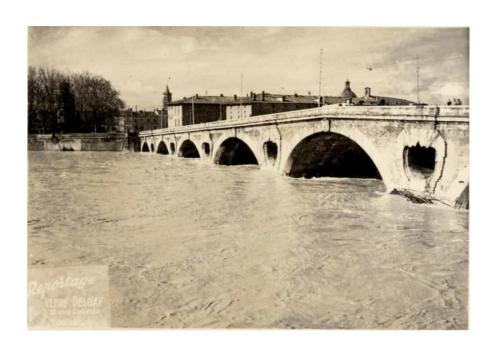




Projet modélisation janvier/février 2011 :

Assimilation de données sur un modèle d'ondes de crues



Crédit photo: Henry Delgay

Encadré par Sophie Ricci (CERFACS) et Olivier Pannekoucke (CNRM/GAME)

Etudiants IENM 2009/2012 Delplanque Alain Guth Jonathan Raynaud Damien

Résumé

La prévision des crues demeure un domaine incontournable lorsque l'on parle de sécurité des biens et des personnes. La multiplication des sites à risques ou à enjeux économiques et environnementaux (les centrales hydroélectriques ou nucléaires) ainsi que le développement de zones urbaines proches de zones inondables ne font que renforcer l'importance d'une bonne prévision. Des efforts dans le domaine de la modélisation et de l'observation permettent d'améliorer la prévision de la propagation d'une crue et d'en limiter les risques humains et technologiques. En France, c'est le Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations (SCHAPI), rattaché au ministère de l'écologie, du Développement Durable, des Transports et du Logement, en lien avec le Service de Prévision de crues (SPC) qui assure une surveillance des crues ¹. Les enjeux sont également économiques. En effet, une prévision fiable de la hauteur d'eau permet une gestion optimale des ressources en eau notamment pour la production électrique par Électricité de France (EDF). Ce projet s'inscrit dans le cadre de la mise en œuvre de la prévision hydraulique en temps réel. Il s'agit ici de travailler sur l'assimilation des observations de hauteurs d'eau et de débits afin d'améliorer les prévisions.

Lors de l'implémentation d'un algorithme d'assimilation de données, il est important de bien modéliser la matrice de covariance des erreurs de prévisions que l'on ne sait pas estimer exactement. En effet, ce sont les covariances d'erreurs d'ébauche qui déterminent l'amplitude et la forme de l'incrément d'analyse. A l'aide d'un modèle physique simple codé à partir de l'équation d'ondes de crues diffusantes on cherche à déterminer les caractéristiques de la matrice des erreurs de prévision.

¹http://www.vigicrues.ecologie.gouv.fr/

Table des matières

1	L'as	ssimilation de données	5
	1.1	Généralités	5
	1.2	Méthode d'assimilation BLUE (Best Linear Unbiaised Estimator)	6
	1.3	Rôle de B	7
	1.4	Diagnostic de la longueur de portée	9
	1.5	Filtre de Kalman	10
	1.6	Filtre de Kalman d'ensemble	10
2	Mo	dèle d'onde de crue	13
	2.1	Physique du problème	13
	2.2	Schéma numérique de l'équation d'advection-diffusion	15
		2.2.1 Schéma de discrétisation	15
		2.2.2 Stabilité du schéma	16
		2.2.3 Équation équivalente	16
		2.2.4 Impact des termes diffusifs et dispersifs sur la propagation des covariances	17
	2.3	Construction du forçage amont	18
3	Dyr	namique des covariances : évolution des longueurs de portée	21
	3.1	Dynamique des covariances sans observations	21
		3.1.1 Cas de l'advection pure	22
		3.1.2 Cas de l'advection diffusion	23
		3.1.3 Bilan de la dynamique des longueurs de portée sans assimilation	24
	3.2	Dynamique des covariances avec observation	25
4	Exp	périences numériques	27
	4.1	Diagnostic de B et des longueurs de portée sans assimilation	27
		4.1.1 Calcul de la matrice de covariances d'erreurs d'ébauche	27
		4.1.2 Calcul des longueurs de portée	28
	4.2		28
		4.2.1 Modèle avec advection seule	28
		4.2.2 Modèle avec advection et diffusion	29
	4.3	Assimilation avec un BLUE	30
	4.4	Implémentation du filtre de Kalman	33
		4.4.1 Cadre de la simulation	33
		4.4.2 Évolution des fonctions de corrélation	36
		4.4.3 Méthode d'estimation des longueurs de portée	36
		4.4.4 Diagnostic des longueurs de portée	38
		4.4.5 Influence du quotient σ_b/σ_o	39
		·	40

TABLE DES MATIÈRES

4.5	Implémentation filtre de Kalman d'ensemble	40
 Ann		44
A.1	Condition de stabilité du schéma numérique	44
A.2	Influence du schéma numérique	45

Introduction

La prévision de phénomènes naturels complexes nécessite une modélisation mathématique précise. Les incertitudes sur les modèles numériques proviennent des conditions aux limites (état initial, forçages extérieurs) et des simplifications et paramétrisations de la physique (paramètres numériques, géométriques, physiques). En résolvant un problème inverse, l'algorithme d'assimilation de données combine les informations issues de l'observation et de la simulation pour estimer des paramètres du modèle ou corriger l'état initial. Ainsi l'assimilation de données permet de réduire l'incertitude sur la simulation et sur la prévision. Cette technique est communément utilisée en météorologie, océanographie, et tend à être utiliser dans d'autres domaines des géosciences ou même dans des domaines plus industriels tels que la neutronique.

Les phénomènes hydrauliques font partie des phénomènes naturels complexes à prévoir. Le Centre d'Etudes Techniques Maritimes et Fluviales (CETMEF) et le Laboratoire national d'Hydraulique et Environnement (LNHE, service R&D d'EDF) ont mis au point un code numérique visant à modéliser ces phénomènes hydrauliques : MASCARET. Ce code numérique 1D, basé sur les équations de Saint-Venant, est capable de représenter une large gamme de phénomènes hydrauliques (ondes de crues, ondes de marées, etc). Cette étude se place dans le cadre d'une collaboration entre le Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS), le SCHAPI, le CETMEF et le LNHE. Cette collaboration vise à construire un système d'assimilation pour le modèle MASCARET [Ricci et al., 2010].

Une difficulté rencontrée lors de la mise en place de la plupart des systèmes d'assimilation de données est de spécifier les statistiques des erreurs associées à une prévision : variance et covariances des erreurs. L'estimation directe des statistiques de l'erreur de prévision est impossible et la propagation explicite des covariances par la dynamique d'un modèle est généralement difficile à mettre en œuvre. C'est le cas pour le modèle MASCARET. Une manière efficace mais coûteuse en temps de calcul est d'utiliser un filtre de Kalman. Dans le cadre de cette étude, une hypothèse forte est alors émise. Les processus opérés par le modèle sont considérés comme identiques ou proches de ceux faits par un modèle plus simple. Ce modèle simplifié est basé sur l'équation de propagation des ondes de crues diffusantes. Le coût de calcul d'un filtre de Kalman devient alors plus raisonnable. Dans le travail proposé ici, on utilise donc cet algorithme d'assimilation sur le modèle d'onde de crues pour estimer une forme paramétrée des fonctions de covariance spatiale. Cette paramétrisation est alors utilisée pour modèliser la matrice de covariance d'erreur de prévision dans un algorithme de type BLUE.

Le premier chapitre présente un rappel sur l'assimilation de données. Le chapitre 2 présente le modèle numérique utilisé. Puis les chapitres 3 et 4 rapportent les résultats de l'étude. Le chapitre 3 correspond à l'étude théorique. Le chapitre 4 s'intéresse aux résultats numériques. Les conclusions et perspectives sont au chapitre 5.

Chapitre 1

L'assimilation de données

1.1 Généralités

L'assimilation de données est un processus mathématique qui permet d'obtenir la représentation la plus probable de l'état d'un système à partir de toutes les sources d'informations disponibles. En pratique, sous l'hypothèse de Gaussiannité, il s'agit de trouver la meilleure combinaison entre un état à priori du système que l'on appelle l'ébauche et les observations. L'utilisation de l'ébauche permet de palier le problème de sous détermination dans le cas où le nombre d'observations est inférieur à la dimension du problème. Le critère d'optimalité pour obtenir le meilleur estimateur statistique de cette combinaison est la détermination du minimum de variance d'erreur d'estimation.

Les informations issues de l'ébauche et des observations sont entachées d'erreur. La combinaison des deux par le biais de divers algorithmes d'assimilation (BLUE, filtre de Kalman...) permet de se rapprocher de l'état réel de notre système. L'état vrai du système, généralement inconnu, est représenté par le vecteur \mathbf{X}^t . L'ébauche est elle notée \mathbf{X}^b . Ces deux vecteurs sont de dimension n. \mathbf{Y}^o désigne le vecteur d'observations, il est de dimension p. Le processus d'assimilation de données formule l'analyse \mathbf{X}^a de dimension p. L'erreur d'ébauche est définie par le vecteur ε^b telle que $\mathbf{X}^b = \mathbf{X}^t + \varepsilon^b$.

Les observations diffèrent également de l'état réel. L'erreur d'observation est définie par ε^o où $\mathbf{Y}^o = \mathcal{H}\mathbf{X}^t + \varepsilon^o$, avec \mathcal{H} est l'opérateur permettant de passer de l'espace du modèle à celui des observations. Par exemple si \mathbf{Y}^o contient des mesures de débit et que les observations ne sont pas faites en un point de la discrétisation de la rivière, \mathcal{H} correspond à une interpolation spatiale du vecteur d'état vers les points d'observations. Cet opérateur peut être non-linéaire. Par exemple en météorologie, l'assimilation des mesures de radiances satellitaires implique un opérateur \mathcal{H} non linéaire lié à l'équation de transfert radiatif.

Toutes les erreurs, d'observations ou d'ébauche, sont considérées comme non-biaisées. Les matrices de covariances d'erreurs sont notées $\mathbf{B} = \mathbb{E}\left[\varepsilon^b(\varepsilon^b)^T\right]$ pour l'ébauche et $\mathbf{R} = \mathbb{E}\left[\varepsilon^o(\varepsilon^o)^T\right]$ pour les observations. Par construction, ces matrices sont symétriques, définies et positives. Elles permettent de quantifier la confiance que l'on a en l'ébauche et les observations. Enfin on considère qu'il n'y a pas de corrélation entre erreurs d'ébauche et erreurs d'observation.

$$\mathbb{E}\left[\varepsilon^o(\varepsilon^b)^T\right] = \mathbf{0}_{n \times p}$$

Dans le cadre de cette étude, les vecteurs \mathbf{X}^t , \mathbf{X}^b et \mathbf{X}^a contiennent respectivement les hauteurs d'eau réelles (connues dans la cadre d'expériences jumelles ¹), simulées, et analysées. \mathbf{Y}^o

¹Simulations numériques où l'état vrai et l'ébauche sont connus et synthétiques.

1.2. MÉTHODE D'ASSIMILATION BLUE (BEST LINEAR UNBIAISED ESTIMATOR)

représente les mesures faites le long de la rivière. Il s'agit également de hauteurs d'eau. Le lien entre vecteur d'état et observations est simple; il ne s'agit que d'une sélection de points. L'opérateur \mathcal{H} est linéaire et sera à présent noté $\mathbf{H} = (0 \cdots 0 \ 1 \ 0 \cdots 0)$. Enfin dans des cas complexes faisant intervenir différentes mesures faites par un même instrument (satellites...), les erreurs d'observations sont corrélées en espace et en temps. Si, comme dans le cadre de ce projet, il y a indépendance entre erreurs d'observations, alors la matrice \mathbf{R} est diagonale.

1.2 Méthode d'assimilation BLUE (Best Linear Unbiaised Estimator)

Dans le problème traité, on recherche le meilleur estimateur linéaire non biaisé de l'état vrai du système comme combinaison linéaire de l'ébauche et des observations. Pour construire un estimateur linéaire, deux matrices $\mathbf L$ et $\mathbf K$ de tailles respectives $n \times n$ et $n \times p$ doivent être déterminées. Elles sont telles que :

$$\mathbf{X}^a = \mathbf{L}\mathbf{X}^b + \mathbf{K}\mathbf{Y}^o.$$

Comme l'estimateur linéaire recherché est non biaisé, cela introduit des contraintes sur ${\bf L}$ et ${\bf K}$:

$$\mathbb{E}\left[\varepsilon^{a}\right] = \mathbb{E}\left[\left(\mathbf{X}^{a} - \mathbf{X}^{t}\right)\right]$$

$$= \mathbb{E}\left[\mathbf{L}\mathbf{X}^{t} + \mathbf{L}\varepsilon^{b} + \mathbf{K}\mathbf{H}\mathbf{X}^{t} + \mathbf{K}\varepsilon^{o} - \mathbf{X}^{t}\right]$$

$$= 0.$$
(1.1)

De plus les erreurs d'observation et d'ébauche sont supposées non biaisées, l'equation 1.1 devient alors

$$(\mathbf{L} - \mathbf{I}_{n \times n}) \mathbb{E} \left[\mathbf{X}^{t} \right] + \mathbf{K} \mathbf{H} \mathbb{E} \left[\mathbf{X}^{t} \right] = (\mathbf{L} - \mathbf{I}_{n \times n} + \mathbf{K} \mathbf{H}) \mathbb{E} \left[\mathbf{X}^{t} \right]$$

$$= 0.$$
(1.2)

La condition pour que l'estimateur X^a soit non biaisé devient donc :

$$L + KH = I_{n \times n}$$
.

L'estimateur optimal, minimise la trace de la matrice de covariance d'erreurs d'analyse $\mathbf{A} = E(\varepsilon^a \varepsilon^{aT})$. Elle s'exprime [Bouttier and Courtier, 1999] :

$$\mathbf{A} = (\mathbf{I}_{n \times n} - \mathbf{K}\mathbf{H})\mathbf{B}((\mathbf{I}_{n \times n} - \mathbf{K}\mathbf{H}))^{T} + \mathbf{K}\mathbf{R}\mathbf{K}^{T}.$$

La minimisation permet d'établir une expression de la matrice K :

$$min(Tr(\mathbf{A})) \Longrightarrow \mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}.$$

Grâce aux conditions trouvées à partir des deux hypothèses du BLUE, X^a s'écrit :

$$\mathbf{X}^{a} = \mathbf{X}^{b} + \mathbf{K}(\mathbf{Y}^{o} - H\mathbf{X}^{b})$$

$$= \mathbf{X}^{b} + \mathbf{B}\mathbf{H}^{T}(\mathbf{H}\mathbf{B}\mathbf{H}^{T} + \mathbf{R})^{-1}(\mathbf{Y}^{o} - H\mathbf{X}^{b}).$$
(1.3)

La matrice K, appelée matrice de gain, permet de définir l'état assimilé comme la somme de l'ébauche et de l'incrément d'analyse

$$\delta \mathbf{X}^a = \mathbf{K}(\mathbf{Y}^o - \mathbf{H}\mathbf{X}^b) = \mathbf{K}d,\tag{1.4}$$

avec d le vecteur d'innovation.

