

Titre: Interpolation de la concentration de contaminant par krigage avec covariance non stationnaire en coordonnées d'écoulement
Title: covariance non stationnaire en coordonnées d'écoulement

Auteur: Martine Rivest
Author:

Date: 2012

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Rivest, M. (2012). Interpolation de la concentration de contaminant par krigage avec covariance non stationnaire en coordonnées d'écoulement [Thèse de doctorat, École Polytechnique de Montréal]. PolyPublie.
Citation: <https://publications.polymtl.ca/1005/>

Document en libre accès dans PolyPublie

Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/1005/>
PolyPublie URL:

Directeurs de recherche: Denis Marcotte, & René Lefebvre
Advisors:

Programme: Génie minéral
Program:

UNIVERSITÉ DE MONTRÉAL

INTERPOLATION DE LA CONCENTRATION DE CONTAMINANT PAR KRIGEAGE
AVEC COVARIANCE NON STATIONNAIRE EN COORDONNÉES D'ÉCOULEMENT

MARTINE RIVEST
DÉPARTEMENT DES GÉNIES CIVIL, GÉOLOGIQUE ET DES MINES
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION
DU DIPLÔME DE PHILOSOPHIÆ DOCTOR
(GÉNIE MINÉRAL)
NOVEMBRE 2012

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée :

INTERPOLATION DE LA CONCENTRATION DE CONTAMINANT PAR KRIGEAGE
AVEC COVARIANCE NON STATIONNAIRE EN COORDONNÉES D'ÉCOULEMENT

présentée par : RIVEST Martine
en vue de l'obtention du diplôme de : Philosophiæ Doctor
a été dûment acceptée par le jury d'examen constitué de :

M. CHOUTEAU Michel, Ph.D., président
M. MARCOTTE Denis, Ph.D., membre et directeur de recherche
M. LEFEBVRE René, Ph.D., membre et codirecteur de recherche
M. MOLSON John, Ph.D., membre
Mme MICHALAK Anna M., Ph.D., membre

REMERCIEMENTS

Je tiens avant tout à remercier mon directeur de recherche Denis Marcotte pour son dévouement, sa compréhension et sa grande générosité. Denis a toujours été prêt à démontrer patiemment des équations nébuleuses pour la visuelle que je suis. Je n'aurais pas pu mener ce projet à bien sans sa confiance et ses encouragements.

Je remercie mon co-directeur René Lefebvre pour l'intérêt manifesté pour mon projet. Les échanges avec René m'ont permis d'avoir un regard nouveau sur mon projet et ses applications possibles.

Mes remerciements vont également aux membres du jury, madame Anna M. Michalak et messieurs Michel Chouteau, Mario Lefebvre et John Molson pour avoir accepté de lire et d'évaluer cette thèse.

Il m'est difficile d'être brève dans mes remerciements lorsque je pense aux gens qui m'ont entourée durant ces années à Poly. Merci aux étudiants, professeurs, chercheurs, techniciens et associés de recherche du laboratoire de Géophysique et Géostatistique pour tous ces midis partagés ensemble au labo. Une mention spéciale aux étudiants ayant séjourné au C-616.5 : Julie, Maude, Camille, Pier-Olivier, Véronique, Min, Clémence, Alain et Louis. Merci pour toutes ces conversations hétéroclites sur les covariances croisées, la bouffe, le théorème de Bayes, Bibi et Geneviève, la politique, les bugs L^AT_EX... Merci à Philippe Pasquier pour les *pep talks* et l'aide avec COMSOL et *Hydrogeo_Inv*. Je remercie également Erwan Gloaguen pour de nombreuses discussions enrichissantes lors de mes visites à Québec, ainsi que les étudiants du LIAMG pour m'avoir accueillie dans leur local. En espérant n'oublier personne, merci à vous tous !

Un gros merci à tous mes amis pour avoir été là, que ce soit pour célébrer les victoires ou m'appuyer durant les périodes moroses. Merci Taïka, pour ta grande sagesse à discerner les moments qui nécessitent de l'écoute de ceux qui requièrent un « rentre-dedans » en bonne et due forme. Les meilleurs amis n'ont pas peur de nous secouer lorsque requis. Merci Camille, ton point de vue et tes conseils m'ont tellement souvent éclairée et aidée à aller de l'avant.

Je remercie également mes parents, mes frères et ma soeur pour leur appui et leur compréhension, pour le bon temps passé au chalet et d'autres choses simples qui ont fait la différence

durant ces longues années d'études.

Et surtout. Merci Jérémie pour avoir vécu ces moments avec moi, pour y avoir pris part aussi intensément que moi. Je te suis reconnaissante pour avoir su me sortir de mon projet par moments, pour m'avoir entraînée dans d'autres aventures et invitée à relever d'autres défis. Je te promets des projets communs à l'avenir.

La réalisation de ce projet a été rendue possible grâce à l'octroi d'une bourse d'études supérieures du Conseil de recherches en sciences naturelles et génie du Canada (CRSNG) et d'un soutien financier provenant de fonds de recherche du même organisme.

RÉSUMÉ

Sur un site contaminé, la concentration d'un contaminant dans l'eau souterraine est connue localement grâce à des échantillons provenant de puits existants et/ou de puits d'observation aménagés à cette fin. L'interpolation des valeurs de concentration observées est requise pour évaluer l'étendue d'un panache de contamination lors de la caractérisation du site. Elle est également nécessaire pour des applications dans les phases subséquentes, telles que la conception et le suivi des mesures de réhabilitation ou de confinement, ou pour l'évaluation de l'atténuation naturelle. La qualité des estimés obtenus par interpolation est largement tributaire du nombre de données disponibles, lequel est limité en raison des coûts associés à l'installation de puits d'observations, à l'échantillonnage et aux analyses chimiques. Les processus physiques qui régissent le transport de contaminants compliquent également cette tâche puisqu'ils entraînent une répartition anisotrope et statistiquement non stationnaire des valeurs de concentration. La faible performance des outils d'interpolation conventionnels rend nécessaire le développement d'approches qui permettent d'utiliser l'information concernant le modèle physique du problème, en particulier les directions d'écoulement.

La méthode de krigeage développée dans cette thèse prend en compte le modèle d'écoulement par l'intermédiaire d'un modèle de covariance non stationnaire, lequel est basé sur l'utilisation conjointe d'une transformation en coordonnées d'écoulement et d'une fonction de covariance qui varie le long des lignes de courant. Dans un premier temps, les coordonnées naturelles d'écoulement, définies par la charge hydraulique et la fonction de courant, permettent de tenir compte de l'anisotropie locale induite par l'advection, telle que décrite dans un modèle conceptuel du problème d'écoulement. En second lieu, le modèle de covariance non stationnaire est paramétrisé pour tenir compte approximativement des effets de la dispersion sur la corrélation spatiale et la variance de la concentration. Différentes paramétrisations de la fonction de covariance non stationnaire sont proposées. Les paramètres du modèle de covariance sont estimés par maximum de vraisemblance. Ceci permet d'utiliser le test du rapport de vraisemblance et le critère d'Akaike (AIC) pour sélectionner le modèle le mieux adapté aux données parmi les modèles candidats. Face aux difficultés entourant la spécification des fonctions de courant en 3D, la présente thèse propose également l'utilisation de coordonnées d'écoulement simplifiées (désignées coordonnées alternatives) basées sur le traçage des lignes de courant.

L'approche proposée est testée sur des modèles numériques 2D et 3D représentant des

panaches continus dans des milieux hétérogènes, isotropes et saturés. Bien qu'elle nécessite la modélisation de l'écoulement, l'approche ne dépend pas de la modélisation du transport et ainsi, la position et l'historique des émissions de la source ne sont pas requis. Pour les exemples présentés, les solutions d'écoulement sont obtenues par la méthode des éléments finis. L'utilisation d'exemples numériques permet la comparaison directe entre les valeurs de concentration estimées et de référence. La qualité des estimés est évaluée à l'aide d'un ensemble de statistiques, tel que l'erreur sur la masse de soluté et sur le volume du panache pour certain seuils, de même que la corrélation entre les valeurs estimées et les valeurs de référence (ϕ et r de Pearson).

Dans un premier temps, on utilise les réseaux d'écoulement de référence pour la transformation de coordonnée. Les résultats montrent que l'utilisation conjointe des coordonnées d'écoulement et des modèles de covariance non stationnaires permet de mieux expliquer les données selon le maximum de vraisemblance. Les cartes de concentration obtenues apparaissent nettement plus réalistes que celles qui proviennent de l'approche conventionnelle de krigeage stationnaire en coordonnées cartésiennes. Ceci est confirmé par l'amélioration de la corrélation (ϕ de Pearson) avec le modèle de référence. Les erreurs sur la masse et le volume du panache sont également réduites comparativement à l'approche conventionnelle. En utilisant un échantillonnage relativement restreint, les exemples présentés montrent que l'approche proposée permet une meilleure utilisation des données. Bien que les paramètres estimés soient en général plausibles, la procédure de sélection basée sur le maximum de vraisemblance ne parvient pas toujours à identifier le modèle qui engendre le meilleur estimé de la concentration. Les résultats obtenus montrent que la validation croisée semble être une option plus robuste pour obtenir les paramètres et déterminer le meilleur modèle.

L'approche proposée a des effets notables sur la variance d'estimation, laquelle hérite du caractère non stationnaire de la fonction de covariance. Pour les modèles de covariance où la variance décroît avec le transport, la variance d'estimation indique une incertitude plus importante à la zone située près de la source.

Les coordonnées alternatives d'écoulement proposées constituent une alternative valable aux coordonnées naturelles. L'exemple en 3D montre qu'elles permettent d'incorporer les caractéristiques de l'écoulement dans les valeurs de concentration estimées, et ainsi d'améliorer l'interpolation. Un exemple utilisant des coordonnées analogues aux coordonnées lagagiennes (utilisation du temps comme coordonnée longitudinale) montre qu'il est possible de produire une interpolation plus réaliste d'un panache de contamination présentant des

« doigts » distincts, sans recourir à un échantillonnage exhaustif.

On présente également une étude de cas où des représentations simplifiées du modèle de référence, calibrés à des observations de charge hydraulique, sont utilisées pour simuler une connaissance partielle du modèle d'écoulement. Cet exemple montre qu'une certaine imprécision sur la transformation de coordonnées n'engendre pas de différences importantes pour les statistiques compilées et que les cartes de concentration obtenues tiennent tout de même compte des caractéristiques principales de l'écoulement.

Des modèles d'écoulement déterministes ont été utilisés dans les exemples d'application et ainsi, dans sa forme actuelle, l'approche proposée ne permet pas de quantifier l'effet de l'incertitude liée à l'écoulement sur l'estimation de la concentration. Par contre, en plus d'améliorer l'estimation des valeurs de concentration, l'approche développée peut s'avérer un outil intéressant pour évaluer la cohérence entre le modèle conceptuel d'écoulement et les données de concentration.

ABSTRACT

Contaminant concentration is known locally by sampling from on-site observation wells. Interpolation of such concentration data often yields disappointing results, especially when it fails to incorporate existing knowledge relative to the underlying physics of groundwater flow and solute transport. Concentration maps, however, are required in several applications such as plume monitoring, design of containment or treatment systems, evaluation of total dissolved mass, and natural attenuation assessment. Interpolation results highly depend on the number of available data, which is often limited by the costs of well installation and sampling. Further, the physical processes governing transport complicate this task by creating an anisotropic and nonstationary concentration distribution. The poor performance of conventional interpolation tools justifies the development of approaches which account for flow and transport.

This thesis presents a flexible kriging approach that incorporates flow knowledge through a nonstationary covariance model that is based on the joint use of natural flow coordinates, defined by the hydraulic head and the stream function, and a class of nonstationary covariance functions that can be parameterized to account for the evolution of concentration correlation and variance with travel time/distance. On one hand, the coordinate transformation makes it possible to account for the spatially-varying anisotropy produced by advection, as described by a given conceptual flow model. On the other hand, the nonstationary covariance model accounts only approximately for the effect of dispersion on the spatial correlation and variance of the concentration. The nonstationary covariance model parameters are obtained by maximum likelihood estimation. This enables the use of the likelihood ratio test and the computation of the Akaike Information Criterion to select between competing model parameterizations. Considering the difficulties associated with the use of the stream function in 3D, this thesis also proposes approximate flow coordinates, termed *alternative flow coordinates*, which are computed using streamline tracking.

The approach is tested on simple 2-D and 3-D synthetic plumes shaped by heterogeneous flow and local dispersion. Although it requires flow modelling, the proposed approach does not depend on transport modelling; thus, the source position and release history need not to be known. Synthetic model testing allows direct comparison of the estimated and reference concentration values. The performance of the interpolation approach is evaluated using a set of statistics, such as the error on contaminant mass and contaminated volume, and the

correlation between estimated and reference concentration values (Pearson's ϕ and r).

In the first series of examples, the reference flow fields are used to compute the flow coordinates. Results show that the joint use of the flow coordinates and nonstationary covariance yields higher likelihood values, compared to models defined in Cartesian coordinates. Concentration maps appear more realistic than those obtained by conventional kriging with Cartesian coordinates and a stationary covariance. This is also confirmed by the improvement of Pearson's ϕ . Errors on estimated mass and volumes are also reduced compared to the conventional approach. These results indicate that the proposed kriging approach makes a more efficient use of limited concentration measurements. Although the estimated covariance parameters are plausible, the model selection based on maximum likelihood is not able to identify the model that yields the best estimate in terms of precision and realism. Cross-validation appears to be more robust for parameter estimation and model selection.

The approach has notable effects on the estimation variance, as it inherits the nonstationarity from the covariance model. For covariance models in which the variance decreases with distance, the estimation variance is higher upstream in the flow model, near the source. Moreover, the proposed alternative coordinates constitute a reasonable approximation to natural coordinates and allow notable improvement for 3D interpolation. We also test the use of a coordinate system analogous to Lagrangian coordinates where travel time is used as the longitudinal coordinate. This coordinate transformation allows for a realistic interpolation of a finger plume, without recourse to exhaustive sampling.

In another synthetic example, simplified version of the flow model are used to simulate the effect of partial knowledge. Hydraulic head data are used to calibrate the model. This example shows that the approach still performs well despite flow model imprecision.

In this work, deterministic models are used to compute the flow coordinates and, in its actual implementation, the approach does not allow for the quantification of the effects of flow uncertainty on the estimated concentration. In addition to concentration estimation, the approach may also be seen as a tool for assessing the agreement between the conceptual model and the concentration data.

TABLE DES MATIÈRES

REMERCIEMENTS	iii
RÉSUMÉ	v
ABSTRACT	viii
TABLE DES MATIÈRES	x
LISTE DES TABLEAUX	xiii
LISTE DES FIGURES	xvi
LISTE DES SIGLES ET ACRONYMES	xix
LISTE DES ANNEXES	xxiii
CHAPITRE 1 INTRODUCTION	1
CHAPITRE 2 REVUE DE LITTÉRATURE	4
2.1 Interpolation de la concentration	4
2.1.1 Application des méthodes déterministes et du krigeage à l'interpolation de la concentration	6
2.1.2 Approches d'interpolation associées à des méthodes inverses	8
2.2 Moments statistiques de la concentration	9
2.2.1 Approches de modélisation stochastique du transport	10
2.2.2 Description qualitative des moments de la concentration	12
2.2.3 Moments conditionnels	13
2.3 Modélisation de fonctions aléatoires non stationnaires	14
2.3.1 Moyenne non stationnaire (dérive)	14
2.3.2 Fonctions aléatoires intrinsèques généralisées	15
2.3.3 Déformation spatiale et transformation de coordonnées	16
2.3.4 Fonctions de covariance non stationnaires	18
2.4 Synthèse	19

CHAPITRE 3 KRIGEAGE NON STATIONNAIRE DE LA CONCENTRATION DANS L'EAU SOUTERRAINE	21
3.1 Transformation de coordonnées	24
3.1.1 Formulation duale de l'écoulement et coordonnées naturelles	24
3.1.2 Coordonnées alternatives d'écoulement	28
3.1.3 Méthode des éléments finis	32
3.1.4 Traçage des lignes de courant	35
3.2 Fonction de covariance	37
3.2.1 Covariance non stationnaire par convolution de noyaux gaussiens	37
3.2.2 Paramétrisation	39
3.3 Krigeage	46
3.3.1 Système de krigeage ordinaire	46
3.3.2 Krigeage universel	48
3.4 Normes d'évaluation	49
3.5 Synthèse	51
CHAPITRE 4 ÉTUDES DE CAS AVEC MODÈLES D'ÉCOULEMENT CONNUS .	53
4.1 Exemple 1. Cas 2D	54
4.1.1 Description du modèle de référence	54
4.1.2 Aperçu des fonctions de covariance non stationnaires	55
4.1.3 Sélection du modèle de covariance	57
4.1.4 Estimation de la concentration	60
4.1.5 Utilisation d'une dérive linéaire	61
4.1.6 Impacts sur la variance de krigeage	65
4.2 Exemple 2. Cas 3D	67
4.2.1 Description du modèle de référence	67
4.2.2 Sélection du modèle de covariance	68
4.2.3 Estimation de la concentration	70
4.3 Exemple 3. Utilisation du temps de parcours comme coordonnée longitudinale	73
4.3.1 Description du modèle de référence	73
4.3.2 Sélection du modèle de covariance	73
4.3.3 Estimation de la concentration	74
4.4 Discussion	79
4.4.1 Coordonnées d'écoulement	79
4.4.2 Modèles de covariance non stationnaires	80
4.4.3 Estimation des paramètres par maximum de vraisemblance	81

4.4.4 Résultats du krigeage	81
CHAPITRE 5 EXEMPLE D'APPLICATION À UN MODÈLE D'ÉCOULEMENT PARTIELLEMENT CONNU	85
5.1 Description du modèle de référence et intégration des données	85
5.2 Sélection du modèle de covariance	88
5.3 Estimation de la concentration	90
5.4 Discussion	92
CHAPITRE 6 CONCLUSION	95
RÉFÉRENCES	101
ANNEXES	111

LISTE DES TABLEAUX

Tableau 4.1	Exemple 1. Paramètres du modèle synthétique	56
Tableau 4.2	Résumé des modèles candidats pour la paramétrisation du modèle de covariance non stationnaire. A : Anisotrope, S : Stationnaire, NS : Non stationnaire	57
Tableau 4.3	Exemple 1. Maximum de vraisemblance (MV), <i>p-value</i> du test du rapport de vraisemblance ($\alpha = 0.05$) et AIC_c pour les modèles en coordonnées cartésiennes	59
Tableau 4.4	Exemple 1. Maximum de vraisemblance (MV), <i>p-value</i> du test du rapport de vraisemblance ($\alpha = 0.05$) et AIC_c pour les modèles en coordonnées naturelles d'écoulement (CNE)	59
Tableau 4.5	Exemple 1. Maximum de vraisemblance (MV), <i>p-value</i> du test du rapport de vraisemblance ($\alpha = 0.05$) et AIC_c pour les modèles en coordonnées alternatives d'écoulement (CAE _{<i>h</i>})	59
Tableau 4.6	Exemple 1. Comparaison des résultats pour les différentes normes d'évaluation	62
Tableau 4.7	Exemple 1. Maximum de vraisemblance MV, <i>p-value</i> du test du rapport de vraisemblance ($\alpha = 0.05$) et AIC_c pour les modèles en coordonnées naturelles d'écoulement (CNE) avec dérive	63
Tableau 4.8	Exemple 1. Comparaison des résultats pour le modèle CNE/NS2 avec et sans dérive	64
Tableau 4.9	Exemple 2. Paramètres du modèle synthétique	67
Tableau 4.10	Exemple 2. Maximum de vraisemblance (MV), <i>p-value</i> du test du rapport de vraisemblance ($\alpha = 0.05$) et AIC_c pour les modèles en coordonnées cartésiennes	69
Tableau 4.11	Exemple 2. Maximum de vraisemblance maximum (MV), <i>p-value</i> du test du rapport de vraisemblance ($\alpha = 0.05$) et AIC_c pour les modèles en coordonnées alternatives d'écoulement	70
Tableau 4.12	Exemple 2. Comparaison des résultats pour les différentes normes d'évaluation	72
Tableau 4.13	Exemple 3. Paramètres du modèle synthétique	75
Tableau 4.14	Exemple 3. Maximum de vraisemblance (MV), <i>p-value</i> du test du rapport de vraisemblance ($\alpha = 0.05$) et AIC_c pour les modèles en coordonnées alternative CAE _{<i>t</i>}	75

