems and should thus be reduced for instance with a Karhunen-Loève approach that consists in separating the random process from the time (or space) dimension. When the observation contains significant information, the PDF established from the previous assimilation cycle is no longer representative of the current PDF for the errors. The surrogate should be improved (in this case, the order of the PC-expansion is increased). In the framework of EnKF, the errors are supposed to follow gaussian PDFs, and only the first and second moments

of the PDF should be updated along the assimilation cycles. The validity of the previously-established surrogate is then linked to the flow dynamics and the temporal evolution of the errors in the model. To some extent, it may be necessary to adapt the polynomial basis along the assimilation cycles.

An additional perspective lies in the assimilation of spatially distributed observations such as remote sensing data, for instance from the upcoming SWOT mission that will provide a high resolution and global coverage of water elevation adapted to the observation of rivers wider than 250 m. Water elevation maps will be available every 4 to 5 days at mid-latitude. The assimilation of these data in complement with in situ data seems very promising for model calibration and improvement of prediction capability. As the data will be spatially distributed and the computation of the Kalman gain for the EnKF requires the estimation of the parameter/model state error statistics, the surrogate model should thus be formulated at different locations of the hydraulic network, where the nature of the uncertainties may differ. Hypothesis on the error distributions should be made eventually dissociating the random process issued from the spatial issue as previously mentioned regarding time dimension. Finally, advanced strategy should be implemented when the dimension of the uncertain space increases. In hydraulics, the bathymetry is a major source of uncertainty and the spatial field should be corrected. PC-expansion methods are limited by the quadrature rule that requires too many direct model evaluations for its crude implementation. Several studies have investigated the use of gPC for high-dimensional problem; the merits of such solutions should be demonstrated in the field of open-channel flows.

Acknowledgements The financial support provided by CNES and EDF R&D is greatly appreciated. The authors acknowledge Michael Baudin, Anne Dutfoy, Anne-Laure Popelin, Géraud Blatman and Bertrand Iooss from MRI (EDF R&D) for helpful discussions on uncertainty quantification and support on OpenTURNS.

References

1. Bates, P., Horrit, M., Fewtrell, T.: A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. Journal of Hydrology **387**, 33–45 (2010)
2. Berveiller, M.: Eléments finis stochastiques : approches intrusive et non intrusive pour des analyses de fiabilité. Ph.D. thesis, Université Blaise Pascal, Clermont-Ferrand (2005)
3. Besnard, A., Goutal, N.: Comparaison de modèles 1D à casiers et 2D pour la modélisation hydraulique d'une plaine d'inondation–Cas de la Garonne entre Tonneins et La Réole. La Houille Blanche **3**, 42–47 (2011)
4. Blatman, G.: Adaptative sparse Polynomial Chaos expansions for uncertainty propagation and sensitivity analysis. Ph.D. thesis, Université Blaise Pascal, Clermont-Ferrand (2009)
5. Blatman, G., Sudret, B.: Adaptative sparse polynomial chaos expansion based on Least Angle Regression. Journal of Computational Physics **230**(6), 2345–2367 (2011)
6. Bozzi, S., Passoni, G., Bernardara, P., Goutal, N., Arnaud, A.: Roughness and Discharge Uncertainty in 1D Water Level Calculations. Environmental Modeling & Assessment **4** (2014). DOI 10.1007/s10666-014-9430-6

7. Camacho, R., Martin, J., Diaz-Ramirez, J., McAnally, W., Rodriguez, H., Suscy, P., Zhang, S.: Uncertainty analysis of estuarine hydrodynamic models: an evaluation of input data uncertainty in the weeks bay estuary, alabama. Applied Ocean Research **47**, 138–153 (2014)
8. Camacho, R., Martin, J., McAnally, W., Diaz-Ramirez, J., Rodriguez, H., Suscy, P., Zhang, S.: A comparison of bayesian methods for uncertainty analysis in hydraulic and hydrodynamic modelling. Journal of the American Water Resources Association **51**(5), 1372–1393 (2015)
9. Deman, G., Konakli, K., Sudret, B., Kerrou, J., Perrochet, P., Benabderrahmane, H.: Using sparse polynomial chaos expansions for the global sensitivity analysis of groundwater lifetime expectancy in a multi-layered hydrogeological model. Reliability Engineering and System Safety **147**, 156–169 (2015)
10. Dubreuil, S., Berveiller, M., Petitjean, F., Salan, M.: Construction of bootstrap confidence intervals on sensitivity indices computed by polynomial chaos expansion. Reliability Engineering and System Safety **121**, 263 – 275 (2014). DOI http://dx.doi.org/10.1016/j.ress.2013.09.011
11. Dutka-Malen, I., Lebrun, R., Saassouh, B., Sudret, B.: Implementation of a polynomial chaos toolbox in openturns with test-case application. In: Conference: Proc. 10th Int. Conf. Struct. Safety and Reliability (ICOSSAR'2009), Osaka, Japan (2009)
12. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. The Annals of Statistics **32**(2), 407–499 (2004)
13. Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research **99**(C5), 10,143–10,162 (1994). DOI 10.1029/94JC00572
14. Goutal, N.: Développements autour de la simulation des écoulements à surface libre en rivière. Equations de Saint-Venant - Couplage de modèles - Incertitudes - Application aux ondes de submersion. Habilitation à diriger des recherches (HDR), Université Paris Est (2014)
15. Goutal, N., Maurel, F.: A finite volume solver for 1D shallow-water equations applied to an actual river. International Journal for Numerical Methods in Fluids **38**(1), 1–19 (2002)
16. Habert, J., Ricci, S., LePape, E., Thual, O., Piacentini, A., Goutal, N., Jonville, G., Rochoux, M.: Reduction of the uncertainties in the water level-discharge relation of a 1D hydraulic model in the context of operational flood forecasting. Journal of Hydrology **532**, 52–64 (2016). DOI 10.1016/j.jhydrol.2015.11.023
17. Horritt, M., Bates, P.: Evaluation of 1d and 2d numerical models for predicting river flood inundation. Journal of Hydrology **268**, 87–99 (2002)
18. Hosder S. Perez, R., Walters, R.: A non-intrusive polynomial chaos method for uncertainty propagation in cfd simulations. In: 48th AIAA Aerospace Sciences Meeting and Exhibit, AIAA-2010-0129. The American Institute of Aeronautics and Astronautics, Inc (2006)
19. Le Maitre, O., Knio, O.: Spectral Methods for Uncertainty Quantification. Springer (2010)
20. Li, J., Xiu, D.: On numerical properties of the ensemble Kalman filter for data assimilation. Comput. Methods Appl. Mech. Engrg. **197**, 3574–3583 (2008)
21. Li, J., Xiu, D.: A generalized polynomial chaos based ensemble kalman filter with high accuracy. Journal of Computational Physics **228**(15), 5454 – 5469 (2009). DOI http://dx.doi.org/10.1016/j.jcp.2009.04.029
22. Liang, G., Kwok Fai, C., Kobayashi, M.H.: Stochastic solution for uncertainty propagation in nonlinear shallow-water equations. Journal of Hydraulic Engineering **134**(12), 1732–1743 (2008)
23. Matheron, G.: Traité de géostatistique appliqué, II: Le Krigeage, brgm edn. Mémoires du Bureau de Recherches Géologiques et Minières, Paris (1963)
24. Migliorati, G., Nobile, F., Von Schwerin, E., Tempone, R.: Approximation of quantities of interest in stochastic PDEs by the random Discret L2 Projection on polynomial spaces. SIAM J. Sci Comput. **35**(3), A1440–A1460 (2013)
25. Moradkhani, H., Sorooshian, S., Gupta, H.V., Houser, P.R.: Dual state-parameter estimation of hydrological models using ensemble Kalman filter. Advances in Water Resources **28**, 135–147 (2005). DOI 10.1016/j.advwatres.2004.09.002

26. Resmini, A., Peter, J., Lucor, D.: Sparse grids-based stochastic approximations with applications to aerodynamics sensitivity analysis. International Journal for Numerical Methods in Engineering **106**(1), 32–57 (2016). DOI 10.1002/nme.5005. Nme.5005
27. Rochoux, M.: Vers une meilleure prévision de la propagation d'incendies de forêt : évaluation de modèles et assimilation de données. Ph.D. thesis, Ecole Centrale Paris (2014)
28. Rochoux, M.C., Ricci, S., Lucor D.and Cuenot, B., Trouvé, A.: Towards predictive data-driven simulations of wildfire spread - Part 1: Reduced-cost Ensemble Kalman Filter based on a Polynomial Chaos surrogate model for parameter estimation. Nat. Hazards and Earth Syst. Sci. **14**(11), 2951–2973 (2014)
29. Spanos, P., Ghanem, R.: Stochastic Finite Elements: A Spectral Approach. Springer (1991)
30. Sudret, B.: Global sensitivity analysis using polynomial chaos expansions. Reliab Eng Syst Safety **93**(7), 964–979 (2008)
31. Thual, O.: Hydrodynamique de l'environnement. Ecole polytechnique (2010)
32. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. Roy. Statist. Soc. Ser. B. **58**, 267 – 288 (1996)
33. Van Langenhove, J., Lucor, D., Belme, A.: Robust uncertainy quantification using pre-oconditioned least-squares polynomial with l1-regularization. Int. Journal for UQ **6**, 57–77 (2016)
34. Ven, T.C.: Open-channel hydraulics. Mc Graw Hill (1959)
35. Wiener, N.: The homogeneous chaos. Am.J.Math **60**(4), 897–936 (1938)
36. Winokur, G.: Adaptive sparse grid approaches to polynomial chaos expansions for uncertainty quantification. Ph.D. thesis, Duke University, Department of Mechanical Engineering & Materials Science (2015)
37. Xiang, S.: Asymptotics on laguerre or hermite polynomial expansions and their applications in gauss quadrature. J.Math. Anal . Appl. **393**(2), 434–444 (2012)
38. Xiu, D.: Numerical Methods for Stochastic Computations: A Spectral Method Approach. Princeton University Press (2010)
39. Xiu, D., Karniadakis, G.E.: The wiener–askey polynomial chaos for stochastic differential equations. SIAM Journal on Scientific Computing **24**(2), 619–644 (2002). DOI 10.1137/S1064827501387826
40. Xiu, D., Karniadakis, G.E.: Modeling uncertainty in flow simulations via generalized polynomial chaos. Journal of Computational Physics **187**(1), 137–167 (2003)
41. Zhou, J., Causon, D., Mingham, C., Ingram, D.: Numerical Prediction of Dam Break Flows in General Geometries with Complex Bed Topography. Journal of Hydraulic Engireering **130**(4), 332–340 (2004)
42. Zokagoa, J.M., Soulaïmani, A.: A POD-based reduced order model for free surface shallow water flows over real bathymetries for Monte-Carlo-type applications. Comput. Methods Appl. Mech. Engrg. **221-222**, 1–23 (2012)

9.1.3 Discussion

On propose dans cette partie une discussion plus approfondie avec une analyse plus conséquente des résultats obtenus dans le papier.

Sensibilité à la résolution numérique

Afin de construire une représentation en polynômes du chaos, on a crée un ensemble de simulations, selon une description de la loi de probabilité, la hauteur d'eau à l'aide du modèle MASCARET. Ce modèle étant de nature déterministe, est utilisé comme une "*boite noire*" afin de mener une étude probabiliste. Dans l'article, on a regardé la sensibilité de la construction polynomiale par rapport au degré et à la taille de l'ensemble. Dans ce paragraphe, on propose d'étudier la sensibilité de la représentation par rapport à des paramètres intrinsèques au modèle numérique, notamment le pas d'espace utilisé. Pour cela, on utilise le modèle simplifié du canal rectangulaire utilisé dans l'article qui simule des courbes de remous. Sur la Fig.6 de l'article, on remarque que l'augmentation du degré polynomial, particulièrement pour les degré 8,9 et 10, dégrade la représentation polynomial pour la zone physique ou se fait le passage entre le régime uniforme et le régime de remous. Vu que le calcul des coefficients polynomiaux est fait à partir d'un ensemble de simulations Monte-Carlo de taille 100 000 membres, cet effet ne peut être expliqué par une erreur d'échantillonnage. On émet alors l'hypothèse que la précision de la représentation polynomiale est contrôlée par la précision de l'approximation numérique. On propose alors de raffiner le maillage. Les calculs présentés dans l'article on été fait à l'aide d'un pas d'espace égal à $100m$, on regarde ici l'approximation polynomiale à l'aide d'un maillage construit avec un pas d'espace égal à $10m$. On calcule l'erreur LH, telle que décrite par l'équation (32) de l'article, le long du canal, à l'aide de 100 000 simulations Monte-Carlo pour la simulation d'un polynôme de degré 10. On trace les deux erreurs en fonction de l'abscisse curviline sur la Fig.9.1. On remarque que le raffinement du maillage a permis de réduire l'erreur de la représentation polynomiale. L'erreur estimée, pour le cas où le maillage est raffiné, dépasse l'équivalent de l'erreur machine ce qui revient à dire que la représentation polynomiale est exacte en ces points.

Vers une réduction des appels au modèle

La figure (Fig. 4) du papier, qui exprime le spectre des coefficients polynomiaux dans l'espace des variables incertaines, montre que le modèle est additif. C'est à dire que la variance du modèle est, majoritairement, expliquée par les termes du premier ordre. Les termes représentants les interactions entre les variables aléatoires n'ont pas beaucoup d'influence sur la variance du modèle. Un autre moyen de voir cet effet, est de regarder les différence entre les estimations des indices de Sobol' et les indices totaux. La figure (Fig. 9.2) représente les indices de Sobol' et les indices totaux pour le même cas. On remarque que pour chaque variable aléatoire, la différence entre les deux indices est négligeable. Ce qui confirme que le modèle est additif. En tenant

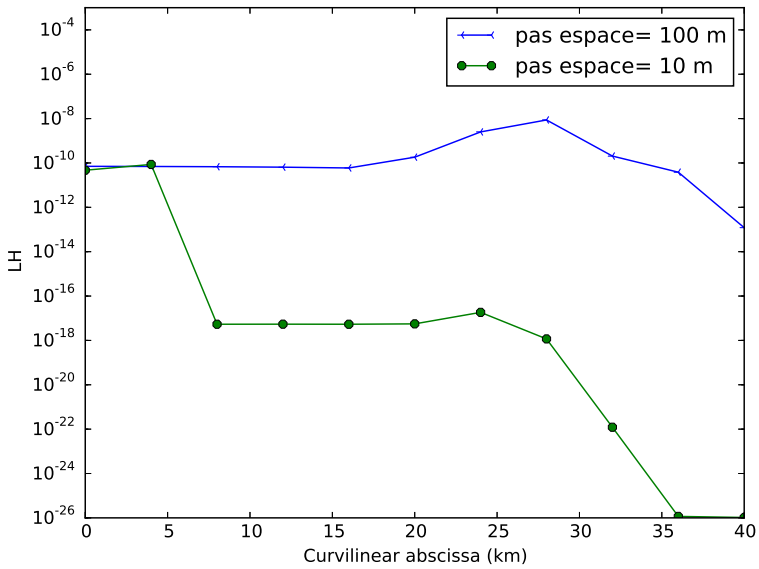


FIGURE 9.1 – *Sensibilité de l'approximation polynomial par rapport à discrétisation numérique.*

compte de cette information, il n'est donc pas nécessaire d'utiliser une représentation polynomiale multivariée afin d'avoir une bonne estimation de la variance. On propose dans cette partie d'utiliser deux représentations polynomiales univariées. Les deux représentations sont estimées à l'aide d'une méthode de quadrature. La représentation univariée permettra de relâcher le fléau de dimensionnalité (c.f. la partie 6.5.1).

On réecrit donc l'équation 14 du papier sous la forme univariée pour chaque variable aléatoire :

$$h_K s = \sum_{i=0}^{P} a_{Ksi}(s)\mathcal{L}_i(\xi_{Ks}), \tag{9.1}$$

$$h_Q = \sum_{i=0}^{P} a_{Q_i}(s)\mathcal{H}_i(\xi_Q) \tag{9.2}$$

où : $\mathcal{L}$ représentent les polynômes de Legendre, et $\mathcal{H}$ sont les polynômes d'Hermite. On remarque que dans cette expression, la somme comporte exactement $P+1$ termes, soit $2(P+1)$ termes à estimer à la place de $\frac{(P+2)!}{P!2!}$. La figure 9.3 montre les modes spectraux des deux décompositions pour $P = 7$. Les modes correspondant à la représentation polynomiale de la hauteur d'eau en fonction du débit sont nuls à partir du coefficient a_5, la part de la variance de la hauteur d'eau expliquée par le débit est totalement expliquée par 4 modes (en enlevant le mode a_0). Les modes spectraux correspondants à la variation du coefficient de frottement montrent la dépendance de ce dernier par rapport à l'écoulement (dans le cas présent la courbe de remous). On remarque notamment que la première partie du coefficient (de l'abscisse 0 à $23km$) est

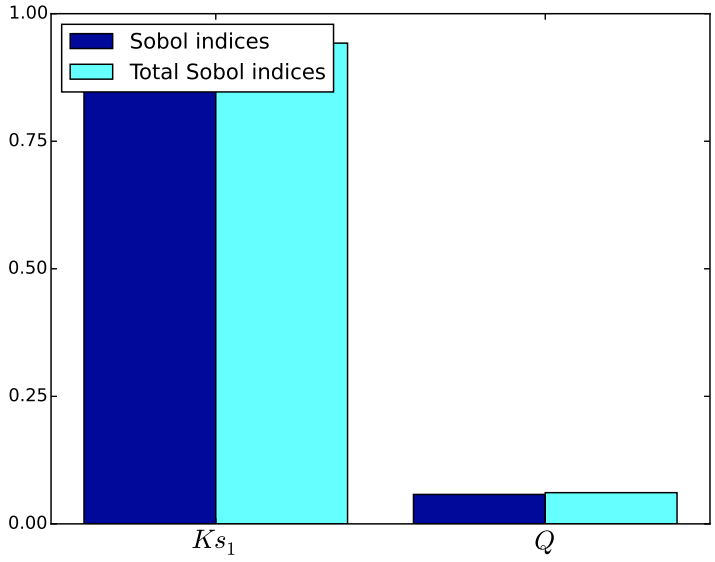


FIGURE 9.2 – *Indices de Sobol', et indices totaux pour le cas de l'équations de Manning.*

constante. De plus, on remarque que la valeur du coefficient a_7 est inférieure à 10^{-5} ce qui correspond à ce que l'on observe sur la figure 4 du papier. Afin d'obtenir les deux représentations polynomiales on a utilisé une méthode de quadrature en utilisant une règle de Gauss à l'aide de 9 nœuds. Ce qui correspond à 18 simulations de Mascaret. La figure 9.4 montre l'estimation de la variance de la hauteur d'eau totale en sommant la variance obtenue des deux représentations univariées. On montre que l'on a une bonne estimation de la variance, ce qui montre que la variabilité totale de la hauteur d'eau n'est pas due aux interactions entre le coefficients de frottement et le débit. la figure 9.4[b] montre que l'erreur entre les deux variances (estimés à l'aide de la somme de deux représentations univariées et celle avec la représentation bivariée) est de l'ordre de 10^{-2}. Cette erreur correspond évidemment aux termes croisées qui existent dans la représentation bivariée.

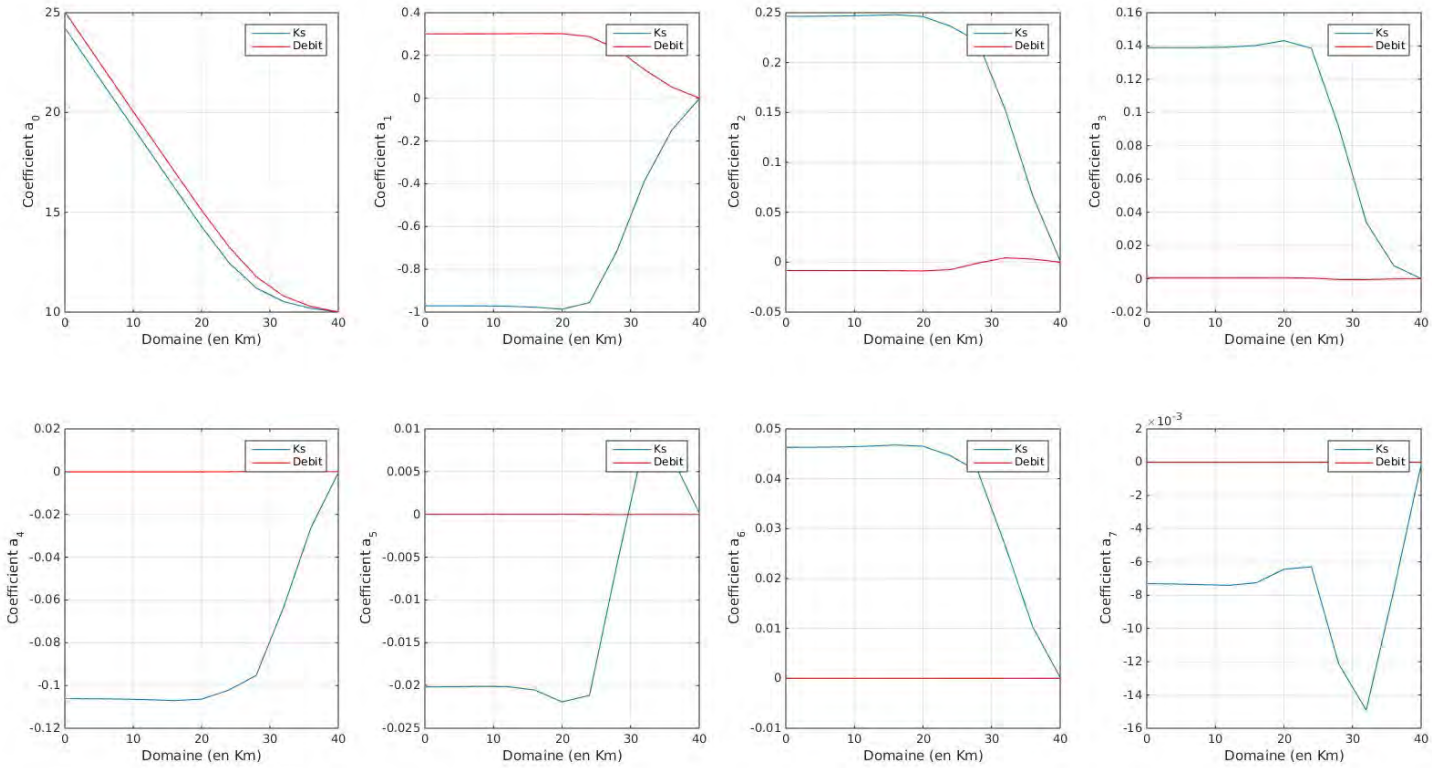
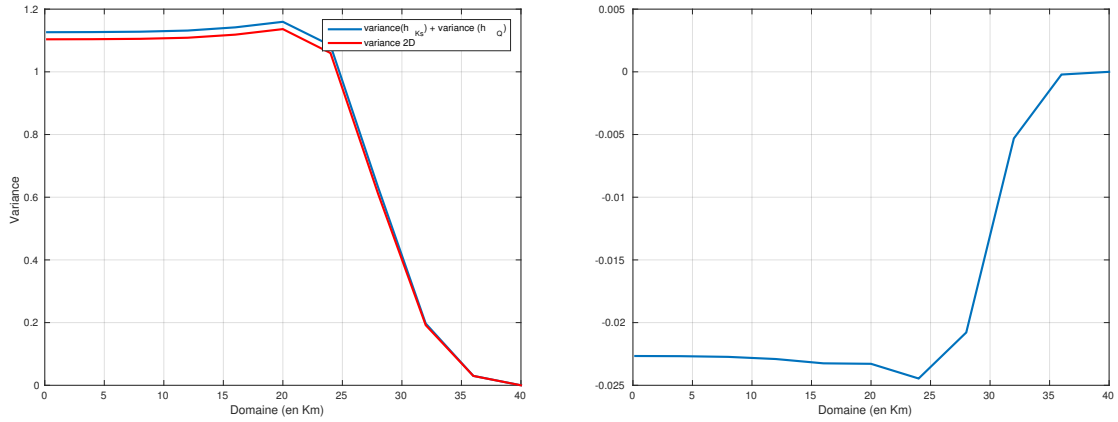


FIGURE 9.3 – *Les modes spectraux correspondants aux deux représentations polynomiales*



[a] [b]

FIGURE 9.4 – *L'estimation de la variance à l'aide des deux représentations polynomiales et la comparaison avec la variance estimée avec une représentation bivariée [a], ainsi que l'estimation de l'erreur entre les deux variances [b]*

coefficient	K_{s_1}	K_{s_2}	K_{s_3}	Débit
Valeur	35	35	35	$4100(m^3/s)$

TABLE 9.1 – *Tableau récapitulatif des valeurs utilisées pour la génération des observations*

9.2 PC-EnKF pour l'assimilation de données SWOT

9.2.1 Introduction

Dans cette partie, on s'intéresse à l'apport des données du type de la future mission satellite SWOT, pour l'estimation des coefficients de frottements (coefficients de Manning) pour des rivières qui ne sont pas jaugées. Dans un algorithme d'assimilation de données, on utilise les hauteurs d'eau, moyennées sur des zones de 10 km afin de réduire l'erreur des observations, simulées sur la Garonne. On se met dans des conditions de régime permanent, car le temps caractéristique des rivières observés par SWOT est supposé négligeable par rapport aux temps de re-visite du satellite. Afin de corriger les coefficients de Manning, on utilise un algorithme d'assimilation de données PC-EnKF. On utilisera notamment les résultats sur l'estimation des matrices de covariances d'erreur d'ébauche à l'aide de l'approximation en polynôme du chaos (9.1.2). Le modèle hydraulique fournit par EDF R&D (LNHE), simule le cas de la Garonne entre Tonneins et La Réole à l'aide de trois coefficients de frottements différents tel que décrit dans l'article (9.1.2). On simule l'écoulement, décrit par les équations de courbe de remous, à l'aide du code Mascaret. Le but de l'assimilation de données est de corriger les coefficients de frottement en gardant la même modélisation, on ne s'intéresse pas à changer les zones des Ks ou à en construire de nouvelles zones de Ks.
Dans un premier temps, on s'intéresse à comparer deux algorithmes d'assimilation de données ensemblistes. Le premier est l'EnKF classique alors que le second est le PC-EnKF qui se base sur la décomposition en chaos polynomial pour l'estimation des statistiques. Dans un second temps, on étudie la sensibilité du moyennage spatial des observations de type SWOT vis à vis de la correction des paramètres.

9.2.2 Génération des observations de type SWOT

Afin de générer des observations de type SWOT, on se placera dans le cadre des expériences jumelles, c'est à dire que la réalité sera une simulation de Mascaret. Pour la première partie de l'exercice, on simule à l'aide de Mascaret la hauteur d'eau le long de la rivière, en utilisant les valeurs décrites dans le tableau 9.1 pour les différents paramètres. Durant la première partie de cet exercice, l'observation est le résultat direct de la simulation. L'erreur que l'on affectera aux observations ici seront très faibles (de l'ordre 10^{-4}, de manière à ce que a réanalyse s'approche de cette simulation. Dans la deuxième partie, on s'intéresse on moyennage spatiale. Par conséquent, on divise la rivière en sections de $10km$ et la hauteur affectée pour chaque section est la moyenne de la simulation sur la section.

9.2.3 Correction de paramètres à l'aide de PC-EnKF

Comme décrit dans la section 7.5.3, lorsque l'on souhaite corriger les paramètres d'un modèles, les matrices $\mathbf{C}_{xy}$ et $\mathbf{C}_{yy}$ (voir 4.3.3) sont estimées stochastiquement avec une approche EnKF. L'utilisation des polynômes du chaos permet d'estimer les moments statistiques telle que la matrice de covariance en utilisant les coefficients spectraux. L'estimation de la matrice $\mathbf{C}_{yy}$ est donc directe et peu couteûse. Le cadre probabiliste de la décomposition polynomiale (c.f 6.2.1) offre une bonne base pour l'estimation de la matrice $\mathbf{C}_{xy}$. En effet si on suppose que x est une variable aléatoire gaussienne de moyenne μ_x de variance σ_x, on peut écrire la décomposition dans la base polynôme du chaos de la manière suivante :

$$x = \mu_x \mathcal{H}_0(\zeta) + \sigma_x \mathcal{H}_1(\zeta), \tag{9.3}$$

où $\mathcal{H}_0(\zeta)$ et $\mathcal{H}_1(\zeta)$ sont respectivement les polynômes d'Hermite d'ordre 0 et d'ordre 1.
De cette manière, on peut estimer la matrice $\mathbf{C}_{xy}$ à l'aide des coefficients de l'approximation polynomiale. En effet, dans le cadre de l'assimilation de donnée, on suppose que les variables aléatoires sont gaussiennes, donc on écrit $y = \sum_{i=0}^{N_{PC}} a_i \mathcal{H}_i(\zeta)$. Ce qui permet d'écrire :

$$\begin{aligned} \mathbf{C}_{xy} &= \mathrm{E}\big[(\text{x-}\mu_x)(y - \mathbf{E}(y))\big] \\ &= \mathrm{E}\big[(\sigma_x \mathcal{H}_1(\zeta))(\textstyle\sum_{i=1}^{N_{PC}} a_i H_i(\zeta))\big] \\ &= \sigma_x.a_1 \end{aligned} \tag{9.4}$$

9.2.4 Algorithme de correction de paramètres

On présente dans la suite un algorithme pour la correction des paramètres à l'aide de l'approximation polynomiale. On suppose que l'on observe la hauteur d'eau sur tout le long de la rivière (une observation SWOT par exemple) que l'on noter $h^{obs}(s)$et dont la matrice de covariance d'erreur d'observation est supposée connue, notée $\mathcal{O}$. Les paramètres que l'on cherche à estimer dans cet exercice sont les trois coefficients de frottement K_{s_1}, K_{s_2} et K_{s_3} ainsi que le débit Q. Ces quatre paramètres sont considérés comme des variables aléatoire gaussienne de moyenne respectivement $\mu_{K_{s_1}} = 30, \mu_{K_{s_2}} = 30, \mu_{K_{s_3}} = 30$ et $\mu_Q = 4031$, et d'écart-type respectivement $\sigma_{K_{s_1}} = 5, \sigma_{K_{s_2}} = 5, \sigma_{K_{s_3}} = 5$ et $\sigma_Q = 400$. Ce contexte permet d'écrire une approximation polynomiale de la hauteur d'eau (spatialisée) de la façon suivante :

$$h^{PC}(s, K_{s_1}(\zeta), K_{s_2}(\zeta), K_{s_3}(\zeta), Q(\zeta)) = \sum_{(i_1,i_2,i_3,i_Q)\in I_P} a_{(i_1,i_2,i_3,i_Q)}(s) \mathcal{H}^1_{i_1}(\zeta) \mathcal{H}^2_{i_2}(\zeta) \mathcal{H}^3_{i_3}(\zeta) \mathcal{H}^Q_{i_Q}(\zeta), \tag{9.5}$$

où $I_P = \{((i_1, i_2, i_3, i_Q) \in \mathbf{N}^4, i_1 + i_2 + i_3 + i_Q \leq P\}$, H sont les polynômes d'Hermite vu que toutes les variables sont supposées gaussiennes.