1.3 Rôle de B

 ${f B}$ représente les statistiques sur l'incertitude sur l'état du système prévu à un instant donné. L'équation (1.4) de $\delta {f X}^a$ met en évidence le rôle primordial de la matrice de covariances d'erreurs d'ébauche :

$$\delta \mathbf{X}^a = \mathbf{B} \underbrace{\mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1} (\mathbf{Y}^o - \mathbf{H} \mathbf{X}^b)}_{\delta \hat{\mathbf{x}}} = \mathbf{B} \, \delta \hat{\mathbf{X}}.$$

Le vecteur $\delta \hat{\mathbf{X}}$ contient des 0 sauf aux points où sont réalisées des observations. C'est la seule matrice \mathbf{B} qui répartit l'information contenue dans ce vecteur sur tout le domaine. Dans le cas multivarié 2 elle permet également de corriger l'ensemble des variables du vecteur de contrôle. \mathbf{B} contient sur sa diagonale les variances d'erreurs d'ébauche à chaque point de grille. Les termes non-diagonaux représentent les covariances d'erreurs entre les différents points. En raisonnant sur la k^{ieme} ligne (ou colonne) de la matrice \mathbf{B} on observe la fonction de covariance du k^{ieme} point du domaine.

Plus concrètement, en considérant une assimilation faite au point d'observation, la variance détermine quel "poids" donner à l'observation face à la hauteur d'eau simulée par l'ébauche. Les covariances permettent elles de répartir spatialement l'information du point d'observation vers les autres points de grille.

La figure 1.1 montre, à partir d'une ébauche (a), et de deux fonctions de covariances (une "étroite" (b) et l'autre plus "large" (c)) la correction faite sur l'analyse. Les états analysés (d) et (e) montrent que plus la fonction de covariance est "étalée" plus la correction apportée au point d'observation se répercute sur des points éloignés. L'étalement des fonctions de covariance peut être caractérisé par une longueur appelée longueur de portée définie au paragraphe 1.4. La figure 1.2.(a) montre une fonction de covariance gaussienne. Celle-ci étant isotrope, l'information apportée par une assimilation au k^{ieme} point sera répartie de manière identique de part est d'autre du point. Il peut parfois être judicieux de rendre ces fonctions de covariance anisotropes. C'est ce qu'illustre la figure 1.2.(b). En confrontant ce profil de fonction de covariance à ce qui a été dit sur la figure 1.1, on s'attend à ce que la correction faite sur l'analyse se répercute plus avant le point d'observation qu'après. La caractérisation de l'anisotropie se fait à partir de la définition de deux longueurs de portée dites amont et aval.

²Le vecteur de contrôle comporte des variables de différentes natures.

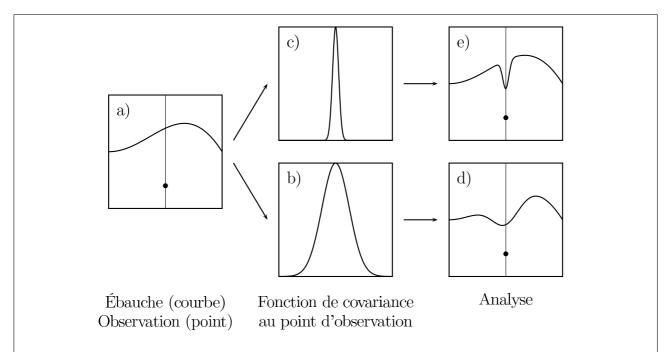


FIG. 1.1 – Influence du profil de la fonction de covariance au point d'observation sur l'élaboration d'une analyse avec un domaine discrétisé.

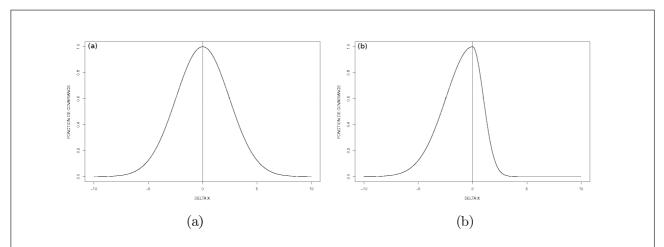


FIG. 1.2 – Fonctions de corrélation isotrope à gauche (a) et anisotrope à droite (b), pour le point central du domaine.

1.4 Diagnostic de la longueur de portée

Une façon d'estimer la distance de corrélation significative est d'introduire le diagnostic de la longueur de portée. Elle caractérise la corrélation d'une erreur en un point avec les erreurs aux points voisins. Daley [Daley, 1991] en donne une formule générale pour les problèmes à une dimension :

$$L_p = \sqrt{\frac{-1}{\nabla^2 \rho(0)}} \Longleftrightarrow L_p = \sqrt{\frac{-1}{\frac{d^2 \rho}{dx^2}(0)}} \Longleftrightarrow \frac{d^2 \rho}{dx^2}(0) = \frac{-1}{L_p^2}.$$
 (1.5)

En effectuant un developpement de Taylor à l'ordre 2 en δx d'une fonction de corrélation $\rho(x)$ on obtient

$$\rho(\delta x) \approx \rho(0) + \frac{\delta x^2}{2} \frac{d^2 \rho(0)}{dx^2}.$$
 (1.6)

L'expression (1.5) reportée dans (1.6) donne alors, en notant que $\rho(0) = 1$, car ρ est une fonction de corrélation :

$$\rho(\delta x) \approx 1 - \frac{\delta x^2}{2L_p^2}.$$

D'où l'expression de L_p :

$$L_p = \frac{\delta x}{\sqrt{2(1 - \rho(\delta x))}}.$$

La figure 1.3 montre en trait plein une demi-fonction de corrélation gaussienne. La courbe en pointillés correspond au développement de Taylor à l'ordre 2 de cette fonction. Cette courbe, appelée parabole osculatrice à l'origine permet de définir la longueur de portée. Cette dernière correspond à la demi-largeur de la parabole lorsque sa hauteur est égale à 0, 5.

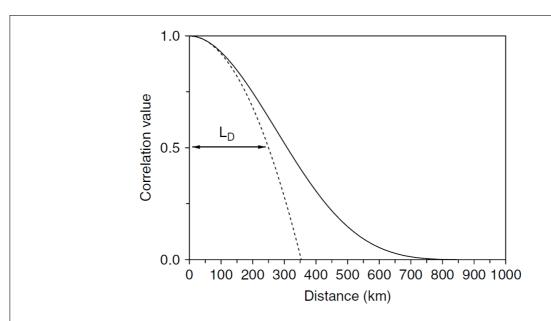


FIG. 1.3 – Illustration graphique de la longueur de portée d'une fonction de corrélation (courbe noire) pour laquelle la parabole osculatrice à l'origine (courbe en tireté) vaut 0.5 [Pannekoucke et al., 2007].

1.5 Filtre de Kalman

Dans le filtre de Kalman le calcul de l'analyse est similaire à celui fait dans l'algorithme du BLUE. Entre autre, il fait également intervenir la notion d'incrément d'analyse

$$\mathbf{d}_i = \mathbf{Y}_i^o - \mathbf{H}_i \mathbf{X}_i^b, \tag{1.7}$$

$$\mathbf{K}_{i} = \mathbf{B}_{i} \mathbf{H}_{i}^{T} \left(\mathbf{H}_{i} \mathbf{B}_{i} \mathbf{H}_{i}^{T} + \mathbf{R}_{i} \right)^{-1}, \tag{1.8}$$

$$\delta \mathbf{X}_{i}^{a} = \mathbf{K}_{i} (\mathbf{Y}_{i}^{o} - \mathbf{H}_{i} \mathbf{X}_{i}^{b}) = \mathbf{K}_{i} \mathbf{d}_{i}. \tag{1.9}$$

En revanche on remarque une différence importante dans la suite des algorithmes : le code du BLUE paramètre la matrice ${\bf B}$ au début du processus. Celle ci est statique durant toute la phase d'assimilation (Par. 1.2). Le filtre de Kalman permet lui, l'évolution à chaque cycle d'assimilation de l'ensemble des matrices intervenant dans l'algorithme. Ceci est vrai en particulier pour ${\bf B}$:

$$\mathbf{A}_i = (\mathbf{I} - \mathbf{K}_i \mathbf{H}_i) \, \mathbf{B_i} \tag{1.10}$$

$$\mathbf{B}_{i+1} = \mathbf{M}\mathbf{A}_i\mathbf{M}^T + \mathbf{Q}_i, \tag{1.11}$$

où \mathbf{Q}_i est la matrice de covariance d'erreurs modèle, \mathbf{A}_i est la matrice de covariance d'erreurs de prévision et $\mathbf{M} = \mathbf{M}_{i,i+1}$ est l'opérateur modèle de prévision depuis l'instant i jusqu'à l'instant i+1. Cette méthode permet en particulier aux différentes matrices entrant en jeu de "suivre" la dynamique du problème. La dernière étape de l'algorithme est à nouveau identique à celui du BLUE : l'analyse sert d'ébauche au cycle d'assimilation suivant.

$$\mathbf{X}_{i+1}^b = \mathbf{M}\mathbf{X}_i^a. \tag{1.12}$$

Dans le cadre de notre étude la matrice de covariance d'erreurs modèle ${\bf Q}$ est considérée comme nulle. Cette hypothèse est forte et entraine une diminution progressive des fonctions de covariances contenues dans ${\bf B}$. Concrètement l'erreur d'ébauche diminue peu à peu. On prend donc de moins en moins en compte les observations . Cependant, dans la mesure où cette étude s'intéresse aux longueurs de portée, un travail sur les fonctions de corrélation (qui elles sont normalisées) est satisfaisant.

1.6 Filtre de Kalman d'ensemble

L'objectif du filtre de Kalman d'ensemble [Evensen, 1994] reste identique au filtre de Kalman vu précédemment, soit prendre en compte dans l'assimilation de données la dynamique d'évolution des erreurs associées à une prévision. Pour atteindre cet objectif, le filtre de Kalman applique les équations du modèle caractérisées par l'opérateur ${\bf M}$ à la matrice ${\bf B}$, au travers de l'équation 1.11.

Le principe du filtre de Kalman d'ensemble est différent, il consiste à établir la matrice **B** de covariance d'erreurs débauche d'un point de vue statistique à partir d'un ensemble de couples ébauche—analyse représentant une même réalité et évoluant parallèlement.

Il opère sur un ensemble comportant N analyses, $\mathcal{E}^a = \{\mathbf{X}^{a,k}, 1 \leq k \leq N\}$, dont chacun des membres représente une même réalité perturbée par une erreur propre supposée non biaisée. Selon le même principe, un ensemble d'ébauches, $\mathcal{E}^b = \{\mathbf{X}^{b,k}, 1 \leq k \leq N\}$, est associé aux analyses.

Ensemble Kalman Filter

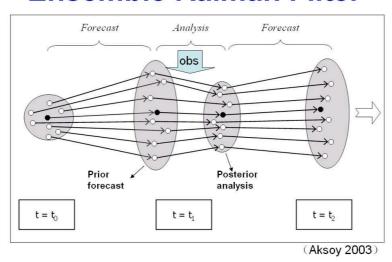


FIG. 1.4 – Principe de fonctionnement d'un filtre de Kalman d'ensemble, les ellipses autour des états schématisent la variance des différents membres.

Un cycle d'assimilation au temps t_i se déroule tel qu'illustré par la figure 1.5. Tout d'abord, un ensemble d'ébauches est obtenu en appliquant le modèle à chacun des membres de l'ensemble d'analyses du pas de temps précédent. Ensuite la matrice de covariance d'erreurs d'ébauche, $\mathbf{B}_i = \mathbb{E}(\mathbf{X}_i^b - \overline{\mathbf{X}_i^b})(\mathbf{X}_i^b - \overline{\mathbf{X}_i^b})^T$, est estimée statistiquement;

$$\mathbf{B}_{i} = \frac{1}{N} \sum_{k=1}^{N} (\mathbf{X}_{i}^{b,k} - \overline{\mathbf{X}}_{i}^{b}) (\mathbf{X}_{i}^{b,k} - \overline{\mathbf{X}}_{i}^{b})^{T}; \quad \text{avec} \quad \overline{\mathbf{X}}_{i}^{b} = \frac{1}{N} \sum_{k=1}^{N} \mathbf{X}_{i}^{b,k}.$$

De même que dans le BLUE et le filtre de Kalman, cette matrice fournit la matrice de gain selon l'équation 1.8. Enfin, l'ensemble d'analyses au temps t_i est obtenu, en assimilant pour chaque membre de l'ensemble d'ébauches l'observation \mathbf{Y}_i^o perturbée par un bruit $\boldsymbol{\varepsilon}_i^k$, selon l'équation (1.9). On oberve alors une baisse de la variance, ces états analysés servent alors après propagation par le modèle d'ébauche pour l'analyse suivante (Fig. 1.4).

La convergence de la matrice \mathbf{B} s'effectue en $1/\sqrt{N}$ selon le théorème central limite. La taille de l'ensemble est déterminée par la précision souhaitée, ainsi que par la complexité des calculs d'ébauche et d'analyse de chacun des membres.

Le filtre de Kalman d'ensemble, en permettant le calcul séparé des différentes analyses et ébauches, est particulièrement bien adapté aux machines massivement parallèles.

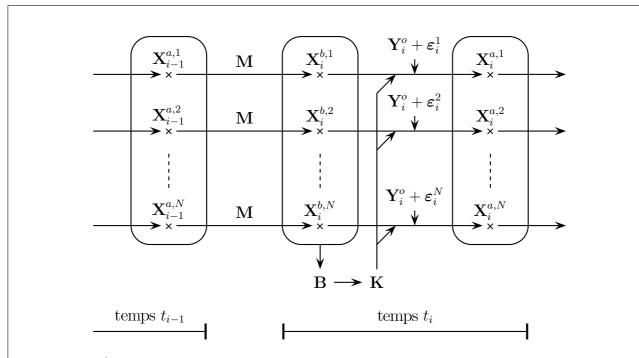


FIG. 1.5 – Évolution du filtre de Kalman d'ensemble. Le modèle appliqué à l'analyse de chacun des membres au temps t_{i-1} permet d'obtenir l'ébauche au temps t_i . L'ensemble des ébauches \mathbf{X}_i^b fournit la matrice de covariance d'erreurs \mathbf{B} , puis la matrice de gain \mathbf{K} . L'observation \mathbf{Y}_i^o , perturbée d'un bruit $\boldsymbol{\varepsilon}_i^o$ est alors assimilée pour chacune des ébauches.