Tableau 4.15	Exemple 3. Maximum de vraisemblance (MV), <i>p-value</i> du test du rapport de vraisemblance ($\alpha = 0.05$) et AIC_c pour les modèles en coordonnées alternative CAE_h	75
Tableau 4.16	Exemple 3. Comparaison des résultats pour les différentes normes d'évaluation	76
Tableau 4.17	Exemple 3. RMSE optimisées pour l'estimation des paramètres par validation croisée	77
Tableau 4.18	Exemple 3. Comparaison des résultats pour les différentes normes d'évaluation pour le modèle CAE_t avec les paramètres estimés par validation croisée	78
Tableau 5.1	Paramètres du modèle synthétique	89
Tableau 5.2	Maximum de vraisemblance (MV), <i>p-value</i> du test du rapport de vraisemblance et AIC_c pour les modèles en coordonnées cartésiennes	90
Tableau 5.3	Scénario 1. Maximum de vraisemblance (MV), <i>p-value</i> du test du rapport de vraisemblance et AIC_c pour les modèles en coordonnées naturelles	90
Tableau 5.4	Scénario 2. Maximum de vraisemblance (MV), <i>p-value</i> du test du rapport de vraisemblance et AIC_c pour les modèles en coordonnées naturelles	90
Tableau 5.5	Coordonnées issues du modèle d'écoulement de référence. Maximum de vraisemblance (MV), <i>p-value</i> du test du rapport de vraisemblance et AIC_c	91
Tableau 5.6	Statistiques comparatives entre les concentrations krigées et de référence. c_s : seuil de contamination.	91
Tableau A.1	Erreur absolue moyenne (MAE) pour le krigeage en coordonnées cartésiennes (CCK) et le krigeage en coordonnées naturelles d'écoulement (NCK).	115
Tableau C.1	Exemple 1. Résultats obtenus avec le modèle $CAE_h/NS2$, pour différents espacements des lignes de courant utilisées lors du calcul de la coordonnées u	118
Tableau D.1	Résultats pour le modèle utilisé à l'exemple 1 avec une autre configuration des données. c_s : seuil de contamination.	119
Tableau E.1	Paramètres du modèle de covariance non-stationnaire : Exemple 1 - coordonnées cartésiennes. Coordonnées maximales : $x_{max} = 400$ m, $y_{max} = 50$ m	121
Tableau E.2	Paramètres du modèle de covariance non-stationnaire : Exemple 1 - coordonnées naturelles d'écoulement. Coordonnées maximales : $\Delta h_{max} = 0.3$ m, $\psi_{max} = 6.658 \times 10^{-6}$ m/s	121

Tableau E.3	Paramètres du modèle de covariance non-stationnaire : Exemple 1 - coordonnées alternatives d'écoulement. Coordonnées maximales : $\Delta h_{max} = 0.3 \text{ m}$, $u_{max} = 200 \text{ m}$	122
Tableau E.4	Paramètres du modèle de covariance non-stationnaire : Exemple 1 - coordonnées naturelles d'écoulement (avec dérive linéaire). Coordonnées maximales : $\Delta h_{max} = 0.3 \text{ m}$, $\psi_{max} = 6.658 \times 10^{-6} \text{ m/s}$	122
Tableau E.5	Paramètres du modèle de covariance non-stationnaire : Exemple 2 - coordonnées cartésiennes). Coordonnées maximales : $x_{max} = 200$, $y_{max} = 100 \text{ m}$, $y_{max} = 50 \text{ m}$	123
Tableau E.6	Paramètres du modèle de covariance non-stationnaire : Exemple 2 - coordonnées alternatives d'écoulement). Coordonnées maximales : $\Delta h_{max} = 0.3 \text{ m}$, $u_{max} = 100 \text{ m}$, $v_{max} = 50 \text{ m}$	123
Tableau E.7	Paramètres du modèle de covariance non-stationnaire (Validation-croisée) : Exemple 2 - coordonnées alternatives d'écoulement). Coordonnées maximales : $\Delta h_{max} = 0.3 \text{ m}$, $u_{max} = 100 \text{ m}$, $v_{max} = 50 \text{ m}$	123
Tableau E.8	Paramètres du modèle de covariance non-stationnaire : Exemple 3 - coordonnées naturelles d'écoulement CAE _{<i>h</i>} . Coordonnées maximales : $\Delta h_{max} = 0.3 \text{ m}$, $u_{max} = 338 \text{ m}$	124
Tableau E.9	Paramètres du modèle de covariance non-stationnaire : Exemple 3 - coordonnées naturelles d'écoulement CAE _{<i>t</i>} . Coordonnées maximales : $\frac{t_{max}}{1 \times 10^7} = 6346 \text{ s}/1 \times 10^7$, $u_{max} = 338 \text{ m}$	124
Tableau E.10	Paramètres du modèle de covariance non-stationnaire (Validation-croisée) : Exemple 3 - coordonnées naturelles d'écoulement CAE _{<i>t</i>} . Coordonnées maximales : $\frac{t_{max}}{1 \times 10^7} = 6346 \text{ s}/1 \times 10^7$, $u_{max} = 338 \text{ m}$	124
Tableau E.11	Paramètres du modèle de covariance non-stationnaire : Exemple chapitre 5 - coordonnées cartésiennes. Coordonnées maximales : $x_{max} = 400 \text{ m}$, $y_{max} = 50 \text{ m}$	125
Tableau E.12	Paramètres du modèle de covariance non-stationnaire : Exemple chapitre 5 - coordonnées naturelles d'écoulement pour le scénario 1. Coordonnées maximales : $\Delta h_{max} = 0.3 \text{ m}$, $\psi_{max} = 7.948 \times 10^{-6} \text{ m/s}$	125
Tableau E.13	Paramètres du modèle de covariance non-stationnaire : Exemple chapitre 5 - coordonnées naturelles d'écoulement pour le scénario 2. Coordonnées maximales : $\Delta h_{max} = 0.3 \text{ m}$, $\psi_{max} = 1.043 \times 10^{-5} \text{ m/s}$	125
Tableau E.14	Paramètres du modèle de covariance non-stationnaire : Exemple chapitre 5 - coordonnées naturelles d'écoulement de référence. Coordonnées maximales : $\Delta h_{max} = 0.3 \text{ m}$, $\psi_{max} = 7.397 \times 10^{-6} \text{ m/s}$	125

LISTE DES FIGURES

Figure 3.1	Schématisation de différentes approches de krigeage.	23
Figure 3.2	Illustration de la transformation en coordonnées naturelles et de la correspondance entre les conditions frontières de la charge hydraulique et de la fonction de courant.	27
Figure 3.3	Illustration des coordonnées alternatives d'écoulement	29
Figure 3.4	Fonction de covariance stationnaire de type Matérn $\nu = 0.5$ (ou covariance exponentielle)	44
Figure 3.5	Une matrice de confusion	50
Figure 3.6	Schématisation de l'approche proposée.	52
Figure 4.1	Exemple 1. Modèle de référence	55
Figure 4.2	Exemple 1. Comparaison des coordonnées naturelles et alternatives d'écoulement.	56
Figure 4.3	Fonction de covariance NS1 en coordonnées cartésiennes (CC/NS1) pour les points (a) (50,50), (b) (150,25) et (c) (350,25).	57
Figure 4.4	Fonction de covariance NS2 en coordonnées cartésiennes (CC/NS2) pour les points (a) (50,50), (b) (150,25) et (c) (350,25).	58
Figure 4.5	Fonction de covariance S1 en coordonnées alternatives d'écoulement (CAE _h /S1) pour les points (a) (50,50), (b) (150,25) et (c) (350,25).	58
Figure 4.6	Cartes de concentration krigées et de référence pour l'exemple 1. (a) CC/S1, (b) CC/NS2, (c) CNE/S1, (d) CNE/NS2, (e) CAE _h /S1, (f) CAE _h /NS2 et (g) Concentrations de référence.	63
Figure 4.7	Exemple 1. Résultats du modèle NS3 en coordonnées d'écoulement. (a) CNE/NS3, (b) CAE _h /NS3.	63
Figure 4.8	Cartes de concentration krigées pour l'exemple 1 avec dérive linéaire. (a) CNE/S1, (b) CNE/NS2, (c) CNE/NS3, (d) Concentrations de référence	64
Figure 4.9	Écarts-types d'estimation associés à certains résultats de krigeage de l'exemple 1. (a) CAE _h /S1, (b) CAE _h /NS2, (c) CAE _h /NS3, (d) CC/S1, (e) CC/NS2, (f) CC/NS3.	66
Figure 4.10	Exemple 2. Modèle de référence (a) Champ de conductivité hydraulique ($\log_{10}K$) (b) Concentration de référence.	68

Figure 4.11	Exemple 2. Surfaces h (rouge), u (vert) et v (bleu). u et v sont obtenues par l'approche combinée de traçage des lignes de courant et d'interpolation linéaire présentée au chapitre 3. h est obtenu à partir de la solution numérique d'écoulement.	69
Figure 4.12	Exemple 2. Vue en plan ($z = 25$ m) des valeurs de concentration de référence et krigées. (a) Concentrations de référence. (b) $CAE_h/NS3$ - meilleur modèle selon le maximum de vraisemblance. (c) $CC/NS3$ - meilleur modèle en coordonnées cartésiennes et (d) $CAE_h/NS2$ - meilleur modèle selon les statistiques comparatives.	71
Figure 4.13	Exemple 2. Vue en coupe ($y = 50$ m) des valeurs de concentration de référence et krigées. (a) Concentration de référence. (b) $CAE_h/NS3$ - meilleur modèle selon le maximum de vraisemblance. (c) $CC/NS3$ - meilleur modèle en coordonnées cartésiennes et (d) $CAE_h/NS2$ - meilleur modèle selon les statistiques comparatives.	72
Figure 4.14	Exemple 3. (a) Champ de conductivité hydraulique ($\log_{10}K$), (b) Coordonnée u (divisée par u_{max}), (c) Panache de référence.	74
Figure 4.15	Exemple 3. Cartes de concentration krigées. (a) $CAE_t/NS2$, (b) $CAE_t/NS3$, (c) $CAE_h/NS2$, (d) $CAE_h/NS3$, (e) Concentrations de référence.	77
Figure 4.16	Exemple 3. Cartes de concentration krigées avec les paramètres obtenus par validation croisée. (a) $CAE_t/NS2$, (b) $CAE_t/NS3$, (c) Concentrations de référence.	78
Figure 5.1	Modèle de référence et modèle simplifié du scénario 2. (a) Champ de conductivité hydraulique de référence ($\log_{10}K$), (b) Concentrations de référence, (c) Interprétation à partir de forages fictifs (illustrés par les lignes noires), (d) Modèle conceptuel du scénario 2.	86
Figure 5.2	Schématisation de l'approche d'intégration des données	88
Figure 5.3	Réseaux d'écoulement associés aux différents modèles (a) Référence, (b) Scénario 1, (c) Scénario 2. Les réseaux d'écoulement illustrés en (b) et (c) ont été calés à 12 observations de charge hydraulique par l'approche d'estimation successive des flux et sont associés à des conductivités hydrauliques modifiées par rapport à celles des modèles conceptuels.	89
Figure 5.4	Cartes krigées et de référence. (a) $CC/NS2$, (b) $CNE/NS2$ pour le scénario 1, (c) $CNE/NS2$ pour le scénario 2, (d) $CNE/NS2$ en utilisant le réseau d'écoulement de référence, (e) Panache de référence.	92

Figure A.1	Modèle de référence. (a) Modèle et conditions frontières, (b) Réseau d'écoulement de référence avec les équipotentielles en noir et les lignes de courant en rouge, (c) Concentration de référence	113
Figure A.2	Résultats du krigage en coordonnées cartésiennes avec (a) 25 données, (b) 18 données, et (c) 8 données.	114
Figure A.3	Résultats du krigage en coordonnées naturelles d'écoulement (a) 25 données, (b) 18 données, et (c) 8 données.	115
Figure B.1	Schématisation de la méthode utilisée pour construire les champs de conductivité hydraulique.	116
Figure C.1	Cartes de concentration krigées en coordonnées alternatives d'écoulement (modèle CAE _h /NS2) pour différentes densités de lignes de courant. (a) $\Delta = 5$ m, (b) $\Delta = 10$ m, (c) $\Delta = 25$ m.	118
Figure D.1	Résultats pour l'exemple 1 avec une configuration différente des données (a) Concentration de référence, (b) CC/NS2, (c) CNE/NS2 . . .	120

LISTE DES SIGLES ET ACRONYMES

Acronymes

AIC	Akaike information criterion
AIC_c	Small sample Akaike information criterion
CAE	Coordonnées alternatives d'écoulement
CC	Coordonnées cartésiennes
CNE	Coordonnées naturelles d'écoulement
EMV	Estimation par maximum de vraisemblance
LRT	Test de vraisemblance (<i>Likelihood ratio test</i>)
MCMC	Markov chain Monte Carlo
MEF	Méthode des éléments finis
MV	Maximum de vraisemblance
NS1	Paramétrisation avec portées croissantes
NS2	Paramétrisation avec variance décroissante
NS3	Paramétrisation avec portée croissante et variance décroissante
RAE_m	Erreur absolue relative sur la masse
RAE_{vol}	Erreur absolue relative sur le volume
RMLE	Estimation par maximum de vraisemblance restreint
RMSE	Racine de la moyenne de l'erreur au carré (<i>root mean square error</i>)
S1	Paramétrisation stationnaire du modèle de covariance
VC	Validation croisée

Alphabet latin

$a_p(s_i)$	Portée spatiale dans la direction principale p du modèle de covariance non-stationnaire au point i
a_l	Coefficient de dérive
a, b, c, d	Quantités définies par la matrice de confusion
c	Concentration
\mathbf{c}_0	Vecteur de covariance au point d'estimation x_0
c_s	Valeur de concentration correspondant à un seuil de contamination s
C	Fonction de covariance

C^{NS}	Fonction de covariance non-stationnaire
$\mathbf{C}(\boldsymbol{\theta})$	Matrice de covariance entre les données
$C[Z_i, Z_j]$	Covariance de la fonction aléatoire Z entre les points i et j
\mathbf{d}	Pas d'avancement
$d\mathbf{A}$	Élément surfacique infinitésimal
dl_i	Distance entre deux points consécutifs sur une ligne de courant
$d\mathbf{r}$	Élément d'arc infinitésimal le long d'une ligne de courant
\mathbf{D}	Tenseur de dispersivité
D_m	Coefficient de diffusion moléculaire
e_i	Résidu de la validation croisée au point x_i
$f^l(x)$	Fonction de base de la dérive
\mathbf{F}	Vecteur des conditions frontières dans la MEF
h	Charge hydraulique
\hat{h}	Fonction test
h_{max}	Charge hydraulique maximale sur le domaine
I	Indicatrice
k	nombre de paramètres d'un modèle
$\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4$	Vitesses intermédiaires dans les méthodes d'Euler et de Runge-Kutta
K	Conductivité hydraulique
\mathbf{K}	Matrice de rigidité
K_{x_i}	Noyau centré en x_i et x_j
L	Logarithme négatif de la vraisemblance (<i>negative log-likelihood</i>)
L	Opérateur différentiel
$L(\lambda)$	Lagrangien
\mathcal{L}	Vraisemblance
$m(x)$	Fonction de dérive
m_0, m_1	Constantes de la dérive linéaire
M^*	Masse de contaminant estimée
M^R	Masse de contaminant de référence
n	Porosité aux équations 3.64 et 3.65
n	Nombre de données
n_e	Porosité effective
N	Nombre de noeuds du maillage (MEF)
q	Flux hydraulique
q_n	Flux hydraulique normal à une frontière
\mathbf{q}	Vecteur de flux hydraulique

Q_{AB}	Débit entre les points A et B
Q_{ij}	Forme quadratique
r	Coefficient de corrélation de Pearson
$R(x, y)$	Résidu dans la méthode des éléments finis
R^S	Fonction de corrélation stationnaire
R^{NS}	Fonction de corrélation non-stationnaire
t_m	Temps de parcours en un point x_m
T	Transmissivité
u	Première coordonnée transversale des CAE
u_i	Coefficient des fonctions de base / solution aux noeuds du maillage MEF
U	Vecteur de la solution (MEF)
$u(x_0^k, y_0^k, z_0^k)$	Coordonnée u du point de départ de la k^{ieme} ligne de courant
v	Deuxième coordonnée transversale des CAE (3D)
v	Vitesse de déplacement de l'eau souterraine
v_u	Vecteur unitaire de la vitesse
V_i	Volume de l'élément i
V_s^R	Volume contaminé de référence pour un seuil s
V_s^*	Volume contaminé estimé, en considérant un seuil s
$v(x_0^k, y_0^k, z_0^k)$	Coordonnée v du point de départ de la k^{ieme} ligne de courant
w_i	Poids d'interpolation linéaire (chap. 2)
w_i	Poids de la méthode des résidus pondérés (chap. 3)
X	Matrice de coordonnées des points d'observation
x_i	Vecteur de coordonnées du point x_i
y_{max}	Limite du modèle en y
z_{max}	Limite du modèle en z
Z	Vecteur d'observations
Z	Fonction aléatoire

Alphabet grec

α_l	Dispersivité longitudinale
α_{tv}	Dispersivité transversale (2D en coupe)
α_{th}, α_{tv}	Dispersivités transversales (3D)
β_q	Facteur de correction pour le flux hydraulique (CAE)

β	Vecteur des coefficients de la dérive
ϕ_i	Fonction de base
ϕ_s	ϕ de Pearson pour un seuil s (<i>Mean square contingency coefficient</i>)
Γ	Ensemble des frontières du modèle d'éléments finis
$\Gamma(\nu)$	Fonction gamma
κ_ν	Fonction de Bessel modifiée de seconde espèce d'ordre ν
λ	Vecteur des poids de krigage
μ	Multiplicateur de Lagrange
ν	Matérn shape parameter
ψ	Fonction de courant
ρ	Portée spatiale
ρ_w	Densité de l'eau
σ_i	Écart-type au point x_i
σ_e^2	Variance d'estimation
σ^2	Variance
$\sigma_i^2, \sigma^2(s_i)$	Variance non-stationnaire au point i
$\sigma_{\ln(K)}^2$	Variance du logarithme naturel de K
Σ_i	Matrice de covariance du noyau gaussien centré en x_i
τ	Distance entre deux points
θ	Vecteur des paramètres du modèle de covariance
$\xi(\xi, \eta, \zeta)$	Coordonnées lagrangiennes
Ω	Domaine d'éléments finis

LISTE DES ANNEXES

Annexe A	ANNEXE A : Exemple simple avec un nombre de données variable	111
Annexe B	ANNEXE B : Construction des champs de conductivité hydraulique hétérogènes	116
Annexe C	ANNEXE C : Évaluation sommaire de la sensibilité des résultats au nombre de lignes de courant utilisées pour obtenir les coordonnées u et v	117
Annexe D	ANNEXE D : Résultats pour l'exemple 1 avec une configuration de données différente	119
Annexe E	ANNEXE E : Compilation des paramètres de covariance utilisés dans les exemples des chapitres 4 et 5	121

CHAPITRE 1

INTRODUCTION

Au Canada, près de neuf millions de personnes dépendent de l'eau souterraine pour leur usage domestique, soit environ le tiers de la population (Statistique Canada, 2003). Bien qu'elle soit généralement une source d'eau potable sécuritaire, cette ressource est toutefois menacée par l'activité humaine. Depuis le début du vingtième siècle, les activités industrielles entraînent l'introduction de contaminants dans l'eau souterraine. Pendant longtemps, l'absence de règlementation relativement à l'entreposage et l'élimination des déchets industriels a largement contribué à ce problème (Fitts, 2002). En dépit des législations, les sources potentielles de contaminants demeurent diverses. Parmi celles-ci, on retrouve notamment les réservoirs enfouis, les sites d'enfouissement sanitaire, les sites d'entraînement militaire, les dépôts à neige, les rejets agricoles (pesticides, fertilisants et rejets animaux) et les haldes de rejets miniers.

Sur un site, la présence de contaminants dans l'eau souterraine est connue localement grâce à des échantillons provenant de puits existants ou de puits d'observation aménagés à cette fin. L'évaluation de l'étendue d'un panache de contamination est un élément primordial pour la caractérisation du site et requiert l'estimation des valeurs de concentration en des points où celles-ci n'ont pas été observées (Reed *et al.*, 2004). En plus de déterminer la taille du panache de contamination, les concentrations estimées peuvent également servir à évaluer la masse totale de contaminant présent. Ces informations sont essentielles pour la conception et le suivi des mesures de réhabilitation ou de confinement hydraulique (Kitanidis et Shen, 1996). Couplées à des valeurs de flux hydraulique, les concentrations estimées peuvent aussi être utilisées pour obtenir les flux de masse (Semprini *et al.*, 1995; Troldborg *et al.*, 2010) et ainsi permettre de déterminer et quantifier l'atténuation naturelle potentiellement en cours sur un site (Bockelmann *et al.*, 2003).