Algorithm 5 PC-EnKF pour la correction des paramètres

1: Décomposer la hauteur d'eau en chaos polynomial selon l'équation 9.5.
2: Calculer la matrice $\mathbf{C}_{yy}$ à l'aide des coefficients polynomiaux.
3: Calculer la matrice
$$\begin{aligned}\mathbf{C}_{yx} &= (\mathbf{C}_{hK_{s_1}}, \mathbf{C}_{hK_{s_2}}, \mathbf{C}_{hK_{s_3}}, \mathbf{C}_{hQ}) \\ &= (\sigma_{Ks_1}.a_{1,0,0,0}(s), \sigma_{Ks_2}.a_{0,1,0,0}(s), \sigma_{Ks_3}.a_{0,0,1,0}(s), \sigma_{KQ}.a_{0,0,0,1}(s).\end{aligned}$$
4: Calculer la matrice de Gain $\mathbf{K} = \mathbf{C}_{xy}(\mathbf{C}_{yy} + \mathcal{O})^{-1}$.
5: Étape d'analyse : $(K^a_{s_1}, K^a_{s_2}, K^a_{s_3}, Q^a) = (K^b_{s_1}, K^b_{s_2}, K^b_{s_3}, Q^b) + \mathbf{K}(h^{obs}(s) - a_{0,0,0,0}(s))$.

9.2.5 Comparaison entre l'EnKF et le PC-EnKF pour la correction des paramètres de frottement et de débit

On compare ici les deux algorithmes d'assimilation de données : EnKF d'une part et le PC-EnKF d'autre part. D'après la figure (Fig. 10) du papier 9.1.2, les indices de Sobol' montrent qu'il pertinent de considérer ces 4 paramètres vu que l'observation de la hauteur d'eau est spatialisée. En effet, l'analyse de sensibilité montre que selon l'endroit où on se place, la hauteur d'eau n'est pas sensible au même paramètres, c'est la raison pour laquelle le vecteur de contrôle est composé des trois coefficients de frottement et du débit. On suppose que les 3 coefficients de frottements sont des variables aléatoires gaussiennes de moyenne 30 et d'écart-type 5, et que le débit suit une loi normale de moyenne 4031 d'écart-type 400.
On utilise l'algorithme 5 décrit dans le paragraphe précédent pour corriger les paramètres. La première étape, étant de construire un modèle polynomial, on fixe le degré polynomial à 6 et on utilise 7 points de quadrature par direction suivant les résultats de l'article 9.1.2. Le coût de la construction du polynôme s'élève donc à 2401 simulations.

On compare les résultats de la correction des paramètres avec l'algorithme proposé ainsi que l'EnKF classique calculé à l'aide de 2500 membres (pour être à coût de calcul égal). La figure 9.5 montre les résultats de cet exercice. L'analyse représente la simulation du cas hydraulique après avoir corrigé les paramètres. On remarque que la modélisation est meilleure lorsque l'on utilise l'algorithme PC-EnKF (courbe en rouge), et plus particulièrement à l'amont de La Garonne. En effet, pour l'amont de la rivière, la variance est plus importante que celle à l'aval (c.f. Fig. 9.7 [a]). 2500 membres ne sont pas suffisants pour estimer la variance à l'amont de la rivière, par conséquent, la précision de la correction, par l'EnKF en comparaison avec le PC-EnKF, est mauvaise à l'amont de la rivière(c.f. Fig. 9.7 [b]).
Le point important de l'utilisation de l'approximation polynomiale est l'utilisation des coefficients polynomiaux (spécialement les coefficients du premier ordre polynomial) afin d'estimer les matrices de covariances croisées ($\mathbf{BH}^T$ où $\mathbf{C}_{xy}$) entre la hauteur d'eau et les paramètres à corriger. La figure 9.6 montre l'estimation de ces matrices à l'aide des coefficients polynomiaux et de manière stochastique. On remarque que pour les 4 paramètres, les deux estimations concordent sauf pour la covariance entre le débit et la hauteur d'eau. Cette différence entre les deux estimations n'affecte pas de manière significative la corrections des paramètres et surtout l'analyse faite après la correction

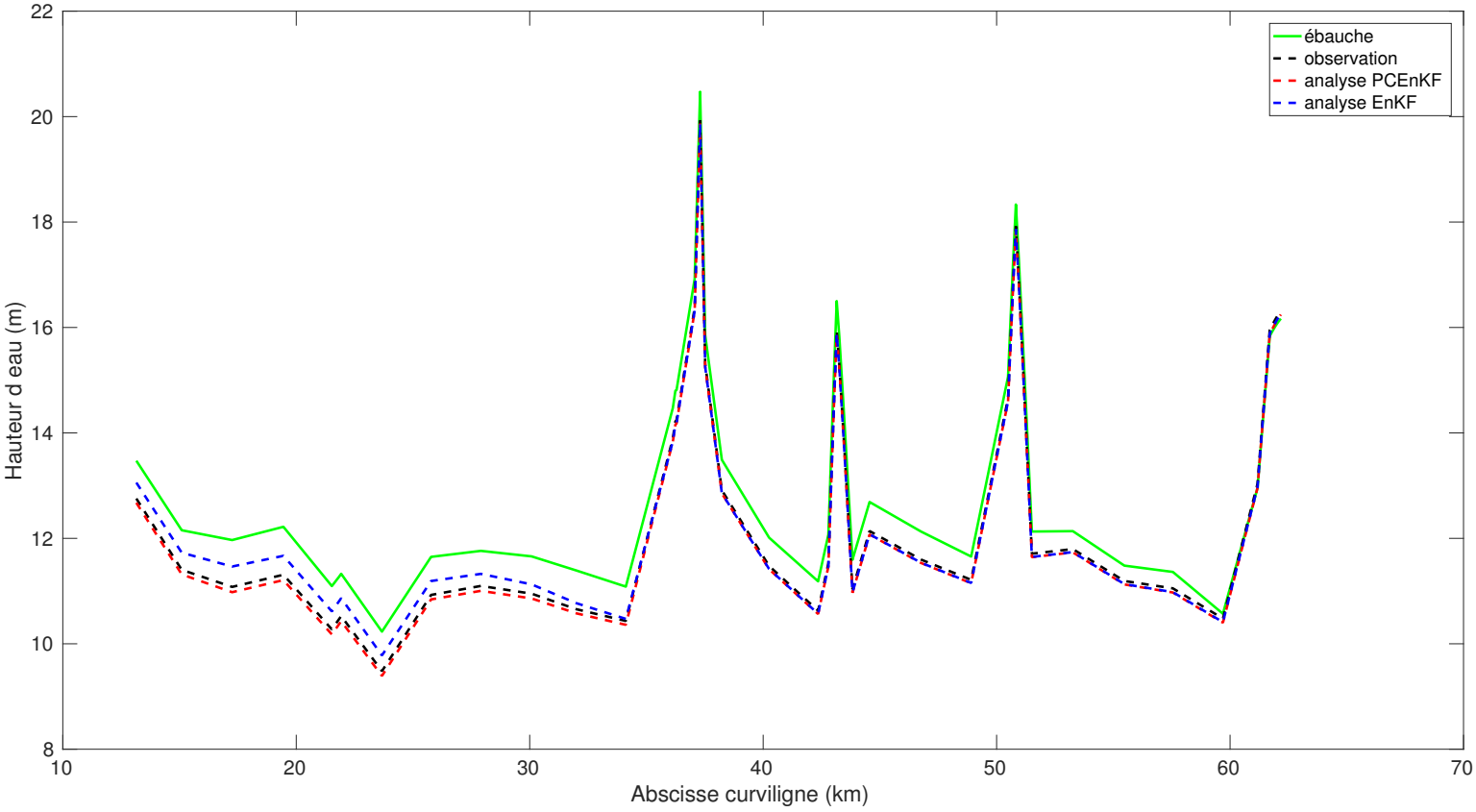


FIGURE 9.5 – *La correction des paramètres de frottements et de débit à l'aide de l'algorithme PCEnKF et l'EnKF*

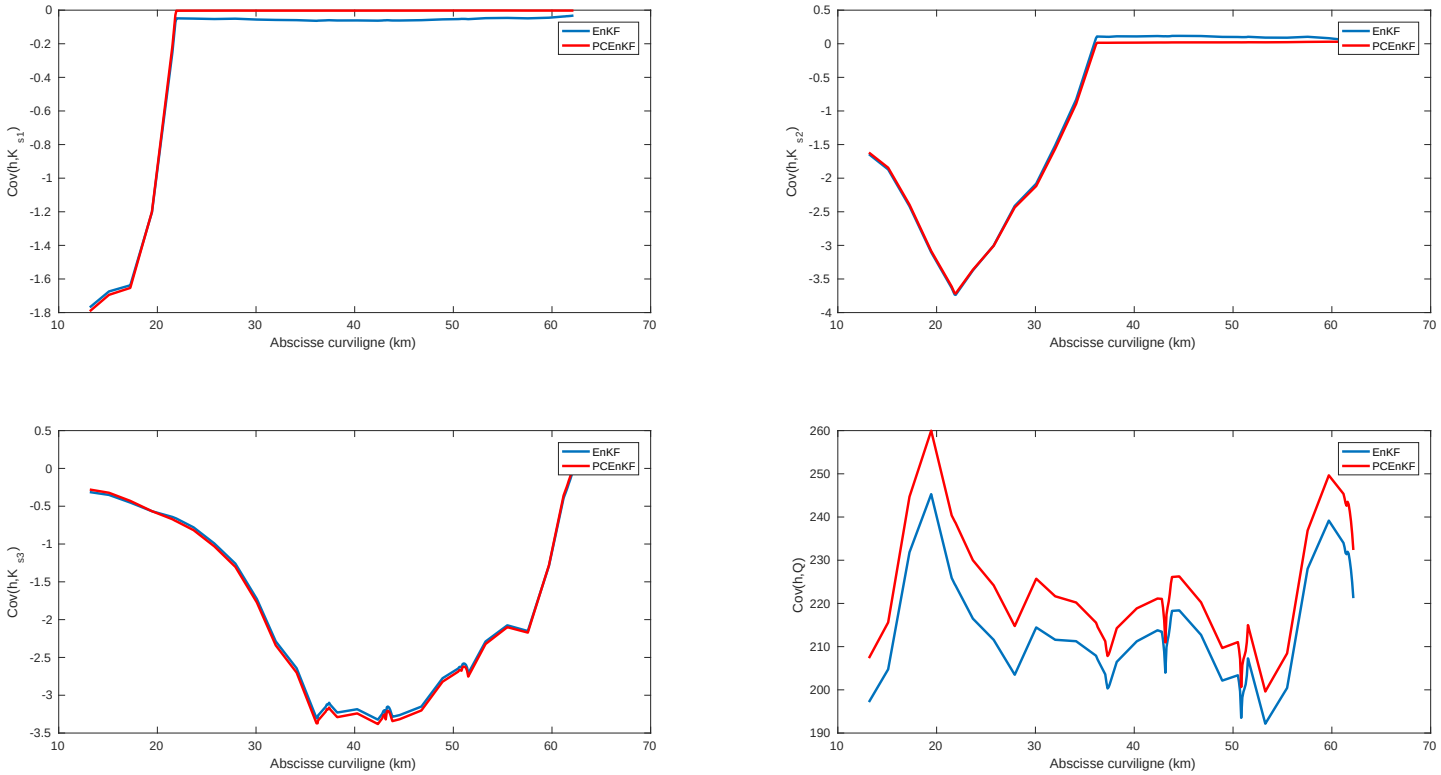
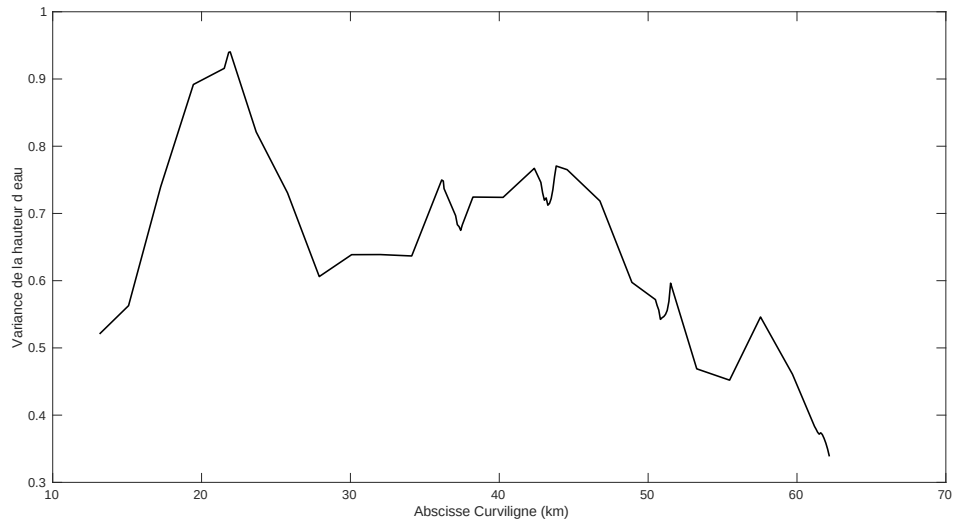
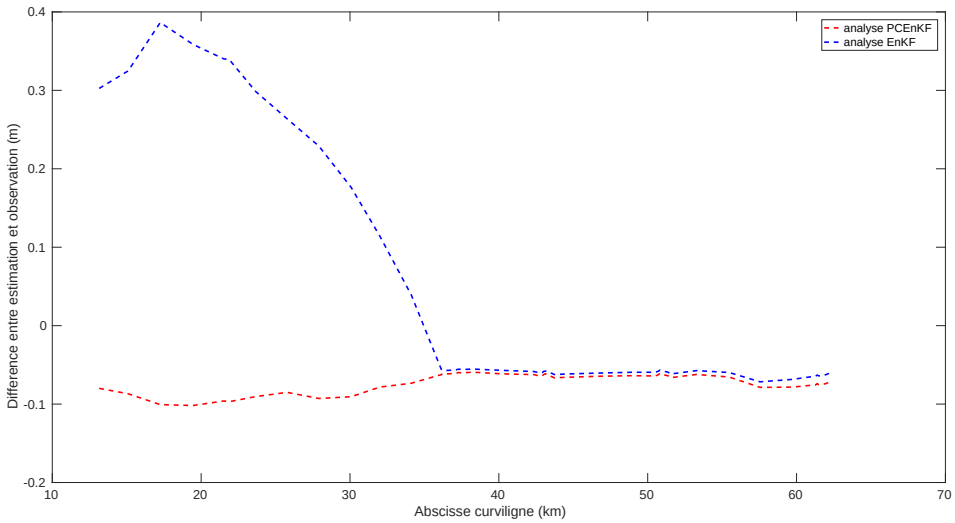


FIGURE 9.6 – *L'estimation de la covariance entre la hauteur d'eau et les coefficients de frottement, et la covariance entre la hauteur d'eau et le débit pour PC-EnKF (rouge) et EnKF(bleu)*



[a]



[b]

FIGURE 9.7 – *L'estimation de la variance sur toute la Garonne [a]. Calcul de l'erreur entre l'analyse et l'observation pour le PC-EnKF et l'EnKF [b]*

car la hauteur d'eau est moins sensible aux variations du débit qu'aux coefficients de frottement (c.f. les indices de Sobol' dans 9.1.2).
Globalement, l'utilisation de l'approche polynôme du chaos permet de réduire le temps de calcul nécessaire, afin d'avoir une bonne estimation des statistiques nécessaires pour un algorithme d'assimilation de données ensembliste (de type EnKF). L'amélioration apportée par l'utilisation de la décomposition polynomiale est d'autant plus significative quand l'estimation d'une statistique, telle que la variance, nécessite un nombre important de simulations. La limite de cette méthode est que dans certain cas, faire 2400 simulations pour un code peut s'avérer coûteux pour construire une approximation polynomiale. Une perspective intéressante pour réduire le coût de cette méthode serait de construire des polynômes univariées afin de réduire encore plus le coût de la méthode, tel que présenté dans le paragraphe 9.1.3 ce qui réduirait le coût de la décomposition polynomiale (28 simulations seulement sont nécessaire dans ce cas). Bien évidement, si la sortie du modèle présente une forte sensibilité aux interactions des paramètres, une méthode astucieuse pour diminuer le coût de calcul serait d'utiliser la méthode "LAR" (c.f. partie 6.5.1).

9.2.6 Impact des caractéristiques des données SWOT sur la réanalyse

On utilise ici la méthode PC-EnKF afin de corriger les paramètres du modèle hydraulique. Dans ce paragraphe, on regarde l'effet du moyennage spatial sur l'estimation des paramètres. On se place dans le cadre des expériences jumelles comme pour le cas précédant, sauf que cette fois l'observation retenue est moyennée au long de la rivière. On compare l'analyse issue de l'observation avant le moyennage, tous les 10 km (c.f. 9.2.2).
La figure 9.8 montre les résultats de cette approche. En trait noir fin, on retrouve l'observation satellitaire moyennée sur un tronçon de 10 km. L'analyse issue de l'observation SWOT se rapproche de l'observation réelle. La correction s'éloigne suffisamment de l'ébauche pour s'approcher de l'observation réelle. La figure 9.9 représente l'erreur sur l'analyse en utilisant les paramètre corrigés à l'aide du PC-EnKF. Elle montre notamment que l'erreur de l'estimation en utilisant des données SWOT est de l'ordre de 6% lorsque l'on utilise des données moyennées sur l'espace.

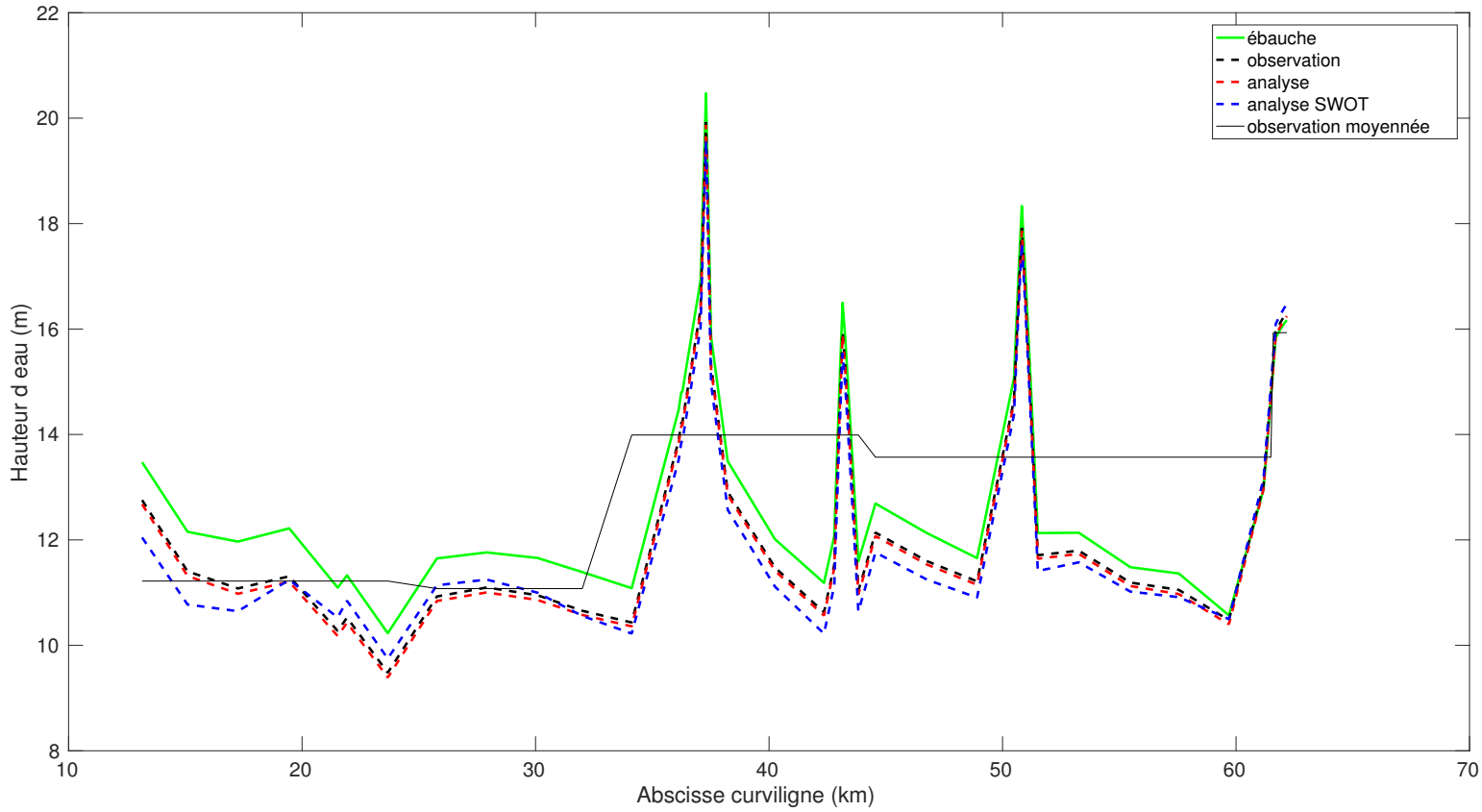


FIGURE 9.8 – *La correction des paramètres de frottements en utilisant des données de type SWOT*

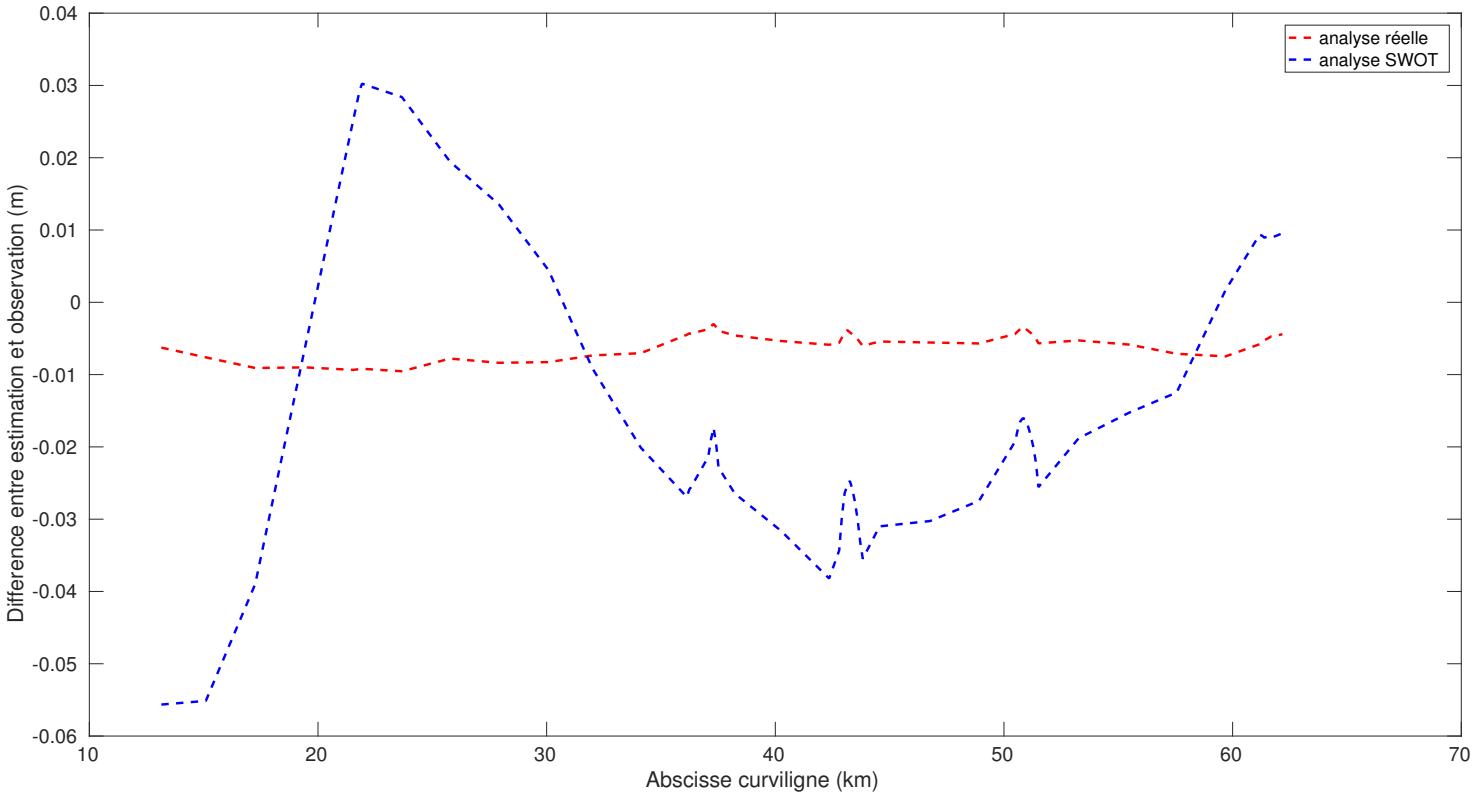


FIGURE 9.9 – *L'erreur relative (à l'observation) due à l'analyse réelle et l'analyse issue des observations moyennées.*

10 PCE dans le cas instationnaire

On s'intéresse ici à la décomposition en polynôme du chaos pour le régime instationnaire. Puisque dans le régime stationnaire le débit est un champ aléatoire, on ne considère dans un premier temps que l'incertitude due aux coefficients de frottement. Puis dans un second temps, on étudie la décomposition de Karhunen-Loève pour la réduction de la dimension du champ incertain.

10.1 Paramètres incertains

10.1.1 Introduction

Dans cette section, on s'intéresse à la correction des coefficients de frottement sur le cas Garonne présenté dans la partie 9.1.2. On utilise l'hydrogramme amont correspondant à La crue de 1981, pour lequel on dispose des mesures à la station de Marmande. L'hydrogramme présente l'évolution horaire, pendant 8 jours, de la hauteur d'eau. Pour des raisons de simplicité , on formule la décomposition polynomiale pour chaque jour, mais uniquement à Marmande où les observations sont disponibles. On étudie notamment la sensibilité de la hauteur d'eau par rapport à la variabilité des trois coefficients de frottement. Puis dans un second temps on utilise l'algorithme PC-EnKF (section 9.2.4).

10.1.2 Description du cadre d'expérience

Contrairement au chapitre précédent où on utilisait des observation de type SWOT, l'observation ici est ponctuelle (à Marmande). D'autre part, on dispose de données d'observations de débit et de la hauteur d'eau pour la crue de 1981, que l'on utilise ici. De plus, on considère que l'incertitude n'est portée que par les coefficients de frottement. Afin de déterminer les coefficients à considérer dans le vecteur de contrôle, on réalise au préalable une analyse de sensibilité.

10.1.3 Analyse de sensibilité

L'analyse de sensibilité dans ce cas là, n'a pas les même buts que dans le cadre classique de la quantification d'incertitudes. Ici, on ne cherche pas seulement à réduire la dimension de l'espace incertain. Mais aussi de déterminer les paramètres influents. En effet, en considérant des paramètres pas (ou peu) influents, comme paramètres d'ébauche pour l'assimilation de donnée, n'améliore pas la prévision du modèle. On se sert de la décomposition polynomiale afin de mener l'analyse de sensibilité (c.f. 6.4). On garde un degré 6 pour l'ordre du polynôme tel que pour le régime stationnaire. La figure 10.1 présente l'évolution temporelle des indices de Sobol' pendant la durée de la crue. On remarque que le seul paramètre auquel la hauteur d'eau à Marmande est sensible le coefficient de frottements K_{s_3}. Ce qui est conforme avec les résultats de la partie stationnaire (c.f section 9.1.2). Par conséquent, on ne gardera que ce coefficient comme paramètre à corriger avec l'assimilation de données.