Chapitre 2

Modèle d'onde de crue

Ce chapitre introduit dans un premier temps le modèle d'écoulement à surface libre dérivé des équations de Saint-Venant, utilisé dans cette étude [Thual, 2010], avant de présenter sa résolution numérique.

2.1 Physique du problème

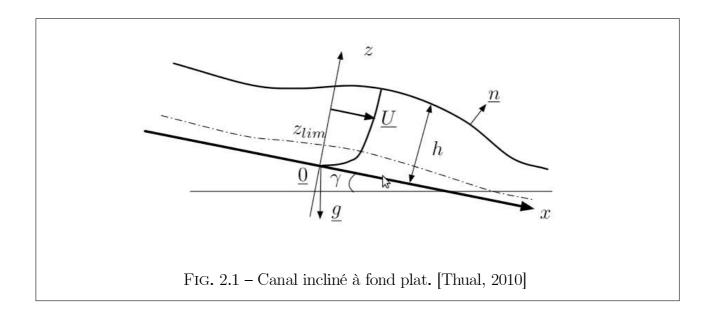
Un bief est un canal à pente faible utilisant la gravité pour acheminer l'eau. On s'intéresse à la modélisation des ondes de crues sur un bief incliné d'un angle γ avec l'horizontale (Fig.2.1). La hauteur d'eau est notée h. On suppose de plus une invariance selon y, c'est à dire $\forall l$, $\frac{\partial l}{\partial y} = 0$. Pour un écoulement turbulent, les équations de Navier-Stokes s'écrivent sous la forme suivante, avec p_t la pression turbulente et ν_t la viscosité turbulente, de façon à traiter la turbulence d'une manière paramétrée

$$\begin{cases}
\frac{\partial u}{\partial x} + \frac{\partial z}{\partial w} = 0, \\
\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + w \frac{\partial u}{\partial z} = -\frac{1}{\rho} \frac{\partial p_t}{\partial x} + g \sin \gamma + \nu_t \Delta u, \\
\frac{\partial w}{\partial t} + u \frac{\partial w}{\partial x} + w \frac{\partial w}{\partial z} = -\frac{1}{\rho} \frac{\partial p_t}{\partial z} - g \cos \gamma + \nu_t \Delta w.
\end{cases} (2.1)$$

A partir des équations 2.1 précédentes, le système d'équations de Saint-Venant se déduit à l'aide de différentes hypothèses :

- l'équation de la surface libre est de la forme F(x, z, t) = z h(x, t) = 0, ce qui permet d'exclure les déformations de type déferlement.
- Au niveau de la couche limite de fond, z_{lim} est considéré comme très petit.
- La profondeur est supposée faible devant l'échelle de variation des phénomènes dans la direction de l'écoulement.
- La vitesse longitudinale moyenne U devient une variable du problème, en intégrant les équations sur la hauteur du canal.
- Les frottements sont paramétrés par la relation : $\tau_{frot} = \frac{1}{2}C_f\rho U|U|$, avec C_f un coefficient de frottement qui dépend de la structure de la rivière (rugosité du fond, hauteur, etc).

¹Il existe plusieurs manières de modéliser ce coefficient. Nous considérons ici l'approximation de Manning-Strickler.



Les équations de Saint-Venant qui est le système d'équations codées dans le modèle MAS-CARET [Ricci et al., 2010] sont alors,

$$\begin{cases}
\frac{\partial h}{\partial t} + U \frac{\partial h}{\partial x} = -h \frac{\partial U}{\partial x}, \\
\frac{\partial U}{\partial t} + U \frac{\partial U}{\partial x} = -g \cos \gamma \frac{\partial h}{\partial x} + g \sin \gamma - \frac{Cf U|U|}{2h}.
\end{cases} (2.2)$$

Dans le cadre de ce projet, ce n'est pas ce système d'équations qui est utilisé. En effet, des approximations supplémentaires sont réalisées afin de simplifier le problème. En particulier, ces simplifications permettent de passer d'un problème multivarié (h, U) à un problème univarié (h).

- La pente est supposée faible, c'est à dire que le terme en $g\sin\gamma$ est négligeable dans le système de Saint-Venant.
- Le nombre de Froude est également supposé faible, ce qui signifie physiquement que la vitesse de l'écoulement est faible devant la vitesse des ondes de surface en milieu peu profond.

Pour de petites perturbations autour d'un état de référence (h_r, U_r) on écrit :

$$h = h_r + \tilde{h}$$
 et $U = U_r + \tilde{U}$.

En réalisant cette approximation et en négligeant les termes d'ordre 2 dans les équations 2.2, apparaît finalement l'équation des ondes de crues diffusantes, avec $\kappa_n = \frac{U_n h_n}{2 \tan \gamma}$, U_n la vitesse moyennée sur la verticale et h_n la hauteur d'eau moyenne :

$$\frac{\partial \tilde{h}}{\partial t} + \frac{5U_n}{3} \frac{\partial \tilde{h}}{\partial x} = \kappa_n \frac{\partial^2 \tilde{h}}{\partial x^2}.$$

L'équation régissant les ondes de crues diffusantes est en fait une équation d'advection/diffusion qui sera noté par la suite 2 :

²Pour plus de simplicité, \tilde{h} est remplacé par h.

$$\frac{\partial h}{\partial t} + c \frac{\partial h}{\partial x} = \kappa \frac{\partial^2 h}{\partial x^2}.$$
 (2.3)

2.2 Schéma numérique de l'équation d'advection-diffusion

L'étude physique précédente permet de modéliser l'écoulement dans un cadre continu : les anomalies de hauteur sont régies par l'équation d'advection-diffusion (Eq. 2.3). Sa modélisation numérique nécessite une discrétisation spatiale et temporelle, puis une approximation des dérivées. La méthode retenue est celle des différences finies, en approximant les dérivées par des taux d'accroissement sur des intervalles de petites dimensions.

2.2.1 Schéma de discrétisation

L'écoulement est modélisé sur un domaine [0; L] discrétisé au moyen d'une grille de pas Δx et comportant n points. Le temps est discrétisé selon un pas de temps Δt . L'écoulement est alors représenté numériquement par l'ensemble discret des valeurs $h_i^j = h(j\Delta x, i\Delta t)$. La modélisation de cet écoulement consiste à établir une relation $\mathbf{h}_{i+1} = \mathbf{M}\mathbf{h}_i$ où $\mathbf{h}_i = \begin{pmatrix} h_0^i & \cdots & h_i^n \end{pmatrix}^T$ est le vecteur des valeurs discrétisées de h à l'instant $i\Delta t$. On utilise un schéma centré d'ordre 2 pour la dérivée spatiale $\frac{\partial h}{\partial x}$, centré d'ordre 1 pour la dérivée $\frac{\partial^2 h}{\partial x^2}$ et un schéma d'Euler explicite d'ordre 1 pour la dérivée temporelle $\frac{\partial h}{\partial t}$. En dehors des extrémités du domaine $(1 \leq j \leq n-1)$, l'équation du modèle numérique s'écrit :

$$\frac{h_{i+1}^j - h_i^j}{\Delta t} + c \frac{h_i^{j+1} - h_i^{j-1}}{2\Delta x} = \kappa \frac{h_i^{j+1} - 2h_i^j + h_i^{j-1}}{\Delta x^2}.$$
 (2.4)

En l'absence de forçage, la condition limite en x=0 est modélisée par $h_i^0=0$. En x=L, elle est modélisée par une sortie libre, $\frac{\partial h}{\partial t}(L,t)+c\frac{\partial h}{\partial x}(L,t)=0$, discrétisée par $\frac{h_{i+1}^j-h_i^j}{\Delta t}+c\frac{h_i^j-h_i^{j-1}}{\Delta x}$. Les équations du modèle du temps i au temps i+1 s'écrivent matriciellement $\mathbf{h}_{i+1}=\mathbf{M}\mathbf{h}_i$ avec la matrice \mathbf{M} du modèle :

$$\mathbf{M} = \begin{pmatrix} 0 & \cdots & 0 & \cdots & 0 \\ a_0 & a_1 & a_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & a_0 & a_1 & a_2 \\ 0 & \cdots & 0 & a_3 & a_4 \end{pmatrix}$$

Les coefficients a_0, \ldots, a_4 dépendent des paramètres de la modélisation physique, c et κ , ainsi que des paramètres de la modélisation numérique, Δx et Δt :

$$a_0 = \frac{c\Delta t}{2\Delta x} + \frac{\kappa \Delta t}{\Delta x^2}; \quad a_1 = 1 - \frac{2\kappa \Delta t}{\Delta x^2}; \quad a_2 = -\frac{c\Delta t}{2\Delta x} + \frac{\kappa \Delta t}{\Delta x^2}; \quad a_3 = \frac{c\Delta t}{dx}; \quad a_4 = 1 - \frac{c\Delta t}{dx}.$$

La prise en compte d'un forçage en x=0, $\mathbf{f}_i=\begin{pmatrix} q(i\Delta t) & 0 & \cdots & 0 \end{pmatrix}^T$ entraı̂ne l'ajout d'un terme non linéaire :

$$\mathbf{h}_{i+1} = \mathbf{M}\mathbf{h}_i + \mathbf{f}_{i+1}.$$

2.2.2 Stabilité du schéma

L'étude de stabilité permet de vérifier que l'énergie, c'est-à-dire la norme carré du signal, reste bornée au cours du temps. Or la discrétisation de l'équation d'advection en utilisant un schéma centré pour la dérivée spatiale est connue pour être instable [Gicquel and Vermorel, 2010]. Il s'agit donc de voir sous quelles conditions la présence du terme de diffusion permet de rendre ce schéma stable.

L'analyse est effectuée selon la méthode de Von Neumann, dont le principe consiste à utiliser une transformation de Fourrier afin d'étudier la stabilité de chaque mode séparément. Plus formellement cette méthode repose sur le théorème de Plancherel, selon lequel une fonction de carré intégrable possède la même norme carrée que sa transformée de Fourrier : $||h||_2 = ||\hat{h}||_2$ [Revuz, 1997]. Les propriétés sur les transformés de Fourier permettent alors de transformer les translations en multiplications :

$$h^{j}(\widehat{x+\Delta x})(\zeta) = e^{-2i\pi\Delta x}\widehat{h^{j}(x)}(\zeta).$$

Ce qui est ensuite appliqué à l'équation du schéma numérique (Éq. 2.4):

$$\frac{\widehat{h^{i+1}} - \widehat{h^{i}}}{\Delta t} + \frac{c}{2\Delta x} \left(e^{2i\pi\zeta\Delta x} - e^{-2i\pi\zeta\Delta x} \right) \widehat{h^{i}} - \frac{\kappa}{\Delta x^{2}} \left(e^{2i\pi\zeta\Delta x} - 2 + e^{-2i\pi\zeta\Delta x} \right) \widehat{h^{i}} = 0$$

Le gain g, quotient de la transformée de Fourrier du signal entre deux instants successifs, est alors :

$$g(\zeta) = \frac{\widehat{h^{i+1}}}{\widehat{h^{i}}} = 1 - \frac{c\Delta t}{\Delta x} i \sin(2\pi\zeta\Delta x) + 2\frac{\kappa\Delta t}{\Delta x^{2}} \left(\cos(2\pi\zeta\Delta x) - 1\right). \tag{2.5}$$

La stabilité du schéma est alors acquise si pour chaque mode, soit $\forall \zeta \in \mathbb{R}$, le module du gain est inférieur à 1. Le détail des calculs figure en annexe A.1, la condition de stabilité prend la forme de trois inéquations :

$$\alpha = c \frac{\Delta t}{\Delta x} \le 1; \quad \beta = 2\kappa \frac{\Delta t}{\Delta x^2} \le 1; \quad \frac{\alpha^2}{\beta} = \frac{c^2 \Delta t}{2\kappa} \le 1.$$

Ce qui conduit à la condition sur Δt , voir figure 2.2 :

$$\Delta t \le \min \left\{ \frac{\Delta x}{c}; \frac{\Delta x^2}{2\kappa}; \frac{2\kappa}{c^2} \right\}.$$

Ces inégalités, nommées condition CFL³, font apparaître la CFL d'un schéma d'advection, $c\frac{\Delta t}{\Delta x} \leq 1$ et la CFL d'un schéma de diffusion, $2\kappa\frac{\Delta t}{\Delta x^2} \leq 1$. Le schéma d'advection seul n'étant pas stable, la dernière inéquation, $\frac{c^2\Delta t}{2\kappa} \leq 1$ indique dans quelle mesure la diffusion permet de le stabiliser. Par exemple, avec les paramètres suivants, c=1 m·s⁻¹, $\kappa=500$ m²·s⁻¹, un bief de longueur L=500 km, discrétisée en n=1000 points, il vient un pas de temps maximum $\Delta t=250$ s.

2.2.3 Équation équivalente

Un développement limité en $(j\Delta x, i\Delta t)$ permet de vérifier que le schéma résout effectivement l'équation d'advection-diffusion (Eq. 2.3). La fonction ν solution du schéma numérique est telle

³Condition Courant-Friedrichs-Lewy, du nom de Richard Courant, Kurt Friedrichs, et Hans Lewy

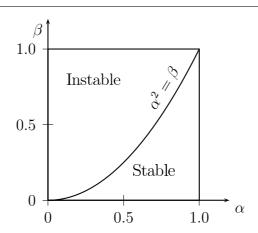


Fig. 2.2 – Condition de stabilité du schéma numérique selon les paramètres α et β .

que:

$$\frac{\partial \nu}{\partial t} + c \frac{\partial \nu}{\partial x} - \kappa \frac{\partial^2 \nu}{\partial x^2} + \underbrace{\frac{\Delta t}{2} \frac{\partial^2 \nu}{\partial t^2} + \Delta x^2 \left(\frac{c}{6} \frac{\partial^3 \nu}{\partial x^3} - \frac{\kappa}{12} \frac{\partial^4 \nu}{\partial x^4} \right) + O(\Delta t^2, \Delta x^3)}_{6} = 0. \tag{2.6}$$

À l'ordre 1 en temps et 2 en espace, le schéma résout l'équation d'advection-diffusion, à l'erreur ε près. D'après l'équation 2.6, le schéma numérique introduit une erreur ε . Cette erreur engendre des comportements du modèle numérique non souhaités dans le modèle physique, en particulier en agissant sur le cœfficient de diffusion, mais également en introduisant de la dispersion.