Les cartes de concentration sont généralement tracées manuellement ou obtenues par interpolation des données disponibles. La qualité des estimés obtenus est largement tributaire du nombre de données utilisées. Or, les coûts associés à l'installation de puits d'observations, à l'échantillonnage et aux analyses chimiques font en sorte qu'en pratique le nombre d'observations est souvent limité (Jones *et al.*, 2003; Schwede et Cirpka, 2010). À ceci s'ajoute également le fait que la concentration présente une distribution anisotrope et statistiquement

non stationnaire, en raison des processus régissant le transport de contaminant dans un aquifère. Les approches conventionnelles d'interpolation, basées uniquement sur les données de concentration disponibles, produisent souvent des résultats qui ne sont pas cohérents avec le modèle déterministe de transport décrit par l'équation d'advection-dispersion. Le krigeage est une approche d'interpolation basée sur des principes probabilistes, laquelle est fréquemment utilisée pour produire des cartes de concentration. Or, les approches de krigeage conventionnelles, en ayant recours à des modèles de covariance stationnaires (i.e. spatialement invariante) ne tiennent pas compte des changements de direction d'anisotropie induits par l'advection en milieu hétérogène, ni des effets de la dispersion sur la variance de la concentration. Il apparaît donc primordial de développer des approches d'interpolation des données de concentration qui permettent de tenir compte de toute l'information disponible, en particulier celle reliée au modèle d'écoulement.

En considérant la conductivité hydraulique comme une fonction aléatoire, les approches de modélisation stochastique permettent d'estimer les moments statistiques de la concentration en tenant compte d'un modèle de transport. Ainsi, de nombreuses études se sont penchées sur la description de ces moments. Or, l'application des approches probabilistes au problème de transport de contaminant est d'autant plus utile lorsqu'elle permet le conditionnement des moments statistiques à des données observées (Graham et McLaughlin, 1989a). Récemment, des méthodes d'inversion géostatistique qui utilisent le conditionnement aux données de concentration ont été proposées pour améliorer l'interpolation de la concentration (Shlomi et Michalak, 2007; Schwede et Cirpka, 2010). Pour ce type d'approches, le recours à des simulateurs d'écoulement et de transport permet d'intégrer de l'information supplémentaire, entre autres, par l'utilisation de fonctions de covariance conditionnées par le modèle d'écoulement et de transport. Ces approches améliorent le réalisme des valeurs de concentration estimées et de leur incertitude. Toutefois, ceci augmente les efforts/temps de calcul nécessaires et requiert généralement des connaissances supplémentaires concernant le modèle sous-jacent (conditions frontières, position de la source).

Objectifs

Cette thèse vise à présenter et à tester une approche de krigeage permettant de tenir compte d'un modèle de d'écoulement, en vue d'améliorer l'interpolation des valeurs de concentration de contaminant dans l'eau souterraine. Ceci est fait en spécifiant directement un modèle de covariance non stationnaire, lequel est basé sur l'utilisation conjointe d'une trans-

formation des coordonnées spatiales et d'une classe particulière de modèles de covariance.

Plus précisément, la présente thèse vise à :

1. présenter le cadre théorique et la justification de l'approche proposée ;
2. proposer une paramétrisation pour le modèle de covariance non stationnaire ;
3. tester l'utilisation conjointe d'une transformation en coordonnées d'écoulement et d'une fonction de covariance non stationnaire pour le krigeage de la concentration ;
4. développer une approche approximative comme alternative aux coordonnées naturelles d'écoulement en 3D.

En ce qui a trait à la complexité et à la précision, la méthode proposée se présente comme un compromis entre les méthodes d'interpolation conventionnelles et les approches récentes basées sur l'inversion. Elle permet de s'affranchir des simulations de transport utilisées dans ces dernières et ainsi d'alléger le temps de calcul, tout en améliorant le réalisme des concentrations estimées par rapport aux méthodes conventionnelles. La méthode proposée ne nécessite pas d'information exhaustive sur la source, telles que sa position, sa taille et l'historique des émissions de contaminants (*release history*). Les travaux présentés dans cette thèse ont fait l'objet de publications dans une revue scientifique (Rivest *et al.*, 2012b; Rivest et Marcotte, 2012) et d'un article de conférence arbitré (Rivest *et al.*, 2012a).

Le chapitre suivant présente les approches actuelles utilisées pour l'interpolation de la concentration. Étant donné leur importance dans le développement de la méthode proposée, ce chapitre traite également des moments statistiques obtenus par modélisation stochastique du transport et de certaines techniques pour la modélisation de fonctions aléatoires non stationnaires. Le troisième chapitre se penche sur les aspects méthodologiques relatifs à l'approche de krigeage non stationnaire, notamment, la transformation de coordonnées, la paramétrisation du modèle de covariance non stationnaire, l'estimation des paramètres et le krigeage. Ce chapitre se termine avec la présentation des normes d'évaluation utilisées dans les exemples d'application. Aux chapitres quatre et cinq, la méthode proposée est testée à l'aide de modèles synthétiques et comparée au krigeage stationnaire en coordonnées cartésiennes. Dans le cas du chapitre quatre, les modèles d'écoulement sont considérés parfaitement connus alors qu'au chapitre cinq, on reprend un des exemples du chapitre précédent, en considérant plutôt des modèles d'écoulement simplifiés et calibrés à des observations de charges hydrauliques provenant du modèle de référence. En conclusion, on présente des améliorations possibles à l'approche développée.

CHAPITRE 2

REVUE DE LITTÉRATURE

L’interpolation de la phase dissoute d’un contaminant dans l’eau souterraine est compliquée par l’influence du transport sur la répartition spatiale de la concentration, combinée à la faible densité d’observations généralement disponible. L’incertitude entourant le modèle de transport et d’écoulement justifie l’emploi de modèles probabilistes, ainsi que des méthodes qui leur sont associées, pour décrire la distribution de la concentration. Ce chapitre couvre donc trois grands thèmes : les méthodes d’interpolation, les moments statistiques de la concentration et la modélisation de fonctions aléatoires non stationnaires.

2.1 Interpolation de la concentration

L’interpolation des données de concentration dans l’eau souterraine permet de dresser un portrait du panache de contamination en utilisant les observations disponibles pour estimer la concentration où celle-ci n’est pas échantillonnée (Reed *et al.*, 2004). La plupart des applications traitant de l’interpolation de la concentration retrouvées dans la littérature utilisent des interpolateurs linéaires de la forme :

$$Z_0^* = \sum_{i=1}^n w_i Z_i \quad (2.1)$$

où n est le nombre de données, Z_i est l’observation de la variables au point x_i et w_i est le poids associé à l’observation i . Les méthodes d’interpolation diffèrent dans la façon de déterminer les poids w_i . Parmi les méthodes les plus couramment utilisées pour l’interpolation des valeurs de concentration dans l’eau souterraine, on distingue deux types d’interpolateurs : les interpolateurs déterministes et les interpolateurs stochastiques (ou géostatistiques) (Reed *et al.*, 2004).

Dans le cas des interpolateurs déterministes, la pondération des données est faite indépendamment des valeurs observées, selon une expression mathématique préétablie. Par exemple, la méthode du plus proche voisin consiste à assigner la valeur de l’observation située la plus près du point interpolé. Dans l’équation 2.1, ceci équivaut à attribuer un poids de 1 à l’observation située la plus près du point à estimer et des poids de 0 aux autres observations. Cette approche est souvent associée aux polygones de Voronoï, lesquels permettent d’identifier la

zone d'influence associée à chaque observation. La méthode du voisin naturel, ou *natural neighbor interpolation* (Sibson, 1981), est également basée sur des polygones de Voronoï. Par contre, la pondération est basée sur la proportion du polygone associé au point à estimer qui se superpose aux polygones de Voronoï déterminés par les observations.

Une méthode classique est l'interpolation linéaire entre les sommets de triangles définis par les données. Dans ce cas, les poids w_i correspondent aux coordonnées barycentriques du point d'interpolation dans un triangle donné. Cette approche est utilisée pour le traçage manuel des lignes de niveau. Pour un même ensemble de points, plusieurs triangulations sont possibles. La triangulation de Delaunay, qui est l'opérateur dual des polygones de Voronoï (Watson, 1992), peut être utilisée à cette fin. Cette méthode produit toutefois une estimation discontinue sur les arêtes des triangles.

La méthode de l'inverse de la distance, quant à elle, utilise des poids inversement proportionnels à la distance euclidienne séparant le point à estimer de chaque observation (Shepard, 1968). Certaines variantes utilisent également des puissances de la distance pour la pondération. Toutefois, cette approche a tendance à imposer un caractère isotrope aux valeurs estimées en raison de l'influence radiale des données et ainsi, elle ne permet pas de reproduire adéquatement des structures linéaires à partir de données dispersées (Watson, 1992).

Les méthodes de type spline visent quant à elles l'ajustement d'une fonction polynomiale de courbure minimale, continue par morceaux et passant par les données. Parmi celles-ci on retrouve les splines bicubiques (Bhattacharyya, 1969) qui garantissent la continuité des deux premières dérivées. Les splines de type Akima (Akima, 1974) sont également basées sur des polynômes d'ordre trois, mais reposent sur un ajustement local de la fonction. Dans le cas des splines de type plaque mince (*thin plate spline*), l'ajustement est basé sur la minimisation d'une fonction d'énergie (Duchon, 1977).

De manière générale, les méthodes d'interpolation déterministes ne tiennent pas compte des processus physiques responsables de la répartition spatiale de la variable interpolée et elles ne permettent pas de quantifier la précision reliée à la méthode d'interpolation (variance d'estimation). Ces approches visent avant tout à automatiser la représentation graphique des données et le traçage de lignes de niveau, ce qui permet d'éviter un traçage manuel éventuellement subjectif et/ou fastidieux. La revue des méthodes déterministes présentée ici ne se veut pas exhaustive ; des ouvrages sur le sujet peuvent être consultés pour des inventaires plus détaillés de ces approches, par exemple Watson (1992).

De leur côté, les approches stochastiques adoptent un modèle probabiliste pour décrire la variable interpolée. Dans le cas de la concentration, le caractère aléatoire est justifié par l'incertitude entourant les conditions d'écoulement et de transport dans l'aquifère (Fetter, 1993). Les observations sont alors assimilées à la réalisation d'une fonction aléatoire. Ayant en main une seule réalisation, l'inférence du modèle sous-jacent est possible en supposant la stationnarité des deux premiers moments de la fonction aléatoire (hypothèse de stationnarité du second ordre) ou de ses accroissements (hypothèse intrinsèque) (Chilès et Delfiner, 1999). Ces prémisses ont permis le développement de la géostatistique et du krigeage, un interpolateur linéaire dont la pondération dépend du modèle de covariance, lequel est déterminé à partir des observations. Contrairement aux approches déterministes, les poids de krigeage s'adaptent à la variabilité du phénomène observé. Le krigeage a également comme propriété d'être un interpolateur sans biais. Les valeurs krigées ont une interprétation mathématique dans le cas où la variable est gaussienne et de moyenne connue (Chilès et Delfiner, 1999), puisqu'elles correspondent alors à l'espérance conditionnelle de la variable étant donné les observations disponibles. Dans le cas non gaussien, les valeurs krigées constituent une bonne approximation de l'espérance conditionnelle. De plus, l'approche géostatistique permet également de fournir une variance d'estimation qui tient compte du modèle de covariance et de la configuration des données. En outre, le krigeage peut également être extrait de son contexte stochastique et être interprété comme un interpolateur déterministe (Gambolati et Volpi, 1979b). Dans ce cas, la covariance peut être vue comme une fonction choisie pour assurer certaines propriétés lors de l'interpolation. L'équivalence formelle entre le krigeage et les splines a d'ailleurs été démontrée par Matheron (1981). Marcotte et David (1988) montrent l'équivalence entre le krigeage et la régression.

2.1.1 Application des méthodes déterministes et du krigeage à l'interpolation de la concentration

Parmi les travaux concernant l'interpolation de la concentration de contaminant dans l'eau souterraine, on note quelques études comparatives entre les approches déterministes et le krigeage (Jones *et al.*, 2003; Reed *et al.*, 2004). Jones *et al.* (2003) comparent la méthode de l'inverse de la distance, la méthode du voisin naturel et le krigeage sur quatre jeux de données réelles en 3D. La validation croisée est utilisée afin de comparer la précision de ces différentes méthodes. Les résultats obtenus indiquent que le krigeage produit en général des erreurs plus faibles. Ceci est également observé dans les travaux de Reed *et al.* (2004), qui ont testé différentes variantes du krigeage et deux versions de la méthode de l'inverse de la

distance sur des données synthétiques en 3D. Ils concluent que le krigeage de quantiles (*quantile kriging*) est la méthode la plus robuste parmi celles testées, mais qu'en raison de leur simplicité d'implémentation les approches déterministes pouvaient être utiles lors d'étapes exploratoires (*screening*). Reed *et al.* (2004) mentionnent toutefois que les résultats du krigeage ordinaire peuvent être fortement biaisés en présence d'un échantillonnage préférentiel et de valeurs de concentration extrêmes.

Différentes variantes du krigeage permettent d'utiliser de l'information supplémentaire ou de tenir compte de certaines contraintes en vue d'améliorer l'interpolation. Par exemple, le cokrigeage permet de tenir compte de la corrélation entre des variables régionalisées afin d'utiliser une variable secondaire pour l'estimation de la variable d'intérêt. Istok *et al.* (1993) présentent une application où les concentrations de nitrate contribuent à l'estimation des concentrations de Dacthal (un herbicide) dans l'eau souterraine. Dans ce cas, l'intérêt provient du fait que les analyses chimiques nécessaires pour déterminer la concentration de nitrates sont moins dispendieuses que celles requises pour le Dacthal. Leur utilisation permet également de réduire la variance d'estimation. Kitanidis et Shen (1996) utilisent un changement de variable, dont la transformation logarithmique est un cas particulier, pour tenir compte de l'asymétrie de la distribution statistique de la concentration et garantir la non négativité des valeurs estimées. Les paramètres du modèle sont obtenus par une variante de l'estimateur du maximum de vraisemblance (*restricted maximum likelihood*). Ces auteurs reconnaissent que l'utilisation de l'information concernant l'écoulement, le transport et les réactions chimiques ayant cours dans l'aquifère permettrait de baser les résultats de l'interpolation sur des considérations physiques. Dans cet ordre d'idée, le krigeage avec dérive externe est utilisé par Saito et Goovaerts (2001) pour l'interpolation de la concentration de métaux lourds dans des sols situés en périphérie d'une fonderie. La direction des vents dominants est utilisée pour définir la dérive et ainsi, incorporer de l'information sur le processus qu'on croit être responsable de la distribution spatiale de la concentration.

Malgré sa grande flexibilité, le krigeage n'a pas donné lieu à des approches qui permettent d'incorporer directement l'information relative au modèle de transport lors de l'interpolation et ainsi les cartes de concentration krigées manquent souvent de réalisme (Shlomi et Michalak, 2007). De plus, les moments statistiques de la concentration sont réputés non stationnaires et sont fortement influencés par les phénomènes liés au transport (Rubin, 1991; Reed *et al.*, 2004). Or, l'ensemble des applications présentées dans cette section utilisent des modèles de covariance stationnaires.

2.1.2 Approches d'interpolation associées à des méthodes inverses

Afin de palier à la faible représentativité des cartes de concentration interpolées, des approches récentes utilisent la modélisation inverse pour tenir compte du modèle de transport lors de l'estimation de la concentration dans l'eau souterraine (Shlomi et Michalak, 2007; Schwede et Cirpka, 2010). Comme pour le krigage, elles utilisent un modèle probabiliste pour décrire la concentration.

Shlomi et Michalak (2007) utilisent l'inversion géostatistique (Kitanidis, 1995) pour estimer l'historique des émissions de contaminant, s . Ces auteurs définissent ainsi deux estimateurs de la concentration, *inverse-forward modeling* (IFM) et *transport-enhanced kriging* (TREK), lesquels tiennent compte de l'incertitude sur s . En supposant un modèle d'écoulement et de transport déterministe, l'IFM permet d'obtenir un estimé de s à partir des observations de la concentration, d'une matrice de sensibilité issue d'un modèle de transport et d'un modèle covariance. La solution du problème direct permet ensuite l'estimation de la concentration sur l'ensemble du domaine. La seconde approche (TREK) intègre également la covariance spatiale associée à la concentration. Dans les deux cas, le modèle de transport intervient par l'intermédiaire de la matrice de sensibilité de la concentration à l'intensité de la source. Celle-ci modifie la covariance et constraint le système d'équations linéaires, semblable au cokrigage, utilisé pour définir l'estimateur. Dans ce système, des valeurs locales de sensibilité introduisent éventuellement des effets (non stationnaires) propres au modèle de transport dans la covariance et la moyenne *a priori* utilisées lors de l'estimation, ce qui contribue à produire des estimés plus réalistes de la concentration.

Schwede et Cirpka (2010) proposent une approche hybride combinant des simulations du champ de conductivité hydraulique, la modélisation numérique de l'écoulement et du transport et l'inversion géostatistique. Le résultat est désigné par *approche de Monte Carlo conditionnelle*. Dans un premier temps, des réalisations du champ de conductivité hydraulique sont utilisées pour estimer les moments non conditionnels de la concentration. L'inversion, sous forme d'une variante de l'approche quasi linéaire (Kitanidis, 1995), est ensuite utilisée pour conditionner les champs de conductivité hydraulique aux données de concentration. Contrairement à l'approche de Shlomi et Michalak (2007), l'historique des émissions est considéré connu et la méthode développée cherche plutôt à tenir compte de l'incertitude liée au champ de conductivité hydraulique. L'approche de Monte Carlo conditionnelle permet d'obtenir la distribution conjointe de la concentration. La concentration (et la variance associée) est estimée en prenant la moyenne d'ensemble des champs de concentration provenant de la modélisation numérique du transport, répétée pour chaque champ de conductivité hydrau-

lique conditionnel. D'ailleurs, Schwede et Cirpka (2010) choisissent plutôt la médiane comme estimateur de la concentration. Les valeurs estimées sont très près des valeurs de référence (modèle synthétique) et respectent les limites physiques (pas de concentrations négatives). Ces résultats sont comparés à ceux d'un krigage qui utilise les covariances non stationnaires calculées à partir des réalisations non conditionnelles de la concentration. Les valeurs krigées et la variance d'estimation sont alors assimilées aux premiers moments d'une distribution bêta, laquelle est utilisée pour le calcul de la médiane. Les résultats obtenus sont inférieurs à ceux issus de l'approche de Monte Carlo conditionnelle, mais ils sont néanmoins supérieurs à ceux obtenus en supposant une distribution gaussienne. Ce résultat montre l'intérêt d'utiliser une covariance non stationnaire basée sur le modèle d'écoulement pour le krigage de la concentration. En raison des étapes de simulation et de conditionnement requises, l'utilisation de l'approche de Monte Carlo conditionnelle comme méthode d'interpolation est toutefois très coûteuse en temps de calcul, en particulier pour des modèles 3D (Schwede et Cirpka, 2010).

La capacité des approches mentionnées plus haut à produire des estimés réalistes est liée au fait qu'elles permettent d'inclure les effets d'un modèle de transport. Toutefois, la détermination de cette influence sur les moments statistiques de la concentration nécessite la modélisation du transport, que ce soit pour calculer la matrice de sensibilité (Shlomi et Michalak, 2007) ou pour calculer les différentes réalisations de la concentration et la covariance non stationnaire (Schwede et Cirpka, 2010).

De leur côté, les travaux de modélisation stochastique du transport de contaminant ont fourni des descriptions des deux premiers moments statistiques et de la corrélation spatiale de la concentration. La prochaine section présente une brève description de ces approches et des principaux résultats concernant la moyenne, la variance et la corrélation de la concentration.

2.2 Moments statistiques de la concentration

En général, les deux processus dominants dans le transport de contaminants sont l'advection (convection) et la dispersion. Dans un premier temps, l'advection consiste en un déplacement induit par le mouvement de l'eau dans laquelle le contaminant est en solution. Le transport purement advectif ne produit pas de zone de mélange entre l'eau transportant le soluté et l'eau environnante ; la distribution du soluté se limite aux tubes de courant dans lesquels il a été initialement introduit (Fetter, 1993). La dispersion hydrodynamique, quant à elle, combine les effets macroscopiques de deux processus distincts : la diffusion et la dis-

persion mécanique (Domenico et Schwartz, 1998). D'une part, la diffusion moléculaire décrit le mélange causé par la migration aléatoire des particules du soluté (Domenico et Schwartz, 1998). D'autre part, la dispersion mécanique décrit un mélange causé par les variations locales des vitesses d'écoulement dans la matrice poreuse. La dispersion mécanique est en fait un phénomène advectif causé par les hétérogénéités du milieu à une échelle inférieure à celle considérée par la composante advective du transport (Gelhar, 1993). La similarité des effets macroscopiques de la dispersion mécanique et de la diffusion moléculaire justifie généralement l'adoption d'un modèle de Fick où ces deux composantes sont combinées en un seul paramètre. La dispersion hydrodynamique entraîne le composé à se répartir dans les tubes de courants adjacents à ceux où il était contenu initialement.