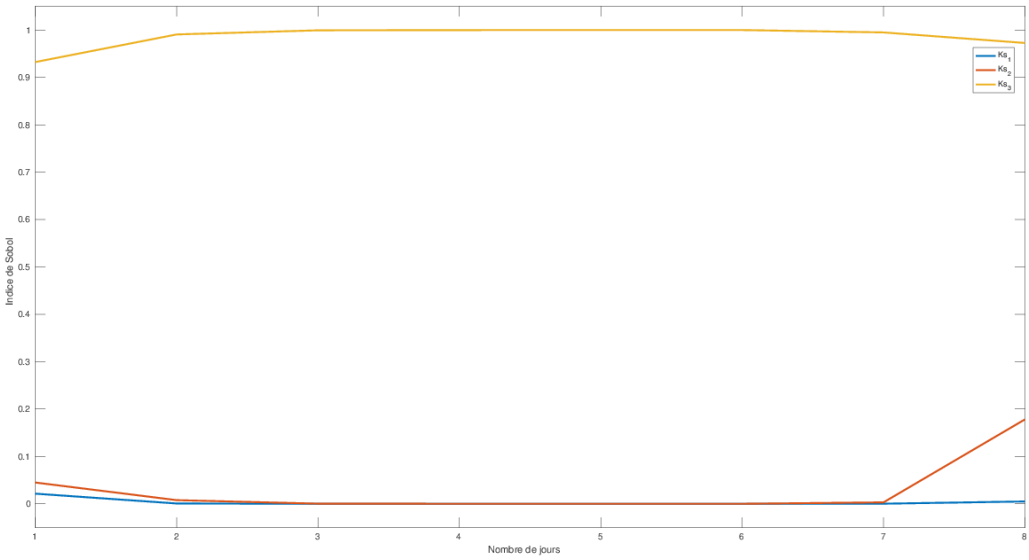


FIGURE 10.1 – *L'évolution temporelle des indices de Sobol*

10.1.4 Décomposition polynomiale

Afin de corriger la valeur du coefficient de frottement, on décompose la hauteur d'eau sous une base de chaos polynomial avec comme variable incertaine que le coefficient de frottement de la zone 3 (Eq. (10.1)). On suppose que le coefficient de frottement suit une loi gaussienne de moyenne $\mu_{K_s} = 30$ et d'écart-type $\sigma_{K_s} = 5$, par conséquent la base polynomiale est une base des polynômes d'Hermite. On garde une méthode de projection basée sur les quadratures de gauss-Hermite pour l'estimation des coefficients polynomiaux. On écrit cette décomposition pour chaque jour de la façon suivante :

$$h_{Marmande}(t) = \sum_{i=0}^{P} \alpha_i(t) H_i(K_s). \tag{10.1}$$

La sensibilité aux degré du polynôme n'est pas guidée par la dimension du problème, car on est en présence d'une seule variable aléatoire incertaine, mais elle est contrainte par la physique du problème. Quand on augmente le degré de troncature, on augmente la taille de l'espace incertain. Par conséquent, comme les variables aléatoires gaussiennes ont un support infini, l'augmentation du degré de troncature induit la prise en compte de solutions qui ne sont pas cohérentes avec la physique. On définit donc le degré de troncature, comme le degré maximal pour lequel tous les points de quadrature correspondants sont en accord avec la physique du problème. Pour le cas présent, le degré polynomial correspondant à la physique est égal à 7.

Quantification de l'incertitude. La figure 10.2 présente l'estimation de la moyenne $\mu_h(t)$ et l'écart-type $\sigma_h(t)$ (que l'on modélise par un domaine d'incertitudes $[\mu_h(t) - \sigma_h(t); \mu_h(t) + \sigma_h(t)]$). On remarque qu'à partir du quatrième jour l'observation dévie de la moyenne simulée à l'aide du modèle. On corrige alors le coefficient du frottement au quatrième jour.
La figure 10.3 présente l'évolution de la variance dans le temps, on remarque que globalement la variance de la hauteur d'eau ne change pas significativement. Cependant, à partir du huitième jour elle commence à augmenter. Physiquement, après le passage d'une crue la description de la physique change, particulièrement pour décrire le frottement de la rivière. D'où l'importance de mettre à jour la valeur du coefficient de frottement.

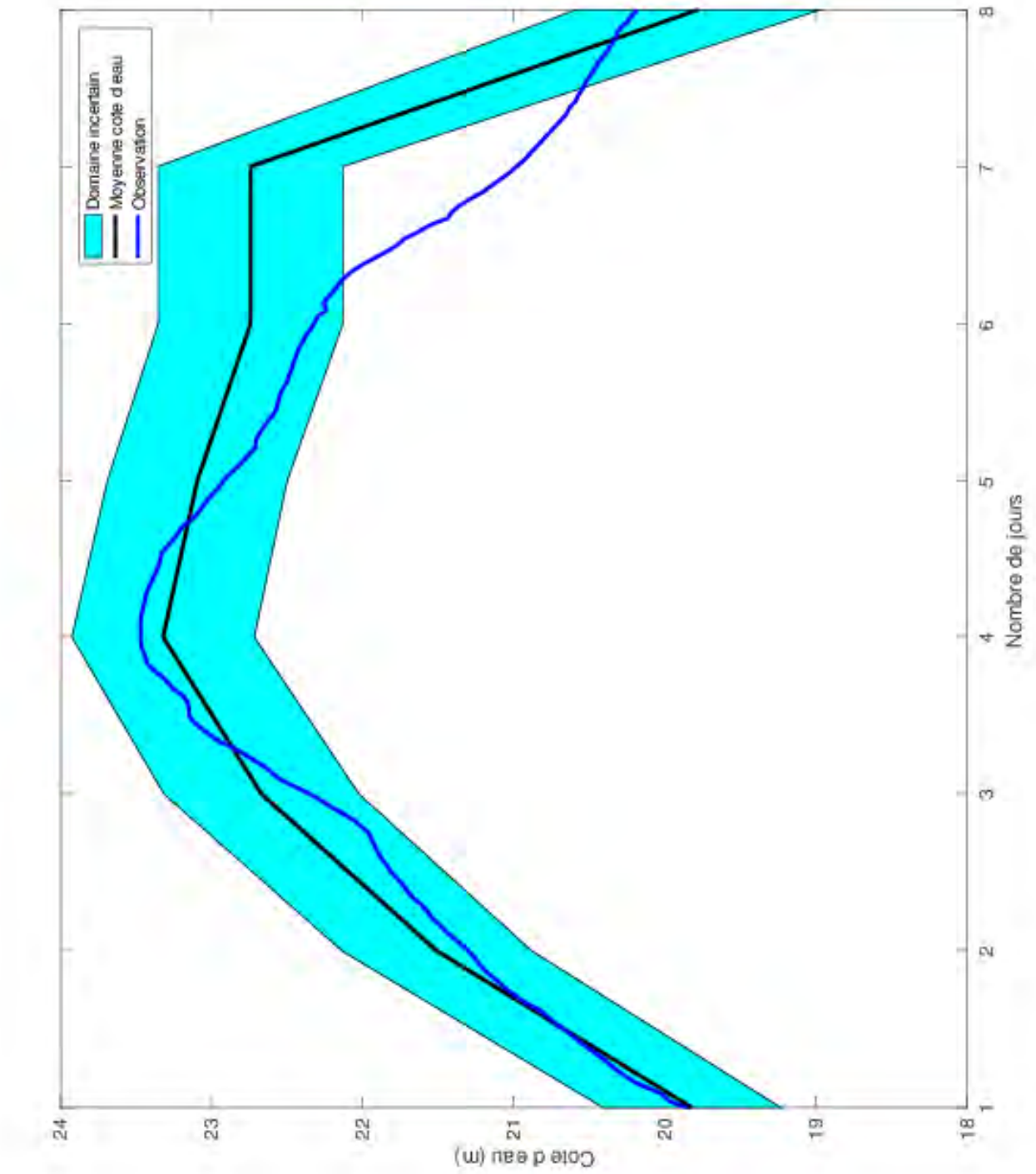


FIGURE 10.2 – *Estimation de la moyenne et l'écart-type de la hauteur d'eau à Marmande et comparaison avec les données d'observation*

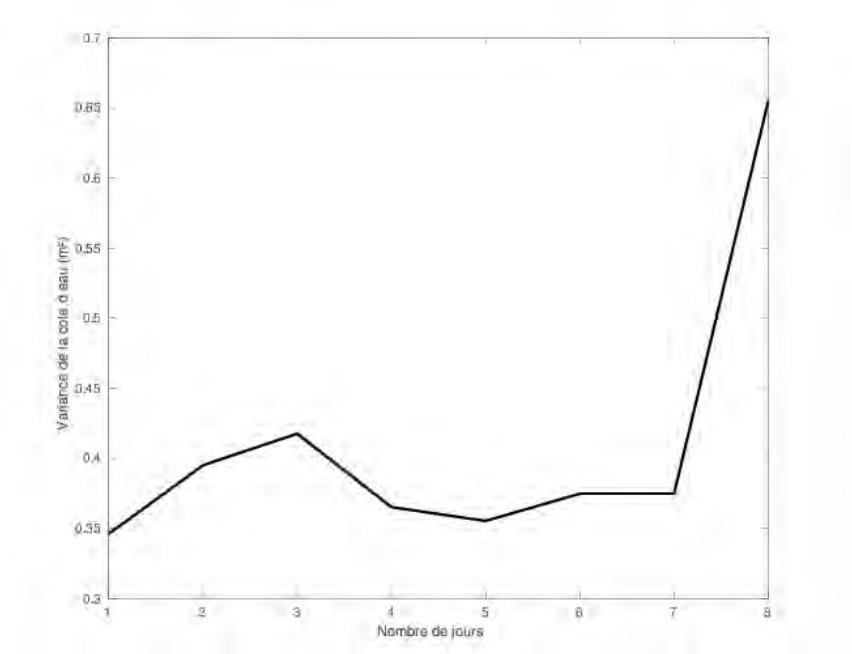


FIGURE 10.3 – *Estimation de la variance à l'aide des coefficients polynomiaux dans le temps*

10.1.5 Vers l'assimilation de données pour la correction du paramètre

On utilise l'algorithme PC-EnKF (c.f. 9.2.4) pour la mise à jour de la valeur du coefficient. Comme indiqué au début, on corrige la valeur du coefficient de frottement au quatrième jour. À l'aide de l'équation d'analyse (Eq. (7.18)), il est possible de retrouver la nouvelle valeur de la variance sur l'estimation du coefficient et d'estimer à l'aide de la décomposition polynomiale la nouvelle variance sur la hauteur d'eau. Puisque l'on corrige que la valeur d'un seul paramètre à l'aide d'une observation localisée, la formule du gain de Kalman Eq. (7.15) couplée avec la décomposition polynomiale Eq. (10.1) se résume à l'équation suivante :

$$\mathbf{K} = \frac{\sigma_{K_s}\alpha_1}{\sum_{i=1}^{P}\alpha_i^2 + \sigma_{obs}} \tag{10.2}$$

où σ_{obs} représente l'écart-type sur l'observation.
La figure 10.4 représente le résultat de la mise à jour de la valeur du coefficient de frottement ainsi que son incertitude. On réécrit la décomposition polynomiale pour chaque jour avec la nouvelle valeur du coefficient de frottement, ainsi que son incertitude. On remarque que l'assimilation de données a permis de réduire l'incertitude autour de la moyenne (le domaine en rose est plus petit que le domaine en bleu). On remarque aussi que l'on a amélioré la prévisibilité du modèle (spécialement autour du

jour 7).

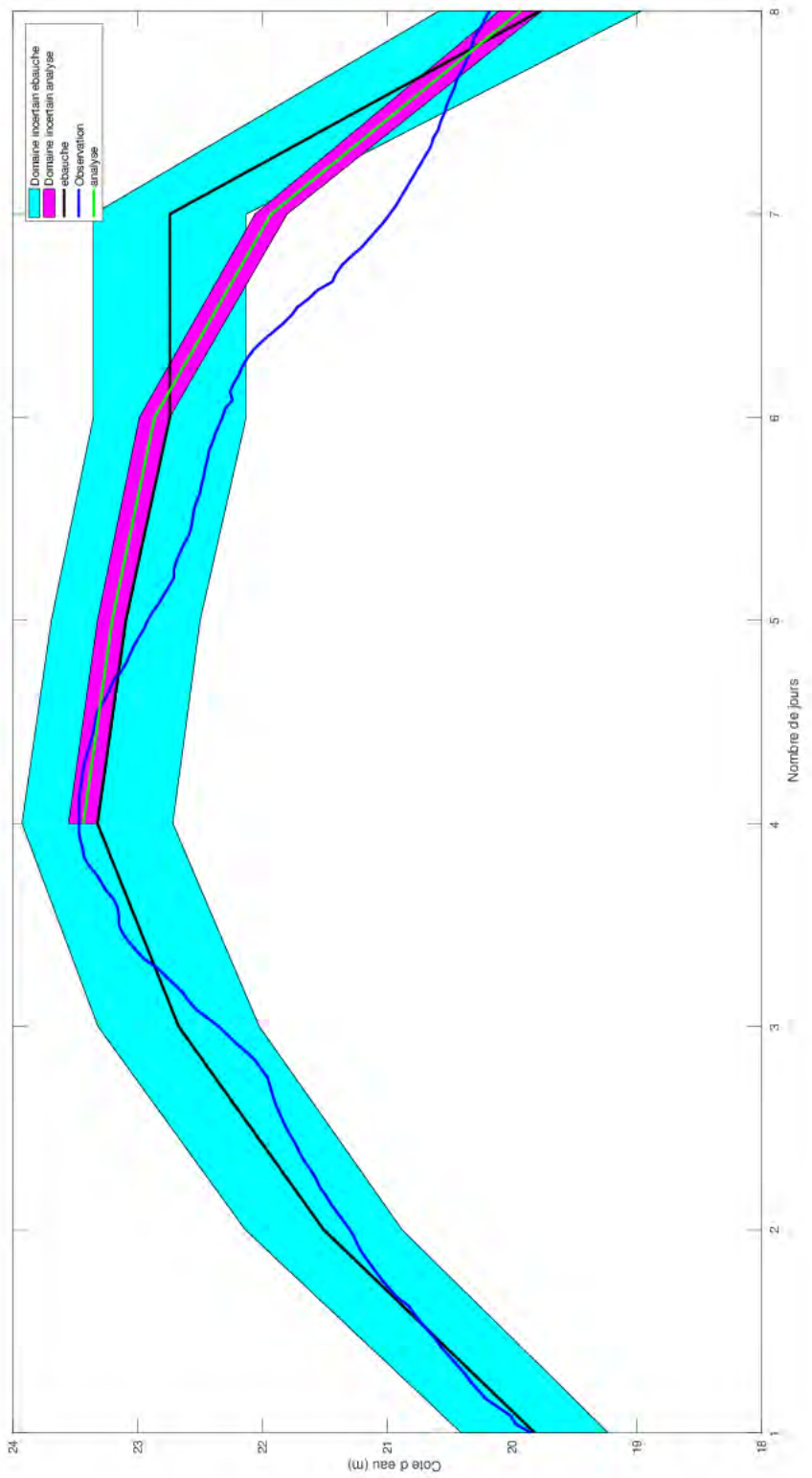


FIGURE 10.4 – *Estimation de l'incertitude de la hauteur d'eau avant et après l'analyse et comparaison avec l'observation*

10.1.6 Prévisibilité du modèle réduit

Dans ce paragraphe, on regarde la prévisibilité du modèle réduit initiale par rapport au modèle réduit établi pour l'étape d'analyse. On s'intéresse notamment aux spectres des coefficients polynomiaux. La figure 10.5 montre le spectre des coefficients pour les jours 4 à 7. La réduction de l'incertitude réduit la valeur des coefficients significativement. Par conséquent, il semble que pour chaque étape d'analyse la mise à jour des coefficients est primordiale. Afin de vérifier cette hypothèse, on regarde l'erreur commise entre l'estimation de la variance à l'aide des coefficients du modèle réduit écrit après l'étape d'analyse et la variance estimé à l'aide d'un échantillonnage Monte-Carlo sur le modèle réduit établit lors de l'étape de propagation. La figure **??** montre que l'estimation de la variance est bonne à l'aide du premier modèle réduit. D'une part Le premier modèle réduit est valable pour un domaine d'incertitude plus grand que celui de l'analyse, d'autre part la correction qu'apporte l'observation ne ramène pas l'analyse en dehors du domaine d'incertitude compris dans l'ébauche. Toutes ces raisons font que les limitations que l'on a décrit dans la partie 7.6.3 ne sont pas valables dans ce cas.

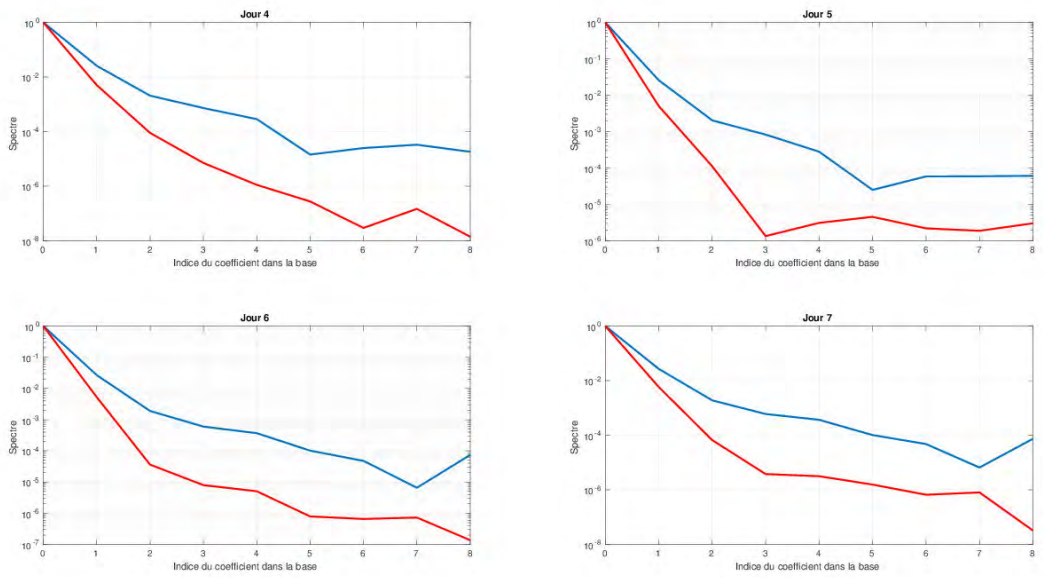


FIGURE 10.5 – *Spectre des coefficients polynomiaux pour l'étape de propagation (en bleu) et l'étape d'analyse (en rouge) pour les jours :* 4, 5, 6 *et* 7

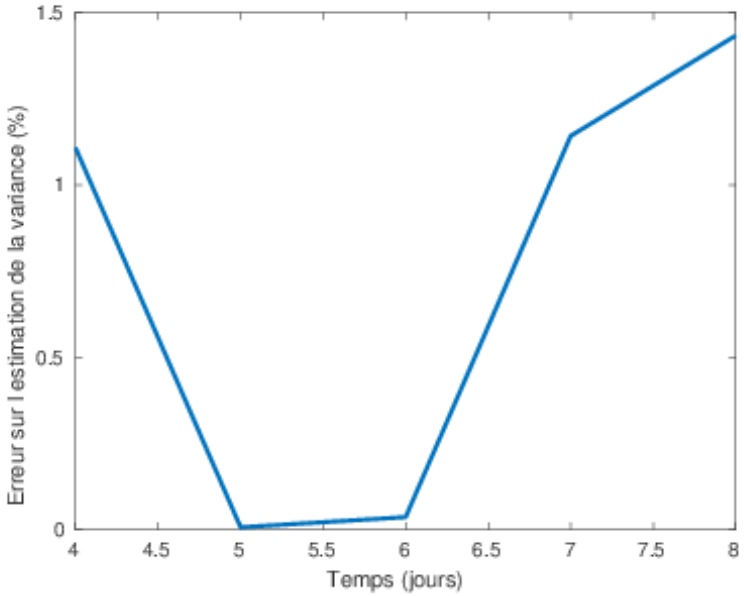


FIGURE 10.6 – *Erreur sur l'estimation de la variance entre le modèle réduit crée après l'analyse et celui avant la propagation*

10.2 Champs incertains

10.2.1 Cadre d'étude

On s'intéresse dans ce cas, à l'étude de la décomposition polynomiale de la hauteur d'eau en fonction d'un débit non-permanent. On se place dans un canal rectangulaire de longueur $20km$ avec un coefficient de frottement uniforme pour tout le canal ($K_s = 35$). Les conditions aux limites sont définies à l'aval par une condition fixe ($h_{out} = 10m$) et à l'amont par un hydrogramme que l'on suppose comme un champ aléatoire ($Q_{in}(t, \theta)$). Les simulations seront menées à l'aide du logiciel Mascaret, et le traitement statistique pour évaluer la décomposition polynomiale sera faite grâce à Openturns. On se placera au milieu du canal, afin d'écrire la hauteur en fonction du temps à l'aide de l'approximation polynôme du chaos.

10.2.2 Représentation stochastique du débit non-permanent

On suppose ici que le débit imposé à l'amont est un champ aléatoire. En effet, le débit est à la fois dépendant du temps (vu que l'on traite un problème instationnaire) et une variable aléatoire. On suppose que l'hydrogramme possède une fonction d'autocorrélation temporelle ($\mathcal{C}(t_1, t_2)$) (que l'on supposera exponentielle pour le cas présent) définie sur la fenêtre temporelle T. On note θ une variable (aléatoire) de l'espace Ω ($\theta \in \Omega$). Étant donné que la représentation stochastique du champ aléatoire est difficile dans ce contexte. On choisit de séparer la variabilité stochastique de la variablité temporelle à l'aide de la décomposition spectrale de Karhunen-Loeve.
Supposons que $\bar{Q}(t)$ est le débit moyen de l'hydrogramme. La décomposition de Karhunen-Loeve de $Q(t, \theta)$ s'écrit de la manière suivante :

$$Q(t, \theta) = \bar{Q}(t) + \sum_{i=1}^{\infty} \sqrt{\lambda_i} f_i(t) \zeta_i(\theta). \tag{10.3}$$

— f_i représentent les vecteurs propres,

— λ_i sont les valeurs propres associées au problème définit par l'équation 10.4.

— $\zeta_i(\theta)$ sont des variables aléatoires indépendantes et pour le cas gaussien sont centrées réduites.

$$\int_{t_2 \in T} \mathcal{C}(t_1, t_2) f_i(t_2) dt_2 = \lambda_i f_i(t_1), t_1 \in T, \tag{10.4}$$

La somme infinie dans l'équation 10.3 pour des raisons pratiques sera tronqué. Afin d'utiliser cet outil pour la représentation stochastique il reste à déterminer l'ordre de troncature appelé d dans l'équation 10.5.

$$Q(t, \theta) \approx \bar{Q}(t) + \sum_{i=1}^{d} \sqrt{\lambda_i} f_i(t) \zeta_i(\theta). \tag{10.5}$$

On choisit pour le cas d'application présent une fonction d'autocorrélation exponentielle. Pour ce choix, une solution analytique de l'intégrale de Fredholm est connue. Une démonstration rigoureuse de cette solution peut se trouver dans [157] ou dans [131]. Soit $\mathcal{C}(t_1, t_2) = exp(-|t_1 - t_2|/b)$ où b est la longueur de corrélation temporelle. Les valeurs propres et les vecteurs propres associés s'écrivent de la manière suivante :

$$\lambda_n = \frac{2b}{1 + (\omega_n b)^2} \tag{10.6}$$

$$f_n(t) = \begin{cases} \frac{cos[\omega_n t(t-1/2)]}{\sqrt{1/2 - \frac{sin(\omega_n)}{2\omega_n}}} & \text{si n est pair,} \\ \frac{sin[\omega_n t(t-1/2)]}{\sqrt{1/2 - \frac{sin(\omega_n)}{2\omega_n}}} & \text{si n est impair.} \end{cases} \tag{10.7}$$

Où les ω_n sont les racines positives ordonnées dans l'ordre croissant de l'équation caractéristique :

$$[1 - b\ \omega\ \tan(\omega/2)][b\ \omega + \tan(\omega/2)] = 0 \tag{10.8}$$

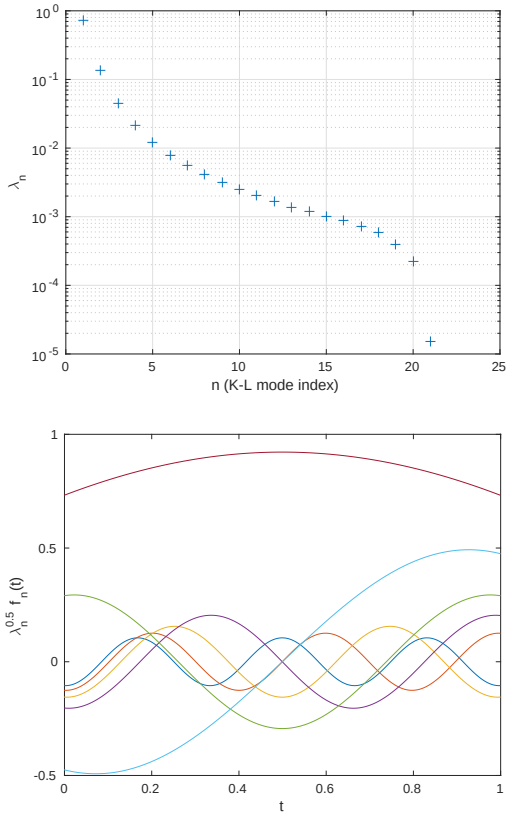


FIGURE 10.7 – *a - Les premières valeurs propres de la fonction de corrélation exponentielle b - Les 7 premiers vecteurs propres*

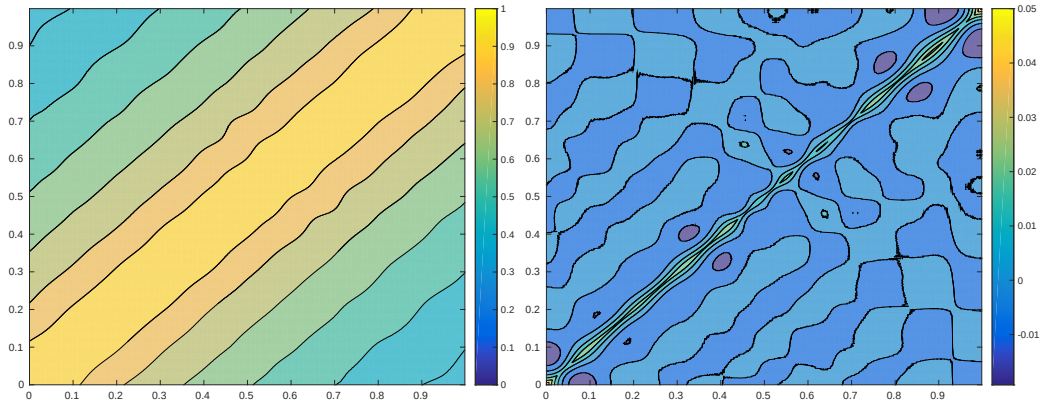


FIGURE 10.8 – *a - La matrice de corrélation estimée avec l'équation 10.9 b -Erreur sur l'estimation*

La figure 10.7 représente les valeurs propres et les vecteurs propres de la fonction d'autocorrélation exponentielle. Les valeurs propres présentent une décroissance rapide en fonction de l'index, cette décroissance aidera à choisir l'ordre de troncature d de la somme dans l'équation 10.5. Cette équation dépend de la longueur de corrélation temporelle (équation 10.6). Plus la longueur de la corrélation est petite plus la décroissance est lente (et donc plus l'ordre de troncature d sera grand). Les premiers vecteurs propres $f_n(t)$ sont présentés dans la figure 10.7. Plus l'index n est grand plus le nombre d'oscillations est important.

Finalement on peut estimer la matrice de corrélation par la formule suivante (une démonstration mathématique peut se trouver dans [76] ou dans [131]) :

$$\mathcal{C}(t_1, t_2) \approx \sum_{i=1}^{d} \lambda_i f_i(t_1) f_i(t_2) \tag{10.9}$$

La figure 10.8 présente la matrice estimé grâce à l'équation 10.9 ainsi que l'erreur sur l'estimation. L'erreur provient de la troncature que l'on a effectué dans l'équation 10.5. Pour cette matrice on a utilisé seulement 10 valeurs propres (i.i $d = 10$). L'erreur d'estimation de la matrice est de l'ordre de 10^{-2}.

10.2.3 Approximation polynomiale

Afin de représenter l'incertitude sur la hauteur d'eau on décompose celle-ci dans une base de chaos polynomial. Dans ce cadre là, on se placera uniquement au milieu du canal, de ce fait la hauteur aura seulement une dépendance stochastique et temporelle (pas de dépendance spatial, toutefois la méthode ne changera pas si on souhaite inclure aussi la variabilité spatiale). On écrira ainsi la hauteur d'eau $h(t, \theta)$ de la manière suivante :

$$h(t, \theta) = \sum_{i=0}^{\infty} a_i(t) \Psi_i(\boldsymbol{\zeta}(\theta)) \tag{10.10}$$

Où

— $\boldsymbol{\zeta} = (\zeta_1, \cdots, \zeta_d)$ représente le vecteur des variables aléatoires (pour ce cas ce sont les modes de la représentation stochastique du débit $Q(t, \theta)$).

— $a_i(t)$ représentent les coefficients (modes spectraux) du polynôme.

En pratique on tronque la somme de l'équation 10.10 jusqu'à un ordre donné P.

$$h(t, \theta) = h_{pc}(t, \theta) \approx \sum_{i=0}^{P} a_i(t) \Psi_i(\boldsymbol{\zeta}(\theta)) \tag{10.11}$$

L'équation 10.10 montre que l'on souhaite projeter dans un espace engendré par les polynômes Ψ_i. Pour des variables aléatoires gaussiennes, la base est construite à partir de la tensorisation des polynômes d'Hermite ($\Psi_i = \prod_{j=1}^{d} H_j(\zeta_j)$, où H_j sont les polynômes d'Hermite) est une base d'approximation complète. L'approximation des variables aléatoires de variance finie dans un tel espace ont une convergence en norme L_2. Ceci mène à réécrire l'équation 10.10 de la manière suivante :

$$h_{pc}(t, \theta) = \alpha_0(t) + \sum_{i=1}^{d} \alpha_i H_1(\zeta_i) + \sum_{j_1+j_2=2} \sum_{i_1=1}^{d} \sum_{i_2=1}^{i_1} \alpha_{i_1,i_2}(t) H_{j_1}(\zeta_{i1}) H_{j_2}(\zeta_{i2})$$

$$+ \sum_{j_1+j_2+j_3=3} \sum_{i_1=1}^{d} \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} \alpha_{i_1,i_2,i_3}(t) H_{j_1}(\zeta_{i1}) H_{j_2}(\zeta_{i2}) H_{j_3}(\zeta_{i3}) + \cdots \tag{10.12}$$

L'équation 10.2.3 donne la forme sous laquelle on projette la hauteurs d'eau dans l'espace généré par le produit tensoriel des polynômes d'Hermite, on peut évidement souligner les relations direct qui existent entre les coefficients $\alpha_i(t)$ et $a_i(t)$. Pour la suite de cette note, on retiendra la formule de l'équation 10.11 pour sa simplicité.
Le calcul direct et complet de toute la base polynomiale étant très coûteux au vu de la dimensionnalité du problème. Le tableau 10.1 permet de montrer le nombre de coefficients à calculer selon le degré de troncature considéré. On utilise une méthode de régression basée sur l'algorithme "LAR-LASSO".