La dérivation de l'équation d'advection-diffusion (Eq. 2.3) permet d'exprimer les dérivées temporelles en fonction des dérivées spatiales.

$$\frac{\partial^2 \nu}{\partial t^2} = c^2 \frac{\partial^2 \nu}{\partial x^2} - 2c\kappa \frac{\partial^3 \nu}{\partial x^3} + \kappa^2 \frac{\partial^4 \nu}{\partial x^4}$$

Cela permet de réécrire l'équation résolue par le schéma numérique, et fait apparaître à l'ordre 2 des termes supplémentaires :

$$\frac{\partial \nu}{\partial t} + c \frac{\partial \nu}{\partial x} - \underbrace{\left(\frac{c\Delta x^2}{6} + c\kappa \Delta t\right) \frac{\partial^3 \nu}{\partial x^3}}_{\text{dispersion}} = \underbrace{\left(\kappa - \frac{c^2 \Delta t}{2}\right)}_{\kappa'} \frac{\partial^2 \nu}{\partial x^2} - \left(\frac{\kappa^2 \Delta t}{2} - \frac{\kappa \Delta x^2}{12}\right) \frac{\partial^4 \nu}{\partial x^4} + O(\Delta t^2, \Delta x^3)$$
(2.7)

2.2.4 Impact des termes diffusifs et dispersifs sur la propagation des covariances

L'équation (2.7), ou équation équivalente résolue par le schéma numérique, fait apparaître un terme de dispersion et deux termes de diffusion supplémentaires. Il faudra prendre en compte ces aspects lors des expériences numériques. En particulier, la diffusion apparente du schéma numérique $\kappa' = \kappa - \frac{c^2 \Delta t}{2}$, doit être prise en compte lors de la comparaison des résultats. En effet le coefficient de diffusion intervient dans la loi d'évolution de la longueur de portée (Par. 3).

En ce qui concerne le terme dispersif, ce dernier déforme les signaux propagés. En effet, la vitesse de phase est alors dépendante de la fréquence, et les différentes composantes d'un signal ne se déplace pas à la même vitesse. D'autant plus que dans certains cas, une des trois conditions CFL impose $\kappa' = 0$. De plus, les deux autre termes, dispersif et diffusif d'ordre 4, affectent particulièrement les petites échelles ce qui peut être gênant lors des simulations. Pour plus de détails, voir l'annexe A.2.

2.3 Construction du forçage amont

Afin d'avoir un signal de covariance spatiale gaussienne en entrée du modèle, la solution choisie est de construire un signal de covariance temporelle gaussienne et de forcer le modèle à son bord à chaque pas de temps avec cette fonction. Il faut construire en amont une condition limite de covariance temporelle Gaussienne [Pannekoucke, 2010]. Pour cela, on peut considérer un bruit blanc défini par sa décomposition de Fourier

$$\zeta = \sum_{n} \zeta_n e_n(t),$$

où les ζ_n sont des variables aléatoires suivant une loi normale centrée réduite. En particulier, l'indépendance deux à deux des variables aléatoires ζ_n impose la propriété suivante :

$$\left\langle \zeta_p \zeta_q^* \right\rangle = \delta_{pq}. \tag{2.8}$$

On applique à ce bruit blanc un opérateur L permettant de passer du bruit blanc au bruit de covariance gaussienne. Cet opérateur est déterminé à partir de coefficients de Fourier. En utilisant le fait que les e_n sont vecteurs propres de L associés aux valeurs propres l_n , il vient que :

$$q = L\zeta = \sum_{n} \zeta_n l_n e_n. \tag{2.9}$$

Il suffit alors de déterminer les l_n pour construire le signal voulu. En effet les ζ_n assurent le caractère aléatoire, tandis que les l_n assurent la coloration Gaussienne (i.e la covariance temporelle Gaussienne) du signal. Une façon d'accéder aux coefficients l_n nécessaires à la fabrication du signal d'entrée q est d'écrire la décomposition de Fourier de la fonction d'auto-covariance :

$$\rho(\delta t) = \langle q(t)q(t+\delta t)^* \rangle.$$

Il vient alors en utilisant l'équation 2.9 et la propriété 2.8 :

$$\rho(\delta t) = \langle q(t)q(t+\delta t)^* \rangle = \sum_{p,q} \langle (\zeta_p l_p e_p(t))(\zeta_q l_q e_q(t+\delta t))^* \rangle
= \sum_{p,q} \langle \zeta_p \zeta_q^* \rangle l_p l_q^* e_p(t) e_q(t+\delta t)^*
= \sum_{p} l_p l_p^* e_{-p}(\delta t)
\rho(\delta t) = \sum_{p} |l_p|^2 e_p(\delta t).$$
(2.10)

2.3. CONSTRUCTION DU FORÇAGE AMONT

La phase permet simplement de décaler le signal dans le temps, ce qui n'est pas important ici. Le signal voulu est alors facilement déductible à partir d'une fonction d'auto-covariance ρ connue. En effet, la décomposition de Fourier selon la base des e_n donne :

$$\rho(\delta t) = \sum_{n} \rho_n e_n(\delta t).$$

Par identification avec l'équation (2.10), il vient alors que $|l_n|^2 = |\rho_n|$ et

$$q(t) = \sum_{n} \zeta_{n} l_{n} e_{n} = \sum_{n} \zeta_{n} \sqrt{|\rho_{n}|} e_{n}(t).$$

Dans le cadre de cette étude, il sera considéré une fonction de covariance gaussienne, avec τ le temps caractéristique d'auto-corrélation :

$$\rho(t) = e^{\frac{-t^2}{2\tau^2}}.$$

L'algorithme permettant de réaliser une fonction d'entrée telle que celle présentée sur la figure (2.3) est détaillé ci-après :

$$> \rho(t) = e^{\frac{-t^2}{2\tau^2}}$$

% On effectue une discrétisation spatiale.

 $> \rho$

% vecteur contenant la discretisation spatiale de ρ ; (ρ_n)

 $> \hat{oldsymbol{
ho}} = \mathrm{fft}(oldsymbol{
ho})$

% Vecteur contenant les coefficients spectraux de $\rho.$

 $> \zeta =$ vecteur de la même taille que ho rempli avec des tirages aléatoires selon une loi normale centrée réduite.

$$> \hat{q} = \sqrt{|\hat{
ho}|} * \zeta$$

% Opération terme à terme.

% Vecteur contenant les coefficients spectraux de q; $(\zeta_n \sqrt{|\rho_n|})$

 $> q = \mathrm{fft}^{-1}(\hat{q})$

% Vecteur de sortie, forçage amont.

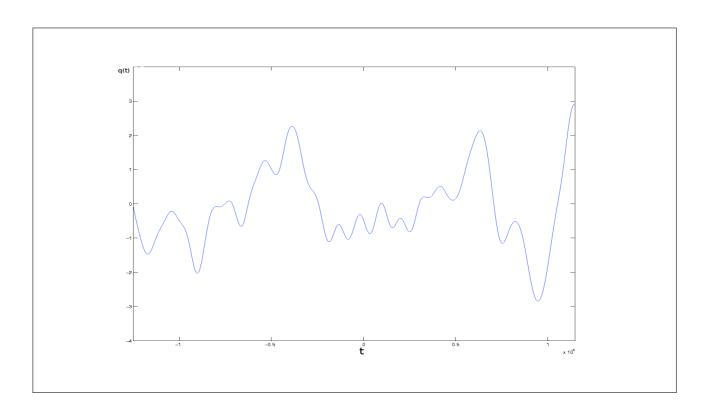


Fig. 2.3 – Exemple de réalisation de forçage amont.

Chapitre 3

Dynamique des covariances : évolution des longueurs de portée

L'étude se tourne désormais vers l'étude de la propagation des fonctions de covariances, et plus particulièrement vers l'évolution des longueurs de portée. Cependant cette étude n'est pas aisée. En effet, alors qu'une fonction de covariance est propagée par le modèle, elle n'est pas directement accessible comme la valeur de la hauteur d'eau. Dans cette partie sera traitée l'aspect théorique de la propagation de fonctions de covariance par la dynamique de l'équation 2.3.

3.1 Dynamique des covariances sans observations

Le but de cette partie est de trouver analytiquement la forme de la covariance spatiale d'un signal propagé par l'équation 2.3. La méthode proposée est de résoudre analytiquement la propagation du signal h(x,t), avant d'écrire la fonction d'auto-covariance spatiale de la solution. Le système à résoudre est le suivant, avec c la vitesse d'advection, κ le coefficient de diffusion, q(t) la condition limite en x=0, et h_0 la condition initiale.

$$\begin{cases}
\frac{\partial h}{\partial t} + c \frac{\partial h}{\partial h} = \kappa \frac{\partial^2 h}{\partial x^2} \\
h(0, t) = q(t) \\
h(x, 0) = h_0(x), x > 0
\end{cases}$$
(3.1)

Le forçage q(t) peut être vu comme une superposition de signaux harmoniques sinusoïdaux :

$$q(t) = \int q_{\omega} e_{\omega}(t) d\omega.$$

Le fait que q(t) soit construit aléatoirement, c'est à dire que les q_{ω} sont des réalisations d'une variable aléatoire (voir 2.3) implique la propriété suivante :

$$\langle q_{\omega 1} q_{\omega 2}^* \rangle = |q_{\omega 1}|^2 \delta_{12}, \tag{3.2}$$

avec δ_{12} le symbole de Kronecker. De plus, q(t) réelle, impose la condition : $|q_{\omega}|^2 = |q_{|\omega|}|^2$. La fonction d'auto-covariance, en utilisant la propriété 3.2 est alors donnée par :

$$\rho(\delta t) = \langle q(t)q(t+\delta t)^* \rangle = \int \rho_{\omega} e^{i\omega\delta t} d\omega = \int |q_{\omega}|^2 e^{i\omega\delta t} d\omega$$
 (3.3)

3.1. DYNAMIQUE DES COVARIANCES SANS OBSERVATIONS

Le signal q(t) possède un bruit Gaussien (i.e. coloration Gaussienne d'un bruit blanc). Ce qui se traduit par :

$$\begin{cases} \rho(\delta t) = e^{-\delta t^2/2\tau^2} \\ |q_{\omega}|^2 = \rho_{\omega} = e^{\omega^2\tau^2/2} \end{cases}$$

La linéarité du problème permet de ne considérer qu'un forçage sinusoïdal $q(t) = q_{\omega}e_{\omega}(t)$ pour la résolution. La solution est reconstruite à l'aide du théorème de superposition. Cette vision revient à chercher une solution de la forme $h(x,t) = q_{\omega}e^{-i\omega t}h_{\omega}(x)$. Dans la suite on cherchera donc des solutions sous cette forme, en posant $h_{\omega}(0) = 1$. La solution générale est dès lors reconstruite avec la relation suivante :

$$h(x,t) = \int q_{\omega} e^{-i\omega t} h_{\omega}(x) d\omega. \tag{3.4}$$

La fonction d'auto-covariance de h(x,t) spatiale est alors calculée en reportant l'équation 3.4 dans :

$$f(x, x + \delta x) = \langle h(x, t)h(x + \delta x, t)^* \rangle, \tag{3.5}$$

ce qui permet d'en déduire des propriétés sur les longueurs de portée.

3.1.1 Cas de l'advection pure

Dans ce cas, le système à résoudre est le suivant :

$$\begin{cases} \frac{\partial h}{\partial t} + c \frac{\partial h}{\partial x} = 0\\ h(x, t) = q_{\omega} e^{-i\omega t} h_{\omega}(x)\\ h_{\omega}(0) = 1 \end{cases}$$

En remplaçant la forme de la solution dans l'équation pronostique, il apparaît une équation différentielle ordinaire en h_{ω} :

$$\frac{\partial h}{\partial t} + c \frac{\partial h}{\partial x} = 0 \Leftrightarrow -i\omega h_{\omega} + ch'_{\omega} = 0 \Leftrightarrow h'_{\omega} = \frac{i\omega}{c} h_{\omega}$$

Les solutions de cette équation différentielle ordinaire, sont de la forme : $h(x,t) = Ae^{i\omega x/c}$, avec A une constante. En utilisant la condition initiale, $h_{\omega}(0) = 1$, il vient en notant $k = \frac{\omega}{c}$

$$h(x,t) = q_{\omega}e^{ik(x-ct)}.$$

On utilise alors la linéarité du problème pour en déduire la solution complète :

$$\begin{cases} q(t) = \int q_{\omega} e^{-i\omega t} d\omega \\ h(x,t) = \int q_{\omega} e^{ik(x-ct)} d\omega. \end{cases}$$

Il ne reste alors plus qu'à déterminer le tenseur des covariances spatiales. Il s'agit donc d'évaluer en un instant donné, la quantité : $f(x, x + \delta x) = \langle h(x, t)h(x + \delta x, t)^* \rangle$. Il vient ainsi que :

$$f(x, x + \delta x) = \langle h(x, t)h(x + \delta x, t)^* \rangle$$

$$= \langle \int_{\omega} q_{\omega} h_{\omega}(x) d\omega \int_{\beta} q_{\beta}^* h_{\beta}(x + \delta x)^* d\beta \rangle$$

$$= \int |q_{\omega}|^2 h_{\omega}(x) h_{\omega}(x + \delta x)^* d\omega$$

$$= \int |q_{\omega}|^2 e^{i\omega \delta x/c} d\omega$$
(3.6)

3.1. DYNAMIQUE DES COVARIANCES SANS OBSERVATIONS

On reconnait alors la forme de la fonction d'auto-covariance de la fonction q(t) en posant $\delta t = \delta x/c$. C'est à dire que l'on a :

$$f(x, x + \delta x) = \rho(\delta x/c)$$

avec $\rho(\delta x/c) = e^{-(\delta x/c)^2/2\tau^2} = e^{-\delta x^2/2Lp^2}$ en posant $Lp = c\tau$. Cette fonction ne dépend ni du temps t, ni de la position x, mais uniquement de la distance entre deux points du domaine. La solution analytique de Lp dans le cadre du forçage considéré, est

$$Lp = Constante = c\tau, (3.7)$$

avec τ le temps caractéristique d'auto-corrélation de la condition limite à gauche.