L'action de ces deux mécanismes sur la répartition de la concentration est décrite par l'équation d'advection-dispersion (Domenico et Schwartz, 1998) :

$$\mathbf{D} \nabla^2 c - \mathbf{v} \cdot \nabla c + c \nabla \cdot \mathbf{v} = \frac{\partial c}{\partial t} \quad (2.2)$$

où \mathbf{D} est le tenseur de dispersivité hydrodynamique, c est la concentration et v , la vitesse linéaire de l'eau souterraine, est donnée par :

$$v = -\frac{K}{n_e} \frac{\partial h}{\partial l} \quad (2.3)$$

où K est la conductivité hydraulique, n_e est la porosité effective et $\frac{\partial h}{\partial l}$ est le gradient hydraulique dans la direction l . Pour un écoulement irrotationnel, l'équation 2.2 devient :

$$\mathbf{D} \nabla^2 c - \mathbf{v} \cdot \nabla c = \frac{\partial c}{\partial t} \quad (2.4)$$

La modélisation stochastique du transport vise à obtenir une solution probabiliste approximative à l'équation d'advection-dispersion. Ceci est fait en modélisant K comme une fonction aléatoire et en étudiant l'effet sur la distribution des variables dépendantes, telles que la concentration. Dans ce type d'approche, la solution probabiliste à l'équation d'advection-dispersion est obtenue sous forme des moments statistiques de la concentration, lesquels résument la fonction de densité de probabilité.

2.2.1 Approches de modélisation stochastique du transport

Parmi les approches les plus utilisées, on retrouve :

- les approches basées sur la méthode des perturbations ;
- l'approche de Monte Carlo.

Les approches basées sur la méthode des perturbations permettent de déduire les moments statistiques des variables hydrogéologiques à partir de formulations probabilistes des équations différentielles décrivant le modèle physique. Ceci est fait en décomposant les variables considérées aléatoires comme la somme d'une moyenne et d'une perturbation d'ordre un. Les équations obtenues sont ensuite exprimées en fonction de la moyenne et de la covariance de la variable considérée incertaine, généralement la conductivité hydraulique (Gelhar, 1993). Les formulations utilisées varient selon que le problème de transport soit dérivé dans un cadre eulérien (Gelhar et Axness, 1983; Graham et McLaughlin, 1989a,b; Vomvoris et Gelhar, 1990; Kapoor et Gelhar, 1994) ou lagrangien (Dagan, 1982, 1984; Rubin, 1991; Vanderborgh, 2001; Fiorotto et Caroni, 2002; Caroni et Fiorotto, 2005), ou encore en coordonnées naturelles d'écoulement (Cirpka *et al.*, 2011a,b). Elles diffèrent également selon les hypothèses formulées et selon les techniques utilisées pour résoudre les expressions analytiques obtenues (ex. solveurs numériques, approches spectrales). En raison de l'approximation d'ordre un, l'application de la méthode des perturbations est limitée à des champs de conductivité hydraulique statistiquement homogènes et faiblement variables, de l'ordre de $\sigma_{\ln(K)}^2 \leq 1$ (Graham et McLaughlin, 1989a; Gelhar, 1993).

La méthode de Monte Carlo a quant à elle été évoquée précédemment au sujet des travaux de Schwede et Cirpka (2010). Cette approche ne vise pas la définition du problème sous forme d'une équation différentielle stochastique, mais procède plutôt à la résolution numérique du problème déterministe pour un grand nombre de réalisations du champ de conductivité hydraulique (Gelhar, 1993). Les résultats sont ensuite compilés pour fournir des estimés de la fonction de densité de probabilité et/ou des deux premiers moments d'ensemble (moyenne et variance). Des estimés des covariances spatiale et temporelle peuvent également être obtenus à partir des résultats. Cette approche est moins restrictive au niveau de la variabilité de la conductivité hydraulique, mais il demeure que des changements trop brusques dans les valeurs de K peuvent limiter la précision des solutions numériques. L'approche de Monte Carlo est souvent utilisée pour valider/comparer les résultats obtenus avec d'autres approches (Graham et McLaughlin, 1989a; Schwede *et al.*, 2008; Cirpka *et al.*, 2011b) ou pour permettre l'analyse avec des champs de conductivité hydraulique de variances plus élevées (Caroni et Fiorotto, 2005). Divers algorithmes de simulation géostatistique peuvent être utilisés pour générer les réalisations de conductivité hydraulique. Dans le cas de formations géologiques complexes et multimodales, il est possible de recourir à la simulation multipoints tels que dans les travaux de Feyen et Caers (2006). De manière générale, l'approche de Monte Carlo est associée à des

temps de calculs importants en raison du calcul des solutions d'écoulement et de transport pour chaque réalisation du champ de conductivité hydraulique.

Dans les deux cas, en postulant un modèle probabiliste pour K , la modélisation stochastique permet de faire ressortir les caractéristiques des moments statistiques de la concentration. Un inventaire détaillé des résultats de la modélisation stochastique dépasse le cadre du présent exposé. Dans ce qui suit, on dresse plutôt un portrait général des moments obtenus.

2.2.2 Description qualitative des moments de la concentration

Les moments statistiques expérimentaux obtenus par modélisation stochastique diffèrent selon les cas particuliers étudiés et le mode d'émission considéré (continu ou instantané). Concernant le mode d'émission, les travaux de Graham et McLaughlin (1991) illustrent bien ces différences. La forte influence du modèle physique sur les moments de la concentration permet toutefois de faire ressortir certaines caractéristiques générales, communes aux deux modes d'émission.

Moyenne La moyenne porte clairement l'empreinte du modèle de transport. Pour un milieu faiblement hétérogène (valeurs de $\sigma_{ln(K)}^2$ de 0.5 et de 1), les travaux de Graham et McLaughlin (1989a) montrent que celle-ci a un aspect semblable à un panache dans un milieu homogène. De Barros et Nowak (2010) présentent des moyennes d'ensemble d'aspect similaire à celles de Graham et McLaughlin (1989a), lesquelles sont obtenues par l'approche de Monte Carlo avec une variance de $\sigma_{ln(K)}^2 = 2$. Pour des champs de conductivité hydraulique de distribution bimodale, Feyen et Caers (2006) obtiennent des moyennes d'ensemble d'aspect moins régulier, reflétant les structures observées sur le champ de conductivité hydraulique moyen.

Variance La variance de la concentration est plus élevée à de faibles distances de la source en raison de l'incertitude sur les chemins d'écoulement empruntés ; elle s'atténue par la suite en raison des processus dispersifs qui entraînent la dilution des valeurs de concentration (Graham et McLaughlin, 1989a; Kapoor et Gelhar, 1994; Kapoor et Kitanidis, 1998; Fiorotto et Caroni, 2002; Caroni et Fiorotto, 2005; Dentz, 2012). Ceci est observé sur les cartes d'écart-type présentées par Graham et McLaughlin (1989a, 1991) et de Barros et Nowak (2010). Près de la source, l'écart-type (et donc la variance) présente souvent une distribution bimodale dans la direction transversale, avec les valeurs élevées situées près de la frange du panache,

lesquelles s'atténuent en s'éloignant de la source. Dans le cas d'une émission instantanée (*pulse plume*), la variance diminue en amont du centre de masse du panache, à mesure que celui se déplace (Graham et McLaughlin, 1991). Pour une source continue, la variance demeure élevée à proximité de celle-ci.

Corrélation Certaines études montrent que la fonction de corrélation spatiale de la concentration est anisotrope, non stationnaire et non symétrique (Vomvoris et Gelhar, 1986; Graham et McLaughlin, 1989a,b, 1991; Vomvoris et Gelhar, 1990; Rubin, 1991). Graham et McLaughlin (1989a) présentent des figures montrant la corrélation spatiale de la concentration pour deux points du domaine d'étude. Celle-ci montre une forte anisotropie avec une persistance élevée dans la direction moyenne d'écoulement, un résultat également mentionné par Vomvoris et Gelhar (1986). Pour les points considérés, la non-stationnarité de la corrélation est moins manifeste, mais tout de même apparente. De faibles corrélations négatives sont retrouvées près de la source. De plus, en étudiant l'évolution de la corrélation spatiale pour deux points dont les coordonnées se déplacent avec l'écoulement, Rubin (1991) montre que la corrélation augmente avec le temps de parcours. Vanderborgh *et al.* (2006) notent que la corrélation, dans la direction perpendiculaire à l'écoulement moyen, augmente avec le transport et avec les valeurs de dispersion. Récemment, Cirpka *et al.* (2011b) ont produit une analyse stochastique en coordonnées naturelles. Ces derniers montrent, qu'en coordonnées naturelles (charge hydraulique et fonction de courant), le carré de la largeur de la frange d'un panache augmente en fonction de la diminution de la charge hydraulique, dont la valeur maximum est définie à partir du point d'entrée de la source. Ces auteurs constatent que le passage aux coordonnées naturelles simplifie la description du processus de dispersion transversale en isolant les effets dus aux variations locales de la direction et de la largeur des tubes de courant.

Il est à noter que dans les études mentionnées précédemment, la taille et position de la source ne sont jamais considérées incertaines. Or, les cartes des statistiques d'ensemble présentées dans certains travaux (Graham et McLaughlin, 1989a; Feyen et Caers, 2006; de Barros et Nowak, 2010) permettent d'apprécier la forte empreinte des caractéristiques de la source sur les statistiques d'ensemble et la corrélation de la concentration.

2.2.3 Moments conditionnels

Le conditionnement permet d'arrimer les moments statistiques obtenus à des données réelles. Selon Graham et McLaughlin (1989a), la moyenne et la variance d'ensemble issues de

la modélisation stochastique et le conditionnement sont des approches complémentaires qui devraient être combinées à chaque fois que cela est possible. Des filtres de Kalman peuvent être utilisés pour conditionner la moyenne d'ensemble de la concentration à des données de concentration, de charge hydraulique et du logarithme de la conductivité hydraulique (Graham et McLaughlin, 1989b). Le (co)krigeage peut également être utilisé à cette fin (Rubin, 1991). Dans ce cas, au lieu d'utiliser un modèle de covariance stationnaire ajusté aux données par variographie, on dispose d'un modèle de covariance qui incorpore l'influence du modèle de transport sur la concentration. C'est ce qui est fait par Schwede et Cirpka (2010) lorsqu'ils utilisent la covariance déduite des réalisations de la concentration dans l'exemple de krigeage discuté précédemment à la section 2.1.2. Par ailleurs, Rubin (1991) indique que la covariance spatio-temporelle peut être utilisée pour la prédiction par cokrigeage des concentrations à un temps donné, à partir de données antérieures.

L'utilisation des moments statistiques issus de la modélisation stochastique permet donc de tenir compte du modèle d'écoulement par l'intermédiaire des covariances propres à celui-ci. Toutefois, l'obtention d'une telle covariance par l'approche de Monte Carlo est très coûteuse en temps de calcul (Graham et McLaughlin, 1989a; Schwede et Cirpka, 2010). De leur côté, les approches basées sur la méthode des perturbations sont limitées à des champs statistiquement homogènes et faiblement variables (Gelhar, 1993; Rubin, 1990).

La modélisation stochastique met clairement en évidence la non-stationnarité des moments statistiques de la concentration. Dans cet ordre d'idées, la section suivante est consacrée aux méthodes qui permettent le krigeage de variables non stationnaires.

2.3 Modélisation de fonctions aléatoires non stationnaires

Différentes approches sont utilisées en géostatistique afin de permettre de tenir compte de fonctions aléatoires non stationnaires. Les moyens employés varient selon le type de non-stationnarité rencontré.

2.3.1 Moyenne non stationnaire (dérive)

En géostatistique, les hypothèses de stationnarité de second ordre et intrinsèque sont requises pour l'estimation du modèle de covariance à partir des données (Chilès et Delfiner, 1999). Toutefois, des variantes du krigeage permettent de tenir compte d'une moyenne non

stationnaire. Le krigeage universel et le krigeage avec dérive externe (Chilès et Delfiner, 1999) modélisent la moyenne sous forme d'une dérive composée de polynômes des coordonnées spatiales ou de variables auxiliaires, respectivement. La variable étudiée est alors vue comme la somme d'une dérive déterministe et d'un résidu stationnaire (Chilès et Delfiner, 1999). Ce type de krigeage a été abondamment utilisé en hydrogéologie (Delhomme, 1978; Gambolati et Volpi, 1979a; Aboufirassi et Marino, 1983; Ahmed et De Marsily, 1987; Tonkin et Larson, 2002; Brochu et Marcotte, 2003; Rivest *et al.*, 2008). Par contre, cette approche requiert l'estimation du modèle de covariance stationnaire des résidus, ce qui est compliqué par le fait que les coefficients associés à la dérive sont inconnus. On note que l'hypothèse intrinsèque mentionnée plus haut permet d'aller plus loin dans la modélisation d'une moyenne non stationnaire grâce aux fonctions aléatoires intrinsèques généralisées (Matheron, 1971; Chilès, 1977).

2.3.2 Fonctions aléatoires intrinsèques généralisées

L'hypothèse intrinsèque consiste à supposer la stationnarité des accroissements (ou incréments) de la fonction aléatoire modélisée, ce qui permet d'en filtrer la moyenne. Ainsi, l'utilisation des différences $Z(x) - Z(x + h)$ (accroissements d'ordre 0) de la variable permet de filtrer une moyenne constante, mais inconnue. Ceci permet également l'identification du modèle de continuité spatiale par l'intermédiaire du variogramme. La théorie des fonctions aléatoires intrinsèques généralisées (Matheron, 1971) permet d'aller plus loin en supposant la stationnarité des accroissements d'ordre k . Ainsi, le fait de travailler avec les accroissements d'ordres supérieurs permet de filtrer des dérives formées de polynômes du même ordre. La fonction qui décrit la corrélation entre les accroissements est appelée covariance généralisée d'ordre k (Chilès, 1977). Par contre, le fait de travailler avec les accroissements complique l'inférence du modèle de covariance pour des données réparties d'une manière quelconque et l'on doit recourir à des procédures d'ajustement automatiques (Chilès et Delfiner, 1999; Reed *et al.*, 2004). Le krigeage généralisé fait partie des méthodes d'interpolation comparées par Reed *et al.* (2004) pour l'estimation de la concentration. Dans cette étude, la fonction de covariance généralisée utilisée est isotrope et filtre une moyenne quadratique. Ces auteurs déplorent que cette approche ne permettent pas l'inférence du modèle par variographie et notent que les résultats du krigeage généralisé sont fortement biaisés en présence d'échantillonnage préférentiel et de données extrêmes. La covariance généralisée demeure une fonction stationnaire, c'est-à-dire invariante sous translation.

2.3.3 Déformation spatiale et transformation de coordonnées

Parmi les approches permettant de tenir compte de variations spatiales d'anisotropie, on note 1) l'utilisation directe de directions d'anisotropie locales, 2) le recours à des distances non euclidiennes calculées le long de la direction principale d'anisotropie, ainsi que 3) les transformation de coordonnées.

En se basant sur les données directionnelles disponibles, certaines approches incorporent directement une anisotropie locale dans la covariance calculée. Ces approches utilisent simultanément une direction d'anisotropie globale, pour le calcul des covariances aux points d'observation, et des directions d'anisotropie locales, pour les points à estimer (ou à simuler). Soares (1990) utilise des observations de l'orientation locale d'une strate minéralisée, obtenue à partir de carottes de forages, pour conditionner la covariance d'une variable indicatrice de la minéralisation. Dans les travaux de Xu (1996), l'anisotropie locale est utilisée pour simuler des faciès. Te Stroet et Snepvangers (2005) utilisent une approche itérative basée sur une méthode du gradient pour estimer les angles et les facteurs d'anisotropie locaux. Dans ces approches, la représentation de l'anisotropie locale requiert soit un échantillonnage suffisamment dense pour l'identification de changements locaux d'orientation, soit la disponibilité de données directionnelles, telles que le pendage, l'azimuth ou le gradient. Une telle situation est plus fréquente dans les domaines pétrolier et minier, mais rarement rencontrée dans le domaine environnemental.

L'utilisation de distances non euclidiennes est sujet au même type de considérations, afin de pourvoir calculer les distances le long de la direction de plus forte continuité spatiale, laquelle doit être préalablement définie. Ainsi, ce type d'approche est surtout utilisé dans des contextes où l'on dispose d'un modèle géographique bien connu (ex. cours d'eau, carte routière, arborescence) pour le calcul des distances. De plus, l'utilisation de distances non euclidiennes est soumise à certaines restrictions au niveau du choix des fonctions de covariance, afin de s'assurer que la distance utilisée produit une matrice de covariance définie positive (Curriero, 1996, 2006).

L'utilisation de transformations de coordonnées, ou de déformations spatiales, permet de contourner les restrictions sur le modèle de covariance rencontrées lors de l'utilisation de distances non euclidiennes. Grâce à une transformation de coordonnées appropriée, la covariance peut être calculée dans un repère spatial où la distance euclidienne décrit adéquatement le degré de similarité entre deux observations (Perrin et Meiring, 1999). Le *multidimensional scaling* (MDS) permet d'organiser les points dans un espace de dimensions plus élevées où les

distances euclidiennes entre les points correspondent approximativement aux distances (non euclidiennes) calculées préalablement dans l'espace original (Sampson et Guttorp, 1992; Monestiez *et al.*, 1993; Loland et Host, 2003; Almendral *et al.*, 2008; Boisvert *et al.*, 2009; Boisvert et Deutsch, 2010). Alternativement, certaines approches modélisent la déformation spatiale comme un processus gaussien et en estiment les paramètres par maximum de vraisemblance (Anderes et Stein, 2008) ou par l'approche de *Markov chain Monte Carlo* (Schmidt, 2001; Damian *et al.*, 2001; Schmidt et O'Hagan, 2003). Clerc et Mallat (2003) proposent une approche pour estimer la déformation associée à un processus stationnaire sous-jacent. Guyon et Perrin (2000) et Perrin et Senoussi (2000) étudient des classes de déformations qui assurent la stationnarité d'une fonction aléatoire. Ces derniers indiquent qu'il est uniquement possible rendre le processus stationnaire par déformation spatiale lorsque la variance est constante.

Des transformations de coordonnées sont utilisées depuis longtemps en géostatistique pour déformer des milieux géologiques plissés en vue d'analyser un gisement selon sa forme originale dans le milieu de déposition (Dagbert *et al.*, 1984). Deutsch et Wang (1996) appliquent une séquence hiérarchique de déformations à un réservoir pétrolier de type fluviatile afin de faciliter la simulation de faciès géologiques. De manière analogue, certaines applications où la variable d'intérêt est liée à l'écoulement de surface utilisent une déformation basée sur la morphologie d'un cours d'eau (Barabas *et al.*, 2001; Legleiter et Kyriakidis, 2006).

Dans tous les cas, les transformations spatiales sont utilisées pour représenter une variable dans un nouvel espace euclidien où la direction d'anisotropie peut être considérée stationnaire. De manière générale, une transformation de coordonnées est valide si elle fournit une « adresse » unique à tous les points dans le domaine d'étude (Christakos *et al.*, 2000; Deutsch, 2002).

Dans le domaine de la modélisation numérique, le recours à une transformation de coordonnées a donné lieu à la méthode des directions principales (*principal direction technique*) qui permet d'améliorer la résolution du problème de transport de contaminants (Frind, 1982) en réduisant la dispersion numérique. Cette méthode consiste à reformuler les équations d'un modèle de transport par éléments finis en fonction des coordonnées naturelles d'écoulement, que constituent la charge hydraulique et la fonction de courant (Bear, 1972). Ce faisant, le problème est simplifié par le fait que cette transformation permet d'orienter les flux de masse selon les directions principales du système de coordonnées et de simplifier l'expression du tenseur de dispersivité, lequel devient alors une matrice diagonale (Cirpka *et al.*, 1999b). Afin de contourner certaines problématiques associées à la fonction de courant, telle que la

modélisation en présence de puits, Cirpka *et al.* (1999b) construisent par traçage des lignes de courant, un maillage 2D orienté selon l'écoulement.

Tel qu'on l'a mentionné à la section 2.2.2, la transformation en coordonnées naturelles d'écoulement a aussi été utilisée dans un contexte stochastique par Cirpka *et al.* (2011b) pour décrire les statistiques du mélange cumulatif (*cumulative mixing*) ayant lieu à la frange d'un panache en régime permanent. Le passage aux coordonnées d'écoulement permet de tenir compte de la trajectoire curviligne du panache, ainsi que des convergences et divergences des tubes de courant. Rivest *et al.* (2012b) utilisent une transformation basée sur les coordonnées d'écoulement pour le krigeage de la concentration dans l'eau souterraine.

2.3.4 Fonctions de covariance non stationnaires

Il existe également des modèles de covariance qui tiennent compte explicitement de la non-stationnarité. Une telle classe de fonctions de covariance est introduite par Higdon *et al.* (1999) et repose sur la convolution de noyaux de type gaussien dont les paramètres varient spatialement. Dans un exemple d'application en 2D, Higdon *et al.* (1999) appliquent le modèle développé à l'estimation de la concentration de dioxine dans le sol. Un modèle hiérarchique est utilisé pour décrire la distribution de la concentration. Dans ce modèle, la variation spatiale de la covariance est paramétrisée à partir de l'ellipse représentant l'écart-type des noyaux gaussiens. La distribution *a posteriori* des paramètres contenus dans le modèle hiérarchique est obtenue par *Markov chain Monte Carlo* (MCMC).