TABLE 10.1 – *Nombre de coefficient polynomiaux à calculer selon le degré de troncature considéré avec* $d = 10$.

P	4	5	6	7	8	9	10
N_{pc}	1001	3003	8008	19448	43758	92378	184756

Erreur d'approximation.

Afin de valider le modèle polynomial, on introduit une erreur de généralisation connue sous le nom de "*leave-one-out error*" (LOO). Classiquement, afin de valider un modèle par un plan d'expériences, on effectue la validation croisée qui consiste à effectuer un tirage Monte-Carlo que l'on divise en deux partie E_1 et E_2, généralement la partie E_1 appelé ensemble d'apprentissage est utilisé par le modèle afin de créer le méta-modèle (le calcul des coefficients pour un modèle polynomial). La deuxième partie E_2, appelé ensemble de validation,est utilisée afin de calculer l'erreur de généralisation pour valider le méta-modèle. Vu la dimensionnalité du problème, le tirage Monte-Carlo est assez coûteux d'où la raison pour laquelle on choisit d'utiliser une erreur LOO. Une étude de [97] montre que cette méthode est bien adaptée pour l'estimation du biais ou de l'erreur quadratique moyenne. La technique pour l'estimation de l'erreur LOO consiste à choisir un plan d'échantillonnage E_{LOO} de taille N_{LOO}. On utilise tout le plan d'expérience sauf un point $E_{LOO}\backslash\{X^i\}$ pour l'estimation des coefficients polynomiaux et on calcul l'écart entre le modèle et le méta-modèle seulement en ce point $\Delta_i = h(t, \boldsymbol{\zeta_i}(\theta)) - h_{pc}(t, \boldsymbol{\zeta_i}(\theta))$. Finalement l'erreur LOO est la moyenne quadratique des écarts pour tous les points de E_{LOO} comme indiqué par l'équation 10.13. Afin de calculer l'erreur il faut estimer le métamodèle N_{LOO} fois en enlevant à chaque fois un point du plan d'expérience. [15] montre que l'on peut estimer l'erreur en faisant une seule fois l'approximation polynomiale à l'aide de la matrice d'information $\boldsymbol{\Psi}\boldsymbol{\Psi}^T$, où$\boldsymbol{\Psi}_{i,j} = \Psi_i(\boldsymbol{\zeta}_j)$.

$$\mathcal{E}_{LOO} = \frac{1}{N_{LOO}} \sum_{i=1}^{N_{LOO}} \Delta_i^2 \tag{10.13}$$

Résultats

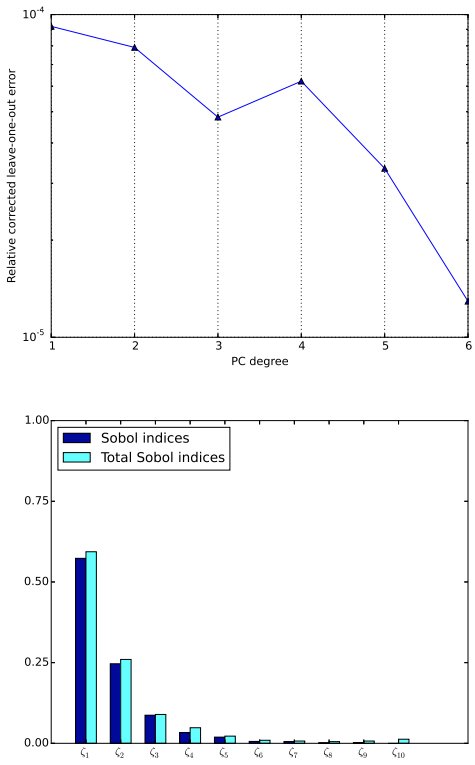


FIGURE 10.9 – *a - La sensibilité par rapport au degré polynomial b -Indices de Sobol pour le premier pas de temps*

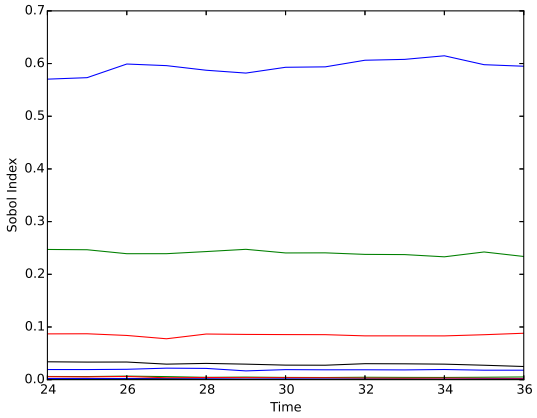


FIGURE 10.10 – *Evolution temporelle des indices de Sobol*

On montre ici les résultats de l'approximation polynomiale de la hauteur d'eau dans l'espace d'approximation généré avec la tensorisation des polynômes d'Hermite. On suppose que la variable aléatoire est l'hydrogramme à l'entrée du canal, que l'on décomposera sous la forme de Karhunen-Loève (équation 10.5). On utilise une fonction d'autocorrélation exponentielle afin d'utiliser les résultats montrés dans la section 10.2.2. On choisit de tronquer la somme de Karhunen-loève à $d = 10$. On simule la hauteur d'eau à l'aide de Mascaret sur un canal avec les spécifications données dans la section 10.2.1 sur une fenêtre de temps de 12h et on construit l'aproximation polynomiale toutes les heures. L'hydrogramme moyen $\bar{Q}(t)$ est supposé constant $\bar{Q}(t) = 4031m^3/s$. La figure 10.9-a montre la décroissance de l'erreur LOO en fonction du degré polynomial considéré. On choisit de tronquer le polynôme au degré 6. La figure 10.9-b montre les indices de Sobol' et les indices totaux pour les dix variables aléatoires considérées. On remarque que les trois premiers modes expliquent 90% de la variance totale. On remarque aussi la présence des effets croisés importants entre les membres (l'indice de Sobol pour ζ_{10} est nul alors que l'indice total est non nul). La figure 10.10 montre l'évolution temporel des indices de Sobol. Globalement, la variance de la hauteur est expliquée, pour tout l'intervalle temporel considéré, par les mêmes variables aléatoires $(\zeta_1, \zeta_2 et \zeta_3)$.
Ces résultats montrent les points suivants :

— Le problème discrétisé en temps comportait plus de 700 valeurs temporelles est réduit à un problème de dimension 10.

— Dans un cadre d'assimilation de données, la correction du débit peut être faite en corrigeant seulement la valeur des trois premiers modes.

— La présence d'effets croisés va induire un problème d'équifinalité : Plusieurs solutions du problème inverse sont possibles pour une même observation.

V

Conclusion et perspectives

Conclusion

Ce travail de thèse a porté sur la construction d'un modèle réduit pour les équations de Saint-Venant en hydraulique fluviale avec une méthode de décomposition en polynôme du chaos. Ce modèle réduit est utilisé à la place du modèle direct afin de réduire le coût de calcul lié aux méthodes ensemblistes en quantification d'incertitudes et assimilation de données. Dans la partie théorique dans le cadre de la correction de paramètres à l'aide de l'assimilation de données, on a prouvé que toutes les matrices, dont on a besoin pour formuler la matrice de gain de Kalman, peuvent être formulées à l'aide des coefficients polynomiaux.
La première partie de ce travail a été dédiée à formuler un polynôme du chaos pour le régime stationnaire. On a validé le modèle réduit pour un canal à pente constante. Puis on a prouvé, à l'aide d'expériences numériques, que le modèle polynomiale est capable d'estimer les statistiques du modèle, dont on a besoin pour le filtre de Kalman d'Ensemble, à moindre coût (49 simulations qui correspond à CPU $\approx$ 5 min) tout en gardant une bonne précision. Le modèle réduit permet aussi d'effectuer une analyse de sensibilité qui permet de réduire le nombre de paramètres à considérer pour l'assimilation de données. Cette étude a pu aussi pointer quelques limitations de l'approximation polynomiale pour les problèmes inverses en général :

— Quand la physique du modèle est assez compliquée à modéliser, l'approximation polynomiale requiert d'augmenter le degré afin d'estimer la densité de probabilité.

— L'approximation polynomiale est sensible à la résolution numérique du modèle.

On a utilisé la décomposition polynomiale formulée dans la partie précédente dans un algorithme d'assimilation de données de type SWOT pour corriger les paramètres de frottements ainsi que le débit permanent. On a montré sur ce cas d'application l'apport de la décomposition polynomiale accélérer l'EnKF et améliorer la précision de son estimation. Finalement, on a profité de cet exercice pour estimer la sensibilité de l'algorithme au moyennage spatial des observations de type SWOT. Les résultats montrent que le maximum de l'erreur dues au moyennage spatial est de l'ordre de 4%. Ce qui est assez prometteur pour l'utilisation des données satellitaires.
On s'est intéressé aussi à l'application du PC-EnKF en régime instationnaire. L'approximation polynomiale est formulée chaque jour. Dans un premier temps, on s'est intéressé à corriger seulement les paramètres de frottements. La décomposition en chaos polynomial a été utilisée pour :

- mener une analyse de sensibilité qui permet de créer un vecteur de contrôle composé que des variables les plus influentes ;
- étudier l'incertitude temporelle (en utilisant une décomposition où les coefficients polynomiaux varient dans le temps) due à la variabilité du vecteur de contrôle ;
- accélérer l'algorithme d'assimilation de donnée.

Finalement, on s'est intéressé à corriger le débit non-permanent. Cette étape a nécessité de réduire la dimension du champ stochastique par la décomposition de

Karhunen-Loève afin de formuler des ensembles. La décomposition de Karhunen-Loève a permis la réduction de l'espace incertain, qui est continue dans le temps. Cette décomposition permet aussi de passer d'un vecteur aléatoire , corrélé dans le temps, à un vecteur de variables aléatoires indépendantes. La décomposition polynomiale montre que la variance de la hauteur d'eau est expliquée par la variance des trois premiers modes. Cette information est utile pour la correction des débits à l'aide de l'assimilation des données sur la hauteur d'eau, car le vecteur de contrôle ne sera composé que des trois premiers modes (au lieu du vecteur correspondant au débit) ce qui amène à réduire le coût de l'algorithme d'assimilation de donnés.
Dans un cadre plus générale, cette thèse a contribué à enrichir les méthodes d'assimilation de données ensemblistes par des méthodes connues dans le milieu de la quantification d'incertitudes et d'analyse de sensibilité pour des applications en hydraulique fluviale. L'analyse de sensibilité est un atout indispensable à l'assimilation de données pour permettre d'identifier les bons paramètres à corriger. L'utilisation des méthodes spectrales telles que la décomposition en chaos polynomial a permis de réduire le coût assez élevé des méthodes ensemblistes d'assimilation de données, notamment grâce aux coefficients spectraux. Ces coefficients permettent d'estimer la matrice de gain de Kalman ce qui accélère l'algorithme de l'assimilation de données de manière significative. Dans le cas où les paramètres sont des champs stochastiques, la décomposition Karhunen-Loève permet de réduire la dimension du problème pour pouvoir le corriger.

Perspectives

Naturellement ce travail ouvre l'accès à des perspectives intéressante. On regroupe ces perspectives selon trois axes :

La future mission SWOT

la future mission SWOT représente certainement une révolution scientifique en matière de l'observation des rivières. On a pu constaté l'atout majeur des données de type SWOT pour l'amélioration de la modélisation hydraulique, à travers la correction des paramètres de frottements et de débit. Cependant, lors de ce travail de thèse, une modélisation de la Garonne était disponible et par conséquent, on possédait une connaissance à priori de la modélisation. Une perspective pour étendre ce travail serait de voir l'apport des observations de type SWOT pour la modélisation des rivières dont la connaissances des paramètres n'est disponible à priori.
L'observation satellitaire est surement très importante pour l'amélioration de la connaissance hydrodynamique des rivières dans le monde. Cependant, lors de ce travail on s'est pas intéressé à la modélisation des erreurs qui proviennent de la mesure satellitaire. Cette information qui est importante en assimilation de données pour pouvoir formuler la matrice de covariances des erreurs d'observations. On a constaté, que la stratégie pour la correction des paramètres à l'aide de l'assimilation de données

de type SWOT est différente de celle utilisée pour les observations *in-situ*. Une perspective assez intéressante serait de formuler une stratégie qui puisse prendre en compte les deux sortes de données en respectant les contraintes de chacune des observations.

Polynômes du chaos en hydraulique

La décomposition en polynôme du chaos a été formulée dans ce travail de thèse pour une modélisation hydraulique 1D dans le régime fluvial. Une prochaine étape, serait de considérer l'approximation polynomiale pour une modélisation 1D mais pour des cas avec la présence de discontinuité, pour traiter le cas de rivière avec la présence de mascaret par exemple. On pourrait profiter alors des travaux de [107] pour gérer les discontinuités. D'autres cas qui restent à traiter en hydraulique sont les cas de la décomposition polynomiale pour un code 2D tel que Telemac-2D. En effet, la présence de méandres et de confluences impliquent une physique 2D. D'où la nécessité de considérer une physique plus compliqué. La difficulté serait de pouvoir formuler une hauteur d'eau qui a une double dépendance spatiale. On pourrait par exemple ne calculer la décomposition polynomiale que pour des points spécifiques de a rivières et utiliser une modèle d'interpolation pour avoir la sortie polynomiale spatialisée en s'inspirant des travaux de [123] par exemple. Pour le long terme, on pourra aussi s'intéresser à la sortie d'un modèle couplée 1D-2D.
Une autre perspective possible, serait de proposer d'autres modèles réduits pour la quantification d'incertitudes et l'analyse de sensibilité. Une publication en cours de rédaction, a été mené conjointement avec le travail de thèse de Pamphile Roy. Elle consiste à comparer les résultats de l'estimation de la matrice de covariance estimée à l'aide d'une méthode basée sur le polynôme du chaos et une méthode basés sur la POD couplée avec le krigeage.

Assimilation de données

Dans le cadre de l'assimilation de données, on a pas pu voir la sensibilité de la prévision par rapport à la fenêtre temporelle. Une perspective serait d'étudier le coût de l'approximation polynomiale au fur et à mesure des étape d'analyses.
Concernant la correction de champs stochastique, la décomposition de Karhunen-Loève a permis de réduire la dimension du problème. Cependant, d'autres études restent en suspens :

— Vu que l'étude était menée sur un canal idéalisé avec un hydrogramme moyen constant dans le temps, il serait intéressant de poursuivre ce travail dans le cadre d'une rivière réelle et avec des données d'hydrogrammes réelle.

— La fonction d'autocorrélation, que l'on a imposé à l'hydrogramme, est défini à l'aide d'une fonction analytique dont on connait les solutions de l'équation de Fredholm. Une possibilité d'améliorer cette étude, pour s'approcher de donner réelle est d'estimer la covariance temporelle à l'aide de données d'observations.

— Dans le cadre de l'assimilation de données, une perspective de ce travail serait de corriger soit les modes les plus influents à l'aide d'observations de la hauteur d'eau, soit les caractéristiques de la fonction d'autocorrélation.

— On pourrait estimer la corrélation temporelle à l'aide d'un semi-variogramme que l'on établira à l'aide des estimations de débits données par SWOT.

On a pu voir dans ce travail l'apport de la décomposition polynomiale pour accélérer les méthodes ensemblistes d'assimilation de données, notamment l'EnKF. Cependant, il serait intéressant de comparer les résultats obtenus ici avec des méthodes hybrides utilisées en assimilation de données, où l'adjoint du modèle est remplacé par la dérivé du modèle polynomiale qui est assez simple à estimer.
Finalement, on pourrait s'intéresser à la correction de champs de la bathymétrie. Des premiers résultats que l'on a mené au cours de cette thèse, montrent que la correction de bathymétrie souffre du problème d'équifinalité (plusieurs solutions sont possibles pour une même observation). Il serait intéressant de décomposer le champ de bathymétrie à l'aide de la méthode de Karhunen-Loève et de corriger les modes de la décomposition. La fonction d'autocorrélation pourra alors être prescrite à l'aide d'un semi-variogramme à l'aide de données SWOT.

VI

Annexes

.1 Rapport Technique : Quantification des incertitudes liées aux paramètres hydrauliques : Application aux coefficients de frottement avec le modèle Mascaret Marne Amont Global

CERFACS

Nabil EL MOCAYD, Johan HABERT, Sophie RICCI

Quantification des incertitudes liées aux paramètres des modèles hydrauliques

Application aux coefficients de frottement avec le modèle Mascaret Marne Amont Global

Mai 2014

Sommaire

1. Introduction....1
2. Méthodologie....1
3. Protocole de test des différents paramètres....2
3.1. La loi de distribution des Ks....3
3.2. La taille de l'échantillon....3
3.3. La condition initiale....4
3.4. Impact du débit sur la sensibilité de la hauteur d'eau au Ks en régime permanent....5
4. Analyse des distributions de perturbations de hauteurs d'eau liées à la perturbation des Ks....8
4.1. Cas Permanent....8
4.2. Cas non-permanent....11
5. Conclusion....15
Annexe :....15

1. Introduction

Les écoulements dans les rivières sont influencés par la granulométrie et la turbulence du fond des rivières. Le transport solide change le débit et par conséquent la nature de l'écoulement d'où la nécessité d'une bonne paramétrisation des coefficients de frottement et de la bathymétrie. Cette dynamique interne dans l'écoulement due aux frottements pariétaux, le transport solide et la granulométrie font que la modélisation du coefficient de frottements devrait être un paramètre variable dans le temps et dans l'espace. Cependant la modélisation de ce coefficient est en réalité faite sous la forme d'un coefficient constant dans le temps et dans l'espace. Par conséquent, il est important de quantifier les erreurs dues aux incertitudes sur la paramétrisation du coefficient de frottement dans les équations de Saint-Venant. On se propose ici, de réduire l'erreur liée au coefficient de frottement en le considérant comme vecteur de contrôle d'un problème d'assimilation de données.

L'objet plus général de ces travaux est de disposer d'un outil permettant de quantifier les erreurs sur les hauteurs et les débits simulés liées à la paramétrisation des modèles hydrauliques. L'outil est dans un premier temps testé pour quantifier les erreurs relatives aux coefficients de Strickler (Ks) puis sera testé pour quantifier l'incertitude liée à la bathymétrie dans une étude ultérieure.

2. Méthodologie

Le Service de Prévision des Crues Seine Amont Marne Amont (SPC SAMA) utilise le code hydraulique 1D MASCARET pour effectuer en temps réel des prévisions sur l'amont de la rivière Marne. Il

a été montré que l'ajout local de données bathymétriques a permis de résoudre une erreur systématique de modélisation des hauteurs d'eau à la station de Mussey alors que les débits étaient correctement simulés. Toutefois, dans le cas où l'on ne dispose pas de données supplémentaires, il est aussi possible de corriger localement les Ks afin de modifier la ligne d'eau calculée par le modèle sans dégrader les débits modélisés. L'écart entre les hauteurs observées et modélisées à Mussey variant dans le temps, la correction locale du Ks doit elle aussi être adaptée temporellement via l'algorithme de l'assimilation de données de hauteurs d'eau. Afin de tester cette méthode, des tests préliminaires de sensibilité aux Ks des lits mineurs et majeurs ont été effectués pour mesurer l'impact sur les hauteurs et les débits d'une correction locale des Ks à Mussey. Il en ressort d'une part que seule la hauteur d'eau est impactée et d'autre part que la sensibilité aux Ks de la hauteur simulée est non-linéaire, les deux points ci-dessous sont des manifestations de cette non-linéarité :

- l'impact d'une perturbation positive diffère d'une négative : quand on perturbe une valeur moyenne de Ks par une petite perturbation dKs positivement on diminue alors la hauteur d'eau d'une valeur dh. Par contre, on n'augmente pas de la même valeur dh la hauteur d'eau lorsque l'on perturbe négativement la valeur moyenne de Ks par le même valeur de la perturbation dKs.
- l'impact est fonction du débit au moment de la perturbation, plus le débit est fort plus l'impact de la perturbation sur la hauteur d'eau est fort.

Une première quantification de ces non-linéarités a été effectuée "manuellement" pour la station de Mussey avec un nombre limité de tests. Afin de pouvoir renouveler l'opération en d'autres points du modèle et pour un plus grand nombre de perturbations, une maquette de quantification des incertitudes a été développée. Cette maquette permet de faire un tirage Monte-Carlo des Ks selon une distribution de probabilité donnée, ce tirage est suivi par l'intégration d'autant de simulations St-Venant à l'aide du code MASCARET en utilisant la fonctionnalité PARASOL. Cette maquette, développée à l'aide du coupleur de codes O-Palm, se décompose en deux parties : la première permet de tirer selon une distribution les perturbations des Ks pour une variance donnée puis pour chaque tirage, de faire l'intégration MASCARET en incluant la bonne valeur du Ks, cette partie a fait l'objet d'un développement informatique à l'aide du coupleur O-Palm et de la fonctionnalité PARASOL. La seconde partie, écrite en Octave, permet de tracer les densités de probabilité (PDF) ainsi que les fonctions de répartition et de calculer les moments des hauteurs d'eau à la sortie des différentes intégrations. Ces étapes sont enchaînées au sein d'un canevas O-Palm. Les intégrations de MASCARET sont faites en parallèle, sur l'ensemble des ressources (le nombre de processeurs) disponibles.

En résumé, nous souhaitons étudier les incertitudes du code MASCARET pour la modélisation de la Marne par rapport aux variations de Ks. Nous allons, dans un premier temps faire tourner la maquette O-Palm pour en régime permanent puis dans un second temps en régime non-permanent. Ceci nous permettra de faire une analyse à la réponse à la perturbation et quantifier l'incertitude sur la hauteur simulée par rapport aux coefficients de frottements.

3. Protocole de test des différents paramètres

On fait l'hypothèse que le coefficient de frottement est une variable aléatoire qui suit une loi gaussienne. On propose donc de travailler sur les perturbations de Ks autour d'une valeur de référence, représentées par une loi gaussienne centrée réduite. Le tirage stochastique des perturbations de Ks dépend de :

- la variance de la loi normale.

- le nombre de membres de l'ensemble.

Deux paramètres peuvent influencer la réponse à la perturbation :

- le niveau de l'eau au moment de la perturbation (la condition limite amont).
- le temps de simulation nécessaire pour que la condition initiale n'influe plus sur les résultats.

3.1. La loi de distribution des Ks

Les Ks du modèle Mascaret Marne Amont ont été calés de façon à minimiser l'écart entre la simulation et l'observation aux 6 stations de mesure du modèle pour 10 crues indépendamment les unes des autres, on obtient alors pour les différentes zones un Ks moyen, une variance et un écart-type. Cette variance issue du processus de calibration est utilisée pour décrire la loi pour la variable aléatoire de perturbation du Ks. On tire les perturbations des Ks mineur et majeur suivant la même loi de distribution, mais cela peut être amélioré au besoin.A Mussey, les Ks des lits mineurs et majeurs retenus après calage sont respectivement de 20 et 13, leurs écarts-types de 3 et 4 ce qui nous fait un écart-type moyen de 3,5 soit approximativement une variance de 12. La variance de la distribution de la loi normale retenue pour la suite de cette étude est de 12 soit un écart-type de perturbation des Ks des lits mineurs et majeurs de 3,464.

L'intervalle de confiance (IC) à 68 % des perturbations est théoriquement [-3.464;+3.64], l'IC théorique à 95 % est [-6.928;+6.928], et enfin l'IC théorique à 97.7 % est [-10.392;+10.392]. Le calcul s'arrête (passage à sec) lorsque le Ks du lit mineur est inférieur à 10, un seuil minimal de 10 est donc appliqué dans le fichier ParametreMascaretAssim.cas.

3.2. La taille de l'échantillon

Lors du tirage stochastique et du traitement statistique la variance des perturbations de Ks doit converger vers la variance théorique renseignée lors du tirage (Figure 1). Plus le tirage est grand, plus la variance de la répartition tend vers la variance théorique mais plus le calcul sera long. Il faut trouver le compromis idéal entre temps de calcul et vraisemblance des résultats. Les calculs (Tableau 1) ont été réalisés avec une variance de 12, pour une simulation durant 86400s avec un débit constant de 150m3/s, dans cette configuration une simulation dure approximativement 4s. Le tableau 1 et la Figure 1 représentent les valeurs des variances des Ks dans le lit majeur et le lit mineur en fonction du nombre de membres choisit. On déterminera la taille de l'échantillon à partir de la Figure 1 et le tableau 1.

Nombre de tirage	Variance du Ks lit mineur σ^2_{Ksmin}	Variance du Ks lit majeur σ^2_{Ksmaj}
10	30.98	7.60
20	18.42	9
50	12.42	13.27
100	12.76	13
200	12.84	13.31
500	12.84	11.22
1 000	12.02	12.40
2 000	12.21	12.41

5 000	12.08	12.33
10 000	11.85	12.21

Tableau 1 : Évolution de la variance des Ks des lits mineurs et majeurs (σ^2_{Ksmin} σ^2_{Ksmaj} en m²) en fonction de la taille de l'ensemble.

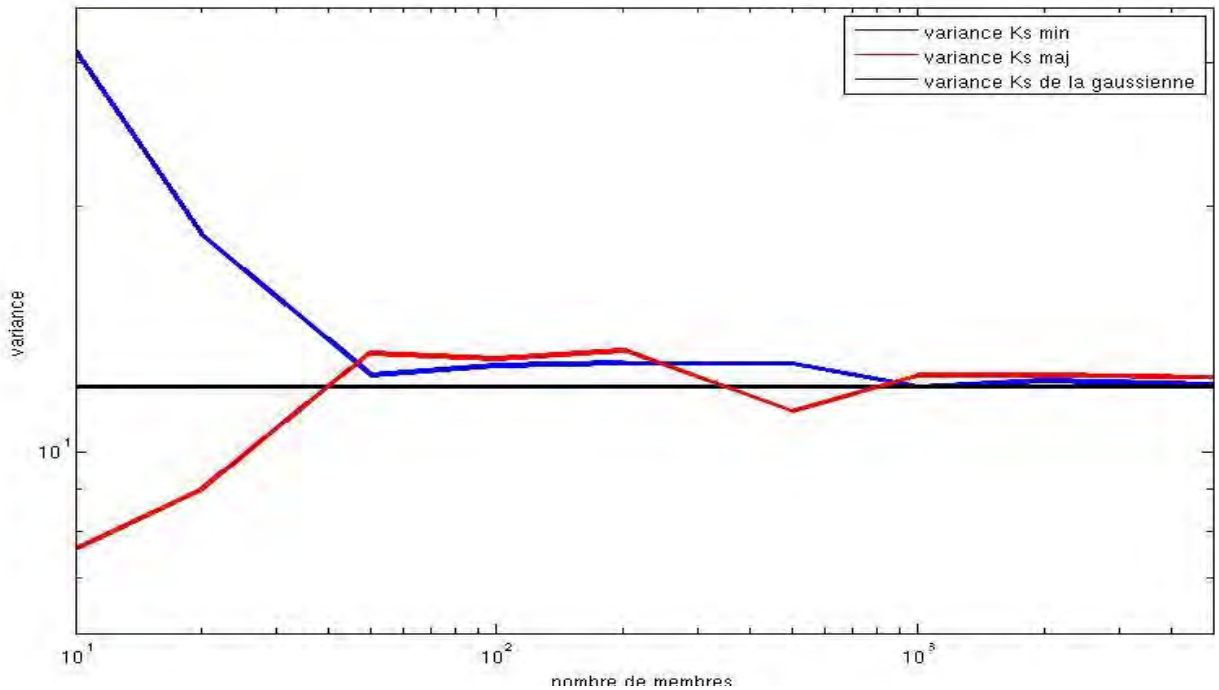


Figure 1: Comparaison de la variance de la gaussienne (droite noire) avec les variances des Ks des lits mineurs (courbe bleu) et majeurs (courbe rouge) obtenus pour les différentes tailles de l'ensemble.

On conclut que 1000 membres est un bon compromis pour représenter correctement la loi de distribution gaussienne supposée pour les Ks (soit 2000 tirages si l'on perturbe les Ks mineurs et majeurs) ce qui représente à peu près 1 heure de simulation. Ce temps de calcul peut être largement réduit si on mène les calculs perturbés sur plusieurs processeurs, les simulations étant indépendantes les unes des autres.

3.3. La condition initiale

Pour tous les tests que l'on va présenter ici, la condition initiale est calculée de façon à assurer la cohérence avec la condition à la limite. Il est nécessaire d'estimer le temps minimum de simulation pour que la condition initiale n'influe plus les résultats. L'écart-type des hauteurs perturbées a été calculé au temps final de la simulation avec un débit permanent de 150m3/s avec 100 perturbations et une variance de 12 pour différents temps de simulation :

Temps de Simulation	Écart-type de la hauteur d'eau σh dans le lit mineur en m	Écart-type de la hauteur d'eau σh dans le lit majeur en m
43 200	0.0690	0.0213
86 400	0.0694	0.0214

129 600	0.0692	0.0214
172 800	0.0692	0.0214

Tableau 2 : Évolution de l'écart-type de la hauteur d'eau simulée (σh en m) en perturbant les Ks mineurs et majeurs en fonction du temps de simulation.

La perturbation s'atténue rapidement entre T0 et T+12h (43200s). Pour avoir une marge de sécurité, nous analyserons la réponse à la perturbation au temps final d'une simulation de 86400s.