3.1.2 Cas de l'advection diffusion

Dans ce cas, le système à résoudre est le suivant :

$$\begin{cases} \frac{\partial h}{\partial t} + c \frac{\partial h}{\partial x} = k \frac{\partial^2 h}{\partial x^2} \\ h(x,t) = q_{\omega} e^{-i\omega t} h_{\omega}(x) \\ h_{\omega}(0) = 1 \end{cases}$$

En remplaçant la forme de la solution dans l'équation pronostique, il apparaît une équation différentielle en h_{ω} :

$$\frac{\partial h}{\partial t} + c \frac{\partial h}{\partial x} = k \frac{\partial^2 h}{\partial x^2} \Leftrightarrow -i\omega h_\omega + ch'_\omega = kh''_\omega \Leftrightarrow kh''_\omega + \frac{i\omega}{c} h_\omega - h'_\omega = 0$$

Le polynôme caractéristique associé à cette équation différentielle est :

$$kZ^2 - cZ + i\omega = 0$$

Les racines de ce polynôme sont : $z^{+/-} = \frac{c}{2k} \pm \frac{\sqrt{c^2 - 4i\omega k}}{2k}$ ce qui permet d'en déduire la forme des solutions de l'équation différentielle :

$$h_{\omega}(x) = Ae^{z^+x} + Be^{z^-x}$$

On déduit ensuite les constantes A et B des données du problème. En particulier, si $A \neq 0$, alors $\lim_{x\to\infty} |h_{\omega}(x)| = \infty$. Cette solution étant physiquement absurde, on en déduit que A = 0. De plus, on a $h_{\omega}(0) = 1$, ce qui donne B = 1. La solution de l'équation différentielle ordinaire est donc pour un signal harmonique :

$$h_{\omega}(x) = e^{\left(\frac{c}{2k} - \frac{\sqrt{c^2 - 4i\omega k}}{2k}\right)x}$$

On utilise alors la linéarité du problème pour en déduire la solution complète :

$$\begin{cases} q(t) = \int q_{\omega} e^{-i\omega t} d\omega \\ h(x,t) = \int q_{\omega} e^{(\frac{c}{2k} - \frac{\sqrt{c^2 - 4i\omega k}}{2k})x - i\omega t} d\omega \end{cases}$$

On note qu'en considérant un cas avec une diffusion peu intense, $k \ll 1$, la racine du polynôme caractéristique se simplifie : $z \approx i \frac{\omega}{c} - \frac{\omega^2 k}{c^3}$. L'ordre zéro, permet de retrouver la solution de l'advection pure, et l'ordre un permet de corriger cette solution par un terme d'atténuation en $e^{-\frac{\omega^2 k}{c^3}x}$.

3.1. DYNAMIQUE DES COVARIANCES SANS OBSERVATIONS

Il ne reste alors plus qu'à déterminer le tenseur des covariances spatiales. Il s'agit donc d'évaluer en un instant donné, la quantité : $f(x, x + \delta x) = \langle h(x, t)h(x + \delta x, t)^* \rangle$. Il vient ainsi que :

$$f(x, x + \delta x) = \langle h(x, t)h(x + \delta x, t)^* \rangle$$

$$= \langle \int_{\omega} q_{\omega} h_{\omega}(x) d\omega \int_{\beta} q_{\beta}^* h_{\beta}(x + \delta x)^* d\beta \rangle$$

$$= \int |q_{\omega}|^2 h_{\omega}(x) h_{\omega}(x + \delta x)^* d\omega$$
(3.8)

En considérant alors l'approximation à l'ordre 1 de la diffusion, il vient alors que :

$$f(x, x + \delta x) = \int \left(|q_{\omega}|^2 e^{-\omega^2 k(2x + \delta x)/c^3} \right) e^{i\omega \delta x/c} d\omega$$

La première remarque à faire est que le tenseur des covariances spatiales, ne dépend toujours pas du temps, mais dépend cependant cette fois-ci de la position x. De plus, en remplaçant $|q_{\omega}|^2$ par sa valeur.

$$f(x, x + \delta x) = \int \left(e^{-\omega^2/2(\tau^2 + 4kx/c^3 + 2k\delta x/c^3)}\right) e^{i\omega\delta x/c} d\omega$$

$$\underset{\delta x \ll 1}{\approx} \int \left(e^{-\omega^2/2(\tau^2 + 4kx/c^3)}\right) e^{i\omega\delta x/c} d\omega$$

$$\underset{\delta x \ll 1}{\approx} \int e^{-\omega^2 L p_x^2/2} e^{i\omega\delta x/c} d\omega$$

$$\underset{\delta x \ll 1}{\approx} e^{-\delta x^2/2L p_x^2}$$
(3.9)

Ce calcul nous donne donc une estimation de Lp donnée par la relation :

$$Lp_x^2 = c^2(\tau^2 + 4kx/c^3) = Lp_0^2 + 4kx/c.$$
(3.10)

3.1.3 Bilan de la dynamique des longueurs de portée sans assimilation

Dans le cas d'une advection pure, la solution analytique de Lp dans le cadre du forçage considéré, est (Fig. 3.1) :

$$Lp = Constante = c\tau,$$
 (3.11)

avec τ le temps caractéristique d'auto-corrélation de la condition limite à gauche.

Dans le cas d'une advection diffusion, une estimation de Lp en supposant que la fonction de covariance est localement gaussienne (on se place dans le cas ou $\delta x \ll 1$). Cette estimation est donnée par la relation (Fig. 3.1) :

$$Lp_x^2 = c^2(\tau^2 + 4kx/c^3) = Lp_0^2 + 4kx/c.$$
 (3.12)

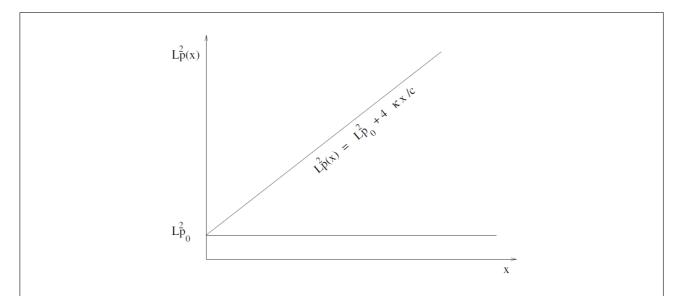


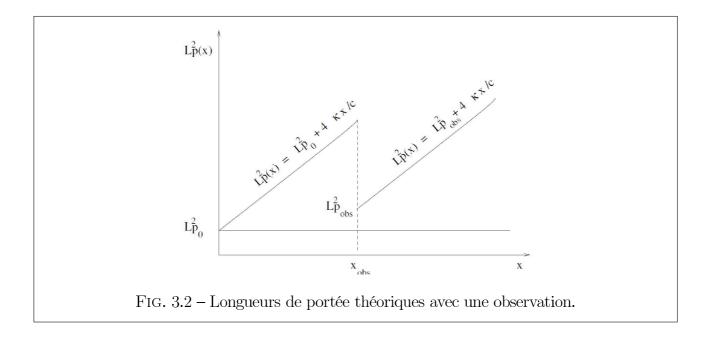
Fig. 3.1 – Représentation de la longueur de portée théorique, dans le cas d'une advection diffusion (droite croissante) et d'une advection pure (droite horizontale).

3.2 Dynamique des covariances avec observation

Le cas avec observation est beaucoup plus compliqué à mettre en œuvre analytiquement. L'assimilation de données implique une diminution de la variance au point d'observation. La propagation des covariances vers l'aval entraine alors une diminution de la longueur de portée en aval du point d'observation (Fig. 3.2). La diminution de cette variance étant dépendante du rapport σ_o^2/σ_b^2 , la diminution de la longueur de portée l'est également .

On se propose alors d'établir un abaque du rapport des longueurs de portée entre l'aval et l'amont au point d'observation en fonction du rapport des variances. Cet abaque a pour but de tenter de paramétrer la matrice de covariance d'erreurs d'ébauche de l'assimilation dans un BLUE pour le modèle MASCARET.

 $^{^1\}sigma_o^2,$ la variance des erreurs d'observations et σ_b^2 la variance des erreurs sur l'ébauche



Chapitre 4

Expériences numériques

Le paragraphe précédent résout mathématiquement la propagation des fonctions de covariance pour le schéma numérique utilisé. Cette partie traite de la vérification expérimentale de ces résultats mathématiques. Pour cela on diagnostique les longueurs de portée dans le cas d'une advection seule et dans celui d'une advection diffusion sans assimilation. Ensuite en se concentrant sur le modèle d'advection/diffusion, les mêmes diagnostics sont effectués dans le cas d'une assimilation à l'aide d'un algorithme de type BLUE ou d'un filtre de Kalman.

4.1 Diagnostic de B et des longueurs de portée sans assimilation

La réalisation de ce diagnostic nécessite la détermination de la matrice ${\bf B}$. Une méthode ensembliste permet d'en obtenir une estimation. Désormais, les matrices de covariances d'erreurs déterminées statistiquement seront indicé par e.

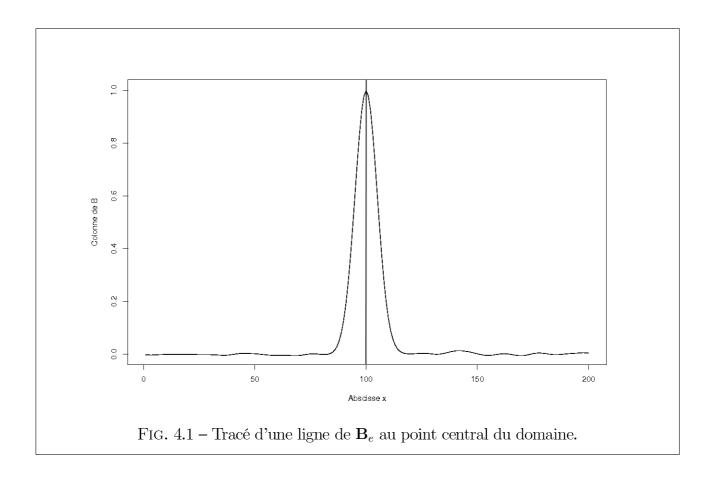
4.1.1 Calcul de la matrice de covariances d'erreurs d'ébauche

L'estimation de la matrice \mathbf{B}_e est faite statistiquement à partir de la réalisation d'un ensemble de propagations de signaux d'entrée q(t) différents. Afin d'obtenir une matrice sans bruit d'échantillonnage, ce travail est réalisé sur un échantillon de 10^5 membres. Chaque forçage est propagé le long de la rivière. Le temps de propagation est choisi de façon à ce que le signal q ait pu atteindre la limite aval du domaine. Les différents vecteurs (\mathbf{X}_k) , $k \in [1, 10^5]$ sont gardés en mémoire (vecteur de taille n contenant la perturbation de la hauteur d'eau en chaque point de discrétisation spatiale). La matrice de covariances d'erreurs spatiales \mathbf{B}_e est alors directement déterminée à partir des vecteurs (\mathbf{X}_k) en calculant les covariances entre les différents vecteurs :

$$\mathbf{B}_e = \frac{1}{N} \sum_{k=1}^{N} ((\mathbf{X}_k - \overline{\mathbf{X}})(\mathbf{X}_k - \overline{\mathbf{X}})^T),$$

en notant $\overline{\mathbf{X}}$ = Vecteur des moyennes en chaque point sur l'ensemble des membres.

Par construction \mathbf{B}_e est carrée, de dimension n^2 et symétrique positive, mais pas nécessairement définie. En effet, son rang est au plus N. La figure 4.1 montre le tracé d'un ligne de \mathbf{B}_e quelconque. On y voit l'allure de la fonction de covariance au point central du domaine.



4.1.2 Calcul des longueurs de portée

A partir de cette matrice \mathbf{B}_e qui contient une estimation statistique des fonctions de covariances, on peut calculer les longueurs de portée amont et aval au point courant. La formule utilisée pour la longueur de portée est celle décrite dans le paragraphe 1.4. Le choix a également été fait de travailler sur les fonctions de corrélation plutôt que sur les fonctions de covariance étant donné que la définition de longueur de portée est valable pour les fonctions de corrélations. Finalement,

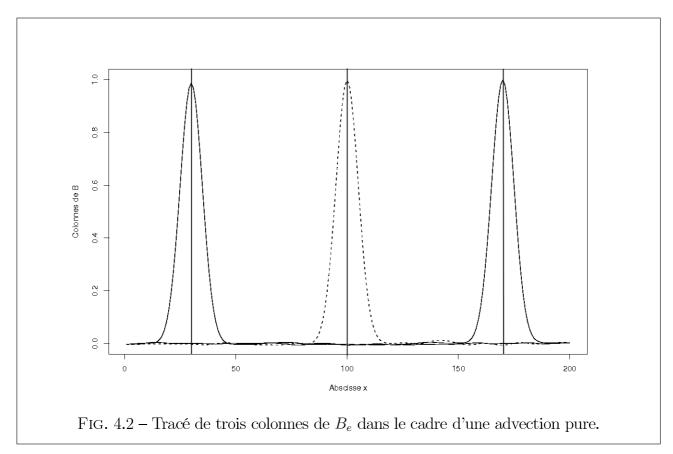
$$\rho(0) = 1, \quad \rho(\delta x) = \frac{\mathbf{B}_e(i, i+1)}{\sqrt{\mathbf{B}_e(i, i)\mathbf{B}_e(i+1, i+1)}} \quad \text{et} \quad \rho(-\delta x) = \frac{\mathbf{B}_e(i, i-1)}{\sqrt{\mathbf{B}_e(i, i)\mathbf{B}_e(i-1, i-1)}}.$$

4.2 Résultats des expériences sans assimilation

4.2.1 Modèle avec advection seule

Dans le cadre de l'advection seule, le signal d'entrée q est simplement propagé au cours du temps vers l'aval sans qu'il ne subisse aucune modification. D'après les calculs réalisés en 3 on s'attend à obtenir une longueur de portée constante par rapport à x sur l'ensemble du domaine. De plus $L_p^- = L_p^+$ implique que la fonction de covariance spatiale est identique (par rapport au point considéré) d'une limite à l'autre du domaine. Les résultats obtenus sont présentés dans les figures 4.2 et 4.3. Sur la figure 4.2 sont superposées les trois fonctions de covariances des points d'abscisses $x = \frac{Lmax}{6}$, $x = \frac{Lmax}{2}$ et $x = \frac{5Lmax}{6}$. La fonction de covariance est donc, au même titre que le signal lui même, simplement translatée par la schéma numérique sans être déformée. La figure 4.3 montre l'évolution du carré de la longueur de portée (leurs courbes

étant identiques, aucune distinction L_p^+ - L_p^- n'est représentée) avec l'abscisse. Une régression linéaire 1 (trait plein) permet de confronter l'expérience et la théorie. Avec une ordonnée à l'origine de $2494km^2$ on est proche des $2500km^2$ attendus. Le coefficient directeur estimé est lui d'un ordre de grandeur négligeable. On est là aussi en accord avec avec la théorie.