Paciorek (2003) et Paciorek et Schervish (2006) généralisent la forme du modèle proposé par Higdon *et al.* (1999) pour l'utilisation de fonctions de covariance de type Matérn et Cauchy non stationnaires. Dans leur exemple d'application, ils emploient une méthode similaire à celle de Higdon *et al.* (1999) (c'est-à-dire un modèle hiérarchique et la méthode MCMC), mais emploient une paramétrisation différente basée sur la décomposition en valeurs et vecteurs propres de la matrice de covariance des noyaux gaussiens.

Dans le modèle de covariance de Higdon *et al.* (1999), l'anisotropie locale (portées spatiales et rotation) est définie par la matrice de covariance du noyau gaussien centré en un point donné. Les termes sur la diagonale de cette matrice peuvent être assimilés aux carrés des portées spatiales et les termes en dehors de la diagonale définissent la rotation, s'il y a lieu. Dans un exemple simple, Paciorek et Schervish (2006) utilisent la présence d'un fort contraste topographique pour justifier l'ajustement de deux structures spatiales distinctes,

lesquelles sont ensuite combinées par le modèle de covariance non stationnaire.

On note également d'autres modèles non stationnaires basés sur la convolution de noyaux (Fuentes, 2001), sur les ondelettes (Nychka et Royle, 2002), ainsi que des extensions aux travaux de Paciorek (2003) par Jun et Stein (2008); Stein (2005); Mateu *et al.* (2011); Anderes et Stein (2011); Kleiber et Nychka (2012). Stein (2005) montre que les expressions fournies par Paciorek (2003) peuvent tenir compte d'une variance et d'une différentiabilité spatialement variables.

2.4 Synthèse

Ainsi, les méthodes conventionnelles d'interpolation ne permettent pas de tenir compte de l'information relative au modèle d'écoulement et de transport. Or, les approches d'interpolation associées à des méthodes inverses (Shlomi et Michalak, 2007; Schwede et Cirpka, 2010) illustrent bien l'intérêt d'utiliser le modèle physique lors de l'estimation de la concentration. Ceci est d'autant plus vrai puisque les études de modélisation stochastique montrent que le modèle physique a une forte empreinte sur les moments statistiques obtenus lorsque la conductivité hydraulique du milieu est considérée incertaine.

Ainsi, l'intégration d'information concernant le modèle de transport apparaît primordiale pour l'amélioration des approches d'interpolation directes, telle que le krigage. Un défi se pose car les résultats de la modélisation stochastique du transport montrent que le modèle probabiliste de la concentration est non stationnaire. Autrement dit, la fonction de densité de probabilité de la concentration, et par conséquent ses moments statistiques, varient spatialement. La modélisation d'une fonction de covariance non stationnaire par des approches basées sur la méthode des perturbations est limitée à des champs de conductivité hydraulique statistiquement homogènes et peu variables. L'approche de Monte Carlo requiert quant à elle des temps de calcul qui peuvent être difficilement justifiables pour la production de cartes de concentration lors de phases préliminaires de caractérisation de site ou pour des projets de petite envergure. Les techniques présentées à la section 2.3 montrent qu'il est possible de spécifier directement un modèle non stationnaire pour l'estimation d'une fonction aléatoire. D'une part, une transformation spatiale basée sur les coordonnées naturelles d'écoulement permet de simplifier l'anisotropie d'un modèle de transport. D'autre part, les modèles de covariance basés sur la convolution de noyaux gaussiens peuvent être facilement paramétrisés afin de représenter la variation spatiale d'un phénomène donné. Ainsi, en s'appuyant sur un

modèle conceptuel de l'écoulement et sur certains résultats concernant les moments statistiques de la concentration, on propose la modélisation directe d'une fonction de covariance non stationnaire approximative basée sur le modèle d'écoulement afin d'améliorer le krigage de la concentration de contaminant dans l'eau souterraine. Le chapitre suivant se penche sur les aspects méthodologiques d'une telle approche.

CHAPITRE 3

KRIGEAGE NON STATIONNAIRE DE LA CONCENTRATION DANS L'EAU SOUTERRAINE

La méthodologie présentée dans ce chapitre vise à spécifier directement un modèle de covariance non stationnaire approximatif pour la concentration de contaminant dans l'eau souterraine. Ce modèle est basé sur l'utilisation conjointe d'une transformation de coordonnées basée sur l'écoulement et de la classe de modèles de covariance non stationnaires proposée par Higdon *et al.* (1999).

On cherche ainsi à contourner les restrictions liées à l'obtention d'un modèle de covariance non stationnaire par la méthode des perturbations (faible variabilité de la conductivité hydraulique) ou par l'approche de Monte Carlo (temps de calcul). L'approche proposée vise à être applicable dans un contexte où l'on dispose de peu d'information concernant la position de la source et son intensité, car elle ne requiert aucune modélisation du transport.

Le calcul des coordonnées d'écoulement nécessite la disponibilité d'un modèle hydrogéologique. Ce modèle est construit à l'aide de l'information recueillie sur le site et de l'interprétation de l'hydrogéologue. On considère donc que la géologie, les propriétés du milieu et les conditions frontières sont connues partiellement et que la synthèse de l'information disponible permet la construction d'un modèle conceptuel. Une partie de l'incertitude provient ainsi de l'écart entre ce modèle et la réalité, lequel n'est pas quantifiable pour des systèmes naturels (Oreskes *et al.*, 1994). Selon l'information disponible et les besoins de l'étude, différentes options peuvent être envisagées :

1. utilisation d'un modèle d'écoulement calé sur des observations de charge hydraulique ;
2. spécification de modèles d'écoulement qui correspondent à des interprétations géologiques concurrentes ;
3. simulation de plusieurs champs de conductivité hydraulique équiprobables.

Il est à noter que le dernier cas constitue une approche de Monte Carlo. Toutefois, contrairement aux approches évoquées au chapitre 2, cette variante ne nécessite pas la simulation du transport puisque les concentrations sont estimées par krigeage.

Dans l'approche proposée ici, le passage à des coordonnées d'écoulement facilite la modélisation de l'anisotropie, en permettant de tenir compte des variations de la direction d'écoulement, ainsi que de la convergence et la divergence des tubes de courant. Une telle transformation en coordonnées d'écoulement est valable pour un système où les conditions d'écoulement sont en régime permanent, de manière à ce que qu'il y ait effectivement une continuité spatiale le long des tubes de courant. De son côté, le type de fonction de covariance utilisé permet de faire varier la portée spatiale et la variance de la concentration en fonction des coordonnées d'écoulement. Un aspect important est que la paramétrisation de la fonction de covariance se voit simplifiée par la transformation de coordonnées.

La figure 3.1 illustre l'approche proposée comparativement aux approches conventionnelles. On y voit que le krigage en coordonnées cartésiennes ne permet pas de tenir compte d'une anisotropie variable. De plus, une simple transformation de coordonnées tient uniquement compte de la non-stationnarité associée à l'advection et non d'éventuelles variations de la portée spatiale ou de la variance. L'approche proposée (en encadré) permet de tenir compte conjointement des variations de la direction d'anisotropie et des variations de la corrélation spatiale.

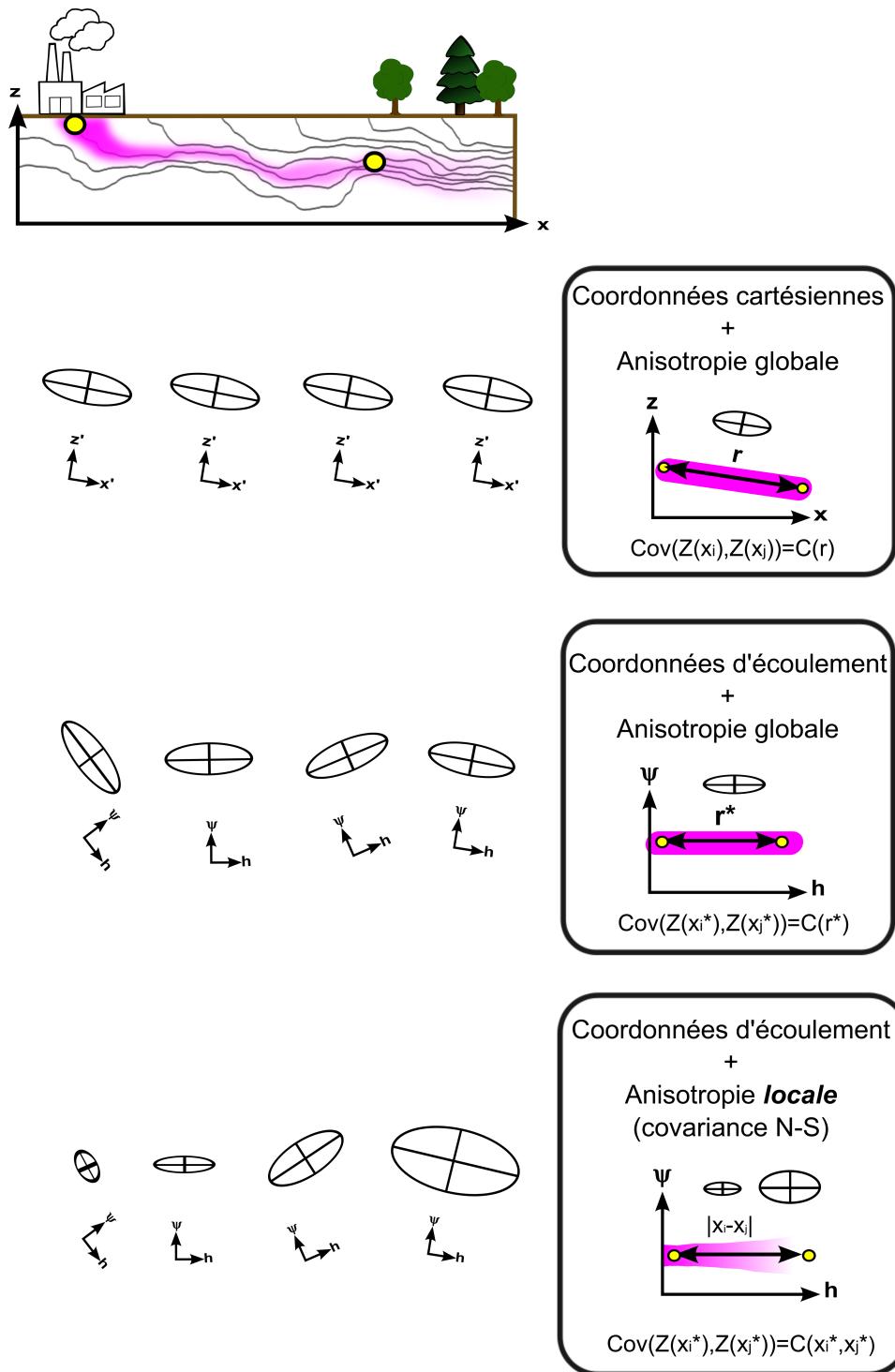


Figure 3.1 Schématisation de différentes approches de krigeage.

Dans les sections suivantes, on se penche d'abord sur divers aspects théoriques et opérationnels liés à la méthodologie proposée. À la section 3.1, on présente les coordonnées d'écoulement et les outils numériques utilisés pour les obtenir. La section 3.2 est dédiée à la

présentation du modèle de covariance non stationnaire, à la paramétrisation proposée pour celui-ci, de même qu'à la méthode utilisée pour estimer les paramètres. La section 3.3 détaille les équations de krigeage. Finalement, la section 3.4 présente les normes d'évaluation utilisées pour évaluer la performance de l'approche lors des études de cas synthétiques aux chapitres 4 et 5.

3.1 Transformation de coordonnées

On décrit ici les principaux éléments méthodologiques liés à l'obtention des coordonnées d'écoulement.

3.1.1 Formulation duale de l'écoulement et coordonnées naturelles

Pour un écoulement bidimensionnel en régime permanent, la formulation duale permet de décrire l'écoulement de l'eau souterraine en terme de la charge hydraulique h et de la fonction de courant ψ . Ces deux variables définissent un système de coordonnées naturelles où l'axe h est tangent au flux hydraulique (conductivités hydrauliques isotropes). L'essentiel de ce qui suit est tiré de Bear (1972) et de Domenico et Schwartz (1998).

Pour un écoulement en milieu poreux, la loi de Darcy énonce que le flux est proportionnel au gradient hydraulique :

$$q = -K \frac{\partial h}{\partial l} \quad (3.1)$$

où q est le flux hydraulique (m/s), K est la conductivité hydraulique (m/s) et $\frac{\partial h}{\partial l}$ est le gradient hydraulique (m/m). Le principe de conservation de la masse indique que le bilan entre la masse entrant dans un élément de volume et celle sortant de cet élément, doit être égale au taux de variation de la masse emmagasinée dans le temps. Lorsqu'on suppose que la densité de l'eau est constante (fluide incompressible), ceci peut être exprimé par :

$$-\left(\frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} + \frac{\partial q_z}{\partial z}\right) = \frac{1}{\rho_w} \frac{\partial(\rho_w n)}{\partial t} \quad (3.2)$$

où q_x , q_y et q_z sont les composantes du vecteur de flux hydraulique, ρ_w est la densité de l'eau (kg/m³) et n est la porosité. En substituant q par son expression dans la loi de Darcy, et en supposant un milieu isotrope et hétérogène, ainsi qu'un écoulement en régime permanent, on obtient :

$$\frac{\partial}{\partial x} \left(K \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left(K \frac{\partial h}{\partial y} \right) + \frac{\partial}{\partial z} \left(K \frac{\partial h}{\partial z} \right) = 0 \quad (3.3)$$

ou, écrit autrement,

$$\nabla \cdot K \nabla h = 0 \quad (3.4)$$

La charge hydraulique est obtenue en résolvant l'équation 3.3 pour un ensemble de conditions frontières et un champ de conductivité hydraulique donné.

De son côté, la fonction de courant peut être déduite de l'expression mathématique décrivant les lignes de courant. Celles-ci sont définies comme des courbes tangentes au flux hydraulique en tout point de l'espace, ce qui est exprimé par :

$$\mathbf{q} \times d\mathbf{r} = 0 \quad (3.5)$$

où \mathbf{q} est le vecteur de flux et $d\mathbf{r}$ est un élément d'arc le long d'une ligne de courant. En 2D, l'équation 3.5 s'exprime comme :

$$q_y dx - q_x dy = 0 \quad (3.6)$$

Pour un écoulement décrit par l'équation 3.4, la solution de l'équation 3.6 est donnée par :

$$\psi = \psi(x, y) = const \quad (3.7)$$

où ψ est la fonction de courant, laquelle prend des valeurs constantes le long des lignes de courant. L'équation qui décrit la fonction de courant a la même forme que celle de la charge hydraulique (i.e. équation 3.4) :

$$\nabla \cdot \frac{1}{K} \nabla \psi = 0 \quad (3.8)$$

La charge hydraulique h et la fonction de courant ψ sont des fonctions harmoniques conjuguées (Haitjema et Kelson, 1996) et obéissent ainsi à l'équation de Cauchy-Riemann, laquelle permet de relier les dérivées partielles de h et de ψ :

$$K \frac{\partial h}{\partial x} = \frac{\partial \psi}{\partial y}; \quad K \frac{\partial h}{\partial y} = -\frac{\partial \psi}{\partial x} \quad (3.9)$$

On note également que la fonction de courant est liée à la valeur du flux passant dans le tube de courant défini par deux lignes de courant et permet ainsi de calculer le débit passant entre deux points (Bear, 1972) :

$$Q_{AB} = \int_A^B \mathbf{q} \cdot d\mathbf{A} = \psi_A - \psi_B \quad (3.10)$$

où Q_{AB} est le débit entre les points A et B, \mathbf{q} est le flux, $d\mathbf{A}$ représente un élément de largeur unitaire et dont la longueur est égale au chemin (arbitraire) connectant les points A et B, ψ_A et ψ_B sont les valeurs de la fonction de courant aux points A et B, respectivement.

Conjointement, h et ψ peuvent être utilisées pour la représentation du réseau d'écoulement (Frind et Matanga, 1985; Fogg et Senger, 1985; Bramlett et Borden, 1990; Buxton et Modica, 1992) ou pour définir un système de coordonnées curvilignes : les *coordonnées naturelles d'écoulement* (Bear, 1972). Comme les équations décrivant h et ψ ont la même forme (équations 3.4 et 3.8), il est possible d'obtenir la solution de la fonction de courant à l'aide des mêmes solveurs numériques que pour la charge hydraulique (Frind et Matanga, 1985; Anderson et Woessner, 2002). Les conditions frontières du problème d'écoulement en ψ sont déduites à partir de l'équation de Cauchy-Riemann. De manière générale, une condition frontière de type Dirichlet sur la charge hydraulique est une condition de type Neumann sur la fonction de courant, et inversement. L'attribution des conditions frontières est discutée en détail par Frind et Matanga (1985) et Anderson et Woessner (2002). La transformation en coordonnées naturelles d'écoulement, de même que la correspondance entre les conditions frontières en h et en ψ , est illustrée à la figure 3.2.

On note que la fonction de courant n'est pas définie lorsqu'un puits ou une source est retrouvé à l'intérieur du modèle, puisque la solution de ψ n'est pas unique dans ce cas. Cette difficulté peut être contournée à l'aide d'une technique qui vise à rattacher le puits (ou la source) à une frontière (Frind et Matanga, 1985; Fogg et Senger, 1985). Toutefois, ceci exige des modifications importantes à la géométrie du modèle.

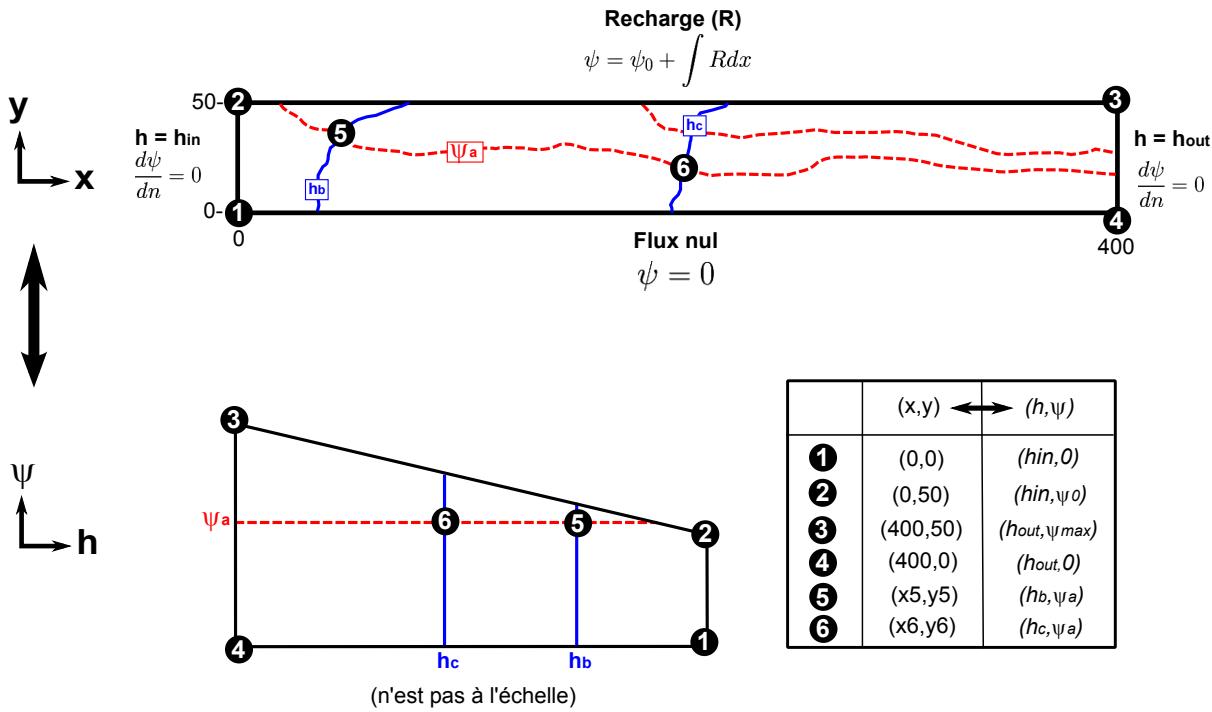


Figure 3.2 Illustration de la transformation en coordonnées naturelles et de la correspondance entre les conditions frontières de la charge hydraulique (**en gras**) et de la fonction de courant (*en italique*).

Certains auteurs présentent des développements théoriques pour définir des fonctions de courant en trois dimensions (Bear, 1972; Matanga, 1993). Dans ce cas, les lignes de courant sont obtenues par l'intersection de deux fonctions de courant, χ et λ . Seul Matanga (1993) présente une application du concept de fonction de courant en 3D. Dans l'exemple présenté, une indétermination lors de la spécification des conditions frontières pour les solutions de χ et λ nécessite des techniques de résolution itératives. Ainsi, très peu de travaux présentent ou utilisent des solutions aux fonctions de courant 3D décrites par Bear (1972). L'applicabilité des fonctions de courant 3D, pour un milieu hétérogène et anisotrope, est questionnée par Cirpka *et al.* (1999b, 2011b). De plus, Narasimhan (2008) indique que les conditions de Cauchy-Riemann sont seulement définies en 2D. Steward (1998) considère qu'une des hypothèses utilisées pour les développements présentés par Bear (1972) et Matanga (1993) est erronée.