3.4. Impact du débit sur la sensibilité de la hauteur d'eau au Ks en régime permanent

La réponse à la perturbation du Ks varie en fonction de la hauteur d'eau dans le modèle. A Mussey, l'écoulement se cantonne dans le lit mineur pour des débits inférieurs à 70m3/s, le lit majeur est impacté à partir de 100m3/s (on a entre les deux une zone de transition difficilement déterminable). La station existe depuis une dizaine d'années, le plus haut débit mesuré est de 270 m3/s en février 2012. Nous testerons la réponse à la perturbation des Ks en permanent en injectant dans le modèle des débits constants compris entre 25 et 500m3/s avec un pas de 25m3/s. Le modèle a 9 conditions limites amont, on impose donc 9 valeurs de débits constantes. Pour 1000 perturbations avec une variance de 12 pour une simulation de 86400s on obtient les écart-types des hauteurs en perturbant les Ks des lits mineurs et majeurs suivants :

Débits à la station de Mussey en m3/s	Écart-type de la hauteur d'eau en perturbant le Ks mineur en m	Écart-type de la hauteur d'eau en perturbant le Ks majeur en m
25	0.045	0
50	0.05	0.003
75	0,05	0.01
100	0.05	0.01
125	0.06	0.02
150	0.06	0.02
175	0.07	0.02
200	0.08	0.03
225	0.08	0.03
250	0.09	0.04
275	0.10	0.04
300	0.11	0.04
325	0.11	0.05
350	0.12	0.05
375	0.13	0.05
400	0.13	0.06

Tableau 3 : Évolution de l'écart-type de la hauteur d'eau (σh en m) en perturbant les Ks mineurs et majeurs

en fonction du débit injecté.

On a pour des perturbations du Ks du lit mineur faibles un arrêt du calcul avec 500 m3/s injecté dans le modèle (passage local à sec en aval).

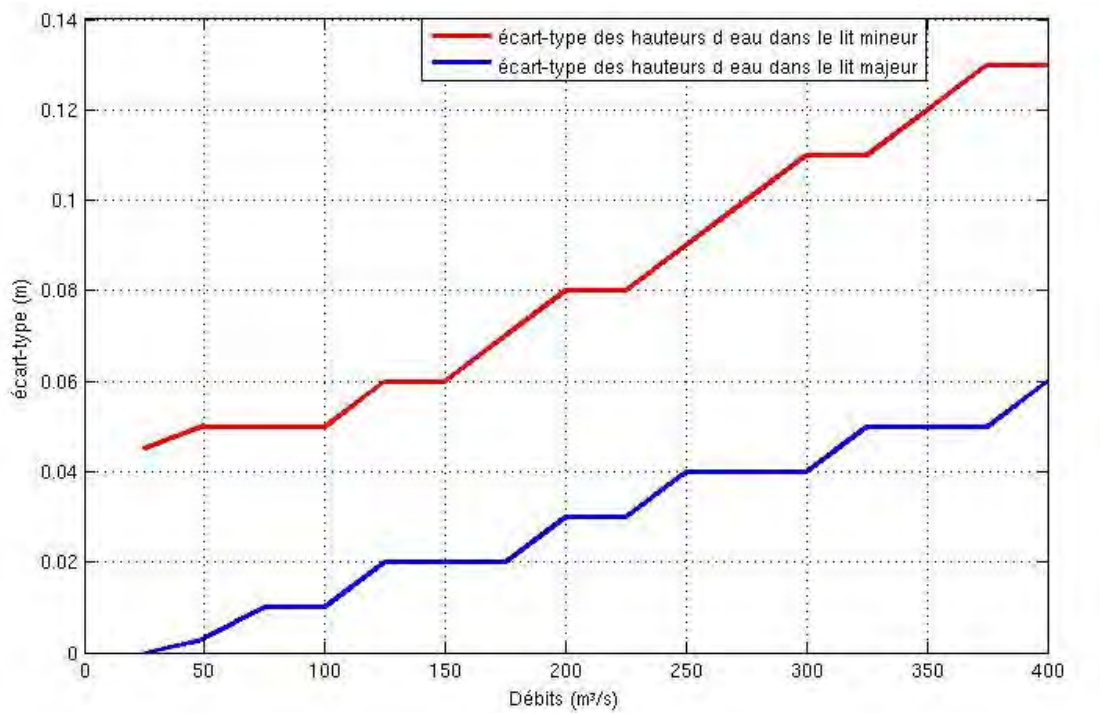


Figure 2: Evolution des écarts-types des hauteurs d'eau perturbées par les Ks des lits mineurs (courbe rouge) et majeurs (courbe bleue) en fonction du débit injecté dans le modèle.

Étant donné que l'augmentation de débit influe aussi sur la hauteur d'eau donc on trace l'évolution du coefficient de variation $C.V = \frac{\sigma}{\mu}$ avec σ l'écart type et μ la moyenne de l'échantillon.

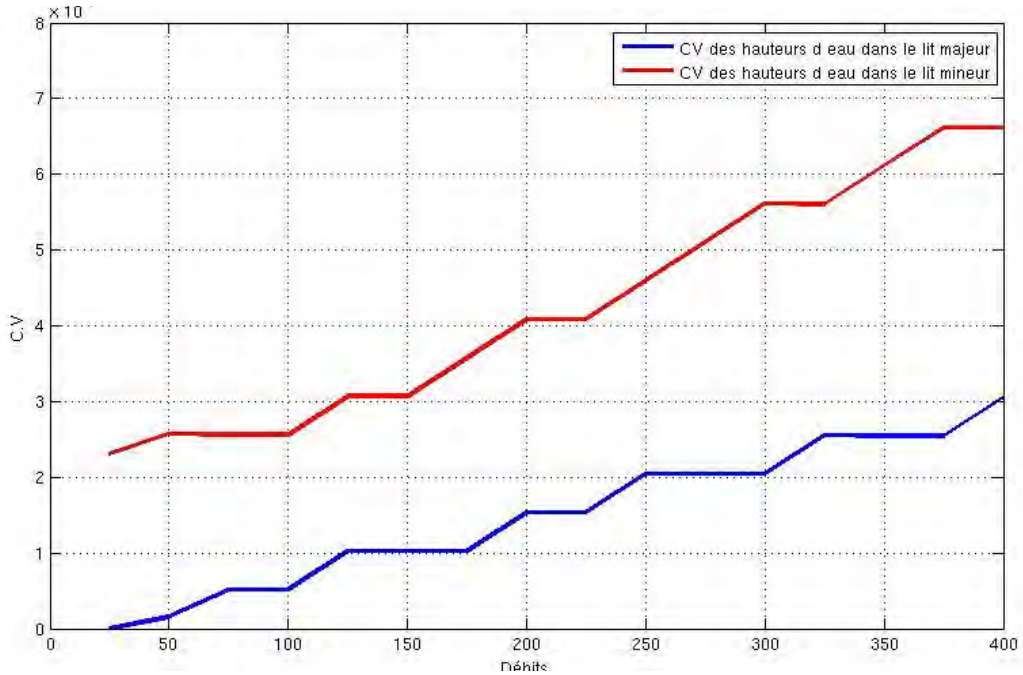


Figure 3: Evolution du coefficient de variation des hauteurs d'eau perturbées par les Ks des lits mineurs (courbe rouge) et majeurs (courbe bleue) en fonction du débit injecté dans le modèle.

La perturbation du Ks du lit majeur impacte le CV de la hauteur d'eau dans une moindre mesure (la pente de la courbe bleue est plus faible que celle de la courbe rouge) et ceci pour des débits plus importants. En dessous de 100m3/s, la perturbation du Ks du lit majeur a peu d'effet voir aucun en dessous de 25m3/s. Pour conclure la sensibilité de la hauteur d'eau au Ks est d'autant plus forte que le débit est grand.

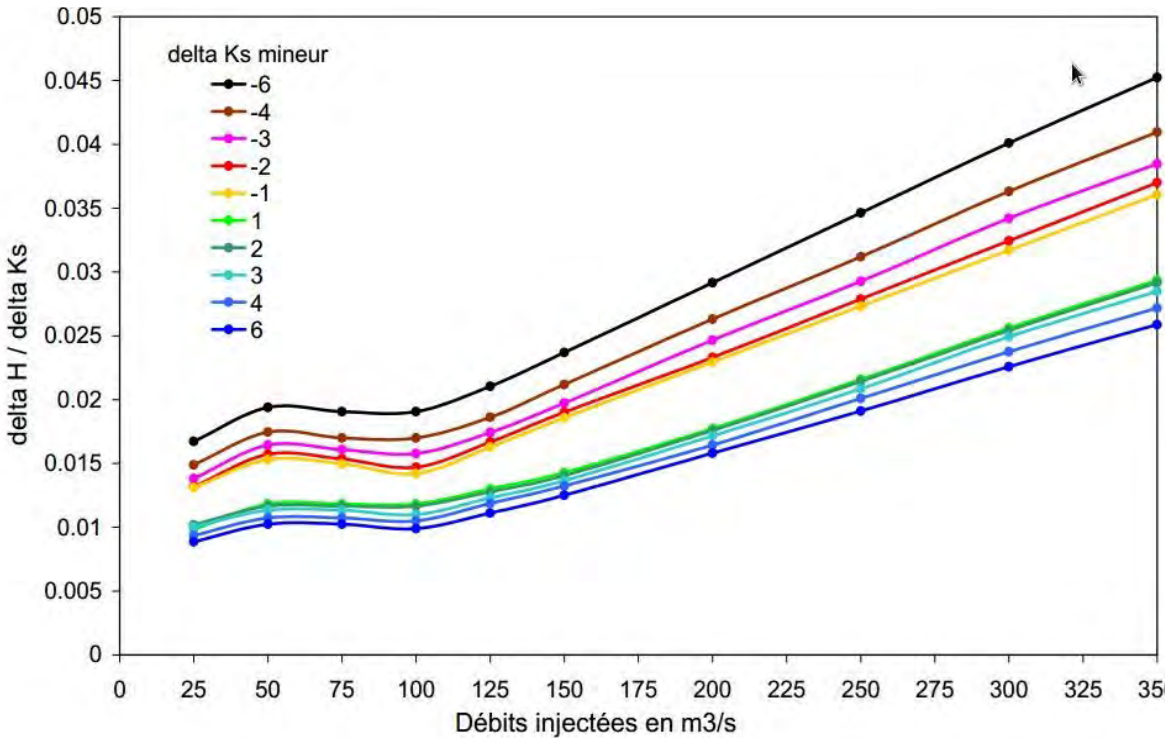


Figure 4: Evolution de l'impact sur la hauteur en fonction de la perturbation du Ks mineur et du débit.

À partir de la figure 4, on observe à partir de 100m3/s une évolution linéaire des courbes en fonction du débit. Pour comprendre ce phénomène, on se place dans le régime permanent et dans le lit mineur, et on regarde l'équation résolue par MASCARET.

Soit l'équation de conservation de la masse :

$$\frac{1}{A}\frac{\partial Q}{\partial t}+\frac{1}{A}\frac{\partial Q^2/A}{\partial x}+g\frac{\partial Z}{\partial x}=g(S_0-S_f)$$

en régime permanent l'équation devient :

$$\frac{-Q^2}{A^3}\frac{\partial A}{\partial x}+g\frac{\partial Z}{\partial x}=g(S_0-S_f)$$

$$On\ a\ dA=BdZ$$

$$donc\ -\frac{Q^2}{A^3}\frac{\partial A}{\partial x}+\frac{g}{B}\frac{\partial A}{\partial x}=g(S_0-S_f)$$

$$\frac{g}{B}\frac{\partial A}{\partial x}\left(1-\frac{Q^2B}{gA^3}\right)=g(S_0-S_f)$$

$$\frac{\partial A}{\partial x}=B\frac{(S_0-S_f)}{(1-Fr^2)}$$

$$donc\ on\ retrouve\ \frac{\partial Z}{\partial x}=\frac{S_0-S_f}{1-Fr^2}$$

A : La section du point,

Q : Le débit

Z : La côte NGF,

S_0 : La pente du fond,

S_f : La pente due aux frottements,

Fr : Nombre de Froude,

g : L'accélération de la pesanteur.

L’équation décrit les courbes de remous que l’on retrouve dans les écoulements naturels pour le régime fluvial. On remarque qu’il existe deux sortes de courbes de remous pour le régime fluvial une M1 et M2 (Figure 5). Selon où se situe la cote d’eau par rapport à la hauteur normale. La hauteur normale est une hauteur de référence qui est définie à partir du frottement du débit et de la pente $h_n = \frac{Q^2}{\sqrt{(I)}\, Ks\, L^2}$. Donc pour une pente fixée et une petite perturbation du Ks, la hauteur normale dépend du débit. Il y aura changement dans la forme de la courbe de remous selon le débit injecté, d’où le changement de comportement du coefficient de variation de la hauteur d’eau par rapport au débit d’entrée.

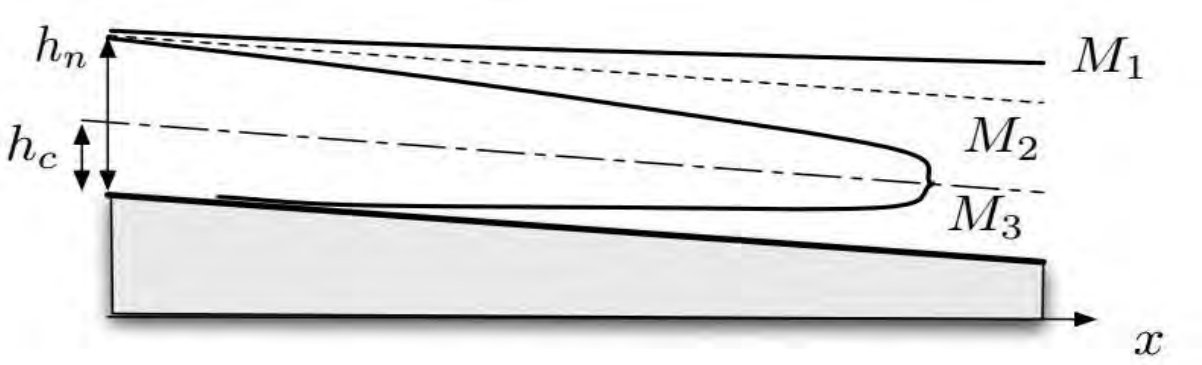


Figure 5: Les différentes courbes de remous dans le cas d'une pente faible (Hydrodynamique de l'environnement, O.Thual).

4. Analyse des distributions de perturbations de hauteurs d’eau liées à la perturbation des Ks

4.1. Cas Permanent

La figure 6 ci-dessous représente la densité de probabilité des hauteurs d’eau pour 1000 perturbations de Ks du lit mineur autour de la valeur moyenne de 20 avec une variance de 12 au temps final d’une simulation de 24h avec un débit de 150m3/s injecté dans le modèle. La courbe verte représente la densité de probabilité gaussienne ajustée par les moments. L’intervalle de perturbation des Ks est compris entre -11,78 et 13,35, on a 501 perturbations négatives, 499 positives.

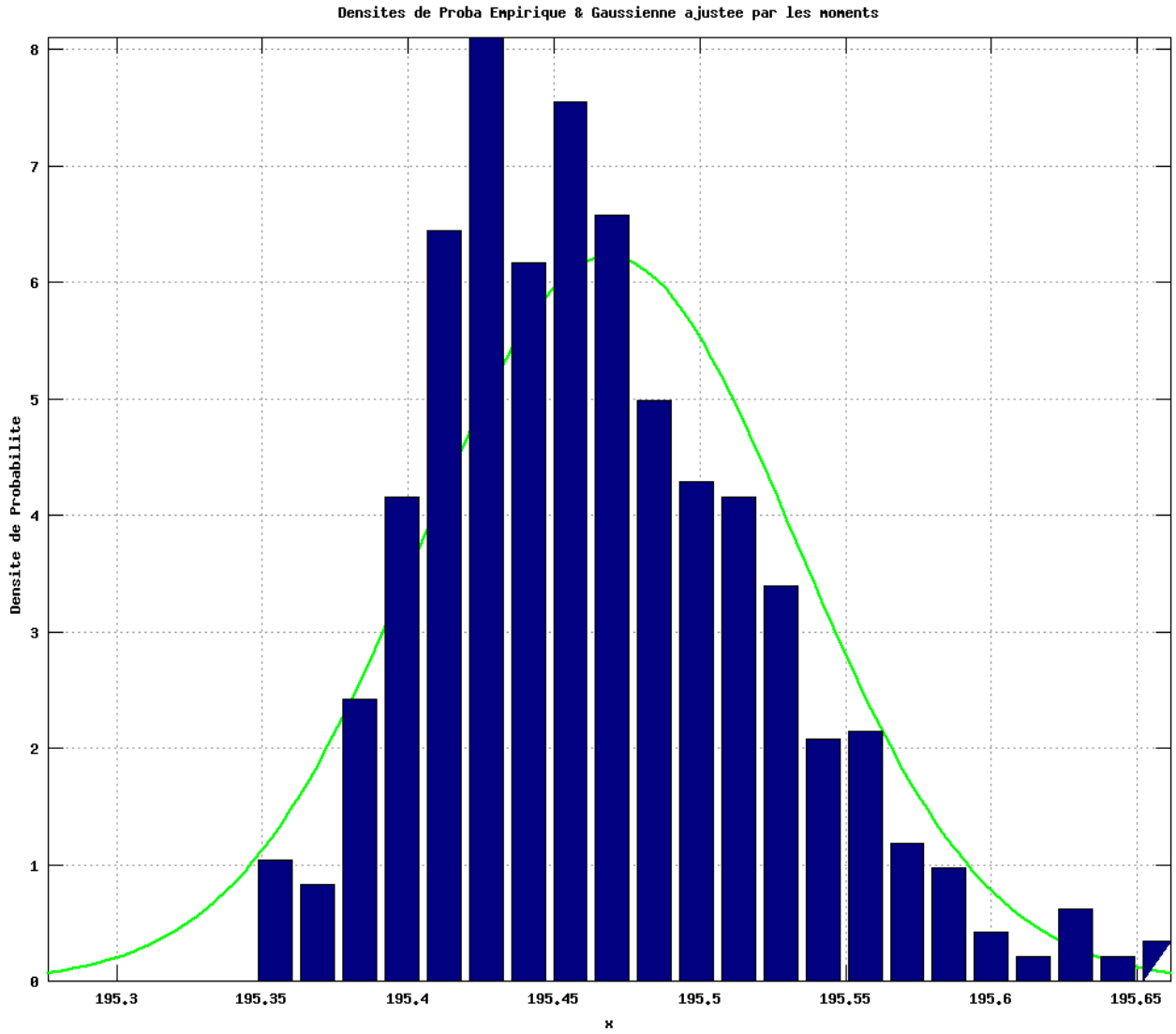


Figure 6: Densité de probabilité de la hauteur d'eau dans le lit mineur.

La figure 7 représente la densité de probabilité des hauteurs d'eau pour 1000 perturbations de Ks du lit majeur autour de la valeur moyenne 13 avec une variance de 12 au temps final d'une simulation de 24h avec un débit de 150m3/s injecté dans le modèle.

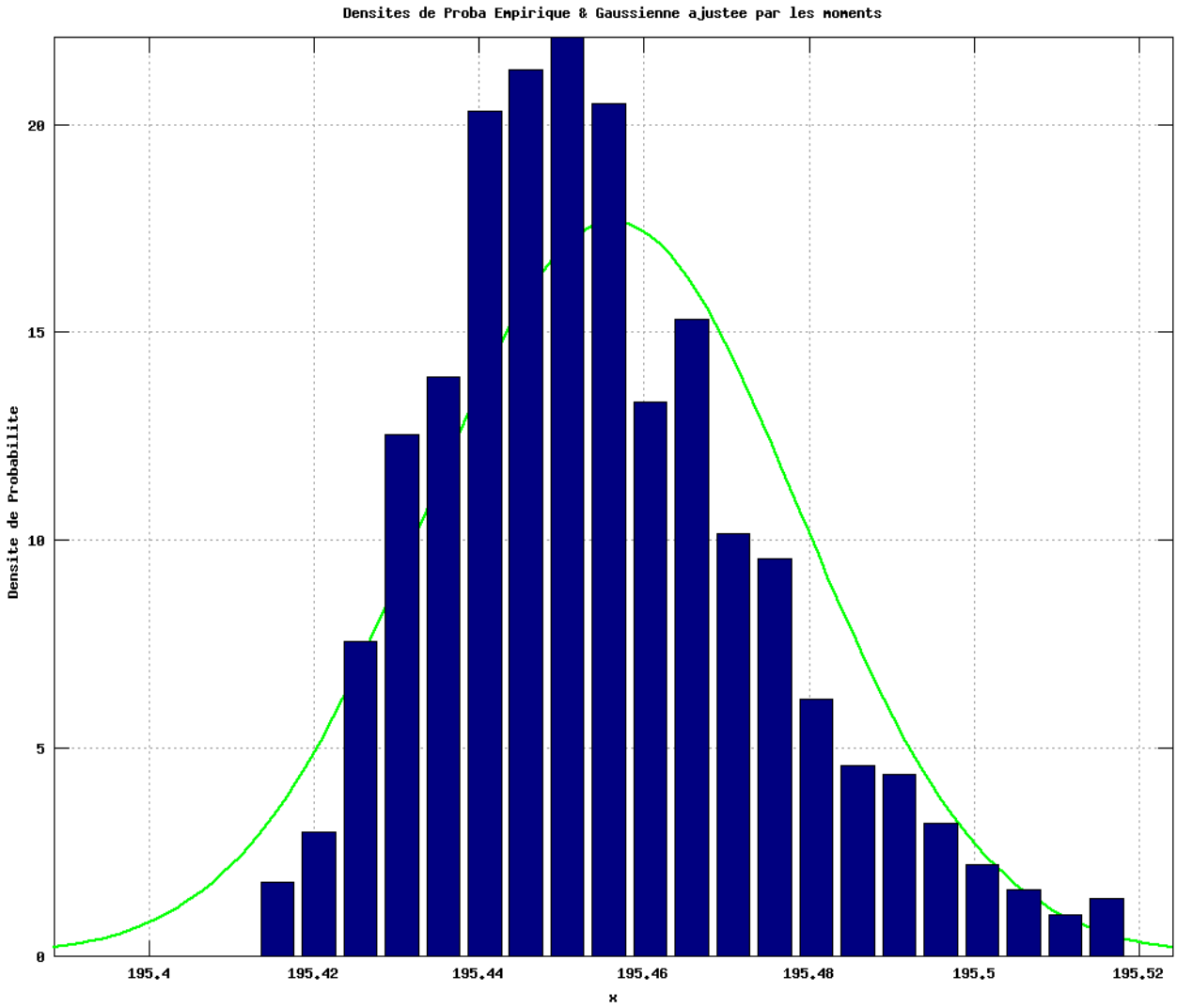


Figure 7: Densité de probabilité de la hauteur d'eau dans le lit majeur.

La hauteur calculée par le modèle sans perturbation est égale à 195.456m alors que la moyenne des hauteurs perturbées est égale à 195.47m, on peut donc en conclure qu'en moyenne les perturbations ont tendance a plus augmenter les hauteurs qu'à les diminuer mais la différence est faible (1.44cm). L'intervalle des hauteurs aux Ks perturbés positivement (donc une baisse de la hauteur) a une amplitude de 13,16cm alors celui des hauteurs aux Ks perturbés négativement (donc une augmentation de la hauteur) a une amplitude de 45,74cm.

La densité de probabilité ajustée par les moments (courbe en vert) donne la réponse linéaire du système. On remarque que la densité de probabilité des hauteurs d'eau est décalée vers la gauche, c'est-à-dire qu'un grand nombre de perturbations est concentré vers des valeurs inférieures mais proches du mode de la PDF. On remarque aussi que les perturbations extrêmes sont décalées vers la droite et la queue de la PDF est plus longue à droite qu'à gauche. Ceci traduit l'effet de la non linéarité du système.

La réponse à la perturbation est non-linéaire : quand on diminue le Ks, la hauteur augmente plus que proportionnellement, quand on augmente le Ks, l'impact de la perturbation sur les hauteurs tend vers 0.

4.2. Cas non-permanent

On va maintenant évaluer la linéarité du modèle MASCARET en régime non-permanent à Mussey en simulant une crue pour différentes valeurs du Ks (Figure 8 et 9) puis on regardera la densité de probabilité de la hauteur d'eau durant la monté de la crue, le pic de crue et la décrue :

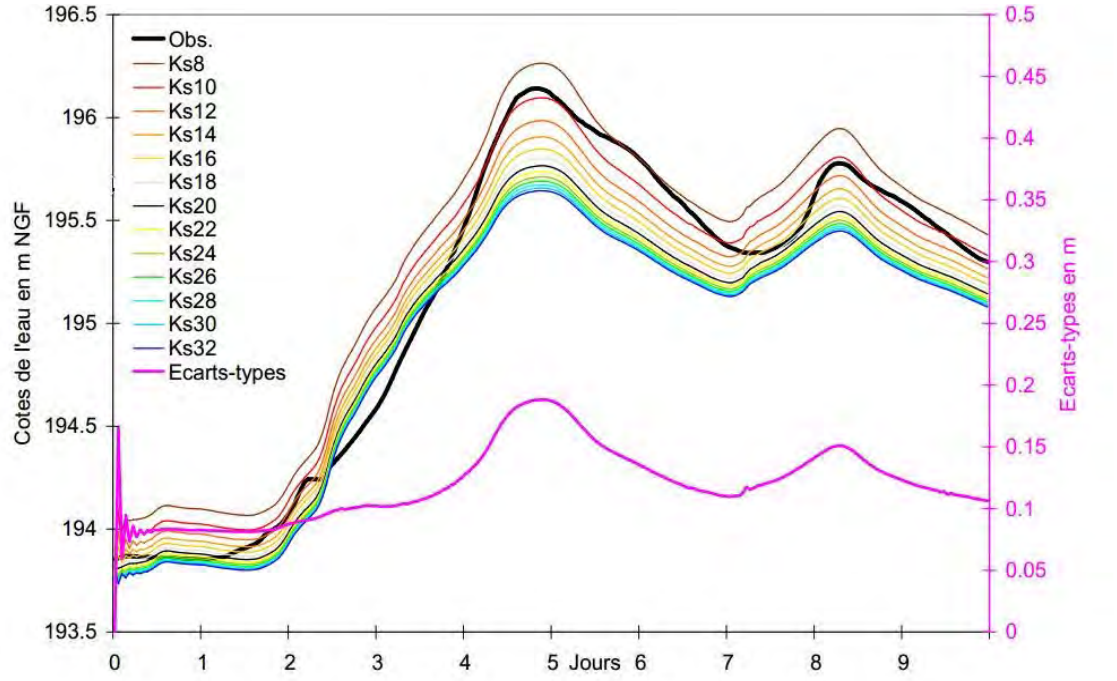


Figure 8: Effet du KS sur la simulation de la hauteur d'eau dans le lit mineur.

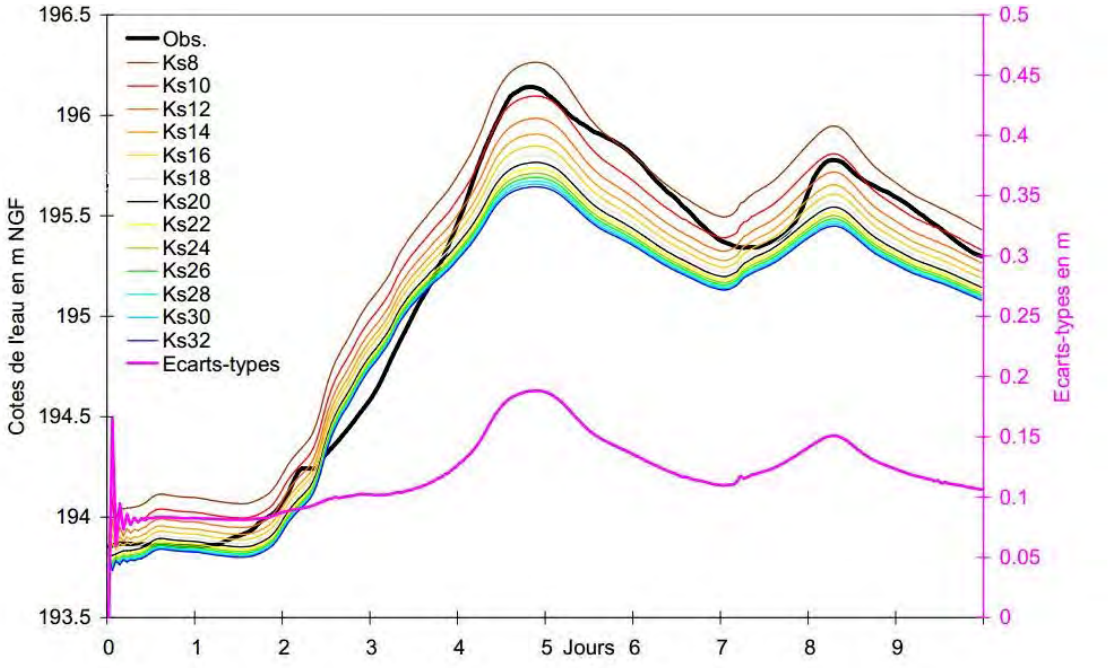


Figure 9: Effet du KS sur la simulation de la hauteur d'eau dans le lit majeur.