4.2.2 Modèle avec advection et diffusion

D'après la théorie (voir chapitre 3), la diffusion doit ici jouer sur "l'étalement" des fonctions de corrélation. En effet si les longueurs de portée augmentent bien linéairement le long de la rivière, on s'attend à un étalement progressif des fonctions de covariance. Expérimentalement la figure 4.4 illustre l'évolution du carré de la longueur de portée en fonction de x. Une régression linéaire a également été faite. Là encore, les coefficients directeurs et ordonnées à l'origine théoriques et expérimentaux sont en accord.

La figure 4.5 superpose plusieurs fonctions de corrélation aux mêmes points qu'en figure 4.2. L'influence de la diffusion qui "écrase et étale" les fonctions de covariance au fur et à mesure de leur propagation est bien visible.

Enfin la figure 4.6 montre un problème numérique auquel on peut être confronté. Le signal injecté en entrée possède un temps caractéristique faible. Autrement dit, ce signal contient beaucoup de "petites échelles". Au début du domaine, ces petites échelles ont pour effet de rendre non négligeable un terme de diffusion en $\frac{\partial^4 h}{\partial x^4}$ (Eq. 2.7) qu'on avait laissé de côté en raison de son ordre de grandeur dans le cadre général. La régression linéaire (en trait plein) réalisée sur les derniers points montre qu'on tend peu à peu vers une solution convenable. Il convient

¹réalisée avec le logiciel R : http://www.r-project.org/

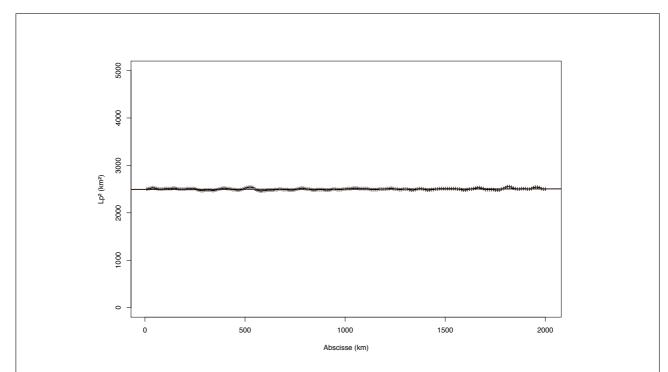


FIG. 4.3 – Représentation des longueurs de portée aval dans le cas purement advectif. Les croix représentent les longueurs de portée expérimentales, la droite correspond à la régression linéaire, et la droite tireté correspond à la droite théorique (les deux droites sont quasiment superposées). Théoriquement, $Lp^2 = 2500km^2$.

donc d'introduire un signal amont possédant un temps caractéristique d'auto-corrélation assez grand pour éviter cette difficulté.

4.3 Assimilation avec un BLUE

On réalise ici le même travail que dans la partie 4.2 mais en effectuant une assimilation de données avec un algorithme de BLUE. La matrice de covariance d'erreur d'ébauche $\bf B$ utilisée est celle déterminée dans la partie 4.1. La matrice de covariance d'erreurs de prévision obtenue est désormais notée $\bf A_e$. Le code numérique implémenté transcrit la formulation du BLUE présentée dans le chapitre 1. Comme précédemment le travail est fait sur 10^5 membres. On attend tout d'abord que le signal ait le temps de se propager sur tout le domaine. Puis la phase d'assimilation démarre. Le point d'observation est placé en milieu de domaine et une observation est assimilée tous les 5 pas de temps. Le temps d'assimilation est assez important pour que l'ajout d'information apporté par les observations ait pu atteindre les limites du domaine.

On considère dans un premier temps la figure ?? représentant les valeurs de la variance de l'erreur d'analyse (en trait plein) le long du cours d'eau. La courbe en tiretés correspond aux valeurs de la variance de l'erreur de prévision dans le cas sans observations. L'assimilation a donc permis une réduction de variance au point d'observation.

Une méthode pour la détermination de la structure de la matrice \mathbf{B}_e serait de boucler le travail précédent; on remplace dans le code du BLUE la matrice \mathbf{B} par la matrice \mathbf{A}_e calculée à partir de l'ensemble des réalisations puis on recommence les calculs pour obtenir une nouvelle

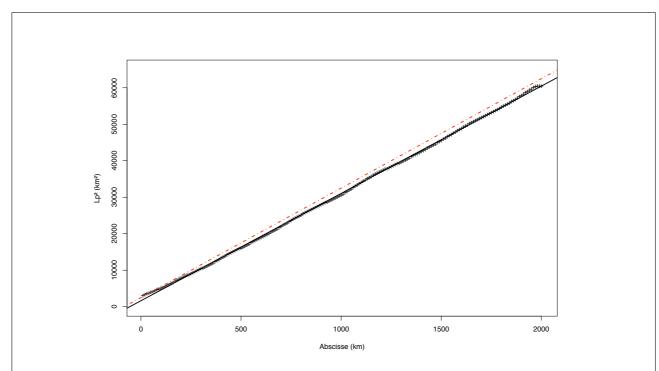


FIG. 4.4 – Représentation des longueurs de portée aval. On y trouve les points expérimentaux, la droite de régression en trait plein, et la droite théorique en tireté. $\kappa=7500m^2/s=7,5.10^{-3}km^2/s,~c=1m/s=10^{-3}km/s,~\frac{4\kappa}{c}=30km$. Coefficient directeur de la droite de régression : 29,4km, ordonnée à l'origine 1.628.10³km² .

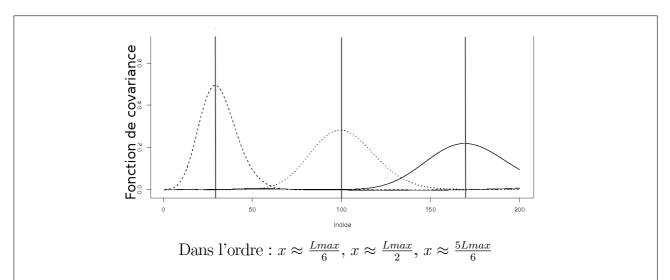


FIG. 4.5 – Représentation de quelques colonnes de \mathbf{B}_e , dans le cas d'une advection diffusion. Les paramètres utilisés pour cette simulation sont : $c = 1m/s, \kappa = 1000m^2/s, N = 100000$.

matrice \mathbf{A}_e qui servira à l'itération suivante. Cependant son coup de calcul est énorme. Le filtre de Kalman est à présent la méthode retenue pour construire une matrice \mathbf{B} convenable.

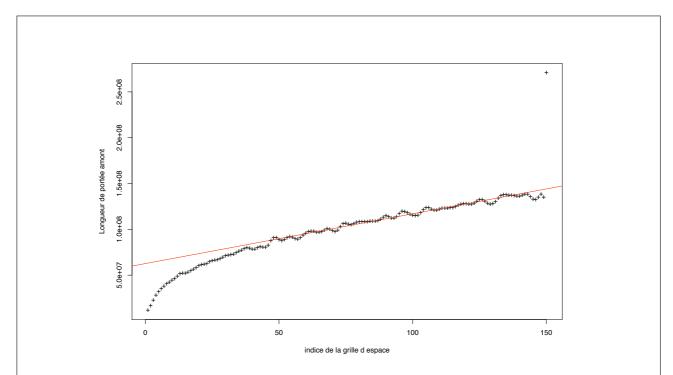


FIG. 4.6 – Représentations des longueurs de portée aval, et régression linéaire, dans le cas d'une hyperdiffusion. Les croix sont les points expérimentaux, la droite correspond à une régression linéaire effectuée en omettant les premiers points expérimentaux.

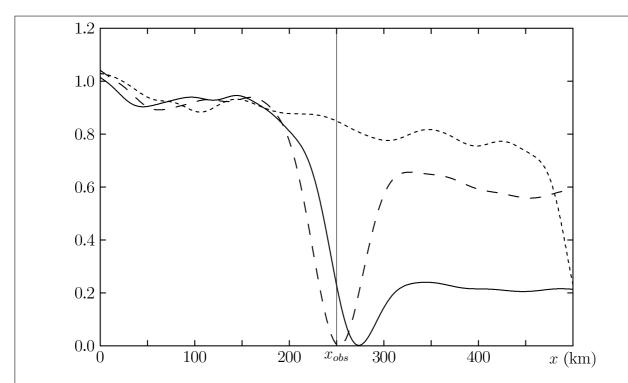


FIG. 4.7 – Représentation des variances de la matrice \mathbf{B}_e sans assimilation (trait pointillé), après un cycle complet d'assimilation tous les 10 pas de temps (tirets), tous les 75 pas de temps (trait plein).

4.4 Implémentation du filtre de Kalman

Dans le BLUE, la matrice **B** de covariance des erreurs d'ébauche est statique. Elle ne tient pas compte de la réduction de l'incertitude engendrée par l'assimilation d'observations. Après quelques itérations, les covariances d'erreurs diffèrent senbiblement des conditions initiales, l'analyse n'est plus optimale.

Le principe du filtre de Kalman a été exposé dans la parte 1.5. Sa mise en œuvre va permettre de déterminer l'effet d'une série d'observations sur les erreurs d'ébauche, en tenant compte de la diffusion et de l'advection. Cela doit permettre un diagnostic des longueurs de portée des fonctions de corrélation au point d'observation.

4.4.1 Cadre de la simulation

Il s'agit de déterminer la matrice de covariance d'erreurs d'ébauche dans un cadre stationnaire (dans le sens où une observation est effectuée à chaque pas de temps) à laquelle est associée une erreur d'écart-type constant. Dans l'état initial, c'est-à-dire avant la période d'observation, les covariances d'erreurs ont été déterminées par méthode ensembliste tel que vu précédemment dans la partie 4.1, avec un écart-type $\sigma_q=1$ m, et un temps caractéristique, $\tau=30\cdot 10^3$ s. La simulation est effectuée sur un domaine de longueur L=500 km, discrétisé sur une grille de n=1000 points. La condition CFL impose un pas de temps de $\Delta t=250$ s. L'observation est affectée d'une erreur d'écart-type 0.5 m. En outre la vitesse d'advection est de c=1 m·s⁻¹, la diffusion de $\kappa=500$ m²·s⁻¹. En vertu des résultats de la partie 2.2.4, la diffusion réellement simulée est alors de $\kappa'=375$ m²·s⁻¹

La figure 4.8 montre les variances d'erreurs d'ébauche tout au long du bief, c'est-à-dire la diagonale principale de la matrice **B**. Les deux courbes correspondent à l'état initial, et à l'état après 400 itérations, soit une advection de 100 km. Dans l'état initial, les variances sont décroissantes. Sur la courbe représentant les variances après 400 iérations, une zone de très faibles variances c'est formée au delà du point d'observation. L'information apportée par l'observation a fortement réduit l'incertitude, au point d'observation, l'advection a ensuite étendue cette zone vers l'aval.

De plus, une zone dans laquelle les variances sont nulles ou quasi-nulles s'est artificiellement propagée à partir de l'amont, sur une distance correspondante à l'advection. En effet, cela provient de l'impossibilité de modéliser l'incertitude entrant en amont dans le système. Celle-ci est négligée et se propage ainsi ensuite. Afin que cette zone n'atteigne pas le point d'observation, la simulation est nécessairement limitée en temps, et les résultats suivants seront donnés sur un domaine restreint par rapport au domaine de calcul, avec un temps d'assimilation limité. Seul l'éloignement du point d'observation permet de limiter ces contraintes. Il est ainsi impossible de parvenir à un état rigoureusement stationnaire, mais les résultats suivants montrent une convergence assez rapide dans le domaine d'intérêt, c'est-à-dire autour du point d'observation. Seul un filtre de Kalman d'ensemble permet de dépasser ces limites.

En dehors de ces deux zones, les variances ne sont que très peu affectées.

La figure 4.9 montre, après 300 pas de temps, l'influence de la fréquence d'observation sur la variance d'erreur d'ébauche. L'augmentation de la fréquence d'observation réduit la variance dans la zone située au dela du point d'observation, jusqu'à la rendre quasiment nulle lorsque les observations sont effectuées à chaque pas de temps.

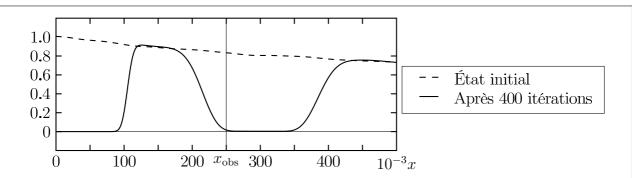


FIG. 4.8 – Propagation par le filtre de Kalman des variances d'erreur données par la diagonale de **B**. Le point d'observation est matérialisé par une ligne verticale à $x_{\rm obs}=250$ km.

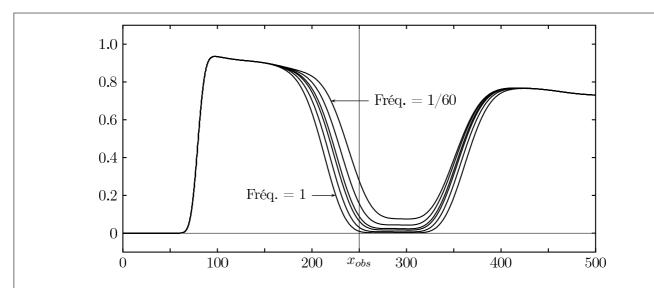


FIG. 4.9 – Variance tout au long du bief après 300 pas de temps, fréquence d'observation de 1 à 1/60, $\sigma_o = 0.5$.

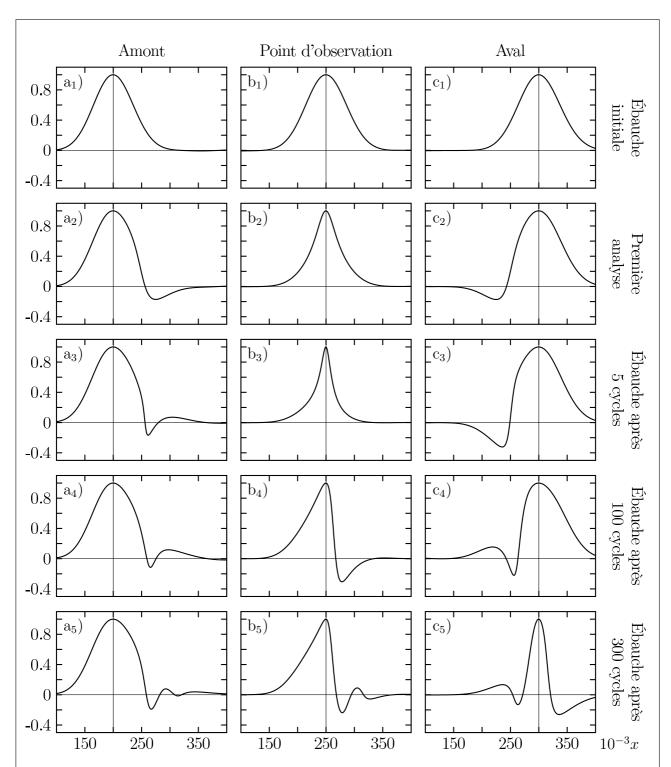


FIG. 4.10 – Fonctions d'autocorrélation représentées en trois points, matérialisés par un trait vertical, en amont du point d'observation (courbes a_1 à a_5), au point d'observation (courbes b_1 à b_5) et en aval (courbes c_1 à c_5). Les fonctions d'autorrélation sont données pour les ébauches à différents instants, ainsi que pour la première analyse.