Ces difficultés, de même que celles entourant la spécification de la fonction de courant en présence de puits (ou de sources) à l'intérieur du domaine, justifient le développement de l'approche décrite à la section suivante. Ainsi, les fonctions de courant 3D ne sont donc pas présentées ici.

3.1.2 Coordonnées alternatives d'écoulement

On présente ici un système de coordonnées curvilignes ne faisant pas appel à la fonction de courant. Les coordonnées transversales visent à ordonner les lignes de courant selon une fonction monotone, à la manière de la fonction de courant. Elles préservent ainsi les relations d'ordre entre les lignes de courant. Toutefois, contrairement à la fonction de courant, ces coordonnées ne sont pas proportionnelles au débit entre deux points.

Pour ce nouveau système, l'orientation de l'axe longitudinal est obtenue en traçant les lignes de courant à partir d'une solution d'écoulement en régime permanent. Comme pour les coordonnées naturelles, la charge hydraulique est utilisée pour définir la coordonnée longitudinale d'écoulement. Pour définir les coordonnées transversales, le point de départ de chaque ligne de courant est associé à des coordonnées définies sur les frontières de recharge (*boundary conforming coordinates*). En 2D, la coordonnée u est définie comme la distance le long des segments contigus qui représentent les frontières de recharge du modèle d'écoulement. En 3D, les coordonnées u et v sont issues d'une application (*mapping*) qui permet la représentation bidimensionnelle des surfaces contiguës définissant les frontières de recharge du modèle. Afin de distinguer des coordonnées naturelles, on désigne par *coordonnées alternatives d'écoulement* le nouveau système de coordonnées.

Dans cette thèse, des modèles à géométrie rectangulaire sont utilisés, ce qui permet une correspondance très simple entre les coordonnées cartésiennes et les coordonnées alternatives d'écoulement. Pour un modèle 2D, la coordonnée u est obtenue en sommant la distance le long des frontières de recharge et peut être aisément calculée à partir des coordonnées cartésiennes du point de départ d'une ligne de courant k donnée :

$$u(x_0^k, y_0^k) = \begin{cases} y_0^k & (x_0^k, y_0^k) \in \Gamma_1, \\ y_{max} + \beta_q x & (x_0^k, y_0^k) \in \Gamma_2. \end{cases} \quad (3.11)$$

où $\beta_q = \frac{q_2}{q_1}$ est le rapport entre les flux hydrauliques moyens sur les frontières Γ_1 et Γ_2 , (x_0^k, y_0^k) sont les coordonnées du point de départ de la k^{ieme} ligne de courant et y_{max} est la limite du modèle en y . Pour le modèle 3D de la figure 3.3, une application similaire est utilisée :

$$u(x_0^k, y_0^k, z_0^k) = y_0^k \quad (x_0, y_0, z_0) \in \Gamma_1 \cup \Gamma_2 \quad (3.12)$$

$$v(x_0^k, y_0^k, z_0^k) = \begin{cases} z_0^k & (x_0, y_0, z_0) \in \Gamma_1, \\ z_{max} + \beta_q x & (x_0, y_0, z_0) \in \Gamma_2. \end{cases} \quad (3.13)$$

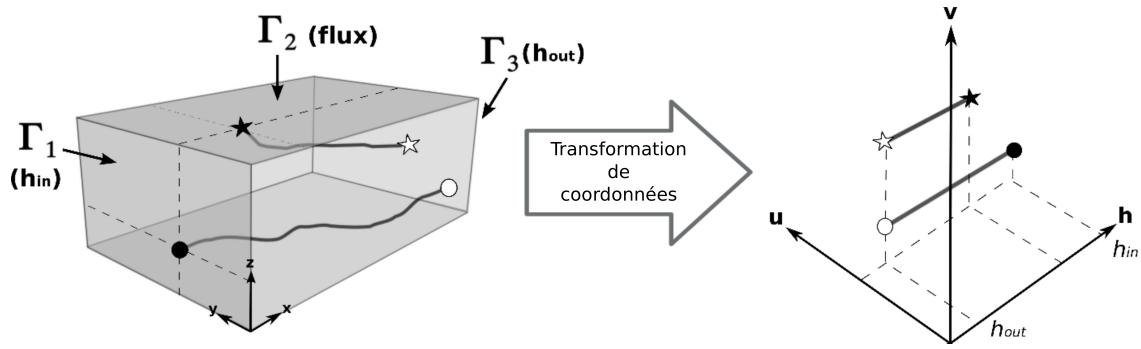


Figure 3.3 Illustration du concept de coordonnées alternatives d'écoulement en 3D. *À gauche* : Modèle d'écoulement en milieu hétérogène présentant des lignes de courant curvilignes. Les symboles en noir correspondent aux points de départ sur les frontières de recharge Γ_1 et Γ_2 . *À droite* : Une représentation 2D des frontières de recharge permet de transformer les coordonnées cartésiennes des points de départ en coordonnées u et v . Ces coordonnées demeurent constantes pour tous les points situés le long d'une même ligne de courant et ainsi, les lignes de courant sont des lignes droites dans l'espace transformé.

Les équations 3.11 à 3.13 permettent uniquement la correspondance entre les deux systèmes de coordonnées aux points situés sur cette ligne de courant. Dans le cadre du krigage, la correspondance entre les deux systèmes doit être disponible pour tous les points d'observation et d'estimation. Or, calculer les lignes de courant pour un nombre élevé de points de départ demande des temps de calcul importants. Ainsi, on propose de calculer uniquement un nombre limité de lignes de courant et d'utiliser une interpolation linéaire, laquelle préserve le caractère monotone de la transformation, pour obtenir les coordonnées u et v en tous points du modèle. L'algorithme suivant résume la procédure utilisée pour l'obtention des coordonnées alternatives d'écoulement :

1. Définir une grille de points de départ sur Γ_1 et Γ_2 .
2. Calculer les lignes de courant par trajectographie pour chaque point de la grille définie en 1.
3. Définir, pour la surface formée par l'ensemble des frontières de recharge (Γ_1 et Γ_2), une application qui permette de transformer les coordonnées (x, y, z) en coordonnées (u, v) .
4. Pour chaque ligne de courant, assigner les coordonnées u et v à tous les points discrétilisant la ligne de courant.
5. Interpoler u et v à tous les points d'observation et de la grille d'estimation en utilisant les valeurs définies en 4.
6. Utiliser la solution numérique de la charge hydraulique pour obtenir h aux points d'observation et aux points de la grille d'estimation.

L'application nécessaire pour assigner les coordonnées u et v dépend du nombre de frontières de recharge présentes et de la complexité de la géométrie du modèle. Une formulation équivalente consiste à définir les coordonnées à partir du point de sortie des lignes de courant, ce qui simplifie le problème si le nombre de frontières de sortie est inférieur au nombre de frontières de recharge. L'idée derrière les coordonnées u et v est avant tout de permettre d'organiser spatialement les lignes de courant de manière à ce que les distances qui les séparent tiennent compte approximativement du débit entre celles-ci. Dans certains cas, la présence de lignes de partage peut éventuellement nécessiter la division du domaine en plusieurs sous-domaines, où la transformation peut être définie de manière locale et où l'estimation est effectuée séparément.

En général, si les frontières de recharge ne sont pas linéaires (ou rectangulaires), des procédures plus complexes sont requises pour calculer u et v . En 2D, il peut être nécessaire de calculer numériquement la distance le long de frontières courbes pour obtenir u . Pour les modèles 3D, des approches numériques peuvent être requises. Thompson *et al.* (1985) présentent une méthode basée sur la résolution d'une équation elliptique avec conditions limites (*elliptic boundary-value problem*) pour générer des grilles sur des surfaces tridimensionnelles. On note également que les approches de placage de textures (*texture mapping* ou *uv mapping*), utilisées pour assigner les pixels issus d'une image 2D à une surface tridimensionnelle, permettent de calculer des coordonnées qui se conforment à la géométrie de cette surface (Levy et Mallet, 1998; Mallet, 2002). Les détails associés à l'implémentation de ces procédures dépassent le cadre de cette thèse. Il est à noter que ce type d'approche est implémentée dans certains logiciels de modélisation géologique, tel que GOCAD[®] (ParadigmTM, 2009).

Il est à noter que, lors du krigage, les paramètres de portée dans la direction transversale permettent de mettre à l'échelle les valeurs de la coordonnée transversale, qu'elle soit u ou ψ . Ces portées sont estimées à partir des données et fournissent un ajustement moyen de la continuité spatiale, lequel tient éventuellement compte d'erreurs de positionnement locales dans la coordonnée transversale. De plus, les erreurs de positionnement peuvent éventuellement être prises en compte par l'ajout d'un effet de pépite au modèle de covariance (voir figure 3.4).

On note que les coordonnées u et v montrent une certaine similarité avec les coordonnées lagrangiennes $\xi(\xi, \eta, \zeta)$ utilisées pour décrire la position initiale d'une particule. Pour un écoulement donné, le point de vue lagrangien décrit l'évolution de la position d'une particule dans le temps à partir de son point de départ à un temps initial t_0 (Bear, 1972) :

$$x = \mathbf{x}(\xi, t); \quad y = \mathbf{y}(\xi, t); \quad z = \mathbf{z}(\xi, t) \quad (3.14)$$

Pour une particule donnée, les coordonnées $\xi(\xi, \eta, \zeta)$ sont constantes pour $t \geq t_0$. Ainsi, on peut définir un autre système de coordonnées curvilignes en assimilant ξ à u et v . La variable t , qui correspond au temps parcouru depuis la position (u, v) située à la frontière de recharge, peut être également obtenue par traçage de particules. Pour un point donné x_m sur une ligne de courant, le temps de parcours est calculé en sommant les temps associés aux pas d'avancement antérieurs à ce point :

$$t_m = \sum_{i=0}^m \frac{dl_i}{v_i} \quad (3.15)$$

où dl_i est la distance entre deux points consécutifs de la ligne de courant, v_i est la vitesse d'écoulement au point i et m est le nombre de pas antérieurs au point x_m . Un algorithme similaire à celui présenté précédemment est utilisé pour déterminer les coordonnées (t, u, v) ; on y inclut les modifications suivantes :

- 4a. Pour chaque point discrétilisant les lignes de courant, calculer le temps de parcours correspondant.
- 6. Interpoler t , u et v à tous les points d'observation et de la grille d'estimation en utilisant les valeurs définies en 4 et 5.

Ainsi, deux systèmes de coordonnées alternatives d'écoulement sont proposés, soit les CAE _{h} , constitués de la charge hydraulique h , et des coordonnées u et v , et les CAE _{t} , constitués du temps de parcours à partir du point de départ sur la frontière t , et des coordonnées u et v .

Comparaison avec la fonction de courant (2D) et limitations

Comme on l'a vu, les coordonnées u et v sont reliées à la géométrie des frontières de recharge et ainsi, elles ne définissent pas nécessairement des tubes de courant aux débits égaux lorsque l'écoulement est hétérogène près de la frontière. Pour un modèle 2D rectangulaire (voir équation 3.11), u est proportionnel à ψ si le flux est constant le long des frontières de recharge. La coordonnée u est donc un substitut approximatif à ψ et le facteur β représente un ajustement moyen. Ceci a pour effet que la distance entre deux points donnés n'est pas toujours équivalente au débit passant entre ces deux points. À des fins de comparaison, on peut utiliser le coefficient de corrélation pour quantifier la ressemblance entre u et ψ . Toutefois, comme les coordonnées alternatives sont proposées pour les cas où ψ n'est pas définie,

on ne pourra généralement pas effectuer cette comparaison.

Comme on l'a mentionné précédemment, le recours à la transformation en coordonnées d'écoulement (autant CNE que CAE) est pertinent pour un système d'écoulement en régime permanent. Cirpka *et al.* (1999a) mentionnent que ceci n'est pas une limitation dans un contexte où le contaminant migre depuis des décennies ou des siècles, étant donné que pour de telles échelles de temps, les variations saisonnières peuvent être négligées. On peut également envisager appliquer l'approche dans des cas transitoires où il y a eu des changements modérés au régime d'écoulement, en gardant à l'esprit que la transformation en coordonnées d'écoulement n'est, dans ce cas, pas complètement cohérente avec l'anisotropie induite par l'advection.

Les procédures utilisées pour calculer les coordonnées d'écoulement comportent leur part d'imprécision numérique. Dans le cas des coordonnées naturelles (CNE), la précision de la solution de charge hydraulique et de la fonction de courant peuvent être améliorées en raffinant le maillage d'éléments finis. Ceci peut également contribuer à améliorer la précision du traçage des lignes de courant et ainsi des coordonnées alternatives u et v . Toutefois, comme la vitesse est obtenue par différentiation numérique de la solution de charge hydraulique, elle demeure moins précise que la solution de charge hydraulique et peu présenter des discontinuités d'un élément à l'autre, affectant ainsi la précision des lignes de courant tracées (Cordes et Kinzelbach, 1992; Srivastava et Brusseau, 1995). De plus, la précision dépend également de l'algorithme utilisé (méthode d'Euler ou de Runge-Kutta).

Malgré leur caractère approximatif, on retient que les coordonnées alternatives d'écoulement (CAE) permettent avant tout de tenir compte de l'anisotropie variable induite par l'écoulement en milieu hétérogène et de préserver les relations d'ordre entre les lignes de courant. Dans les cas où la fonction de courant peut être définie (2D sans puits ou source), on favorise plutôt l'utilisation des coordonnées naturelles d'écoulement, (h, ψ) . Autrement, les CAE constituent une solution acceptable, ce qui est illustré au chapitre 4.

3.1.3 Méthode des éléments finis

L'obtention des coordonnées d'écoulement est basée sur la disponibilité des champs de charge hydraulique, de fonction de courant (pour les coordonnées naturelles) et des vitesses (pour les coordonnées alternatives). En présence d'une géométrie irrégulière et/ou de paramètres hétérogènes, les équations 3.4 et 3.8 ne peuvent être résolues analytiquement et les solutions sont plutôt obtenues en ayant recours à des méthodes numériques. La méthode

des éléments finis, largement répandue pour la modélisation hydrogéologique, est préférée en raison de sa facilité à s'adapter à des géométries complexes. Dans ce qui suit, la présentation de la méthode des éléments finis se limite au cas d'un écoulement bidimensionnel en régime permanent et est tirée en grande partie des ouvrages de Pinder et Gray (1977) et de Pepper et Heinrich (2006), ainsi que de la thèse de Pasquier (2005).

Soit une équation différentielle de la forme :

$$Lh(x, y) = f(x, y) \quad (x, y) \in \Omega \quad (3.16)$$

où L est un opérateur différentiel, h est la fonction recherchée et f est une fonction quelconque et inconnue. La méthode des éléments finis vise à approximer la solution $h(x, y)$ en tout point d'un domaine Ω . À cette fin, la fonction h est remplacée par une fonction test \hat{h} , exprimée comme une combinaison linéaire de fonctions de base $\phi_i(x, y)$:

$$h \approx \hat{h} = \sum_{i=1}^N u_i \phi_i(x, y) \quad i = 1, 2, \dots, N \quad (3.17)$$

où les u_i sont des constantes inconnues et N est le nombre de points de la solution. La méthode des éléments finis tire son nom de la discrétisation du domaine Ω en sous-domaines, où \hat{h} est définie localement. N correspond aux noeuds du maillage d'éléments finis, lesquels assurent la continuité de la solution.

La résolution du problème est basée sur la méthode des résidus pondérés. Ainsi, on définit un résidu $R(x, y)$ tel que :

$$R(x, y) = L\hat{h} - f(x, y) \quad (3.18)$$

La méthode des résidus pondérés consiste à déterminer \hat{h} de manière à ce que le résidu (ou erreur) soit minimisé sur l'ensemble du domaine. Les valeurs des u_i sont obtenues en satisfaisant aux contraintes imposées par les intégrales pondérées :

$$\int_{\Omega} R(x, y) w_i(x, y) d\Omega = 0 \quad i = 1, 2, \dots, N \quad (3.19)$$

où $w_i(x, y)$ est une fonction de pondération non nulle. Les différentes variantes de la méthode des résidus pondérés se distinguent par le choix de la fonction $w(x, y)$. La méthode de Galerkin consiste à poser que $w_i(x, y) = \phi_i(x, y)$.

Pour le problème d'écoulement bidimensionnel (plan) en régime permanent, la fonction

$f(x, y)$ est nulle et l'opérateur différentiel $Lh(x, y)$ correspond au membre de gauche dans l'équation 3.3, soit :

$$\frac{\partial}{\partial x} \left(T \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left(T \frac{\partial h}{\partial y} \right) \quad (3.20)$$

où T est la transmissivité (m^2/s) obtenue en intégrant la conductivité hydraulique sur la profondeur de l'aquifère. L'expression obtenue en substituant l'équation 3.18 dans l'équation 3.19 est généralement simplifiée en appliquant le théorème de Green, lequel permet de réduire l'ordre de l'équation obtenue. Après substitution de l'éq. 3.18 dans l'éq. 3.19, en utilisant l'opérateur différentiel défini à l'éq. 3.20, et application du théorème de Green, on a donc :

$$-\int_{\Omega} T \frac{\partial \hat{h}}{\partial x} \frac{\partial w_i}{\partial x} + T \frac{\partial \hat{h}}{\partial y} \frac{\partial w_i}{\partial y} d\Omega + \int_{\Gamma} T \frac{\partial \hat{h}}{\partial n} w_i d\Gamma = 0 \quad (3.21)$$

En remplaçant \hat{h} par la combinaison linéaire de fonction de base de l'équation 3.17 et w_i par les fonctions de base $\phi_i(x, y)$ (méthode de Galerkin), on obtient l'expression du problème en terme des fonctions de base $\phi_i(x, y)$ et des constantes u_i recherchées :

$$\int_{\Omega} T \sum_{j=1}^N u_j \frac{\partial \phi_j}{\partial x} \cdot \frac{\partial \phi_i}{\partial x} + T \sum_{j=1}^N u_j \frac{\partial \phi_j}{\partial y} \cdot \frac{\partial \phi_i}{\partial y} d\Omega = \int_{\Gamma} T \sum_{j=1}^N u_j \frac{\partial \phi_j}{\partial n} \phi_i d\Gamma \quad (3.22)$$

L'équation 3.22 sous-tend le système d'équations linéaires suivant :

$$\mathbf{KU} = \mathbf{F} \quad (3.23)$$

où $\mathbf{U} = [u_1, u_2, \dots, u_N]^T$ est la solution recherchée en chaque noeud du maillage, et avec la matrice :

$$\mathbf{K} = [k_{i,j}] = \left[\int_{\Omega} T \left(\frac{\partial \phi_i}{\partial x} \frac{\partial \phi_j}{\partial x} + \frac{\partial \phi_i}{\partial y} \frac{\partial \phi_j}{\partial y} \right) d\Omega \right] \quad (3.24)$$

et le vecteur :

$$\mathbf{F} = [f_i] = \left[\int_{\Gamma} T \frac{\partial \phi_i}{\partial n} \phi_i d\Gamma \right] \quad (3.25)$$

La solution au problème d'éléments finis est une fonction continue par parties. En 2D, les éléments les plus rencontrés ont la forme de triangles et de quadrilatères. Les éléments triangulaires ont l'avantage de s'adapter aisément à des surfaces irrégulières, ce qui en fait le type d'élément le plus utilisé (Pepper et Heinrich, 2006). La grille définie par l'ensemble des éléments comporte également des noeuds, dont le nombre et la position dépendent de l'ordre du polynôme utilisé pour la fonction ϕ . Par exemple, dans le cas d'éléments triangulaires de

type linéaires, trois noeuds sont situés sur les sommets du triangle.

Le solveur d'éléments finis COMSOL Multiphysics[®] (Version 3.5a) est utilisé dans les exemples d'application présentés au chapitre suivant pour obtenir les solutions de charge hydraulique et de fonction de courant (coordonnées naturelles d'écoulement, section 3.1.1), de même que les vitesses utilisées pour tracer les lignes de courant (coordonnées alternatives d'écoulement, section 3.1.2). Il est à noter que l'approche proposée ne se limite pas à l'utilisation de solutions obtenues par éléments finis. Des solution analytiques, ou obtenues numériquement par différences finies ou volumes finis peuvent également être utilisées.

3.1.4 Traçage des lignes de courant

Pour les coordonnées alternatives d'écoulement, la détermination des valeurs des coordonnées u et v à l'intérieur du domaine requiert le traçage des lignes de courant à partir des vitesses obtenues lors de la résolution du problème d'écoulement par éléments finis.