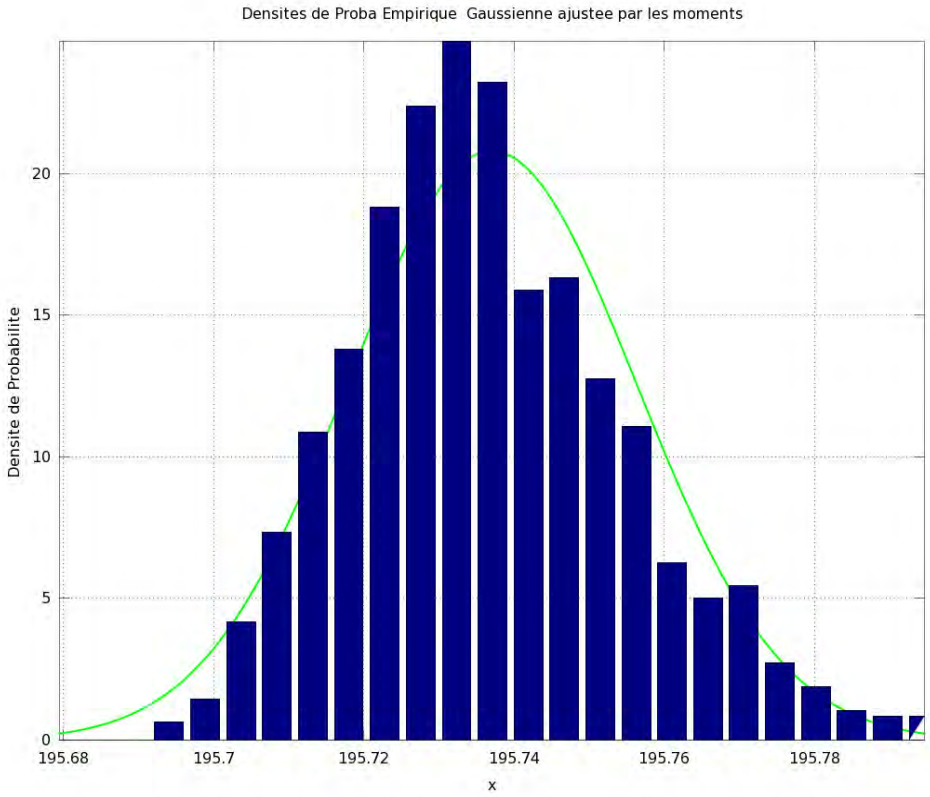


Figure 10: Densité de probabilité pour le lit majeur pendant la montée de crue.

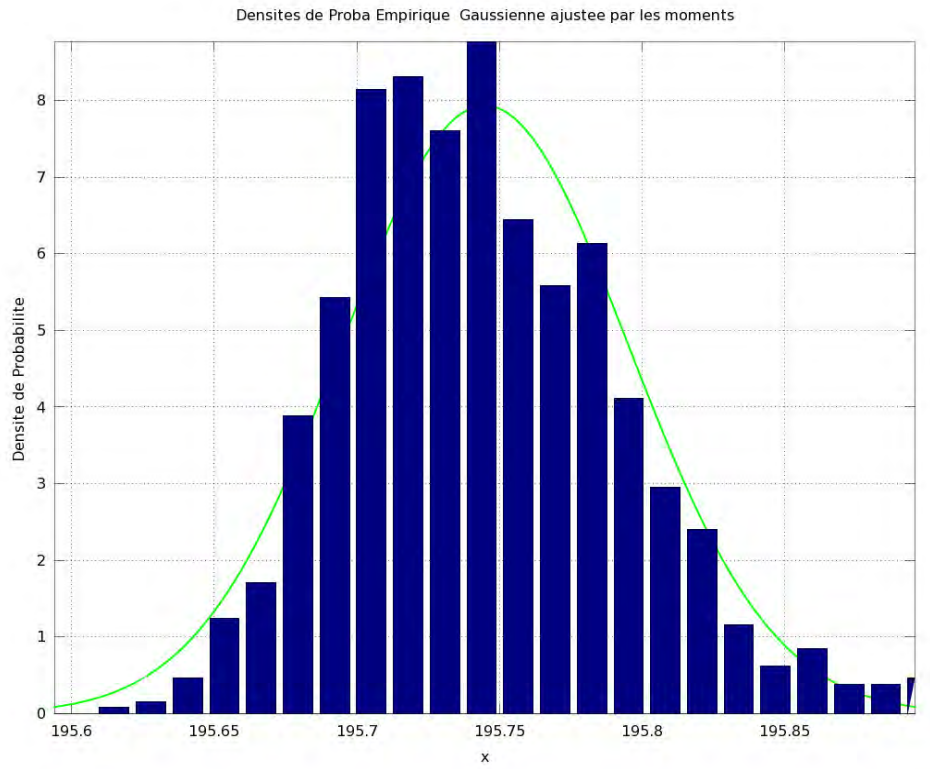


Figure 11: Densité de probabilité pour le lit mineur pendant la montée de crue.

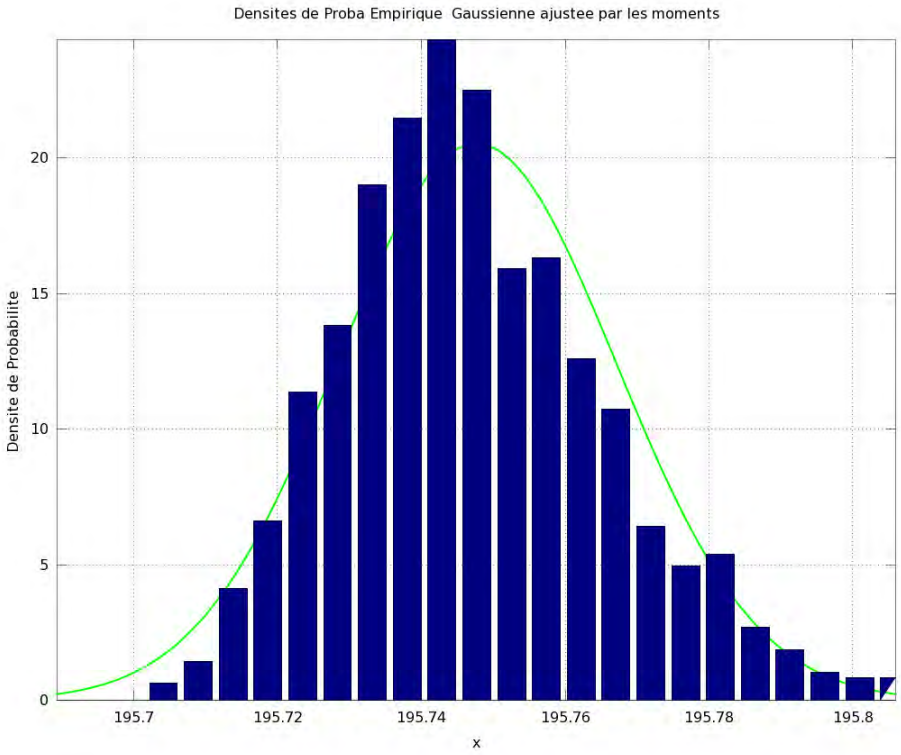


Figure 12: Densité de probabilité pour le lit majeur pendant la décrue.

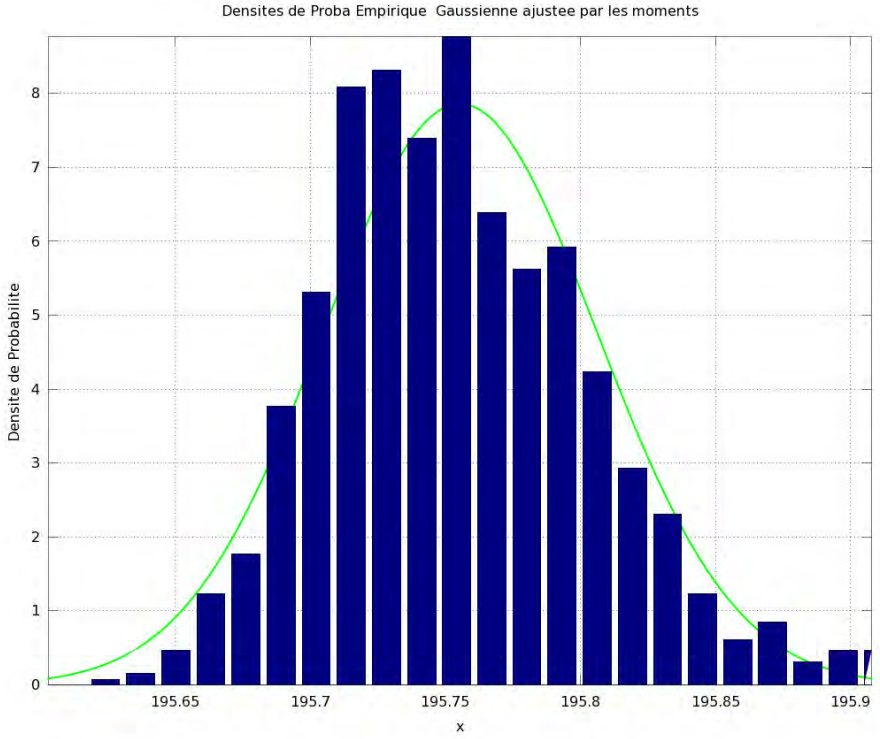


Figure 13: Densité de probabilité pour le lit mineur pendant la décrue.

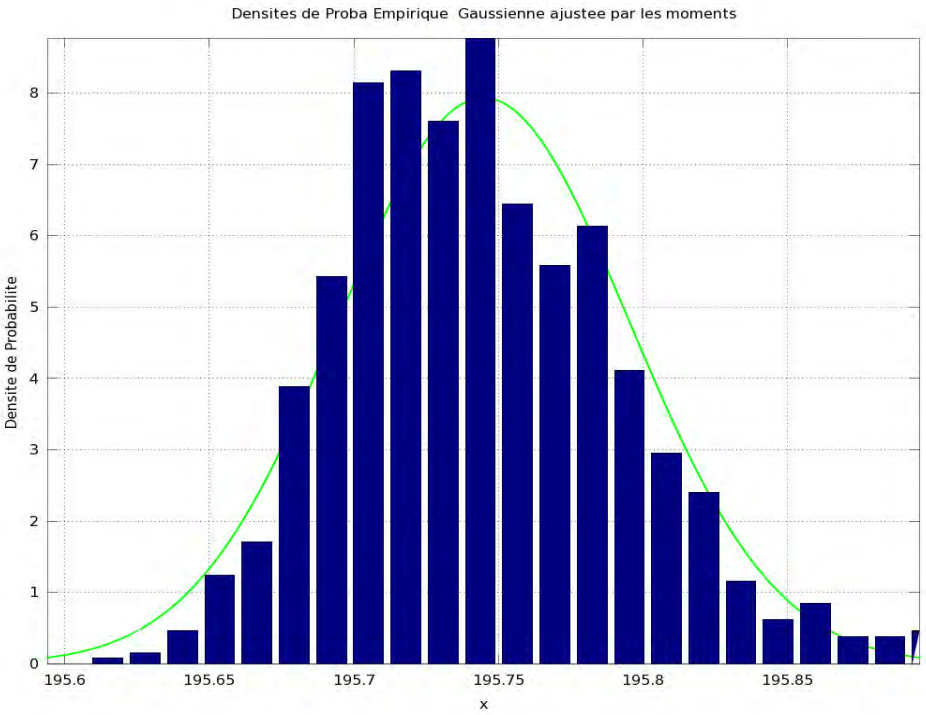


Figure 14: Densité de probabilité pour le lit mineur pendant la crue.

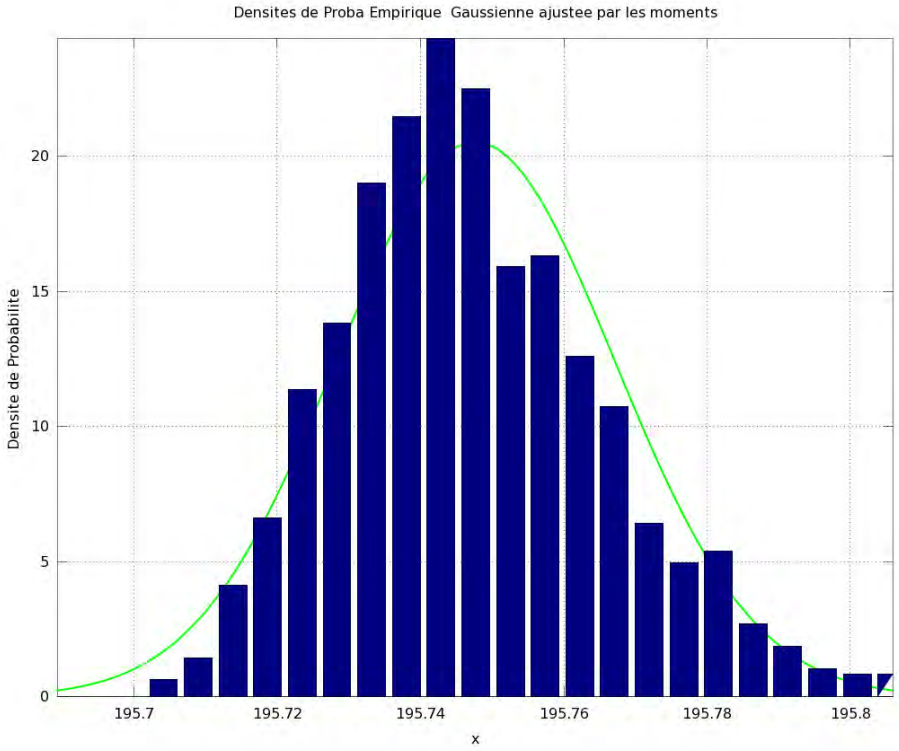


Figure 15: Densité de probabilité pour le lit majeur pendant la crue.

Les figures 10 à 15 montrent que la non-linéarité observée en régime permanent se retrouve aussi en régime non-permanent.

5. Conclusion

Cette étude a permis de mettre en évidence la sensibilité du code MASCARET par rapport aux valeurs de Ks. La non-linéarité des équations de Saint-Venant ainsi que la non-régularité de ces équations rendent le code sensible à la perturbation des Ks (figure 6, 7, 10, 11, 12, 13, 14, 15). On a pu aussi voir qu'il existe une corrélation forte entre les frottements et les conditions limites amont. La variation temporelle des conditions aux limites implique une variation temporelle des Ks, une correction locale de la valeur des Ks par un algorithme d'assimilation de données pourrait améliorer essentiellement les hauteurs d'eau simulées et prévues du modèle. L'étude menée par J.Habert (TR n°1) montre que la correction locale du coefficient de frottement pallie à un manque d'information sur la bathymétrie. Ceci montre qu'une étude semblable devrait être menée pour voir la sensibilité du code MASCARET par rapport à la bathymétrie.

Annexe :

La figure 16 présente la PDF calculée en régime permanent grâce à 10.000 membres. Ceci valide le fait qu'une convergence statistique est obtenue pour 1000 membres, comme expliqué dans la partie 3.2. D'un autre côté on remarque l'asymétrie causée par la non-linéarité de la hauteur d'eau par rapport au Ks.

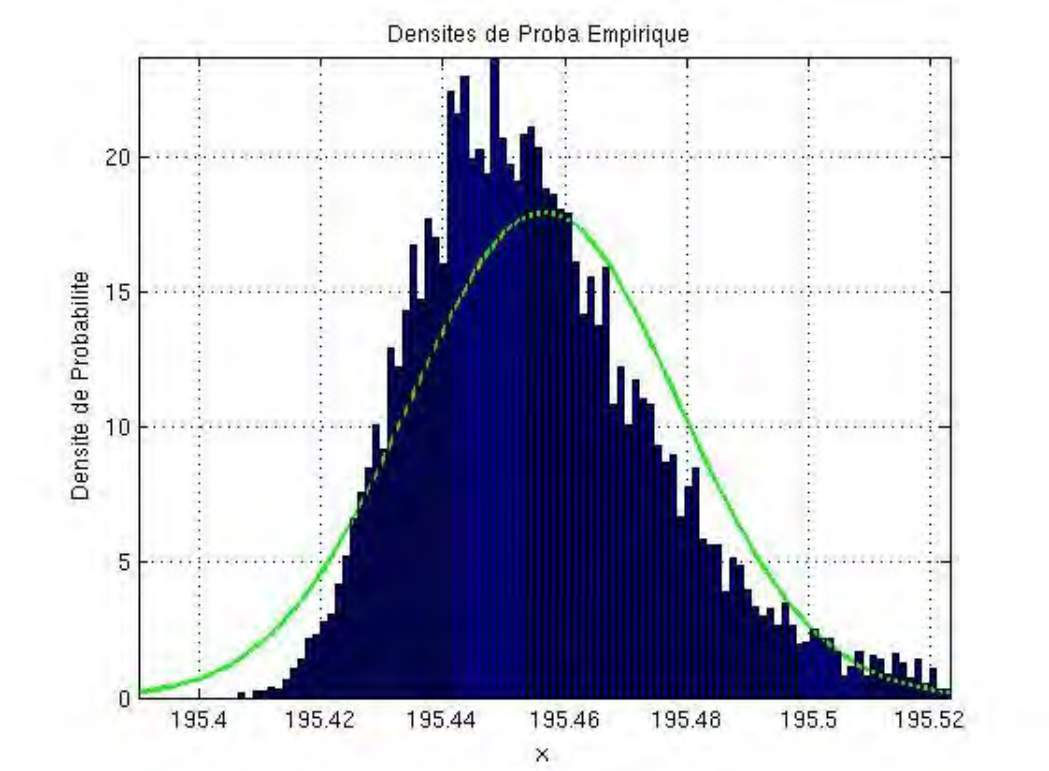


Figure 16: Pdf des hauteurs d'eau calculée en régime permanent avec 10.000 membres.

.2 Rapport Technique : Etude d'incertitudes autour de l'équation de Manning-Strickler

Etude d'incertitude autour de l'équation de Manning-Strickler

4 mai 2015

Table des matières

1 Introduction **1**

2 Dérivation de l'équation de Manning-Strickler **2**

3 Paramètres incertains, représentation statistique de l'incertitude **4**

4 Comparaison entre les résultats théoriques et des simulations Monte Carlo avec Manning-Strickler **5**
- 4.1 Distribution Uniforme 6
- 4.2 Distribution exponentielle 8
- 4.3 Distribution gaussienne 9

5 Comparaison entre les résultats théoriques de l'équation de Manning-Strickler et les résultats de Mascaret pour un cas Canal puis pour la cas de la Marne avec une approche stochastique **12**
- 5.1 Cas avec un Canal satisfaisant les conditions de l'équation de Manning-Strickler 13
- 5.2 Cas avec un Canal régime Uniforme, C.L non satisfaisant l'équation de Manning-Strickler 13
- 5.3 Cas réelle de la Marne-bathymétrie réelle, régime permanent 14

6 Conclusion **17**

1 Introduction

L'étude d'incertitude sur les sorties de modèles tels que Mascaret et Telemac est importante pour les décisions opérationnelles. L'incertitude sur les sorties simulées aux équations Saint-Venant est liée à l'incertitude sur les paramètres d'entrée de ce code. Dans notre cas, on s'intéresse à l'incertitude liée au coefficient de Strickler Ks et à son impact sur la hauteur d'eau. Dans ce rapport, on s'intéresse à l'approximation

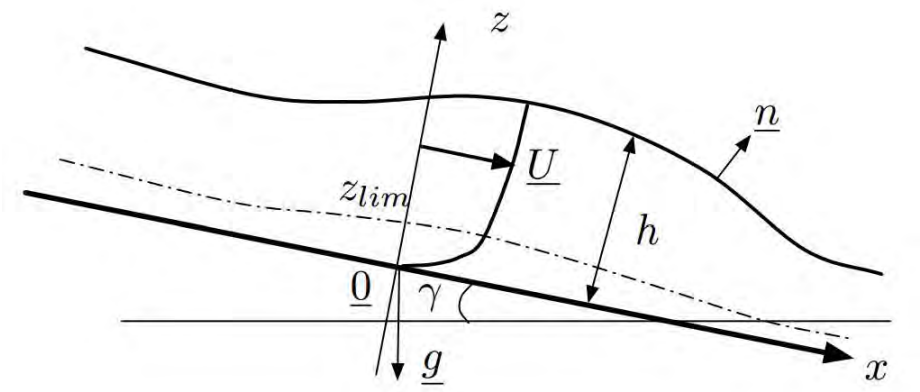
onde de crue, en utilisant l'équation de Manning-Strickler pour un régime permanent et une pente constante pour une géométrie de type "Canal", dont les valeurs numériques approchent les valeurs moyennes du tronçon de la Marne (TR1). Pour ce cas idéalisé, cette équation possède une solution analytique, ce qui permet d'exprimer la forme analytique de la densité de probabilté (PDF) des hauteurs d'eau en fonction de la densité de probabilité des Ks. On comparera dans un premier temps la solution analytique à celle obtenue stochastiquement sur un ensemble de réalisations de l'équation de Manning-Strickler.

2 Dérivation de l'équation de Manning-Strickler

L'équation de Manning-Strickler est utilisée pour paramétriser les frottements dans les équations de Saint-Venant. On considère un écoulement à surface libre sur un plan incliné, le modèle des équations de Saint-Venant s'écrit :

$$\frac{\partial h}{\partial t} + U\frac{\partial h}{\partial x} = -h\frac{\partial U}{\partial x},$$

$$\frac{\partial U}{\partial t} + U\frac{\partial U}{\partial x} = -g'\frac{\partial h}{\partial x} + gI - \frac{C_f}{2}\frac{U|U|}{h}. \quad (1)$$



h	Hauteur d'eau
U	Vitesse de l'écoulement
$\frac{\partial h}{\partial x}$	La pente de la surface libre
g	L'accelération de la pesanteur
I= sin γ	La pente du fond
C_f	Le coefficient de frottement.

L'approximation des ondes de crues consiste à prendre la force de frottements égale au terme due à la pente gravitaire (On néglige les termes d'accélération et de pression). Donc, on retrouve l'équation suivante :

$$gI = \frac{C_f}{2}\frac{U|U|}{h}. \tag{2}$$

Le coefficient C_f est un coefficient adimensionnel qui permet de définir le cisaillement au fond de la rivière. En hydraulique à surface libre on utilise souvent l'équation de Manning-Strickler pour modéliser C_f, qui s'écrit de la manière suivante :

$$C_f = \frac{2g}{K_s^2 h^{1/3}}. \tag{3}$$

Finalement, on remplace C_f par l'équation (3) dans l'équation (2) et on retrouve l'équation de Manning-Strickler :

$$U = K_s h^{2/3} I^{1/2},$$

$$q = Uh = K_s h^{5/3} I^{1/2},$$

$$h = \left(\frac{q}{K_s I^{1/2}} \right)^{3/5}. \tag{4}$$

— On suppose que la rivière à la forme d'un canal donc q=Q/W avec q : le débit linéique en (m^2/s), Q : le débit surfacique en (m^3/s) et W : la largeur de la rivière en (m) que l'on supposera constante.

— On suppose aussi que la pente I est constante sur tout le canal.

— Dans la suite on utilisera l'équation de Manning-Strickler sous la forme de l'équation (4).

3 Paramètres incertains, représentation statistique de l'incertitude

On se place dans le régime permanent i.e q=Uh=cste. Et on cherche à quantifier l'incertitude sur la hauteur d'eau. On suppose que Ks est une variable aléatoire, et f_{Ks} sa densité de probabilité. On cherche à estimer la densité de probabilité de h notée f_h en fonction de celle de f_{Ks} en supposant que h et Ks sont reliés par l'équation de Manning-Strickler.

Propriété 1. *Si X est une variable aléatoire dont la densité de probabilité est $f_X(x)$ et Y une deuxième variable aléatoire telle qu'il existe une fonction g monotone et dérivable avec $g(X) = Y$. Alors la densité de probabilité de Y s'écrit sous la forme suivante :*

$$f_Y(y) = \frac{1}{|g'(g^{-1}(y))|} . f_X(g^{-1}(y)). \tag{5}$$

Dans le cas présent X représente K_s et Y représente h. Selon l'équation (4) on a

$h = (\frac{q}{K_s I^{1/2}})^{3/5}$, on pose a= $(\frac{q}{I^{1/2}})$, puisque l'on est en régime permanent uniforme. La fonction $g(x)$, dans ce cas s'écrit de la forme suivante : $g(x) = (\frac{a}{x})^{3/5}$. La fonction g est décroissante et dérivable sur $\mathbb{R}^{+*}$. De plus on a g(Ks)=h=$(\frac{a}{K_s})^{3/5}$. Donc d'après l'équation (5) on trouve la densité de probabilité de la hauteur d'eau par rapport à celle de la distribution des Ks.

$$g^{-1}(h) = \frac{a}{h^{3/5}},$$
$$g'(K_s) = -\frac{3}{5}\frac{a^{3/5}}{K_s^{8/5}}.$$

Finalement on trouve :

$$f_h(h) = \frac{5}{3}\frac{a}{h^{8/3}} . f_{Ks}(\frac{a}{h^{5/3}}). \tag{6}$$

4 Comparaison entre les résultats théoriques et des simulations Monte Carlo avec Manning-Strickler

On cherche à comparer dans cette partie les résultats théoriques obtenus à l'aide de l'équation (6), et les résultats obtenus avec l'approche stochastique sur un cas idéalisé de type "Canal" à pente constante en régime permanent uniforme dont les propriétés géométriques sont une approximation de celle du tronçon de la Marne. On choisit Q= 150 m^3/s, pour une largeur de canal L=100m. La pente est fixée à la valeur moyenne de celle du tronçon de la Marne, c'est à dire 1.2E-3. On tire un échantillon des Ks selon une loi donnée (uniforme, exponentielle ou normale), puis on calcule les hauteurs d'eau correspondantes via l'équation de Manning-Strickler sur le cas idéalisé. On représente alors la PDF en sortie des hauteurs d'eau sur laquelle on trace celle issue de l'équation théorique (équation 6).

4.1 Distribution Uniforme

On suppose que Ks suit une distribution uniforme. Donc $f_{Ks}(Ks) = \mathbf{1}_{[c\ b]}(K_s)$ avec $\mathbf{1}_{[c\ b]}(x) = \begin{cases} 1 & \text{si } x \text{ appartient à } [c\ b], \\ 0 & \text{sinon} \end{cases}$.

On prend c=7 et b=33. Selon l'équation (6) on peut trouver la forme de la PDF des hauteurs d'eau en sortie de l'équation de Manning-Strickler :

Tout d'abord on cherche l'intervalle de l'indicatrice de la fonction uniforme :

$$\begin{aligned} 7 \leq K_s \leq 33 \quad &\Longleftrightarrow \quad 7 \leq \tfrac{a}{h^{5/3}} \leq 33 \\ &\Longleftrightarrow \quad (\tfrac{a}{33})^{3/5} \leq h \leq (\tfrac{a}{7})^{3/5} \\ &\Longleftrightarrow \quad 1.1 \leq h \leq 2.8. \end{aligned}$$

Finalement on retrouve l'expression analytique de la PDF de la hauteur d'eau pour ce cas là :

$$f_h(h) = \frac{5a}{3h^{8/3}} \cdot \frac{1}{b-c} \mathbf{1}_{[1.1\ 2.8]}. \tag{7}$$

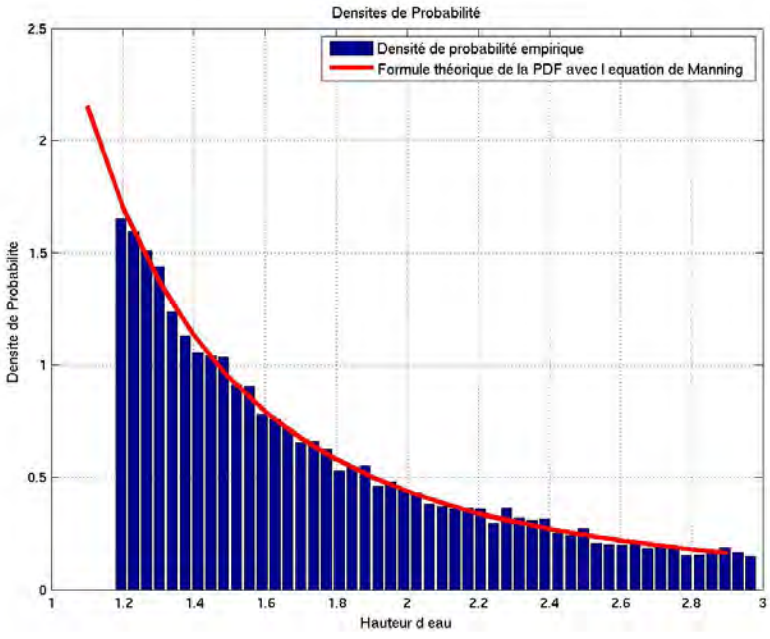


FIGURE 1 – Comparaison entre simulation Monte-Carlo et l'équation 7 pour une distribution en entrée uniforme

La figure (1) représente la PDF en sortie des hauteurs d'eau pour une distribution uniforme des Ks, en utilisant l'équation (7) (courbe rouge) et un tirage Monte Carlo (courbe bleue) de 10000 membres effectué sur les Ks selon une distribution uniforme. Ci dessous on rappelle la définition des premiers moments statistiques calculés stochastiquement et puis à l'aide de l'expression théorique :

— Moyenne calculée avec une loi continue : $E(h) = \int_{\mathbf{R}} h f_h(h) dh$.

— Estimation de la moyenne calculée à l'aide d'un échantillon : $\widehat{E(h)} = \frac{1}{N} \sum_{k=1}^{N} h$

.

— Variance calculée avec une loi continue : $Var(h) = \int_{\mathbf{R}} (h - E(h))^2 f_h(h) dh.$

— Estimation de la variance calculée à l'aide d'un échantillon : $\widehat{Var(h)} = \frac{1}{N} \sum_{k=1}^{N} (h - E(h))^2.$

	moyenne	écart-type
Moment stochastique	1.73	0.47
Moment théorique	1.81	0.5

Les résultats avec l'approche stochastique convergent correctement vers les moments théoriques en utilisant un échantillon de 1000 membres. De plus on remarque que la PDF théorique (rouge) et stochastique (bleue) sont en accord. On note qu'un échantillon de 1000 membres permet la convergence du calcul statistique pour le cas test présent.

On remarque que la PDF des sorties n'est pas uniforme contrairement à celle imposée sur les coefficients Ks. Ceci traduit le comportement non linéaire de l'expression de la hauteur d'eau via l'équation de Manning-Strickler en fonction du Ks. Une pertubation négative du Ks n'a pas le même effet qu'une perturbation positive. Dans un cas linéaire on s'attend à avoir une densité de probabilité uniforme en sortie. L'équation de Manning-Strickler ne réagit pas uniformément à chacune des perturbations du K_s, la

hauteur d'eau réagit plus à une perturbation négative qu'à une perturbation positive du K_s.

La moyenne de l'échantillon vaut 1.73 m alors que la médiane vaut 1.59 m, donc le comportement du système tend à réagir fortement aux perturbations négatives plus qu'aux perturbations positives ce qui représente la non-linéaité décrite par l'équation de Manning-Strickler.