4.4.2 Évolution des fonctions de corrélation

Le graphique 4.10 montre les fonctions d'autocorrélation en trois points du bief, au point d'observation et en deux points amont et aval, chacun distant de 50 km du point d'observation. Comparée aux longeurs de portée initiales, cette distance est relativement faible, et ces deux points sont influencés par la présence des observations.

Dans l'état initial les fonctions d'autocorrélation sont celles de la partie 4.2 caractérisées par des longueurs de portée croissantes en allant vers l'aval. La première assimilation impacte de manière isotrope la fonction d'autocorrélation au point d'observation (courbe b_2) et réduit sensiblement les longueurs de portée. Les points amont et aval (courbes a_2 et c_2) sont impactés de manière symétriques. L'advection n'a pas pu, à ce stade, entraîner un déplacement de l'information vers l'aval. Cependant, dès les premiers cycles d'assimilation (courbes a_3 à c_3), une tendance à l'anisotropie semble se dégager.

Au cours des cycles suivants, les fonctions d'autocorrélation en amont, au point d'observation, et en aval évoluent de manières différentes. Au point amont (courbes a₄ et a₅), l'assimilation de données influence la fonction d'autocorrélation, mais celle-ci conserve des longueurs de portée relativement isotropes et constantes au cours du temps.

Au point d'observation, la fonction d'autocorrélation prend très rapidement un caractère anisotrope qu'elle conserve ensuite. Les longueurs de portée amont et aval diffèrent fortement, mais leur quotient semble converger assez rapidement. Ce point est étudié plus particulièrement dans la suite.

Au point aval, la fonction d'autocorrélation passe par trois étapes : isotrope avec des longueurs de portée importantes, anisotrope, puis isotrope avec des longueurs de portée réduites. Cela résulte de l'advection. Tout d'abord ce point n'est pas concerné par l'information provenant de l'observation, ensuite l'information parvient à ce point entraînant le caractère anisotrope, finalement ce point est traversé par un flot continu d'information. La fonction de corrélation retrouve ainsi un caractère isotrope mais avec des longueurs de portée réduites.

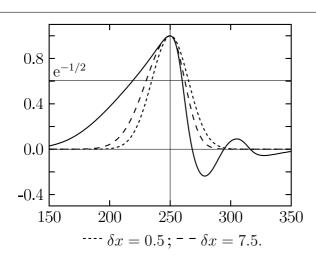
4.4.3 Méthode d'estimation des longueurs de portée

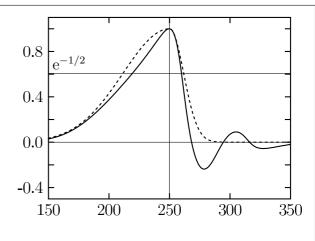
Les graphiques précédents (figure 4.10) ont montré que les fonctions d'autocorrélation au point d'observation ne sont plus gaussiennes. L'estimation des longueurs de portée selon la formule (Eq. 4.1), basée sur l'approximation de fonctions gaussiennes, est fortement dépendante de l'écart au point d'observation, δx , utilisé. La figure 4.11 illustre les approximations gaussiennes obtenues pour différentes valeurs de δx au moyen des fonctions :

$$\rho(dx) = \begin{cases}
\exp\left(-\frac{dx^2}{2Lp^-}\right) & \text{pour } dx \le 0; \\
\exp\left(-\frac{dx^2}{2Lp^+}\right) & \text{pour } dx \ge 0.
\end{cases}$$
(4.1)

Il en ressort que cette méthode d'approximation traduit mal le caractère anisotropique de la fonction d'autorrélation pour des valeurs trop petites de δx , et oblige à un choix arbitraire de δx . L'estimation de l'anisotropie par le quotient Lp^+/Lp^- devient alors dépendante non seulement des paramètres du modèle physique, mais également de ceux du modèle numérique.

La méthode proposée, illustrée par la figure 4.12, consiste à définir Lp^+ et Lp^- comme étant





- (a) Longueurs de portée déterminées selon l'équation (1.5), pour différentes valeur de δx .
- (b) Longueurs de portée déterminées selon la formule (4.2).

Fig. 4.11 – Comparaison de différentes approximations du calcul de la longueur de portée. La fonction d'autocorrélation au point d'observation est représentée en trait plein. L'approximation gausienne selon la formule (4.1) est représentée en tirets.

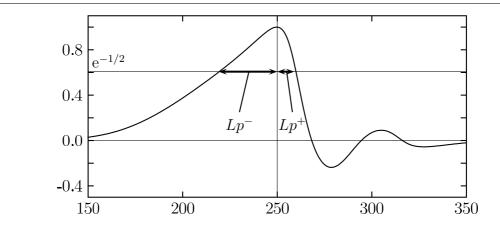


Fig. 4.12 – Définition de la longueur de portée utilisée dans la suite de ce chapitre.

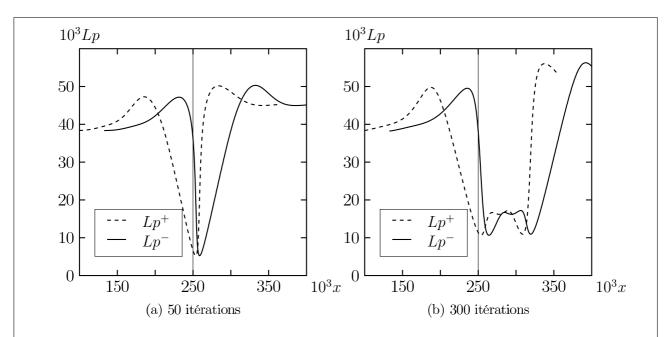


FIG. 4.13 – Longueurs de portée après itérations du filtre de Kalman, selon la position dans le bief. La ligne verticale symbolise la position du point d'observation.

les longueurs vérifiant :

$$\rho(Lp^+) = \rho(-Lp^-) = \exp\left(-\frac{1}{2}\right).$$
(4.2)

Cette définition, aussi arbitraire qu'elle soit, est plus robuste en cas de changement des paramètres du modèle numérique.

4.4.4 Diagnostic des longueurs de portée

Dans l'espace: La figure 4.13 donne les longueurs de portée en fonction de la position dans le bief. Au début de la simulation (figure 4.13(a)), les seuls points impactés se situent autour du point d'observation.

Cette zone s'élargie ensuite vers l'aval, figure 4.13(b). Deux zones présentent alors d'importantes différences entre les longueurs de portée à gauche et à droite; au point d'observation (x=250 km), et en aval (x=325 km). Cette deuxième zone correspond à la frontière jusqu'à laquelle l'advection a tranporté l'information.

Conformément au graphique 4.10, chaque point situé en aval passe par trois étapes. Dans un premier temps l'information n'est pas encore parvenue et les longueurs de portée sont isotropes et importantes. Quand l'information arrive, les longueurs de portée deviennent anisotropes, après quoi elles reprennent un caractère isotrope, mais deviennent beaucoup plus réduites.

Après 300 itérations, le graphique 4.13(b) fait apparaître un intervalle dans lequel les longueurs de portée sont réduites, et semblent présenter une légère pente positive due à la diffusion. L'étendue de cette zone correspond à la longueur d'advection obtenue après 300 itérations, soit 75 km. Cela est conforme à la figure 3.2, mais le nombre limité d'itérations possibles ne permet pas de voir si cette tendance se poursuit.

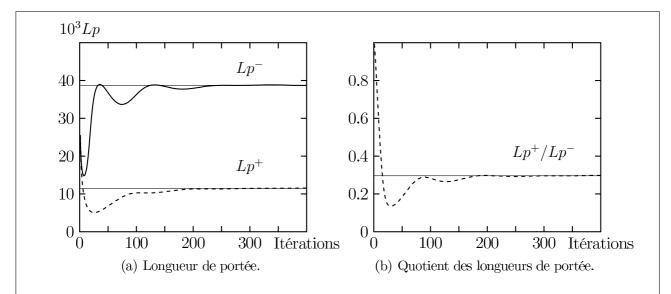


Fig. 4.14 – Évolution de la longueur de portée au point d'observation. Le trait horizontal matérialise la valeur après 300 itérations.

Dans le temps: La figure 4.14 montre l'évolution des longueurs de portée au point d'observation, selon le nombre d'itérations. Compte tenu de l'advection de l'information, les longueurs de portée devraient se stabiliser après un temps de l'ordre de grandeur de la longueur de portée divisée par la vitesse, soit 80 itérations pour une longueur de portée de 40 km. Quelques oscillations sont visibles au delà de 80 itérations, et les longueurs de portée se stabilisent à partir de 150 itérations. Dans la suite, les longueurs de portée au point d'observation sont définies comme celles obtenues après 300 itérations.

4.4.5 Influence du quotient σ_b/σ_o

Dans toutes les simulations précédentes, la précision de l'observation est restée constante. Cette précision est caractérisée par le quotient σ_b/σ_o , avec σ_b l'écart-type initial d'erreur d'ébauche au point d'observation, et σ_o celui de l'observation.

La figure 4.15 donne le quotient Lp^+/Lp^- de la longueur de portée à droite par la longueur de portée à gauche. Ce quotient dépend essentiellement de la longueur de portée aval. La figure 4.15(a) montre une division de la longueur de portée à droite par un facteur supérieur à deux, alors que la longueur de portée à gauche ne diminue que peu. L'information progressant vers l'aval, la longueur de portée amont n'est que peu influencée par l'observation. La figure 4.15(b), en coordonnées logarithmique, permet de vérifier que lorsque la précision de l'observation augmente (fort quotient σ_b/σ_o), la longueur de portée aval diminue (faible quotient Lp^+/Lp^-): Une augmentation de l'erreur d'observation réduit son importance et de ce fait réduit l'anisotropie.

L'objectif est maintenant d'obtenir une relation empirique du quotient Lp^+/Lp^- au point d'observation en fonction du quotient σ_b/σ_o . L'utilisation des coordonnées logarithmiques sur

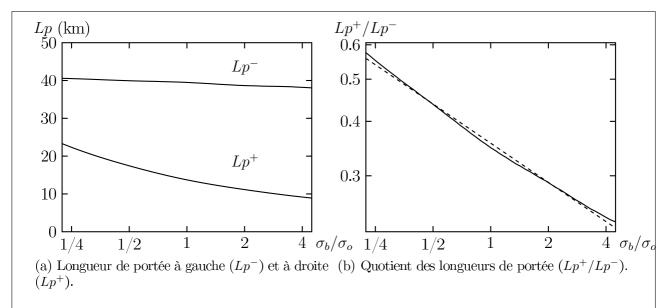


FIG. 4.15 – Longueurs de portée en fonction du quotient σ_b/σ_o en traits pleins, et ajustement affine.

la figure 4.15(b) montre qu'un ajustement affine est justifié.

$$\ln\left(\frac{Lp^{+}}{Lp^{-}}\right) = -1.03 \ln\left(\frac{\sigma_b}{\sigma_o}\right) - 0.30.$$

La paramétrisation des longueurs de portée semble donc envisageable.

4.4.6 Influence du temps caractéristique du signal

Le temps caractéristique du forçage amont détermine les longueurs de portée initiales, cela se répercute ensuite sur le quotient Lp^+/Lp^- . La figure 4.16 reprend les graphiques de la figure 4.15(b), pour différentes valeurs du temps caractéristique, τ . Le graphique 4.16(a), donnant les longueurs de portée, montre une forte dépendance de la longueur de portée amont. Celle-ci est essentiellement déterminée par les conditions initiales. En revanche, la longueur de portée aval est plus faiblement affectée, la précision de l'observation étant là le paramètre déterminant.

Une augmentation du temps caractéristique a alors comme principale conséquence une augmentation des longueurs de portée amont, puis une réduction du quotient Lp^+/Lp^- , tel qu'illustré sur la figure 4.16(b). Cette figure représente les courbes du quotient Lp^+/Lp^- en fonction de σ_b/σ_o pour différentes valeurs du temps caractéristique.

4.5 Implémentation filtre de Kalman d'ensemble

Le principe du filtre de Kalman d'ensemble a été étudié dans la partie 1.6. Sa mise en œuvre va permettre de lever les limitations vues en 4.4.1. Les paramètres de la simulation restent ceux de la partie 4.4.1, les résultats sont obtenus sur un ensemble de N=10000 membres.

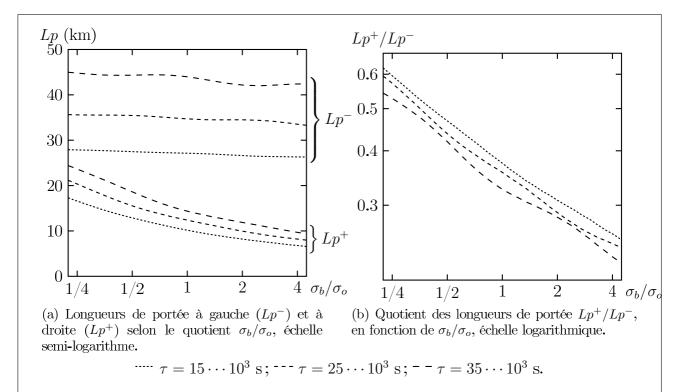


Fig. 4.16 – Influence du temps caractéristique, τ , sur les longueurs de portée, en fonction du quotient σ_b/σ_o .

La figure 4.17 donnant la fonction d'autocorrélation au point d'observation, permet de constater qu'une centaine de cycles suffisent à déterminer les longueurs de portée. Au delà, seules quelques petites oscillations se propagent. Ces résultats sont similaires à ceux du filtre de Kalman (Fig. 4.10). L'utilisation d'un filtre de Kalman est donc suffisante.

La figure 4.18 donne les longueurs de portée, tout au long du bief, selon l'équation (4.2). Les calculs sont effectués après 1997 cycles, soit le temps nécéssaire pour qu'un écoulement traverse complètement le bief. Au niveau du point d'observation il n'y a pas de différences notables avec les résultats du filtre de Kalman (Fig. 4.13). De plus la courbe obtenue est conforme à l'étude réalisée dans la partie 3.2.