Diverses méthodes numériques peuvent être utilisées pour calculer la succession de points définissant une ligne de courant à partir d'un point de départ x_0 et d'un champ de vitesse. Parmi celles-ci, on note la méthode d'Euler et les méthodes de Runge-Kutta. Le fonctionnement général de ce type de méthodes, appliqué au traçage des lignes de courant, est décrit par l'algorithme suivant¹ :

1. initialisation du point de départ de la ligne de courant ;
2. détermination d'un pas d'avancement d_i (dans le cas où il dépend localement de la taille de l'élément où est situé x_i) ;
3. calcul du déplacement et initialisation de la nouvelle position x_{i+1} ;
4. itération sur les étapes 2 et 3, jusqu'à ce qu'une frontière soit atteinte.

En général, la différence entre les méthodes numériques réside dans la détermination du déplacement entre les points x_i et x_{i+1} , ce qui a un impact sur la précision du résultat obtenu. Dans ce qui suit, on présente les équations pour une seule composante de la position. En réalité, on utilise plutôt un système d'équations afin d'obtenir les déplacements dans chaque direction du système d'axe.

Pour la méthode d'Euler, la position du point x_{i+1} est donnée par :

1. Adapté d'un algorithme présenté dans la thèse de Pasquier (2005), lequel s'inspirait de Anderson et Woessner (2002).

$$x_{i+1} = x_i + v_u(x_i)d \quad (3.26)$$

avec $v_u(x_i)$ la pente donnée par la composante unitaire de la vitesse en x_i pour la direction considérée (i.e. x , y , ou z) et d est le pas d'avancement². Une source d'erreur importante dans la méthode d'Euler vient du fait que l'on suppose que la vitesse calculée au point x_i s'applique pour l'ensemble de l'intervalle entre x_i et x_{i+1} (Chapra et Canale, 2002). Dans les méthodes de Runge-Kutta, la pente utilisée dans l'équation 3.26 est estimée à partir des vitesses en plusieurs points afin d'améliorer sa représentativité sur l'intervalle. Par exemple, une variante de la méthode de Runge-Kutta d'ordre deux (point milieu) est obtenue en posant :

$$k_1 = v_u(x_i) \quad (3.27)$$

$$k_2 = v_u(x_i + \frac{1}{2}k_1d) \quad (3.28)$$

(3.29)

La position du point subséquent est ensuite donnée par :

$$x_{i+1} = x_i + k_2d \quad (3.30)$$

Les méthodes de Runge-Kutta d'ordre quatre sont les plus utilisées en pratique (Chapra et Canale, 2002). Dans ce cas, une précision accrue est obtenue en calculant quatre vitesses intermédiaires :

$$k_1 = v_u(x_i) \quad (3.31)$$

$$k_2 = v_u(x_i + \frac{1}{2}k_1d) \quad (3.32)$$

$$k_3 = v_u(x_i + \frac{1}{2}k_2d) \quad (3.33)$$

$$k_4 = v_u(x_i + k_3d) \quad (3.34)$$

(3.35)

La position du point est déterminée en utilisant une moyenne pondérée de ces vitesses (méthode classique) :

$$x_{i+1} = x_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)d \quad (3.36)$$

2. Pour un pas de temps on utilise directement la vitesse.

La version utilisée du logiciel COMSOL Multiphysics[®] effectue le traçage des lignes de courant par une méthode de Runge-Kutta d'ordre deux. Pour les cas hétérogènes considérés aux chapitres 4 et 5 cette implémentation ne permettait pas toujours un traçage adéquat des lignes de courant (lignes incomplètes). Ainsi, il a plutôt été choisi de recourir aux fonctions `stream2` et `stream3` de Matlab[®].

3.2 Fonction de covariance

L'utilisation d'un modèle probabiliste est justifiée par l'incertitude concernant la distribution réelle de la concentration et ne serait évidemment pas requise si les modèles d'écoulement et de transport, de même que l'historique des émissions de contaminants, étaient parfaitement connus. Dans ce qui suit, on présente la fonction de covariance utilisée pour modéliser cette incertitude.

3.2.1 Covariance non stationnaire par convolution de noyaux gaussiens

Tel que mentionné au chapitre précédent, Higdon *et al.* (1999) ont démontré qu'une classe de fonctions de covariance non stationnaires peut être obtenue par la convolution de noyaux dont les paramètres varient spatialement :

$$C^{NS}(x_i, x_j) = \int_{\Re^p} K_{x_i}(\mathbf{u}) K_{x_j}(\mathbf{u}) d\mathbf{u} \quad (3.37)$$

où $C^{NS}(x_i, x_j)$ est une fonction de covariance non stationnaire, \mathbf{x}_i , \mathbf{x}_j et \mathbf{u} sont des points situés dans \Re^p avec $p = 1, 2, \dots$ et K_{x_i} est le noyau centré en \mathbf{x}_i . Cette construction par moyenne mobile assure que la covariance obtenue est définie positive (Mateu *et al.*, 2011). L'utilisation d'un noyau gaussien défini par :

$$K_x = (2\pi)^{-p/2} |\Sigma|^{1/2} \exp(-(1/2)(\mathbf{x} - \mathbf{u})^T \Sigma^{-1} (\mathbf{x} - \mathbf{u})) \quad (3.38)$$

permet d'obtenir une forme explicite pour C^{NS} . En utilisant la notation employée par Paciorek et Schervish (2006), C^{NS} est exprimé comme :

$$C^{NS}(x_i, x_j) = \sigma^2 |\Sigma_i|^{\frac{1}{4}} |\Sigma_j|^{\frac{1}{4}} \left| \frac{\Sigma_i + \Sigma_j}{2} \right|^{-\frac{1}{2}} \exp(-Q_{ij}) \quad (3.39)$$

où $\Sigma_i = \Sigma(\mathbf{x}_i)$ est la matrice de covariance du noyau gaussien centré en \mathbf{x}_i et σ^2 est la variance de la fonction aléatoire modélisée. L'équation 3.39 contient une forme quadratique, Q_{ij} , définie comme :

$$Q_{ij} = (\mathbf{x}_i - \mathbf{x}_j) \left(\frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} (\mathbf{x}_i - \mathbf{x}_j)^T \quad (3.40)$$

où \mathbf{x}_i est le vecteur de dimensions $1 \times p$ des coordonnées du point i . Paciorek et Schervish (2006) démontrent que d'autres expressions peuvent être obtenues pour C^{NS} et proposent une forme plus générale pour la corrélation non stationnaire :

$$R^{NS}(x_i, x_j) = |\Sigma_i|^{\frac{1}{4}} |\Sigma_j|^{\frac{1}{4}} \left| \frac{\Sigma_i + \Sigma_j}{2} \right|^{-\frac{1}{2}} R^S(\sqrt{Q_{ij}}) \quad (3.41)$$

où R^S est une fonction de corrélation stationnaire, valide sur \Re^p pour tout $p = 1, 2, \dots$. Cette condition est remplie par les modèles exponentiel, Cauchy (gravimétrique) et Matérn. La généralisation exprimée par l'équation 3.41 permet donc l'utilisation de modèles présentant des différentiabilités à l'origine moindre que celle du modèle gaussien, laquelle est réputée problématique (Stein, 1999). Les modèles de covariance de type Matérn comportent un paramètre de différentiabilité ν et peuvent être plus ou moins dérivables à l'origine. On note toutefois que dans l'équation 3.41, la matrice Σ correspond toujours à la matrice de covariance d'un noyau de type gaussien (équation 3.38), ceci étant possible puisque les fonctions de corrélation admises pour R^S (modèles exponentiel, Cauchy, Matérn) peuvent être exprimées comme des mélanges (*scale mixtures*) du modèle gaussien (voir la preuve dans Paciorek et Schervish (2006)).

Le caractère non stationnaire des modèles présentés aux équations 3.39 et 3.41 provient de la variation spatiale de Σ_i . Par ailleurs, l'utilisation d'une matrice Σ_i constante dans les équations 3.39 et 3.41 génère une fonction de covariance/corrélation stationnaire. Les valeurs dans la matrice Σ_i sont responsables pour les changements de portées et de direction d'anisotropie de la fonction aléatoire.

De plus, la variance σ^2 peut également varier spatialement (Stein, 2005). En effet, si R^{NS} est définie positive, alors $D_\sigma R^{NS} D_\sigma$ est aussi positive définie, où D_σ est une matrice diagonale où les valeurs sur la diagonale sont les écarts-types σ_i aux points x_i et on a que :

$$C^{NS}(x_i, x_j) = \sigma_i \sigma_j R^{NS}(x_i, x_j) \quad (3.42)$$

Le modèle de covariance non stationnaire est donc obtenu en combinant les structures locales décrites par les noyaux gaussiens. Dans les applications présentées par Higdon *et al.* (1999), Swall (1999) et Paciorek et Schervish (2006), les paramètres de la matrice Σ_i s'insèrent dans un modèle hiérarchique gaussien où leur distribution est reliée à celle de la variable

d'intérêt. La distribution *a posteriori* est ensuite estimée à partir des données par MCMC.

Dans ce qui suit, on propose une paramétrisation simplifiée pour le modèle de covariance non stationnaire de la concentration. Celle-ci s'inspire des moments statistiques observés dans les études de modélisation stochastique mentionnées au chapitre 2.

3.2.2 Paramétrisation

Le modèle de covariance basée sur la convolution de noyaux peut donc s'adapter aux changements dans la structure spatiale d'une fonction aléatoire. Pour ce faire, on doit toutefois paramétriser ces changements dans la matrice de covariance des noyaux, Σ_i , ainsi que dans la variance σ_i^2 , et procéder à l'estimation de leurs paramètres respectifs.

La paramétrisation proposée ici s'inspire du comportement non stationnaire général de la concentration, tel qu'observé dans les études de modélisation stochastique du transport et décrit à la section 2.2. De plus, cette paramétrisation se voit simplifiée par l'emploi de la transformation de coordonnées. L'objectif n'est pas d'utiliser des expressions analytiques spécifiques, mais plutôt un modèle simple reflétant les grandes lignes du comportement de la corrélation et la variance de la concentration, tout en s'adaptant au fait que la position et la taille de la source ne sont pas connues. Une préoccupation est de limiter le nombre de paramètres, puisqu'en pratique on dispose souvent d'un nombre restreint d'observations pour les estimer.

Matrice Σ_i

Les transformations de coordonnées proposées à la section 3.1 ont pour effet d'orienter un des axes selon le flux hydraulique. Cette transformation est analogue à celle utilisée dans la méthode des directions principales pour la modélisation numérique du transport (Frind, 1982; Cirpka *et al.*, 1999b). La méthode des directions principales permet d'utiliser une matrice diagonale pour le tenseur de dispersivité. Dans le cas présent, le fait que la direction de continuité principale, i.e. la direction du flux hydraulique, corresponde à un des axes du système de coordonnées permet d'utiliser une matrice diagonale pour Σ_i , avec comme termes diagonaux les portées locales du modèle de covariance. Les effets de l'advection et de la dispersion longitudinale sont confondus dans la portée spatiale parallèle à l'écoulement. De leur côté, les portées transversales incorporent les effets liés à la largeur de la source et la dispersion transversale.

Tel qu'énoncé précédemment, on suppose que la position et la taille de la source sont inconnues. Or, ces paramètres semblent avoir une influence considérable sur la forme de la fonction de covariance dans la direction transversale, tel qu'illustré dans certains travaux (Graham et McLaughlin, 1989a, 1991). En absence d'information sur les caractéristiques de la source, il apparaît plus prudent de supposer que la fonction de covariance varie uniquement dans la direction longitudinale.

La paramétrisation de la matrice Σ_i se limite donc aux portées situées sur sa diagonale. Étant donné les indications concernant l'augmentation de la corrélation spatiale avec le transport (Graham et McLaughlin, 1989a, 1991; Rubin, 1991; Vanderborght *et al.*, 2006), on choisit de faire croître les portées en fonction de la coordonnée longitudinale. Un modèle linéaire est utilisé à cette fin et les portées dans les différentes directions sont données par :

$$a_p(s_i) = a_{p0} + b_p s_i \quad a_{p0} \geq 0, b_p \geq 0 \quad (3.43)$$

où $a_p(s_i)$ est le paramètre de portée au point i dans la direction $p = 1, 2, 3$ (i.e. les directions principales du système d'axes³), a_{p0} est la portée initiale en amont de la zone d'étude, b_p est le taux d'augmentation de la portée et s_i est la coordonnée longitudinale au point i . Par convention, on choisit d'utiliser $s = h_{max} - h$ de manière à ce que b_p soit positif (ceci n'est pas essentiel). Pour une direction donnée la persistance spatiale (portée) est donc décrite par deux paramètres, a_{p0} et b_p . On note qu'un paramètre de portée stationnaire est obtenu en posant $b_p = 0$. Ainsi, pour un modèle en 2D, la matrice de covariance d'un noyau gaussien au point i est définie comme suit :

$$\Sigma_i = \begin{bmatrix} a_1(s_i)^2 & 0 \\ 0 & a_2(s_i)^2 \end{bmatrix} \quad (3.44)$$

Les portées représentent les écarts-types du noyau gaussien centré au point i .

Variance σ_i^2

Pour la variance, on néglige l'augmentation habituellement retrouvée à de faibles distances de la source. Ceci est justifié en partie par le fait que l'incertitude relative à l'écoulement est diminuée en raison de l'information apportée par le modèle conceptuel. On modélise la variance comme une fonction décroissante.

3. En coordonnées cartésiennes : x, y, z . En coordonnées naturelles : h et ψ . En coordonnées alternatives : h, u et v .

Une fonction exponentielle décroissante assure des valeurs positives sur l'ensemble du domaine et σ_i^2 est invariante dans la direction transversale. Elle est donnée par :

$$\sigma^2(s_i) = \sigma_0^2 \exp(-ds_i) \quad c \geq 0 \quad (3.45)$$

où $\sigma^2(s_i)$ est la variance locale, σ_0^2 est la variance maximale, retrouvée en amont de la zone d'étude, et d est le taux de décroissance de la variance. Une variance stationnaire est obtenue en posant $d = 0$. Les variances locales sont utilisées dans l'équation 3.42 pour calculer la covariance non stationnaire C^{NS} .

Modèles candidats

Les paramètres proposés plus haut sont des simplifications d'un modèle sous-jacent et inconnu. On propose donc de tester différentes combinaisons des paramètres de portées et de la variance. On définit ainsi quatre modèles de covariance candidats :

- S1 : Modèle stationnaire avec portées et variance constantes ;
- NS1 : Modèle non stationnaire avec portées de la forme donnée à l'équation 3.43 et variance constante ;
- NS2 : Modèle non stationnaire avec portées constantes, et variance non stationnaire de la forme donnée à l'équation 3.45 ;
- NS3 : Modèle non stationnaire avec portées et variance des formes données aux équations 3.43 et 3.45, respectivement.

Il est à noter que pour les modèles 3D, des versions plus parcimonieuses avec des portées transversales égales, sont aussi proposées, ce qui porte à huit le nombre de modèles candidats en 3D. Les modèles présentant une isotropie dans la direction transversale sont identifiés S1* et NSX*, avec X = 1, 2, 3.

Les modèles de covariance proposés ne tiennent pas compte explicitement des effets potentiels liés à l'historique des émissions, au mode d'émission (instantané ou continu), à la taille de la source, ni de ceux liés à la présence de dégradation ou à la réactivité du contaminant. On estime que les impacts de ces éléments sur la continuité spatiale et la variabilité de la concentration sont pris en compte lors de l'ajustement des paramètres de portée et de la variance. Le modèle est néanmoins mieux adapté à un mode d'émission de type continu (mais dont l'intensité peut être variable) en raison de la forme choisie pour la variance (équation 3.45). En effet, cette fonction ne prévoit pas de diminution de la variance en amont du centre de masse du panache, tel qu'observé pour des panaches produits par une émission instantanée de contaminant (voir par exemple Graham et McLaughlin (1991)). En effet,

l'absence d'information spécifique sur la moyenne (non stationnaire) de la concentration ne permet pas de modéliser la variance d'un panache issue d'une émission instantanée de manière représentative.

Estimation par maximum de vraisemblance (EMV) et sélection de modèle

Les paramètres des modèles de covariance candidats sont estimés par maximum de vraisemblance. En pratique, plutôt que de maximiser la vraisemblance \mathcal{L} , on procède généralement à la minimisation du logarithme négatif $L = -\ln(\mathcal{L})$ (Cressie, 1993; Kitanidis et Shen, 1996). Sous hypothèse gaussienne, L est donné par :

$$L(\beta, \theta) = \frac{n}{2} \ln(2\pi) + \frac{1}{2} \ln(|\mathbf{C}(\boldsymbol{\theta})|) + \frac{1}{2} (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{C}(\boldsymbol{\theta})^{-1} (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}) \quad (3.46)$$

où n est le nombre de données disponibles, $\mathbf{C}(\boldsymbol{\theta})$ est la matrice de covariance des données de concentration avec comme paramètres $\boldsymbol{\theta}$ ⁴, \mathbf{Z} représente les données de concentration, \mathbf{X} contient les fonctions de base de la dérive, et finalement, $\boldsymbol{\beta}$ est constitué des coefficients de dérive. La minimisation de L est effectuée à l'aide de la fonction `fminsearch` de MATLAB, basée sur la méthode de Nelder-Mead (Nelder et Mead, 1965). Des contraintes sont ajoutées à la fonction minimisée afin d'éviter l'estimation de paramètres de portée excessivement grands relativement aux dimensions du domaine, ou négatifs.

La moyenne de la concentration est représentée par une constante et ainsi, \mathbf{X} est un vecteur de 1 et $\boldsymbol{\beta}$ est un paramètre correspondant à la moyenne. Une moyenne constante est utilisée en dépit du fait qu'il s'agit d'un choix plus ou moins réaliste, compte tenu des résultats observés dans les études stochastiques, où la moyenne a plutôt la forme du panache moyen. Or, la modélisation d'une dérive plus réaliste, qui tiendrait compte de variations transversales dans les valeurs de concentration, nécessite des connaissances (ou hypothèses) sur la position et la taille de la source. Une dérive de type linéaire, décroissante le long de la direction longitudinale, est également considérée à l'exemple 1 du chapitre 4. On verra toutefois que celle-ci introduit un biais plus important que la moyenne constante. Il est à noter qu'il s'agit ici de la moyenne non conditionnelle, ou *a priori*, laquelle est basée sur l'information disponible avant le conditionnement aux données par krigeage.

Dans un contexte où la moyenne est inconnue, l'estimation par maximum de vraisemblance est sujette à un biais potentiellement important sur les paramètres du modèle de covariance (Chilès et Delfiner, 1999). Il est possible de réduire ce biais en ayant recours à l'estimation

4. $\boldsymbol{\theta}$ est constitué des paramètres présentés plus haut pour Σ_i et σ^2 .

par maximum de vraisemblance restreint (*Restricted Maximum Likelihood Estimation*, Kitanidis (1987); Kitanidis et Shen (1996)). Cette approche requiert la transformation de la variable d'intérêt en incrément permettant de filtrer la moyenne $\mathbf{X}\boldsymbol{\beta}$, ainsi que le recours à une fonction de covariance généralisée. Toutefois, dans la mesure où le modèle de covariance non stationnaire proposé dans cette thèse est directement lié aux moments statistiques de la concentration, il n'est guère certain que des incréments généraux conservent les caractéristiques non stationnaires que l'on tente de modéliser. On se limite donc à l'estimation par maximum de vraisemblance présentée à l'équation 3.46. De plus, bien que l'on considère la moyenne inconnue lors du krigeage (krigeage ordinaire), celle-ci est fixée à la moyenne des observations lors de l'estimation des paramètres afin d'utiliser une moyenne commune pour tous les modèles de covariance testés. Ceci est fait afin de permettre la comparaison des différents modèles candidats sur la base de l'ajustement des paramètres du modèle de covariance uniquement (et non sur l'ajustement de la moyenne). Il a été vérifié que cette procédure avait relativement peu d'effet sur les paramètres estimés. On note également que les moyennes estimées par krigeage ordinaire sont très similaires, que l'on utilise les paramètres estimés avec une moyenne fixe ou non.

Un paramètre d'effet de pépite est également inclus dans le modèle ; celui-ci est spécifié comme une fraction du palier de la covariance. Ce paramètre représente la variabilité à petite échelle due aux erreurs de positionnement ou de mesure, ainsi qu'aux structures de dimension inférieure à la résolution possible par l'échantillonnage (Chilès et Delfiner, 1999). À moins d'indications à ce sujet, la fraction d'effet de pépite n'est pas estimée et est fixée de manière *ad hoc* dans les exemples des chapitres suivants, dans le but de réduire le nombre de paramètres à estimer, étant donné le nombre relativement restreint de données utilisées dans chaque cas. Une étude de sensibilité sur ce paramètre a toutefois été effectuée et bien que dans certains cas la modification de la fraction d'effet de pépite mène à l'estimation de paramètres légèrement différents, ceci ne modifie pas les conclusions présentées aux chapitres quatre et cinq.