4.2 Distribution exponentielle

On utilise dans cette partie une distribution exponentielle pour le tirage des Ks. La loi exponentielle est une loi continue avec un paramètre λ qui est l'inverse de la moyenne. La densité de probabilité de la loi exponentielle est : $f(x) = \lambda e^{-\lambda x}$.

Afin d'échantilloner une loi exponentielle, on utilise la propriété suivante : Si u est une variable aléatoire de distribution uniforme, alors la variable $-\frac{1}{\lambda} ln(1-u)$ suit une loi exponentielle de paramètre λ.

Soit Ks suit un loi exponentielle de paramètre λ , à l'aide de l'équation (6) on retrouve la densité de probabilité de la hauteur d'eau :

$$f_h(h) = \frac{5a\lambda}{3h^{8/3}} \exp(-\frac{\lambda a}{h^{5/3}}). \tag{8}$$

La figure (2) représente la PDF en sortie des hauteurs d'eau pour une distribution exponentielle des Ks, en utilisant l'équation (8) (courbe rouge) et un tirage Monte Carlo (100.000 membres) des K_s transformés en hauteurs d'eau grâce à l'équation de Manning-Strickler. On remarque sur la figure 2 que l'équation (6) permet de bien reproduire la PDF des hauteurs d'eau pour l'équation de Manning-Strickler comme pour le cas précédent.

Même si à première vue, la PDF des hauteurs d'eau ressemble à une loi exponentielle, l'expression des moments statistiques (particulièrement les coefficients d'aplatissement et d'asymétrie) montre qu'elle ne l'est pas.

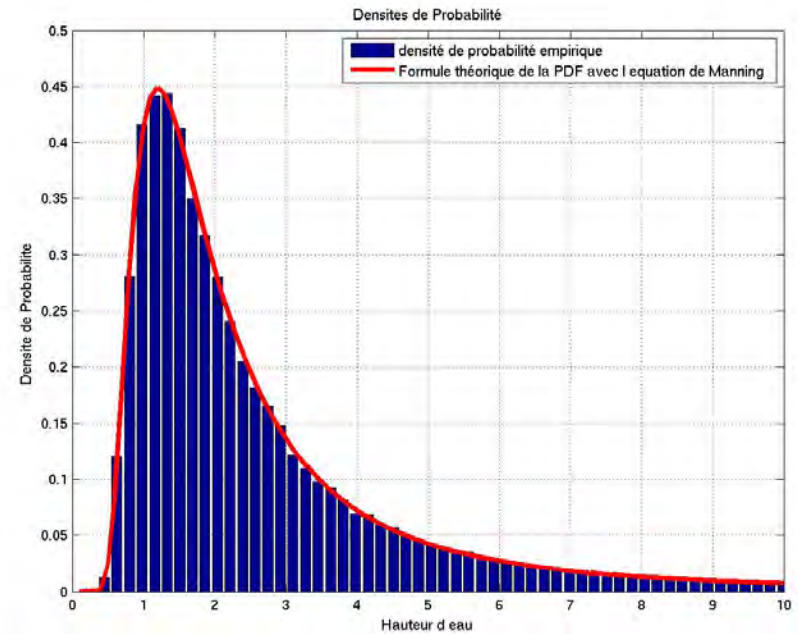


FIGURE 2 – Comparaison entre simulation Monte-Carlo et l'équation 8 pour une distribution en entrée exponentielle

	moyenne	écart-type	asymétrie	aplatissement
PDF des hauteurs d'eau	3.5	0.47	7.12	66.03
PDF exponentielle	$\frac{1}{\lambda}$=3.5	$\frac{1}{\lambda}$=3.5	2	6

On conclut que l'équation de Manning-Strickler rend la PDF des hauteurs d'eau asymétrique positive.

Finalement, à l'aide des moments calculés stochastiquement on montre que la PDF des hauteurs d'eau pour une perturbation exponentielle des K_s n'est pas exponentielle, ceci confirme le comportement non-linéaire de l'équation de Manning-Strickler.

4.3 Distribution gaussienne

On utilise dans cette partie une distribution gaussienne pour le tirage des Ks. La loi normale est une loi continue à deux paramètres : m (la moyenne) et σ (l'écart-type). La densité de probabilité d'une variable aléatoire X qui suit une loi normale s'écrit de la façon suivante :

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{(-\frac{1}{2}.\frac{(x-m)^2}{\sigma^2})}.$$

On suppose que Ks suit une loi normale de paramètres : m(20) et σ (3.64). On retrouve, à l'aide de l'équation (6), l'expression de la densité de probabilité de la hauteur d'eau :

$$f_h(h) = \frac{5a}{3\sigma\sqrt{2\pi}h^{8/3}} \exp((-\frac{1}{2}.\frac{(a-h^{5/3}m)^2}{(h^{5/3}\sigma)^2})). \tag{9}$$

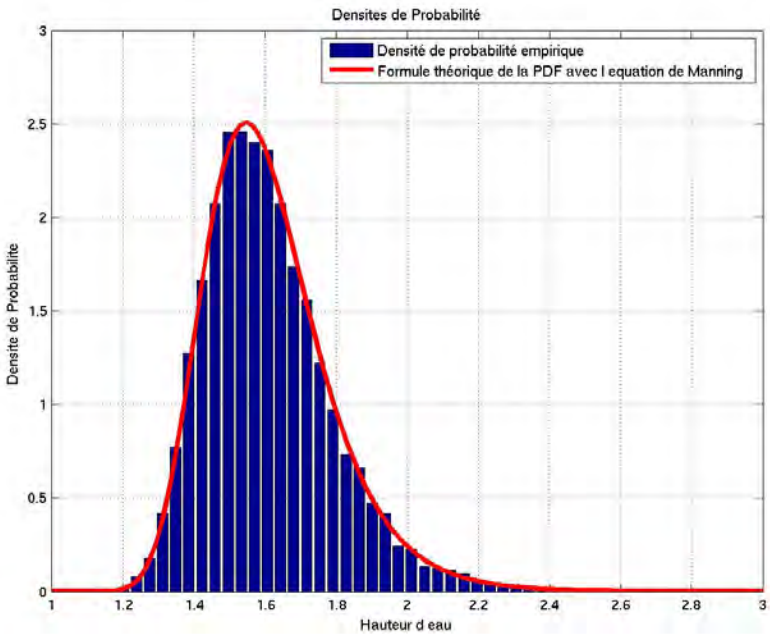


FIGURE 3 – Comparaison entre simulation Monte-Carlo et l'équation (9) pour une distribution en entrée gaussienne

La figure (3) représente la densité de probabilité empirique à partir d'un tirage Monté Carlo des Ks suivant une loi normale, et la densité de probabilité théorique obtenue à l'aide de l'équation (9). La formule théorique représente bien la densité empirique ce qui valide l'équation (6) pour le cas Manning-Strickler. On regarde les moments estimés de manière stochastique de la PDF des hautteur d'eau :

	moyenne	médiane	écart-type	asymétrie	aplatissement
PDF des hauteurs d'eau	1.61	1.58	0.19	0.94	1.63

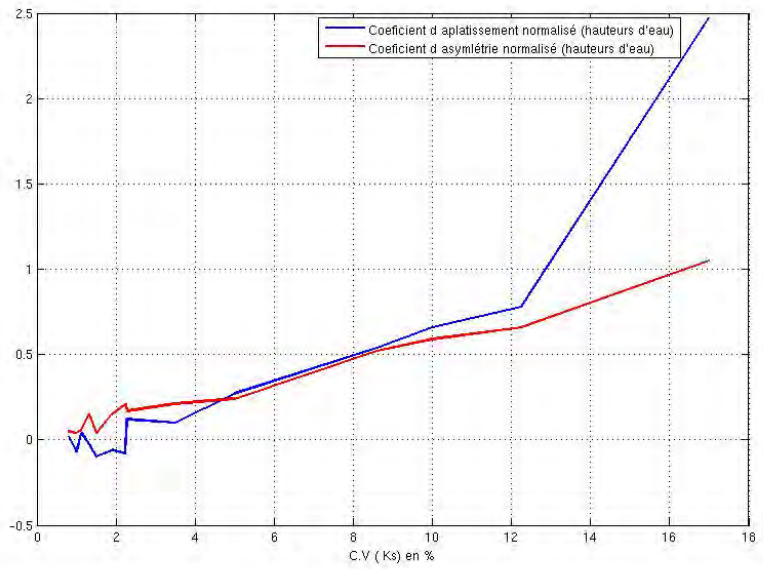


FIGURE 4 – Coefficient d'aplatissement et d'asymétrie de la PDF en sortie en fonction du Coefficient de variation de la PDF en entrée

On remarque que la valeur de l'écart-type est petite, avec un coefficient d'asymétrie positive, ce qui veut dire que la réaction de l'équation Manning-Strickler n'est pas linéaire par rapport à une perturbation positive ou négative du Ks : une perturbation positive du Ks ($+dKs$) induirait une perturbation négative sur la hauteur d'eau($-dh_{pos}$), mais une perturbation négative du Ks (de valeur absolue égale à la perturbation positive $-dKs$) induirait une perturbation positive de la hauteur d'eau (de valeur absolue différente que la précédente $+dh_{neg}$ et $dh_{neg} \neq dh_{pos}$). D'autre part, on remarque un coefficient d'applatissement assez grand, ce qui veut dire que la densité de probabilité est pointue par rapport à une gaussienne. Ceci montre que le comportement du système n'est pas linéaire.

Les trois cas d'étude précédents montrent que l'expression de la hauteur d'eau en fonction du Ks via l'équation de Manning-Strickler est non-linéaire. Cette conclusion

est importante pour la suite des travaux en assimilation de données. En effet, dans le cadre des algorithmes d'assimilation de données, on se place généralement sous l'hypothèse linéaire. Ici, il est donc nécessaire de faire une hypothèse de linéarité locale. Afin de déterminer les limites de cette hypothèse, on s'intéresse à l'ordre de grandeur de la perturbation par rapport à la valeur moyenne du Ks. On trace sur la figure 4 les coefficients d'asymétrie et d'aplatissement de la densité de probabilité des hauteurs d'eau en fonction du coefficient de variation des Ks. On définit le coefficient de variation en statistique par le rapport de l'écart-type par rapport à la moyenne ($C.V = \frac{\sigma}{m}$). Ce coefficient permet de mesurer la dispersion d'une variable aléatoire et présente l'avantage d'être un coefficient adimensionnel. On remarque pour des variations inférieures à 12% les coefficients d'asymétrie et d'aplatissements sont presque nuls, pour ce cas la PDF des hauteurs d'eau est une gaussienne, et donc le système est linéaire. On note donc que pour des perturbations de l'ordre de 12% autour d'une valeur de référence qui vaut ici 20 $m^{1/3}s^{-1}$, la réponse est linéaire.

5 Comparaison entre les résultats théoriques de l'équation de Manning-Strickler et les résultats de Mascaret pour un cas Canal puis pour la cas de la Marne avec une approche stochastique

Dans cette section on compare les résultats de la PDF des sorties du code Mascaret avec l'équation 6, afin de quantifier la non-linéarité due aux frottements. Dans un premier temps, on se mettra dans les conditions de Manning-Strickler de façon à avoir la surface libre parallèle à la pente sur un canal rectangulaire satisfaisant les conditions de l'équation de Manning-Strickler, Puis dans un second temps, on utilisera ce canal en changeant les conditions aux limites de sorte de ne plus satisfaire l'équilibre décrit par l'équation de Manning-Strickler. Puis finalement on s'intéressera aux résultats de Mascaret pour au cas de la Marne.

5.1 Cas avec un Canal satisfaisant les conditions de l'équation de Manning-Strickler

Afin de se placer dans les conditions de l'équations de Manning-Strickler on doit imposer une géométrie uniforme et un régime permanent donc $\frac{\partial h}{\partial t} = 0 et \frac{\partial U}{\partial t} = 0$. Les équation de Saint-Venant se réduisent donc à l'équation suivante : $-g' \frac{\partial h}{\partial x} + gI - \frac{C_f}{2} \frac{U|U|}{h} = 0$. Pour que l'on retrouve l'équation de Manning-Strickler il faut que la hauteur d'eau h soit uniforme sur tout le tronçon. Donc pour éviter d'avoir des remous il faut imposer des conditions aux limites qui respectent les conditions de Manning-Strickler. Dans le cas du canal précédant (géométrie uniforme et régime permanent i.e L=100m et I=2.5E-4), on impose (Q=55,16 m^3/s et h=1m), ce qui permet de simuler un écoulement permanent uniforme i.e (Q(t)=55,16 m^3/s à l'aval et h=1m à l'amont car on est en régime fluvial). On peut noter que pour cette configuration, il est légitime de faire l'hypothèse des grandes largeurs ($Rh = 0.98m$). La figure 6 permet de constater la non-linéarité des équations de Saint-Venant due à la paramétrisation du frottement, ainsi qu'elle valide l'équation 6 pour le code Mascaret dans le cas où on impose des conditions similaires à l'équation de Manning-Strickler.

5.2 Cas avec un Canal régime Uniforme, C.L non satisfaisant l'équation de Manning-Strickler

Dans cette partie, on garde la même géométrie du canal que précédemment et on impose des conditions à l'amont et à l'aval de sorte que ça ne satisfait pas l'équation de Manning-Strickler (Q=1000m^3/s et h=5m). On note aussi que l'approximation grande largeur ($Rh = h$) reste discutable dans ce cas car on commet une erreur de 6% sur l'estimation du débit ($Q_{Rh=h} = 767m^3/s; Q_{Rh\neq h} = 817m^3/s$). On regarde à l'issue la PDF calculée stochastiquement à l'aide du code Mascaret.

La figure 6 représente la PDF estimée stochastiquement des sorties de hauteurs d'eau, on remarque que la courbe de l'équation 6 correspond à la PDF issue de Mascaret malgré le fait que les CL ne soient pas en adéquation avec l'équation de Manning-Strickler. En effet, sur la figure 7 on remarque que la ligne d'eau est parallèle à la pente

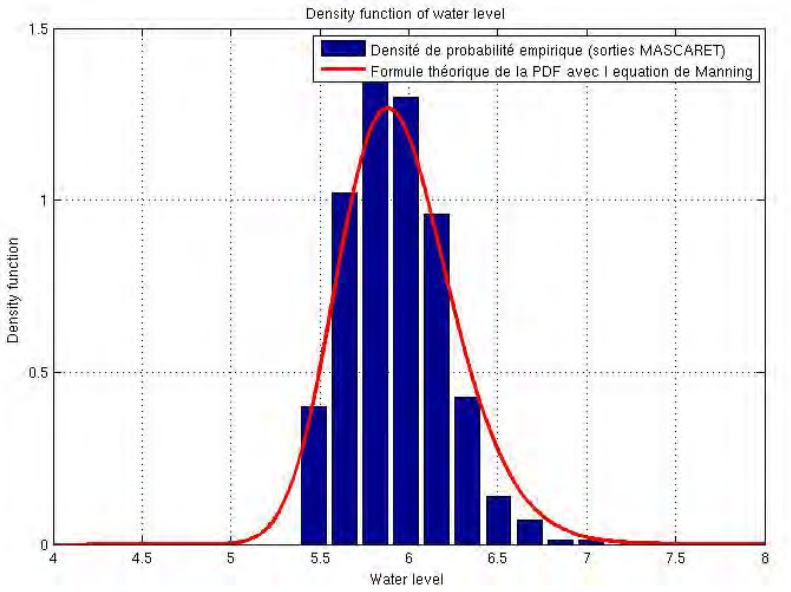


FIGURE 5 – Densité de probabilité pour les sorties Mascaret pour un canal en régime uniforme et l'équation de Manning-Strickler

du canal, sauf pour l'amont ou il existe une zone de remous. Cependant, on calcule la PDF au milieu du domaine, or au centre du canal on remarque que l'écoulement est parallèle à la bathymétrie ce qui justifie la correspondance de la PDF à la sortie de Mascaret avec la PDF analytique calculée avec l'équation de Manning-Strickler.

5.3 Cas réelle de la Marne-bathymétrie réelle, régime permanent

Dans cette partie, on souhaite comparer la PDF issue du MASCARET sur un cas avec une bathymétrie complexe avec un régime permanent (Q=150m^3/s), et la PDF analytique de l'équation de Manning-Strickler (équation 6),.

La figure 8 représente la densité de probabilité que l'on obtient avec Mascaret lorsque l'on perturbe les Ks de manière gaussienne, et on regarde les hauteurs d'eau à la sortie. La courbe en rouge représente la densité de probabilité théorique développée à l'aide de l'équation de Manning-Strickler sachant la distribution des Ks. Dans ce cas les Ks sont tirés de manière gaussienne, donc on utilise l'équation 6 pour caractériser la densité de probabilité de la hauteur d'eau suivant l'équation de Manning-Strickler.

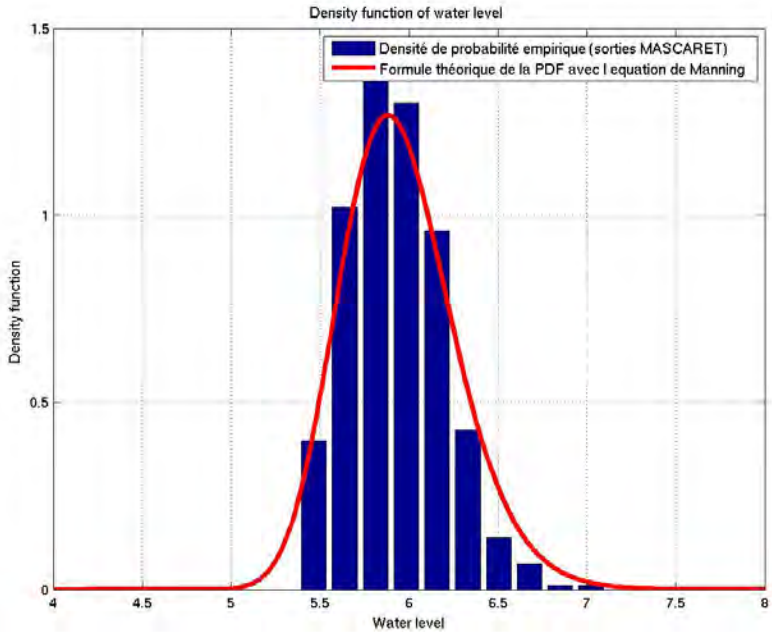


FIGURE 6 – Densité de probabilité pour les sorties Mascaret pour un canal en régime uniforme et l'équation de Manning-Strickler

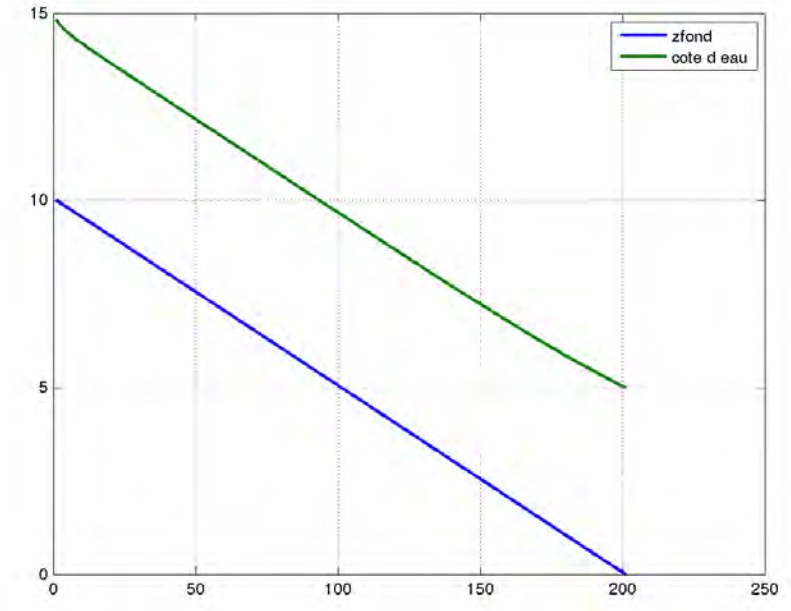


FIGURE 7 – Exemple côte d'eau de la sortie Mascaret

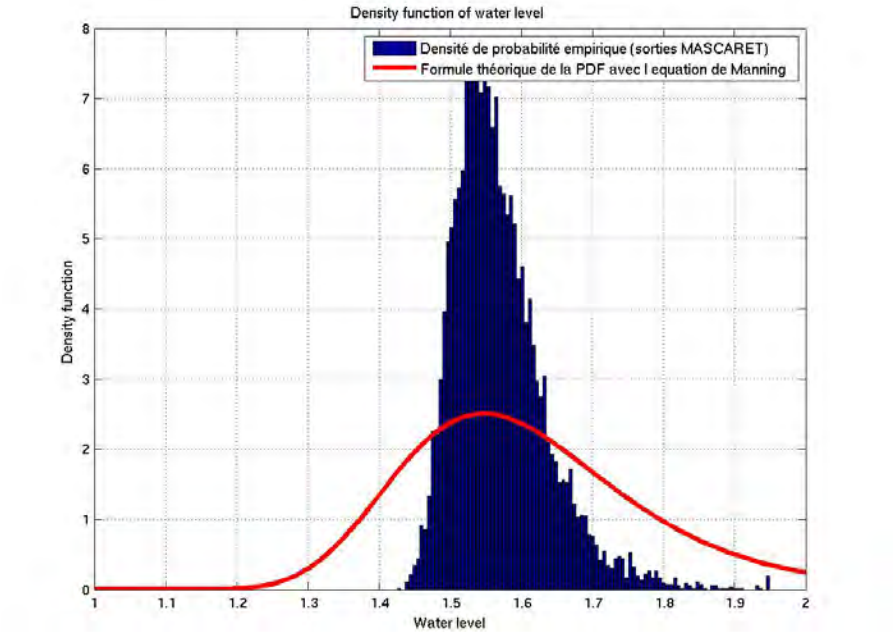


FIGURE 8 – Densité de probabilité pour les sorties Mascaret du cas de la Marne et l'équation de Manning-Strickler

On perturbe les Ks avec une loi normale de moyenne 20 et de variance 12. Les deux courbes sont différentes, le cas de la Marne n'obéit pas forcément aux lois de l'équation de Manning-Strickler. En effet, le cas de la Marne est un cas avec une géométrie complexe due à l'existence d'une bathymétrie complexe. Bien que le régime soit permanent, l'existence de cette bathymétrie va induire des courbes de remous et donc l'écoulement ne sera pas nécessairement parallèle à la pente du fond d'ou la différence avec l'équation de Manning. On remarque aussi l'effet d'asymétrie qui apparait bien sur les sorties de Mascaret, ainsi que sur les courbes théoriques.

Regardons les premier moments statistiques pour les PDF :

	moyenne	écart-type	asymétrie	aplatissement
PDF Mascaret	1.5	0.06	1.29	3.03
PDF Manning-Strickler	1.62	0.18	1.05	2.50

Les deux premiers moments statistiques (la moyenne et l'écart-type) montrent une

ressemblance entre les deux PDF (Manning-Strickler et Mascaret avec le cas Marne). Les deux systèmes sont non-linéaires par rapport à une perturbation de Ks. Par contre les coefficients d'asymétrie et d'aplatissement sont différents (les deux plus grand pour Mascaret pour le cas Marne que pour Manning-Strickler), l'ajout de la bathymétrie réelle à engendrer plus de déformations sur la PDF et donc l'hypothèse de l'équation de Manning-Strickler n'est plus valable. Ceci indique que l'on ne peut pas estimer la PDF à la sortie de Mascaret à l'aide de l'équation 6 et que l'approche stochastique est indispensable pour connaitre la forme de la PDF.

6 Conclusion

L'équation de Manning-Strickler permet de mettre en évidence la non-linéarité du système Saint-Venant. Son écriture théorique permet de trouver une relation entre la PDF des Ks et la PDF de la hauteur d'eau. L'équation (6) donne le lien entre les deux PDF. On remarque qu'il existe un coefficient entre les deux PDF qui n'est pas constant d'où la non-linéarité du système. Finalement, on conclut que pour une perturbation positive ou négative de la valeur du Ks, la réponse à cette perturbation sera différente dans les deux cas. Cette étude permet aussi de dire que l'approximation de la linéarité de l'équation de Manning-Strickler pour un algorithme d'assimilation de données, par exemple dans le cas du filtre de Kalman où l'on souhaite supposer que le code soit linéaire, n'est valable que pour un coefficient de variation autour de 10% pour une valeur moyenne du Ks du 20 $m^{1/3}s^{-1}$. Finalement, la comparaison avec la PDF à la sortie du code de Mascaret permet de dire que l'équation de Manning-Strickler explique en partie la non-linéarité des équations de Saint-Venant par rapport aux variations du Ks (Canal régime uniforme), par contre l'ajout d'une bathymétrie complexe déforme la PDF en sortie de Mascaret et elle ne peut être estimée à l'aide de l'équation 6. Dans ce cas une approche stochastique est nécessaire pour quantifier les incertitudes.

Bibliographie

[1] F. Alcrudo and P. Garcia-Navarro. A high-resolution godunov-type scheme in finite volumes for the 2d shallow-water equations. *International Journal for Numerical Methods in Fluids*, 16(6) :489–505, 1993.

[2] F. Alcrudo, P. Garcia-Navarro, and J.-M. Saviron. Flux difference splitting for 1d open channel flow equations. *International Journal for Numerical Methods in Fluids*, 14(9) :1009–1018, 1992.

[3] A. Alexanderian. A brief note on the karhunen-loève expansion. *arXiv preprint*.

[4] A. Alexanderian, J. Winokur, I. Sraj, A. Srinivasan, M. Iskandarani, W. C. Thacker, and O. M. Knio. Global sensitivity analysis in an ocean general circulation model : a sparse spectral projection approach. *Computational Geosciences*, 16(3) :757–778, 2012.

[5] A. Arakawa, Y. Mintz, and A. Katayama. *Numerical simulation of the general circulation of the atmosphere*. Department of Meteorology, University of California, 1968.

[6] R. Askey and J. A. Wilson. *Some basic hypergeometric orthogonal polynomials that generalize Jacobi polynomials*, volume 319. American Mathematical Soc., 1985.

[7] K. E. Atkinson. *The numerical solution of integral equations of the second kind*, volume 4. Cambridge university press, 1997.

[8] S. Barthélémy. *Assimilation de donnée ensembliste et couplage de modèles hydrauliques 1D-2D pour la prévisiond es crues en temps réel. Application au réseau hydraulique Adour Maritime.* PhD thesis, Institut National Polytechnique de Toulouse, 2015.

[9] A. Bermudez and M. E. Vazquez. Upwind methods for hyperbolic conservation laws with source terms. *Computers Fluids*, 23(8) :1049 – 1071, 1994.

[10] P. Bernardara, E. R. de Rocquigny, N. Goutal, A. Arnaud, and G. Passoni. Uncertainty analysis in flood hazard assessment : hydrological and hydraulic calibrationthis article is one of a selection of papers published in this special issue on hydrotechnical engineering. *Canadian Journal of Civil Engineering*, 37(7) :968–979, 2010.

[11] M. Berveiller. *Eléments finis stochastiques : approches intrusive et non intrusive pour des analyses de fiabilité.* PhD thesis, Université Blaise Pascal, Clermont-Ferrand, 2005.

[12] A. Besnard and N. Goutal. Comparaison de modèles 1D à casiers et 2D pour la modélisation hydraulique d'une plaine d'inondation–Cas de la Garonne entre Tonneins et La Réole. *La Houille Blanche*, (3) :42–47, 2011.

[13] V. Bjerknes. *Das Problem der Wettervorhersage : betrachtet vom Standpunkte der Mechanik und der Physik*. 1904.

[14] A. Björck. *Numerical methods for least squares problems*. Siam, 1996.

[15] G. Blatman. *Adaptative sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis*. PhD thesis, Université BLAISE PASCAL-Clermont, 2009.

[16] G. Blatman and B. Sudret. Anisotropic parcimonious polynomial chaos expansions based on the sparsity-f-effects principle. In *Proc ICOSSAR'09, International Conference in Structural Safety and Relability*, 2009.

[17] G. Blatman and B. Sudret. Adaptative sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics*, 2011.

[18] F. Bouttier and P. Courtier. Data assimilation concepts and methods March 1999. *Meteorological training course lecture series. ECMWF*, 2002.

[19] S. Bozzi, G. Passoni, P. Bernardara, N. Goutal, and A. Arnaud. Roughness and Discharge Uncertainty in 1D Water Level Calculations. *Environmental Modeling & Assessment*, (4), 2014.

[20] T. Braconnier, M. Ferrier, J.-C. Jouhaud, M. Montagnac, and P. Sagaut. Towards an adaptive pod/svd surrogate model for aeronautic design. *Computers & Fluids*, 40(1) :195–209, 2011.

[21] S. F. Bradford and B. F. Sanders. Finite-volume model for shallow-water flooding of arbitrary topography. *Journal of Hydraulic Engineering*, 128(3) :289–298, 2002.

[22] G. W. Brunner. Hec-ras river analysis system. hydraulic reference manual. version 1.0. Technical report, DTIC Document, 1995.

[23] R. H. Cameron and W. T. Martin. The orthogonal development of non-linear functionals in series of fourier-hermite functionals. *Annals of Mathematics*, pages 385–392, 1947.

[24] W. Castaings, F. Le Dimet, and D. Dartus. Assimilation de données en hydrologie. *Société française de photogrammétrie*, 2003.

[25] E. Castillo, A. S. Hadi, A. Conejo, and A. Fernández-Canteli. A general method for local sensitivity analysis with application to regression models and other optimization problems. *Technometrics*, 46(4) :430–444, 2004.

[26] A. Chkifa, A. Cohen, and C. Schwab. Breaking the curse of dimensionality in sparse polynomial approximation of parametric {PDEs}. *Journal de Mathématiques Pures et Appliquées*, 103(2) :400 – 428, 2015.

[27] V.-T. Chow. *Open-channel Hydraulics*. McGraw-Hill Book Company, Inc ; New York, 1959.

[28] S. E. Cohn. An Introduction to Estimation Theory. *Sciences-New York*, 1997.