Le filtre de Kalman d'ensemble n'améliore pas la qualité des résultats au point d'observation, il permet néanmois une vision plus globale des longueurs de portée. Cependant son utilisation pourrait s'avérer nécessaire pour atteindre la convergence des longueurs de portée dans le cas d'une réduction de la fréquence d'observation.

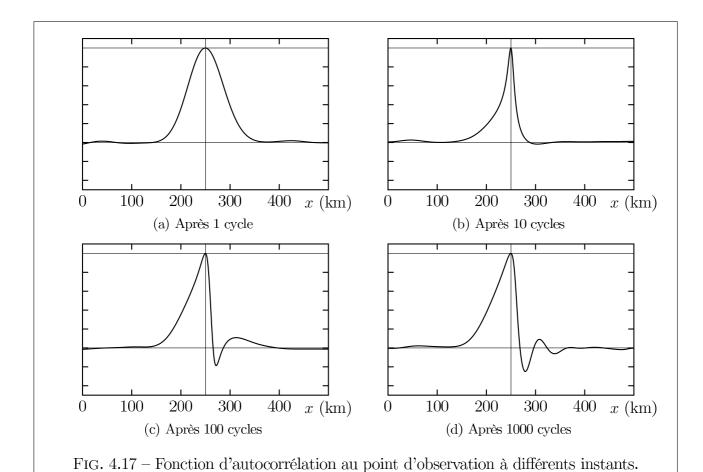


Fig. 4.18 – Longueurs de portée en fonction de la position dans le bief, obtenues par un filtre de Kalman d'ensemble. La ligne verticale matérialise le point d'observation.

Conclusion

Cette étude a permis, par le diagnostic des longueurs de portée, de quantifier le caractère anisotrope des fonctions de covariance d'erreurs d'ébauche en fonction du rapport $\frac{\sigma_b}{\sigma_o}$ au point d'observation. Une partie de l'information sur la dynamique de ces fonctions est représentée par leur caractère anisotrope. Les perspectives immédiates de ces travaux portent sur la modélisation d'une matrice B en but de l'utiliser dans un algorithme de BLUE. Bien que celle-ci reste statique, on peut s'attendre à une amélioration de l'analyse grâce à l'introduction de fonctions de covariance anisotropes correspondantes à la dynamique de l'équation d'onde de crues. Il s'agira donc de vérifier cette hypothèse et de quantifier l'amélioration obtenue dans le modèle d'ondes de crues, avant de tenter une modélisation similaire dans MASCARET.

D'autres pistes auraient également pu être approfondies. En effet l'influence du coefficient de diffusion sur le rapport $\frac{Lp^+}{Lp^-}$ n'a pas été étudiée. De plus il a été décidé d'assimiler une observation à chaque pas de temps. On aurait pu s'intéresser à l'effet d'une fréquence d'assimilation plus faible. Les conséquences peuvent être importantes notamment sur le nombre d'itérations nécessaires pour atteindre un état stationnaire des fonctions de covariance.

Annexe A

Annexe

Condition de stabilité du schéma numérique A.1

En introduisant les coefficients positifs $\alpha=c\frac{\Delta t}{\Delta x}$ et $\beta=2\kappa\frac{\Delta t}{\Delta x^2}$ et le réel $\theta=2\pi\zeta\Delta x$, le gain à chaque cycle du schéma numérique (Eq. 2.5) s'écrit :

$$|g|^2 = \underbrace{\alpha^2 \sin^2 \theta}_A + \underbrace{\left(1 - \beta + \beta \cos \theta\right)^2}_B \tag{A.1}$$

Dans l'équation (A.1) il est nécessaire d'avoir $A \leq 1$, d'où la condition $\alpha \leq 1$. Il est également nécéssaire d'avoir $B \leq 1$, ce qui est vérifié dès que $\beta \leq 1$.

La dernière condition, $\frac{\alpha^2}{\beta} \leq 1$, lie les paramètres α et β , comme illustré par la figure 2.2. Cette condition peut s'obtenir en étudiant le gain comme un polynôme en $X=\cos\theta$:

$$|g|^2 = ((1-\beta) + \beta X)^2 + \alpha^2 (1 - X^2)$$

= $(\beta^2 - \alpha^2) X^2 + 2(1-\beta)\beta X + (1-\beta)^2 + \alpha^2$

- Cas où $\beta^2 - \alpha^2 > 0$:

 $|g|^2$ prend ses valeurs maximales pour X=-1 et X=1. $|g|^2_{X=-1}=\beta^2-\alpha^2-2(1-\beta)\beta+(1-\beta)^2+\alpha^2=4\beta^2-4\beta+1$. La condition $\beta\leq 1$ entraı̂ne bien $|g|_{X=-1}^2 \le 1$. $|g|_{X=1}^2 = \beta^2 - \alpha^2 + 2(1-\beta)\beta + (1-\beta)^2 + \alpha^2 = 1$ Dans ce cas aucune condition supplémentaire n'est nécessaire pour obtenir $|g| \le 1$.

$$|g|_{X=1}^2 = \beta^2 - \alpha^2 + 2(1-\beta)\beta + (1-\beta)^2 + \alpha^2 = 1$$

- Cas où $\beta^2 - \alpha^2 < 0$:

 $|g|^2$ est représenté par une parabole d'axe $X = \frac{(1-\beta)\beta}{\alpha^2 - \beta^2}$.

Or
$$\frac{(1-\beta)\beta}{\alpha^2-\beta^2} \ge 1 \Leftrightarrow \beta \ge \alpha^2$$
.

Donc en prenant $\beta \geq \alpha^2$, $|g|^2$ atteint son maximum en X=1 et le calcul précédent fournit $|g|_{X=1}^2 = 1.$

En prenant $\beta < \alpha^2$, $|g|^2$ atteint son maximum en $X = \frac{(1-\beta)\beta}{\alpha^2 - \beta^2}$.

$$|g|_{X=\frac{(1-\beta)\beta}{\alpha^2-\beta^2}}^2 = \frac{\alpha^2(1-\beta)^2}{\alpha^2-\beta^2} + \alpha^2$$
. On a alors $|g|_{X=\frac{(1-\beta)\beta}{\alpha^2-\beta^2}}^2 \le 1 \Leftrightarrow \beta^2-2\alpha^2\beta+\alpha^4 \le 0$. L'étude

du polynôme $\beta \mapsto \beta^2 - 2\alpha^2\beta + \alpha^4$ montre qu'aucune valeur de β ne permet d'obtenir un schéma stable.

- Cas où $\beta^2 - \alpha^2 = 0$: $|g|^2 = 2(1-\beta)\beta X + (1-\beta)^2 + \alpha^2 \text{ atteint son maximum en } X = 1.$ $|g|^2_{X=1} = 2(1-\beta)\beta + (1-\beta)^2 + \alpha^2 = 1 + \alpha^2 - \beta^2 = 1.$

Dans ce cas aucune condition supplémentaire n'est nécessaire pour obtenir $|g| \le 1$. La stabilité du schéma numérique est définie par trois inéquations :

$$\alpha \le 1; \quad \beta \le 1; \quad \frac{\alpha^2}{\beta} \le 1$$

Soit:

$$c\frac{\Delta t}{\Delta x} \le 1; \quad 2\kappa \frac{\Delta t}{\Delta x^2} \le 1; \quad \frac{c^2 \Delta t}{2\kappa} \le 1$$

A.2 Influence du schéma numérique

L'équation 2.7, ou équation équivalente résolue par le schéma numérique fait apparaître un terme de dispersion. Ce terme, en $\frac{\partial^3 h}{\partial x^3}$, affecte tout particulièrement les petites échèles.

L'équation 2.7 présente également un coefficient de diffusion $\kappa' = \kappa - \frac{c^2 \Delta t}{2}$ cohérent avec les conditions de stabilité établies précédemment et dans lesquelles figure l'inéquation $\kappa - \frac{c^2 \Delta t}{2} \ge 0$. Cette inéquation s'interprète alors comme la nécessité d'avoir un coefficient de diffusion positif.

Le schéma numérique tend donc à sous-estimer la diffusion. Le graphique A.1 montre l'évolution d'une fonction gaussienne h(x,t), en prenant comme condition initiale une fonction de solution analytique connue.

$$h_{\kappa}(x,t) = q_m \frac{l_0}{l^2(t)} \exp\left[-\frac{x - x_0 - ct}{2l^2(t)}\right], \quad l^2(t) = l_0^2 + 2\kappa t.$$

La fonction simulée est moins diffusée qu'attendu par la solution analytique, cela étant d'autant plus vrai que la valeur de κ est petite.

La fonction h simulée reste une gaussienne, advectée à la même vitesse que la solution analytique. À chaque instant, elle coïncide avec une solution analytique de cœfficient de diffusion $\kappa_n(t)$, tel que $h(x,t) = h_{\kappa_n(t)}(x,t)$.

 $\kappa_n(t)$, peut-être considéré comme le coefficient de diffusion effectif du schéma numérique. Son estimation permet de vérifier que le coefficient κ' , introduit dans l'équation 2.7, est bien représentatif de la diffusion réelle du schéma.

Il est déterminé en considérant le maximum $h_m(t)$ de $h(x,t)=h_{\kappa_n(t)}(x,t)$, et vérifie ainsi $h_m^2(t)=q_m^2\frac{l_0^2}{l_0^2+2\kappa_n(t)t}$. D'où une estimation de $\kappa_n(t)$ par :

$$\kappa_n(t) = \left[\left(\frac{q_m}{h_m(t)} \right)^2 - 1 \right] \frac{l_0}{2t}.$$

Le graphique A.2, représentant l'évolution de $\kappa_n(t)$ à chaque pas de temps, montre une convergence de $\kappa_n(t)$. Cette valeur limite, κ_n , peut-être considérée comme le cœfficient de diffusion du schéma numérique. Les fluctuations observées lors des premiers pas de temps sont liées

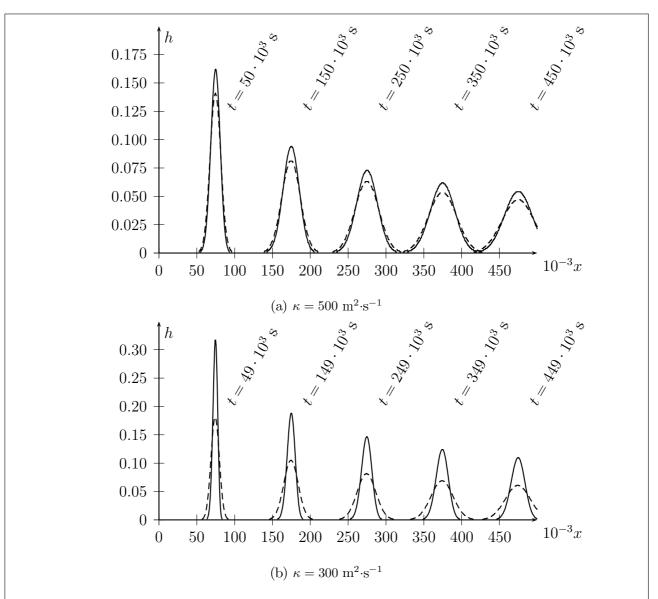


FIG. A.1 – Évolution d'un signal gaussien avec N=1000, $l_0=1$ km. En tirés, la solution analytique de l'équation d'advection-diffusion, en traits pleins, le signal de même condition initiale, simulé par le schéma numérique.

A.2. INFLUENCE DU SCHÉMA NUMÉRIQUE

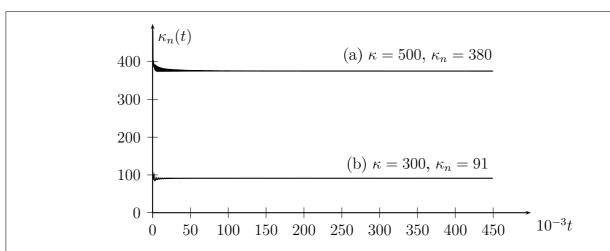


FIG. A.2 – Cœfficient κ_n permettant à chaque itération d'obtenir une courbe théorique identique à la courbe simulée. Les simulations sont effectuées avec les paramètres $\Delta x = 500$ m et $c = 1 \text{ m} \cdot \text{s}^{-1}$.

- (a) Courbe réalisée avec $\kappa = 500 \text{ m}^2 \cdot \text{s}^{-1}$. La condition de stabilité donne $\Delta t = 250 \text{ s}$, donc $\kappa' = \kappa \frac{c^2 \Delta t}{2} = 375 \text{ m}^2 \cdot \text{s}^{-1}$, alors que $\kappa_n = 380 \text{ m}^2 \cdot \text{s}^{-1}$.
- (b) Courbe réalisée avec $\kappa = 300 \text{ m}^2 \cdot \text{s}^{-1}$, La condition de stabilité donne $\Delta t = 416.7 \text{ s}$, donc $\kappa' = 91.7 \text{ m}^2 \cdot \text{s}^{-1}$, alors que $\kappa_n = 91$.

aux petites échelles, particulièrement sensibles aux termes en $\frac{\partial^3 h}{\partial x^3}$ et $\frac{\partial^4 h}{\partial x^4}$. Ces petites échelles sont rapidement diffusées, d'où la convergence de $\kappa_n(t)$.

Dans chacun des cas le coefficient κ' permet d'expliquer la majeure partie de l'écart entre la diffusion souhaitée, κ , et la diffusion effective du schéma numérique. La prise en compte d'une correction du coefficient de diffusion κ devra donc être effectuée lorsque le but d'une simulation est de tester la dépendance d'un résultat par rapport au coefficient de diffusion.

Bibliographie

- F. Bouttier and P. Courtier. Data assimilation concepts and methods, 1999.
- R. Daley. Atmospheric data analysis. Cambridge atmospheric and space science series. Cambridge University Press, 1991.
- G. Desroziers. Assimilation de données, notes de cours. Ecole Nationale de la Météorologie, 2010.
- Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *J. Geophys. Res.*, 99(C5):10143–10162, 1994.
- L.Y.M. Gicquel and O. Vermorel. Computational fluid dynamics (cfd): Discretization, stability, dispersion and dissipation. 2010.
- O. Pannekoucke. Note interne: Dynamique des covariances, 2010.
- O. Pannekoucke, L. Berre, and G. Desroziers. Background-error correlation length-scale estimates and their sampling statistics. *Quarterly Journal of the Royal Meteorological Society*, 2007.
- D. Revuz. Mesure et intégration. Collection Méthodes. Hermann, 1997.
- 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. 2010.
- A. Tarantola. Inverse problem theory: Methods for data fitting and model parameter estimation. Elvesier, 1987.
- O. Thual. Hydrodynamique de l'environnement. Ecole Polytechnique, 2010.