[29] P. R. Conrad and Y. M. Marzouk. Adaptive smolyak pseudospectral approximations. *SIAM Journal on Scientific Computing*, 35(6) :A2643–A2670, 2013.

[30] P. Courtier and O. Talagrand. Variational assimilation of meteorological observations with the adjoint vorticity equation. ii : Numerical results. *Quarterly Journal of the Royal Meteorological Society*, 113(478) :1329–1347, 1987.

[31] M. De Lozzo. Substitution de modèle et approche multifidélité en expérimentation numérique. *Journal de la Société Française de Statistique*, 156(4) :21–55, 2015.

[32] B. J. Debusschere, H. N. Najm, P. P. Pébay, O. M. Knio, R. G. Ghanem, and O. P. Le Maître. Numerical challenges in the use of polynomial chaos representations for stochastic processes. *SIAM journal on scientific computing*, 26(2) :698–719, 2004.

[33] G. Desroziers, J.-T. Camino, and L. Berre. 4denvar : link with 4d state formulation of variational assimilation and different possible implementations. *Quarterly Journal of the Royal Meteorological Society*, 140(684) :2097–2110, 2014.

[34] M. Durand, K. M. Andreadis, D. E. Alsdorf, D. P. Lettenmaier, D. Moller, and M. Wilson. Estimation of bathymetric depth and slope from data assimilation of swath altimetry into a hydrodynamic model. *Geophysical Research Letters*, 35 :1–5, 2008.

[35] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, et al. Least angle regression. *The Annals of statistics*, 32(2) :407–499, 2004.

[36] B. Efron and C. Stein. The jackknife estimate of variance. *The Annals of Statistics*, pages 586–596, 1981.

[37] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5) :10143–10162, 1994.

[38] R. Faivre, B. Iooss, S. Mahévas, D. Makowski, and H. Monod. *Analyse de sensibilité et exploration de modèles : application aux sciences de la nature et de l'environnement*. Editions Quae, 2013.

[39] R. J. Fennema and M. H. Chaudhry. Implicit methods for two-dimensional unsteady free-surface flows. *Journal of Hydraulic Research*, 27(3) :321–332, 1989.

[40] P. Finaud-Guyot. *Macroscopic flood modelling : Taking into account directional flows and main channel–floodplain transfer*. PhD thesis, PhD thesis, Université Montpellier 2, 2009 (in French), 2009.

[41] G. Fishman. *Monte Carlo : concepts, algorithms, and applications*. Springer Science & Business Media, 2013.

[42] B. L. Fox. *Strategies for Quasi-Monte Carlo*, volume 22. Springer Science & Business Media, 1999.

[43] P. Frauenfelder, C. Schwab, and R. A. Todor. Finite elements for elliptic problems with stochastic coefficients. *Computer Methods in Applied Mechanics and Engineering*, 194(2–5) :205 – 228, 2005. Selected papers from the 11th Conference on The Mathematics of Finite Elements and Applications.

[44] P. GarcíaNavarro, F. Alcrudo, and J. M. Savirón. 1d openchannel flow simulation using tvdmccormack scheme. *Journal of Hydraulic Engineering*, 118(10) :1359–1372, 1992.

[45] T. B. Gatski, M. Y. Hussaini, and J. L. Lumley. *Simulation and modeling of turbulent flows*. Oxford University Press, 1996.

[46] A. Gelb. *Applied optimal estimation*. MIT press, 1974.

[47] J.-B. V. P. K. G. Gerritsma, M. Van der Steen. Time-dependant generalized polynomial chaos. *Journal Of Computational Physics*, 2010.

[48] T. Gerstner and M. Griebel. Dimension–adaptive tensor–product quadrature. *Computing*, 71(1) :65–87, 2003.

[49] R. Ghanem and P. Spanos. A stochastic galerkin expansion for nonlinear random vibration analysis. *Probabilistic Engineering Mechanics*, 8(3) :255 – 264, 1993.

[50] G. H. Golub and J. H. Welsch. Calculation of gauss quadrature rules. *Mathematics of computation*, 23(106) :221–230, 1969.

[51] N. Goutal. The malpasset dam failure. an overview and test case definition. In *Proc. of CADAM Zaragoza meeting*, pages 1–7, 1999.

[52] N. Goutal. *Développements autour de la simulation des écoulements à surface libre en rivière. Equations de Saint-Venant - Couplage de modèles - Incertitudes - Application aux ondes de submersion*. Habilitation à diriger des recherches (HDR), Université Paris Est, 2014.

[53] N. Goutal, J. Lacombe, F. Zaoui, and K. El-Kadi-Abderrezzak. Mascaret : a 1-d open-source software for flow hydrodynamic and water quality in open channel networks. pages 6–8, 2012.

[54] N. Goutal and F. Maurel. A finite volume solver for 1D shallow-water equations applied to an actual river. *International Journal for Numerical Methods in Fluids*, 38(1) :1–19, 2002.

[55] N. Goutal and J. Sainte-Marie. A kinetic interpretation of the section-averaged Saint-Venant system for natural river hydraulics. *International Journal for Numerical Methods in Fluids*, 67(7) :914–938, 2011.

[56] P. Gustafson, C. Srinivasan, and L. Wasserman. Local sensitivity analysis. *Bayesian statistics*, 5 :197–210, 1996.

[57] J. Habert. *Prévisions des crues en temps réel sur le bassin de la Marne : assimilation in situ pour la correction du modèle hydraulique mono-dimensionnel Mascaret*. PhD thesis, 2016.

[58] M. Hayoun. La méthode de monte carlo metropolis. *École «Simulation Numérique en Matière Condensée»(Paris, 29-31 mai 2002)*, 2002.

[59] J.-M. Hervouet. Telemac, a hydroinformatic system. *La Houille Blanche*, (3-4) :21–28, 1999.

[60] J.-M. Hervouet. *Hydrodynamics of free surface flows : modelling with the finite element method*. John Wiley & Sons, 2007.

[61] J.-M. Hervouet and J.-M. Janin. Finite element algorithms for modelling flood propagation. In *Modelling of Flood Propagation Over Initially Dry Areas*, pages 102–113. ASCE, 1994.

[62] T. Homma and A. Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1) :1–17, 1996.

[63] M. Honnorat. *Assimilation de données lagrangiennes pour la simulation numérique en hydraulique fluviale*. PhD thesis, Institut National Polytechnique de Grenoble-INPG, 2007.

[64] M. Horritt. Stochastic modelling of 1-d shallow water flows over uncertain topography. *Journal of Computational Physics*, 180(1) :327 – 338, 2002.

[65] T. Y. Hou, W. Luo, B. Rozovskii, and H.-M. Zhou. Wiener chaos expansions and numerical solutions of randomly forced equations of fluid mechanics. *Journal of Computational Physics*, 216(2) :687–706, 2006.

[66] B. Iooss. Contributions au traitement des incertitudes en modélisation numérique. *HDR thesis, Université Toulouse III*, 2009.

[67] R. E. Kalman. A new approach to linear filtering and prediction problems, 1960.

[68] K. Karhunen. Zur spektraltheorie stochastischer prozesse. 1946.

[69] A. Keese, H. G. Matthies, et al. Numerical methods and smolyak quadrature for nonlinear stochastic partial differential equations. 2003.

[70] A. S. Kronrod. Nodes and weights of quadrature formulas. 1965.

[71] C. Lanczos. *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*. United States Governm. Press Office Los Angeles, CA, 1950.

[72] J. V. Langenhove, D. Lucor, and A. Belme. Robust uncertainty quantification using preconditioned least-squares polynomial approximations with l_1-regularization. *International Journal for Uncertainty Quantification*, 6(1) :57–77, 2016.

[73] O. Le Maıtre, O. Knio, H. Najm, and R. Ghanem. Uncertainty propagation using wiener–haar expansions. *Journal of Computational Physics*, 197(1) :28–57, 2004.

[74] D. Le Touzé. *Méthodes spectrales pour la modélisation non-linéaire d'écoulements à surface libre instationnaires*. PhD thesis, Ecole Centrale de Nantes(ECN) ; Université de Nantes, 2003.

[75] H. Le Treut. Climat : Pourquoi les modeles n'nont pas tort. *RECHERCHE-PARIS-*, pages 68–73, 1997.

[76] O. LeMaître and O. Knio. *Spectral Methods for Uncertainty Quantification*. Springer, 2010.

[77] J. K. Levy and J. Hall. Advances in flood risk management under uncertainty. *Stochastic Environmental Research and Risk Assessment*, 19(6) :375–377, 2005.

[78] J. Li and D. Xiu. On numerical properties of the ensemble kalman filter for data assimilation. *Computer Methods in Applied Mechanics and Engineering*, 197(43) :3574–3583, 2008.

[79] J. Li and D. Xiu. A generalized polynomial chaos based ensemble Kalman filter with high accuracy. *Journal of Computational Physics*, 228 :5454–5469, 2009.

[80] F.-S. Lien. A pressure-based unstructured grid method for all-speed flows. *International Journal for Numerical Methods in Fluids*, 33(3) :355–374, 2000.

[81] D. Liu. *Uncertainty quantification with shallow water equations*. PhD thesis, University of Florence, 2010.

[82] M. Loève. Probability theory. 1977, 1977.

[83] W.-L. Loh et al. On latin hypercube sampling. *The annals of statistics*, 24(5) :2058–2080, 1996.

[84] F. Lu, M. Morzfeld, X. Tu, and A. J. Chorin. Limitations of polynomial chaos expansions in the bayesian solution of inverse problems. *Journal of Computational Physics*, 282 :138–147, 2015.

[85] D. Lucor and G. Karniadakis. Stochastic flow-structure interactions. *Computational Fluid and Solid Mechanics 2003*, pages 1426 – 1429, 2003.

[86] H. Madsen and C. Skotner. Adaptive state updating in real-time river flow forecasting - A combined filtering and error forecasting procedure. *Journal of Hydrology*, 308 :302–312, 2005.

[87] P.-O. Malaterre. Le contrôle automatique des canaux d'irrigation : Etat de l'art et perspectives. In *Colloque Automatique et Agronomie, Montpellier*, 2003.

[88] S. Manabe, J. Smagorinsky, and R. F. Strickler. Simulated climatology of a general circulation model with a hydrologic cycle. *Mon. Wea. Rev*, 93(12) :769–798, 1965.

[89] L. Margheri and P. Sagaut. A hybrid anchored-anova – pod/kriging method for uncertainty quantification in unsteady high-fidelity {CFD} simulations. *Journal of Computational Physics*, 324 :137 – 173, 2016.

[90] Y. M. Marzouk and H. N. Najm. Dimensionality reduction and polynomial chaos acceleration of bayesian inference in inverse problems. *Journal of Computational Physics*, 228(6) :1862–1902, 2009.

[91] Y. M. Marzouk, H. N. Najm, and L. A. Rahn. Stochastic spectral methods for efficient bayesian solution of inverse problems. *Journal of Computational Physics*, 224(2) :560–586, 2007.

[92] S. Massart, a. Piacentini, and O. Pannekoucke. Importance of using ensemble estimated background error covariances for the quality of atmospheric ozone analyses. *Quarterly Journal of the Royal Meteorological Society*, 138(April) :889–905, 2012.

[93] P. Matgen, M. Montanari, R. Hostache, L. Pfister, L. Hoffmann, D. Plaza, V. R. N. Pauwels, G. J. M. De Lannoy, R. De Keyser, and H. H. G. Savenije. Towards the sequential assimilation of SAR-derived water stages into hydraulic models using the Particle Filter : Proof of concept. *Hydrology and Earth System Sciences*, 14 :1773–1785, 2010.

[94] B. Merz and A. H. Thieken. Separating natural and epistemic uncertainty in flood frequency analysis. *Journal of Hydrology*, 309(1) :114–132, 2005.

[95] N. Metropolis and S. Ulam. The monte carlo method. *Journal of the American statistical association*, 44(247) :335–341, 1949.

[96] G. Modgil, W. A. Crossley, and D. Xiu. Design optimization of a high-pressure turbine blade using generalized polynomial chaos (gpc). In *10th World Congress on Structural and Multidisciplinary Optimization*, volume 1, pages 1–16, 2013.

[97] A. M. Molinaro, R. Simon, and R. M. Pfeiffer. Prediction error estimation : A comparison of resampling methods. *Bioinformatics*, 21(15) :3301–3307, Aug. 2005.

[98] D. C. Montgomery. *Design and analysis of experiments*. John Wiley & Sons, 2008.

[99] H. Moradkhani, S. Sorooshian, H. V. Gupta, and P. R. Houser. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, 28 :135–147, 2005.

[100] J. C. Neal, P. M. Atkinson, and C. W. Hutton. Flood inundation model updating using an ensemble Kalman filter and spatially distributed measurements. *Journal of Hydrology*, 336 :401–415, 2007.

[101] A. B. Owen. Detecting near linearity in high dimensions. Technical report, Citeseer, 1998.

[102] F. Pappenberger, K. Beven, M. Horritt, and S. Blazkova. Uncertainty in the calibration of effective roughness parameters in hec-ras using inundation and downstream level observations. *Journal of Hydrology*, 302(1) :46–69, 2005.

[103] F. Pappenberger and K. J. Beven. Ignorance is bliss : Or seven reasons not to use uncertainty analysis. *Water Resources Research*, 42(5) :n/a–n/a, 2006. W05302.

[104] V. Pedinotti, A. Boone, S. Ricci, S. Biancamaria, and N. Mognard. Assimilation of satellite data to optimize large scale hydrological model parameters : a case study for the SWOT mission. *Hydrology and Earth System Sciences Discussions*, 11 :4477–4530, 2014.

[105] G. Poëtte. *Propagation d'incertitudes pour les systèmes de lois de conservation, "Méthodes Spectrales Stochastiques"*. PhD thesis, Université Pierre et Marie Curie, 2009.

[106] G. Poëtte, B. Després, and D. Lucor. Uncertainty quantification for systems of conservation laws. *Journal of Computational Physics*, 228(7) :2443–2467, 2009.

[107] G. Poëtte and D. Lucor. Non intrusive iterative stochastic spectral representation with application to compressible gas dynamics. *Journal of Computational Physics*, 231(9) :3587–3609, 2012.

[108] N. Pramanik, R. K. Panda, and D. Sen. One dimensional hydrodynamic modeling of river flow using dem extracted river cross-sections. *Water Resources Management*, 24(5) :835–852, 2010.

[109] A. Resmini, J. Peter, and D. Lucor. Sparse grids-based stochastic approximations with applications to aerodynamics sensitivity analysis. *International Journal for Numerical Methods in Engineering*, 106(1) :32–57, 2016. nme.5005.

[110] M. Ricchiuto, P. M. Congedo, and A. I. Delis. *Runup and uncertainty quantification : sensitivity analysis via ANOVA decomposition*. PhD thesis, INRIA, 2014.

[111] S. Ricci. Assimilation variationnelle océanique : modélisation multivariée de la matrice de covariance d'erreur d'ébauche. 2004.

[112] S. Ricci, a. Piacentini, O. Thual, E. Le Pape, and G. Jonville. Correction of upstream flow and hydraulic state with data assimilation in the context of flood forecasting. *Hydrology and Earth System Sciences*, 15 :3555–3575, 2011.

[113] L. F. Richardson. *Weather prediction by numerical process*. Cambridge University Press, 2007.

[114] F. Riesz. B. sz.-nagy functional analysis, 1990.

[115] M. C. Rochoux, S. Ricci, B. Lucor, D.and Cuenot, and A. Trouvé. Towards predictive data-driven simulations of wildfire spread - part 1 : Reduced-cost ensemble kalman filter based on a polynomial chaos surrogate model for parameter estimation. *Nat. Hazards and Earth Syst. Sci.*, 2014.

[116] O. Roderick, M. Anitescu, and Y. Peet. Proper orthogonal decompositions in multifidelity uncertainty quantification of complex simulation models. *International Journal of Computer Mathematics*, 91(4) :748–769, 2014.

[117] P. Roe. Appoximate riemann solvers, parameters vectors and difference schemes. *Journal of Computational Physics*, 1981.

[118] A. Saint-Venant. Théorie du mouvement non-permanent des eaux, avec application aux crues des rivières et à l'introduction des marées dans leur lit. *Compte rendu à l'Académie des Sciences de Paris*, 73 :147–154, 1871.

[119] A. Saltelli, K. Chan, E. M. Scott, et al. *Sensitivity analysis*, volume 1. Wiley New York, 2000.

[120] A. Saltelli and I. M. Sobol'. About the use of rank transformation in sensitivity analysis of model output. *Reliability Engineering System Safety*, 50(3) :225 – 239, 1995.

[121] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto. *Sensitivity analysis in practice : a guide to assessing scientific models*. John Wiley & Sons, 2004.

[122] A. Saltelli, S. Tarantola, and K.-S. Chan. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41(1) :39–56, 1999.

[123] R. Schobi, B. Sudret, and J. Wiart. Polynomial-chaos-based kriging. *International Journal for Uncertainty Quantification*, 5(2), 2015.

[124] E. Serre, P. Stansby, D. Laurence, and B. Launder. Écoulement en eau peu profonde autour d'un modèle d'île conique. *Revue Européenne des Éléments*, 12(2-3) :361–371, 2003.

[125] S. J. Sheather and M. C. Jones. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 683–690, 1991.

[126] Y.-e. Shi. *Résolution numérique des équations de Saint-Venant par la technique de projection en utilisant une méthode des volumes finis dans un maillage non structuré*. PhD thesis, Université de Caen-Basse Normandie, 2006.

[127] P. Sleigh, P. Gaskell, M. Berzins, and N. Wright. An unstructured finite-volume algorithm for predicting flow in rivers and estuaries. *Computers Fluids*, 27(4) :479 – 508, 1998.

[128] S. A. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. In *Dokl. Akad. Nauk SSSR*, volume 4, page 123, 1963.

[129] I. M. Sobol'. On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe Modelirovanie*, 2(1) :112–118, 1990.

[130] C. Soize and R. Ghanem. Physical systems with random uncertainties : chaos representations with arbitrary probability measure. *SIAM Journal on Scientific Computing*, 26(2) :395–410, 2004.

[131] P. Spanos and R. Ghanem. *Stochastic Finite Elements : A Spectral Approach*. Springer, 1991.

[132] B. Sudret. Global sensitivity analysis using polynomial chaos expansion. *Reliab Eng Syst Safety*, 2007.

[133] B. Sudret. *Uncertainty propagation and sensitivity analysis in mechanical models – Contributions to structural reliability and stochastic spectral methods*, 2007. Habilitation à diriger des recherches, Université Blaise Pascal, Clermont-Ferrand, France.

[134] B. Sudret. Meta-models for structural reliability and uncertainty quantification. *arXiv preprint arXiv :1203.2062*, 2012.

[135] O. Thual. *Introduction à la mécanique des milieux continues déformables*. Cépadués-Editions, 1997.

[136] O. Thual. *Hydrodynamique de l'environnement*. Ecole polytechnique, 2010.

[137] M. K. Tippett, J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker. Ensemble square root filters*. *Monthly Weather Review*, 131(7) :1485–1490, 2003.

[138] L. N. Trefethen. Is gauss quadrature better than clenshaw-curtis ? *SIAM review*, 50(1) :67–87, 2008.

[139] J. Tryoen. *Méthodes de Galerkin stochastiques adaptatives pour la propagation d'incertitudes paramétriques dans les modèles hyperboliques*. PhD thesis, Université Paris-Est, 2011.

[140] T. Turányi. Sensitivity analysis of complex kinetic systems. tools and applications. *Journal of Mathematical Chemistry*, 5(3) :203–248, 1990.

[141] D. Van den Eynde, A. Giardino, J. Portilla, M. Fettweis, F. Francken, and J. Monbaliu. Modelling the effects of sand extraction, on sediment transport due to tides, on the kwinte bank. *Journal of Coastal Research*, pages 101–116, 2010.

[142] F. Vilar. *Utilisation des méthodes Galerkin discontinues pour la résolution de l'hydrodynamique Lagrangienne bi-dimensionnelle*. PhD thesis, Bordeaux 1, 2012.

[143] M. E. Vázquez-Cendón. Improved treatment of source terms in upwind schemes for the shallow water equations in channels with irregular geometry. *Journal of Computational Physics*, 148(2) :497 – 526, 1999.

[144] X. Wan and G. E. Karniadakis. Long-term behavior of polynomial chaos in stochastic flow simulations. *Computer methods in applied mechanics and engineering*, 195(41) :5582–5596, 2006.

[145] M. Wand and M. Jones. Kernel smoothing. 1995. *Chapman&Hall, London*.

[146] I. Warren and H. Bach. Mike 21 : a modelling system for estuaries, coastal waters and seas. *Environmental Software*, 7(4) :229–240, 1992.

[147] N. Wiener. The homogenous chaos. *Am.J.Math*, 1938.

[148] D. Williams. *Diffusions, Markov processes, and martingales. Vol. 1, Foundations*. Wiley, 1979.

[149] G. Winokur. *Adaptive Sparse Grid Approaches to Polynomial Chaos Expansions for Uncertainty Quantification*. PhD thesis, Duke University, 2015.

[150] Y. Xing and C.-W. Shu. High order finite difference weno schemes with the exact conservation property for the shallow water equations. *Journal of Computational Physics*, 208(1) :206–227, 2005.

[151] D. Xiu. Fast numerical methods for stochastic computations : A review. *Commun.Comput.Phys*, 2008.

[152] D. Xiu. *Numerical Methods for Stochastic Computations : A Spectral Method Approach*. Princeton University Press, 2010.

[153] D. Xiu and G. E. Karniadakis. The wiener–askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing*, 24(2) :619–644, 2002.

[154] D. Xiu and G. E. Karniadakis. Modeling uncertainty in flow simulations via generalized polynomial chaos. *Journal of Computational Physics*, 2003.

[155] C. Xu and G. Z. Gertner. Uncertainty and sensitivity analysis for models with correlated parameters. *Reliability Engineering & System Safety*, 93(10) :1563–1573, 2008.

[156] Y. Yoon, M. Durand, C. J. Merry, E. a. Clark, K. M. Andreadis, and D. E. Alsdorf. Estimating river bathymetry from data assimilation of synthetic SWOT measurements. *Journal of Hydrology*, 464-465 :363–375, 2012.

[157] D. Youla. The solution of a homogeneous wiener-hopf integral equation occurring in the expansion of second-order stationary random functions. *IRE Transactions on Information Theory*, 3(3) :187–193, September 1957.

[158] J.-M. Zokagoa and A. Soulaïmani. A pod-based reduced order model for free surface shallow water flows over real bathymetries for monte-carlo-type applications. *Comput. Methods Appl. Mech. Engrg.*, 2012.

Résumé Ce travail de thèse porte sur la construction d'un modèle réduit pour les équations de Saint-Venant en hydraulique fluviale avec une méthode de décomposition en polynôme du chaos. Ce modèle réduit est utilisé à la place du modèle direct afin de réduire le coût de calcul lié aux méthodes ensemblistes en quantification d'incertitudes et assimilation de données. Le contexte de l'étude est la prévision des crues et la gestion de la ressource en eau.

Ce manuscrit est composé de cinq parties, chacune divisée en chapitres.

La première partie présente un état de l'art des travaux en quantification des incertitudes et en assimilation de données dans le domaine de l'hydraulique ainsi que les objectifs de la thèse. On présente le cadre de la prévision des crues, ses enjeux et les outils dont on dispose (numériques et observation) pour prévoir la dynamique des rivières. On présente notamment la future mission SWOT [1] qui a pour but de mesurer les hauteurs d'eau dans les rivières avec un couverture globale à haute résolution. On précise notamment l'apport de ces mesures et leur complémentarité avec les mesures *in-situ*.

La deuxième partie présente les équations de Saint-Venant, qui décrivent les écoulements dans les rivières. On s'intéresse notamment à une représentation 1D des équations. On formule une discrétisation numérique de ces équations, telle qu'implémentée dans le logiciel Mascaret. Le dernier chapitre de cette partie propose quelques simplifications des équations de Saint-Venant.

La troisième partie de ce manuscrit présente les méthodes de quantification et de réduction des incertitudes. On présente notamment le contexte probabiliste qui permet de bien définir un problème de quantification d'incertitudes et d'analyse de sensibilité. On propose ensuite de réduire la dimension d'un problème stochastique quand on traite de champs aléatoires dans le contexte des modèles géophysiques. Les méthodes de décomposition en polynômes du chaos sont ensuite présentées ; on présente notamment les différentes stratégies pour le calcul des coefficients polynomiaux. Cette partie dédiée à la méthodologie s'achève par un chapitre consacré à l'assimilation de données ensemblistes (typiquement le filtre de Kalman d'Ensemble) et à l'utilisation des modèles réduits dans ce cadre.

La quatrième partie de ce manuscrit est dédiée aux résultats. On commence par identifier les sources d'incertitudes en hydraulique que l'on s'attache à quantifier et réduire par la suite. Un article en cours de révision détaille la méthode et la validation d'un modèle réduit polynomial pour les équations de Saint-Venant en régime stationnaire lorsque l'incertitude est majoritairement portée par les coefficients de frottement et le débit d'apport à l'amont. L'étude est menée sur la rivière Garonne. On montre que les moments statistiques, la densité de probabilité et la matrice de covariances spatiales pour la hauteur d'eau sont efficacement et précisément estimés à l'aide du modèle réduit dont la construction ne nécessite que quelques dizaines d'intégrations du modèle direct.

On met à profit l'utilisation du modèle réduit pour réduire le coût de calcul du filtre de Kalman d'Ensemble dans le cadre d'un exercice d'assimilation de données synthétiques de type SWOT. On cherche à reconstruire les coefficients spatialisés de frottement et le débit d'apport à l'amont. On s'intéresse précisément à la représentation spatiale de la données telle que vue par SWOT : couverture globale du réseau, moyennage spatial entre les pixels observés. On montre notamment qu'à budget de calcul donné (2500 simulations du modèle direct) les résultats de l'analyse d'assimilation de données qui repose sur l'utilisation du modèle réduit sont meilleurs que ceux obtenus avec le filtre de Kalman d'Ensemble classique.

On s'intéresse enfin à la construction du modèle réduit en régime instationnaire. On suppose dans un premier temps que l'incertitude est liée aux coefficients de frottement. Il s'agit à présent de juger de la nécessité du recalcul des coefficients polynomiaux au fil du temps et des cycles d'assimilation de données. Pour ce travail seul des données ponctuelles et in-situ ont été considérées. On suppose dans un deuxième temps que l'incertitude est portée par le débit en amont du réseau, qui est un vecteur temporel. On procède à une décomposition de type Karhunen-Loève pour réduire la taille de l'espace incertain aux trois premiers modes. Nous sommes ainsi en mesure de mener à bien un exercice d'assimilation de données.

Pour finir, les conclusions et les perspectives de ce travail sont présentées en cinquième partie.

Mots-clés Hydraulique fluviale, Assimilation de données, Polynômes du chaos, Quantification des incertitudes

1. SWOT : Surface Water and Ocean Topography

Abstract This work deals with the formulation of a surrogate model for the shallow water equations in fluvial hydraulics with a chaos polynomial expansion. This reduced model is used instead of the direct model to reduce the computational cost of the ensemble methods in uncertainty quantification and data assimilation. The context of the study is the flood forecasting and the management of water resources.

This manuscript is composed of five parts, each divided into chapters.

The first part presents a state of art of uncertainty quantification and data assimilation in the field of hydraulics as well as the objectives of this thesis. We present the framework of flood forecasting, its stakes and the tools available (numerical and observation) to predict the dynamics of rivers. In particular, we present the SWOT [2] mission, which aims to measure the height of water in rivers with global coverage at high resolution. We highlight particularty their contribution and their complementarity with the *in-situ* measurements.

The second part presents the shallow water equations, which describe the flows in the rivers. We are particularly interested in a 1D representation of the equations. We formulate a numerical discretization of these equations, as implemented in the Mascaret software. The last chapter of this part proposes some simplifications of the shallow-water equations.

The third part of this manuscript presents the uncertainty quantification and reduced order methods. We present particularly the probabilistic context which makes it possible to define well-defined problem of uncertainty quantification and sensitivity analysis. It is then proposed to reduce the size of a stochastic problem when dealing with random fields in the context of geophysical models. The methods of chaos polynomial expansion are then presented ; we present in particular the different strategies for the computation of the polynomial coefficients. This section devoted to methodology concludes with a chapter devoted to Ensemble based data assimilation (specially the Ensemble Kalman filter) and the use of surrogate models in this framework.

The fourth part of this manuscript is dedicated to the results.

The first step is to identify the sources of uncertainty in hydraulics that should be quantified and subsequently reduced. An article, in the review state, details the method and the validation of a polynomial surrogate model for shallow water equations in steady state when the uncertainty is mainly carried by the friction coefficients and upstream inflow. The study is conducted on the river Garonne. It is shown that the statistical moments, the probability density and the spatial covariance matrice for the water height are efficiently and precisely estimated using the reduced model whose construction requires only a few tens of integrations of the direct model.

The use of the surrogate model is used to reduce the computational cost of the Ensemble Kalman filter in the context of a synthetic SWOT like data assimilation exercise. The aim is to reconstruct the spatialized friction coefficients and the upstream inflow. We are interested precisely in the spatial representation of the data as seen by SWOT : global coverage of the network, spatial averaging between the observed pixels. We show in particular that at the given calculation budget (2500 simulations of the direct model) the results of the data assimilation analysis based on the use of the polynomial surrogate model are better than those obtained with the classical Ensemble Kalman filter.

We are then interested in the construction of the reduced model in unsteady conditions. It is assumed initially that the uncertainty is carried with the friction coefficients. It is now necessary to judge the need for the recalculation of polynomial coefficients over time and data assimilation cycles. For this work only ponctual and in-situ data were considered.

It is assumed in a second step that the uncertainty is carried by the upstream inflow of the network, which is a time vector. A Karhunen-Love decomposition is used to reduce the size of the uncertain space into only the first three modes. We are thus able to carry out a data assimilation exercise.

Finally, the conclusions and perspectives of this work are presented in the fifth part.

Keywords Hydraulics, Data assimilation, Chaos polynomial, Uncertainty Quantification

2. SWOT : Surface Water and Ocean Topography