Dans l'équation 3.46, $\mathbf{C}(\boldsymbol{\theta})$ est une version non stationnaire de la fonction de covariance Matérn, avec comme paramètres les portées et la variance définies précédemment. La version stationnaire de la corrélation de type Matérn est donnée par :

$$R^S(\tau) = \frac{1}{\Gamma(\nu)2^{\nu-1}}(2\sqrt{\nu}\frac{\tau}{\rho})^\nu \kappa_\nu(2\sqrt{\nu}\frac{\tau}{\rho}) \quad (3.47)$$

où τ est la distance, ρ est la portée, κ_ν est une fonction de Bessel modifiée de seconde espèce et d'ordre ν , le paramètre de forme de la covariance Matérn. L'équation 3.47 est utilisée dans

3.41 pour calculer la version non stationnaire et dans ce cas, $\frac{\tau}{\rho}$ est remplacé par la forme quadratique donnée à l'Éq. 3.40. On utilise une valeur de $\nu = 0.5$, ce qui est équivalent à un modèle de covariance de type exponentiel. Le paramètre ν est lié à la différentiabilité du modèle de covariance et peut être estimé à partir des données et inclus dans θ . Ceci est testé dans l'exemple d'application présenté par Paciorek et Schervish (2006) où les résultats obtenus pointent vers l'utilisation de $\nu = 0.5$. La figure 3.4(a) illustre la fonction donnée par l'équation 3.47 pour un paramètre $\nu = 0.5$; elle illustre également le variogramme correspondant et certains paramètres géostatistiques.

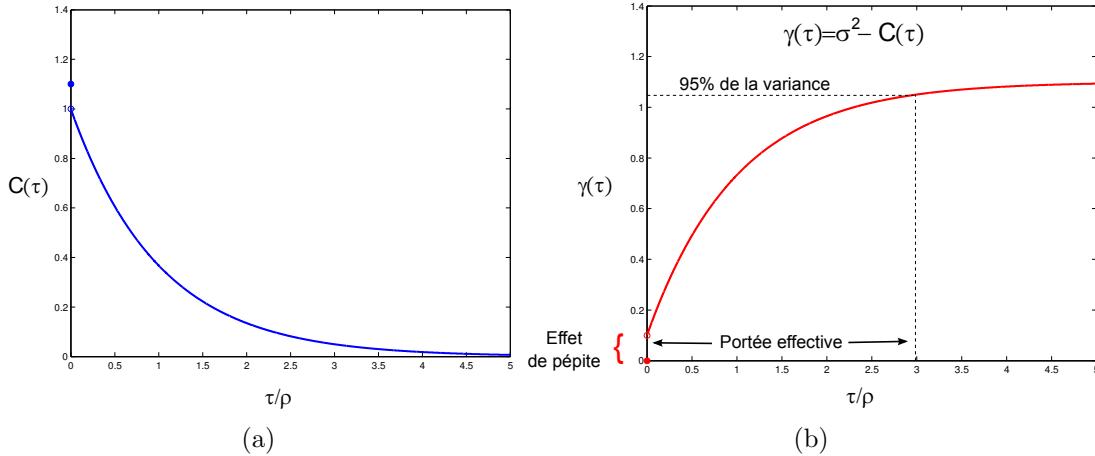


Figure 3.4 Fonction de covariance stationnaire de type Matérn avec $\nu = 0.5$ (équivalente à la fonction de covariance exponentielle) avec effet de pépite (a) Covariance, (b) Variogramme correspondant avec l'illustration de différents paramètres géostatistiques, tels que l'effet de pépite et la portée effective.

L'utilisation de l'équation 3.46 sous-tend que la concentration suit une distribution gaussienne. Or, ce type de distribution est inadéquat pour décrire la concentration car il implique que des probabilités non nulles soient associées à des valeurs de concentration négatives. En dépit de problèmes potentiels de biais, l'équation 3.46 est néanmoins utilisée pour estimer les paramètres du modèle de covariance. Il est fréquent de considérer l'estimation par maximum de vraisemblance comme une procédure robuste à des données non gaussiennes (Dowd et Pardo-Iguzquiza, 2012), la vraisemblance obtenue étant toutefois considérée approximative (ou quasi-maximum de vraisemblance). Une autre approche est de normaliser les données de concentration et de supposer une distribution multinormale. Toutefois, la transformation inverse requise après le krigage peut entraîner un problème de biais sur la moyenne.

Une attention doit être portée à la surparamétrisation des modèles de covariance non

stationnaires, puisque ceux-ci nécessitent davantage de paramètres que des modèles stationnaires (Paciorek et Schervish, 2006). Certains des modèles proposés sont imbriqués (*nested*) et permettent de tester l'ajout des paramètres non stationnaires par le test du rapport de vraisemblance (*likelihood ratio test*, LRT), en utilisant les valeurs de maximum de vraisemblance issues de l'estimation des paramètres (Kendall et Stuart, 1977; Mardia *et al.*, 1979).

Soit θ le vecteur des paramètres décrivant la distribution des données, le LRT teste l'hypothèse $H_0 : \theta \in \Omega_r$ contre l'hypothèse alternative $H_1 : \theta \in \Omega_q$, où Ω_r est un sous-ensemble de Ω_q (Kendall et Stuart, 1977; Mardia *et al.*, 1979). La statistique

$$\lambda = -2\ln \left(\frac{\max \mathcal{L}(\theta_r)}{\max \mathcal{L}(\theta_q)} \right) \quad (3.48)$$

où $\mathcal{L}(\theta_r)$ et $\mathcal{L}(\theta_q)$ sont les fonctions de vraisemblance des modèles avec r et q paramètres, possède une distribution asymptotique en χ^2_{q-r} .

Le modèle stationnaire S1 peut être vu comme une version simplifiée des modèles non stationnaires, dans lesquels certains paramètres sont fixés à 0. Par contre, les modèles NS1 et NS2 ne sont pas imbriqués puisque l'un ne peut pas être exprimé comme une version plus simple de l'autre (mais chacun est imbriqué avec S1). En 3D, les modèles NSX*, ayant des portées transversales égales, sont imbriqués dans leur version à trois portées différentes. Il est à noter que le test du rapport de vraisemblance ne peut être utilisé qu'entre les modèles utilisant un même système de coordonnées spatiales.

Le test de vraisemblance est considéré robuste à un écart à l'hypothèse gaussienne (Kim et Cai, 1993; Yuan et Bentler, 1999). La validation croisée est une alternative à l'estimation des paramètres par maximum de vraisemblance (Marcotte, 1995). Celle-ci ne dépend d'aucune hypothèse sur la distribution des données, mais on ne dispose toutefois pas de test sur les statistiques de la validation croisée pour déterminer si l'ajout de paramètres est significatif.

Le critère d'Akaike corrigé pour les échantillons de petites tailles, AIC_c , est également utilisé pour valider la sélection obtenue à l'aide du LRT. Comme pour le LRT, l' AIC_c est calculé à partir du maximum de vraisemblance obtenu lors de l'estimation des paramètres géostatistiques pour chacun des modèles candidats. Lorsque le nombre de données est limité, comme c'est le cas dans les exemples présentés, Burnham et Anderson (2002) recommandent l'utilisation de l' AIC_c plutôt que de l'AIC. AIC_c est donné par :

$$AIC_c = -2\ln(\mathcal{L}(\theta)) + 2k \left(\frac{n}{n-k-1} \right) \quad (3.49)$$

où $\mathcal{L}(\theta)$ est le maximum de vraisemblance associé à un modèle donné, k est le nombre de paramètres dans le modèle et n est le nombre de données. On choisit le modèle ayant la plus faible valeur d' AIC_c parmi un ensemble de modèles candidats. On note que d'autres critères de sélection auraient également pu être utilisés, tel que le BIC (*Bayesian Information Criterion*, Schwarz (1978)) ou la version de l' AIC_c proposée par Hoeting *et al.* (2006). Il a été testé que, pour les cas présentés aux chapitres quatre et cinq, ces paramètres engendrent la même classification que l' AIC_c choisi ici.

3.3 Krigeage

3.3.1 Système de krigeage ordinaire

Une fois les coordonnées spatiales transformées et le modèle de covariance choisi, l'estimation de concentration est faite par krigeage. Dans le cas d'une moyenne (non conditionnelle) constante, mais inconnue, on parle alors de krigeage ordinaire et l'interpolateur est de la forme :

$$Z_0^* = \sum_i^n \lambda_i Z_i \quad (3.50)$$

où Z_0^* est la concentration estimée au point x_0 , λ_i est le poids associé à l'observation située en x_i et Z_i est la concentration observée au point x_i . La valeur estimée Z_0^* correspond à la moyenne conditionnelle de la concentration au point x_0 (cas gaussien) ou à un bon estimé de celle-ci (cas général).

Les poids de krigeage sont obtenus de manière à minimiser la variance d'estimation σ_e^2 :

$$\sigma_e^2 = Var[Z_0 - Z_0^*] = Var[Z_0] + \sum_i^n \sum_j^n \lambda_i \lambda_j C[Z_i, Z_j] - 2 \sum_i^n \lambda_i C[Z_0, Z_i] \quad (3.51)$$

où $Var[Z_0]$ est la variance de la concentration au point x_0 , $C[Z_i, Z_j]$ est la covariance entre les observations i et j et $C[Z_0, Z_i]$ la covariance entre la concentration au point x_0 et l'observation située en x_i . Afin que l'interpolateur obtenu soit sans biais, i.e. que $E[Z_0^*]$ soit égale à la moyenne, une contrainte (éq. 3.54) doit être incluse au système d'équations. La minimisation

sous contrainte est effectuée à l'aide de la méthode de Lagrange :

$$L(\lambda) = \sigma_e^2 + 2\mu \left(\sum_i^n \lambda_i - 1 \right) \quad (3.52)$$

où $L(\lambda)$ est le lagrangien. Le système d'équations menant à l'obtention des poids (λ_i) recherchés est obtenu en annulant les dérivées partielles de $L(\lambda)$ par rapport aux λ_i et à μ . On obtient donc :

$$\sum_i^n \lambda_i C[Z_i, Z_j] + \mu = C[Z_0, Z_i] \quad (3.53)$$

$$\sum_i^n \lambda_i = 1 \quad (3.54)$$

Ce système d'équations, communément appelé *système de krigeage ordinaire*, peut être réécrit sous forme matricielle :

$$\begin{bmatrix} \mathbf{C} & \mathbf{1} \\ \mathbf{1}' & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \mu \end{bmatrix} = \begin{bmatrix} \mathbf{c}_0 \\ 1 \end{bmatrix} \quad (3.55)$$

où \mathbf{C} est la matrice des covariances entre les observations, $\mathbf{1}$ est un vecteur composé de 1, $\boldsymbol{\lambda}$ est le vecteur des poids de krigeage, μ est le multiplicateur de Lagrange associé à la contrainte de non-biais exprimée à l'équation 3.54 et \mathbf{c}_0 est le vecteur des covariances entre le point à estimer et les points d'observations. Ce système est résolu pour chaque point constituant la grille d'estimation.

Ainsi l'approche proposée dans cette thèse substitue un système de coordonnées d'écoulement aux coordonnées cartésiennes habituelles utilisées pour x_i . De plus, la fonction de covariance présentée à la section 3.2 est utilisée pour calculer la matrice \mathbf{C} et le vecteur \mathbf{c}_0 .

Le système de krigeage permet également d'évaluer l'incertitude associée à l'estimateur sous forme de la variance d'estimation, dont la forme générale est donnée par :

$$\sigma_e^2 = E[(Z^* - Z)^2] \quad (3.56)$$

Tel que mentionné plus haut, le krigeage est un estimateur qui minimise l'équation 3.56. Pour le système de krigeage ordinaire, celle-ci correspond à :

$$\sigma_e^2 = \sigma^2 - \boldsymbol{\lambda}' \mathbf{c}_0 - \mu \quad (3.57)$$

3.3.2 Krigeage universel

Le système de krigeage peut être adapté pour tenir compte d'une moyenne non stationnaire. Dans ce cas, l'estimateur prend alors la forme suivante :

$$Z(x) = m(x) + Y(x) \quad (3.58)$$

La variable est alors divisée en deux composantes : la dérive $m(x)$, laquelle tient compte de la partie déterministe du phénomène, et le résidu $Y(x)$, qui représente les fluctuations aléatoires (Chilès et Delfiner, 1999). La dérive peut être exprimée comme la somme de $L + 1$ fonctions de base :

$$m(x) = \sum_{l=0}^L a_l f^l(x) \quad (3.59)$$

où les $f^l(x)$ sont des fonctions de base connues et les a_l sont des coefficients fixes et inconnus. Ce type de krigeage est désigné par krigeage universel. Il est à noter que $l = 0$ correspond au cas où la moyenne est constante sur l'ensemble du domaine, soit le cas du krigeage ordinaire. Dans le cas du krigeage universel, $L + 1$ conditions d'universalité (contraintes) doivent être respectées afin d'assurer le non-biais de l'estimateur. Celles-ci sont de la forme :

$$\sum_{i=1}^n \lambda_i f_i^l = f_0^l \quad (3.60)$$

où les λ_i sont les poids de krigeage, les f_i^l sont les $L + 1$ fonctions évaluées aux points d'observations et les f_0^l sont les $L + 1$ fonctions évaluées au point à estimer.

Le système de krigeage résultant peut être exprimé de façon matricielle :

$$\begin{bmatrix} \mathbf{C} & \mathbf{F} \\ \mathbf{F}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_0 \\ \mathbf{f}_0 \end{bmatrix} \quad (3.61)$$

où \mathbf{C} est la matrice de covariance entre les observations, \mathbf{F} est la matrice contenant la dérive évaluée aux points d'observations⁵, $\boldsymbol{\lambda}$ est le vecteur des poids de krigeage, $\boldsymbol{\mu}$ est le vecteur des $L + 1$ multiplicateurs de Lagrange, \mathbf{c}_0 est le vecteur de covariance entre le point à estimer et les points d'observations et \mathbf{f}_0 est le vecteur de la dérive évaluée au point que l'on cherche à estimer⁶.

5. La première colonne de cette matrice est composée de 1, soit la valeur de f_i^0 .

6. Le premier élément de ce vecteur est 1, soit la valeur de f_0^0 .

À l'exemple 1 du chapitre 4, une moyenne linéaire décroissante de la forme donnée à l'équation suivante sera également testée :

$$m(x) = m_0 - m_1 s \quad (3.62)$$

où m_0 et m_1 sont des constantes et $s = h_{max} - h$ est la diminution de charge hydraulique par rapport à sa valeur maximum sur le domaine. Dans ce cas le système de krigage utilisé est celui donné à l'équation 3.61 et l'on a la correspondance : $m_0 = a_0$, $s = f^1(x)$ et $m_1 = a_i$.

3.4 Normes d'évaluation

En plus de la vraisemblance maximum, du test du rapport de vraisemblance et de AIC_c , on utilise d'autres statistiques pour évaluer la performance des différents modèles de covariance candidats, ainsi que pour les comparer par rapport au krigage en coordonnées cartésiennes. On définit ici les normes d'évaluation utilisées dans les études de cas.

Dans un premier temps, on calcule la racine de la moyenne des erreurs au carré (*root mean square error*, RMSE) à partir des résidus de validation croisée :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i)^2} \quad (3.63)$$

où n est le nombre de données et les e_i sont les résidus issus de la validation croisée. Lors de la validation croisée, chaque donnée est retirée du jeu et estimée à partir des données restantes. La minimisation des résidus de validation croisée peut également être un critère pour l'estimation des paramètres du modèle (Marcotte, 1995).

Il est à noter que les seules statistiques pouvant être calculées pour un cas réel sont la RMSE, le maximum de vraisemblance et l' AIC_c . Par contre, pour les exemples synthétiques présentés aux chapitres 4 et 5, on considère également des statistiques basées sur la comparaison directe des valeurs de concentration de référence et de celles obtenues par krigage. Ceci permet de calculer l'erreur absolue relative sur la masse de contaminant RAE_m :

$$M = \sum_{i=1}^{n_{ele}} n c_i V_i \quad RAE_m = \frac{|M^R - M^*|}{M^R} \quad (3.64)$$

où c_i est la concentration interpolée au centre de l'élément i , n_{ele} est le nombre d'éléments finis, V_i est le volume de l'élément i , M^R et M^* sont respectivement les masses calculées à partir des cartes de référence et krigées, n est la porosité.

On calcule également l'erreur absolue relative entre le volume de contaminant estimé et réel, RAE_{vol} . Ceci est fait à l'aide d'une variable indicatrice qui identifie les éléments dont la valeur (au point central) dépasse un seuil fixé. Le volume contaminé et l'erreur relative sont obtenus d'une manière similaire à l'équation 3.64 :

$$V_s = \sum_{i=1}^{n_{ele}} n I_i V_i \quad RAE_{vol} = \frac{|V_s^R - V_s^*|}{V_s^R} \quad (3.65)$$

où V_s est le volume au-dessus d'un seuil s et I_i est la variable indicatrice pour l'élément i , V_s^R est le volume de référence, V_s^* est le volume krigé et n est la porosité. Par contre, RAE_{vol} n'indique pas si les concentrations estimées sont bien localisées par rapport au vrai panache. Bien que ceci puisse être évalué en comparant visuellement la carte estimée et la carte de référence, on utilise une matrice de confusion (figure 3.5) afin de quantifier ceci de manière plus objective. La matrice de confusion est un outil utilisé dans le domaine de l'apprentissage automatique (*machine learning*) pour évaluer la performance d'un algorithme de classification (Kohavi et Provost, 1998; Kubat *et al.*, 1998). Dans le cas présent, la matrice de confusion contient les sommes des volumes correspondant aux éléments du maillage où :

- a - la concentration est correctement estimée comme étant sous le seuil (vrai négatif)
- b - la concentration n'est pas correctement estimée comme étant au-dessus du seuil (faux positif)
- c - la concentration n'est pas correctement estimée comme étant sous le seuil (faux négatif)
- d - la concentration est correctement estimée comme étant au-dessus du seuil (vrai positif)

		Estimé	
		Non-C.	Contamin.
		a	b
Réel	Non-C.		
	Contamin.	c	d

Figure 3.5 Une matrice de confusion

Plusieurs normes de performance peuvent être calculées à partir de la matrice de confusion. Parmi celles-ci, on retient le ϕ de Pearson, lequel est utilisé pour évaluer la corrélation entre deux variables binaires :

$$\phi_s = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \quad (3.66)$$

Cette statistique permet d'évaluer si les valeurs de concentration estimées sont effecti-

vement situées dans le panache. La valeur maximale pouvant être atteinte par ϕ est 1. Par contre, la valeur maximale de ϕ dépend de la distribution marginale de la variable et une valeur de 1 est obtenue quand les proportions $\frac{a+b}{a+b+c+d}$ et $\frac{a+c}{a+b+c+d}$ sont égales (Warrens, 2008). Certains auteurs suggèrent de diviser ϕ par sa valeur maximale conditionnelle à la distribution marginale afin de corriger ceci (Warrens, 2008; Davenport et El-Sanhurry, 1991). Cette correction n'est pas utilisée dans les résultats présentés aux chapitres suivants.

Finalement, le coefficient de corrélation r est également calculé pour évaluer la ressemblance entre la carte de concentration krigée et la référence :

$$r = \frac{\sum_{i=1}^m (c_{*i} - \bar{c}_*)(c_{Ri} - \bar{c}_R)}{\sqrt{\sum_{i=1}^m (c_{*i} - \bar{c}_*)^2} \sqrt{\sum_{i=1}^m (c_{Ri} - \bar{c}_R)^2}} \quad (3.67)$$

où c_{*i} est la concentration estimée au point i , \bar{c}_* est la moyenne des valeurs de concentration estimées, c_{Ri} est la concentration au point i du modèle de référence, \bar{c}_R est la moyenne des concentrations du modèle de référence.

Aucune des statistiques présentées ici ne permet à elle seule de décrire la qualité de l'estimation. Une faible erreur sur l'estimé de la masse est peu significative si les concentrations sont estimées en dehors du panache de référence (i.e. faible RAE_m combinée à des faibles valeurs de corrélation ϕ). Inversement, un estimé présentant des valeurs élevées de ϕ indique que la méthode permet de bien délimiter la zone contaminée, mais sera d'une utilité moindre en pratique si les valeurs de concentration associées ne permettent pas d'estimer correctement la masse. Ainsi, l'ensemble des statistiques présentées plus haut est nécessaire pour évaluer la performance des différentes approches. L'appréciation visuelle des cartes de concentration obtenues permet de compléter l'analyse des résultats.

3.5 Synthèse

Les outils nécessaires à l'implémentation de l'approche de krigeage non stationnaire de la concentration ont été présentés dans ce chapitre. La figure 3.6 schématisé l'enchaînement des différentes opérations lors de l'application de la méthode proposée.

Dans les chapitres suivants, l'approche est appliquée à des exemples numériques. Dans un premier temps, le chapitre 4 vise à explorer divers aspects liés à la méthode proposée, dans un contexte où l'écoulement est parfaitement connu. Au chapitre 5, l'exemple proposé utilise plutôt des réseaux d'écoulement basés sur des modèles conceptuels simplifiés de l'écoulement pour le calcul des coordonnées, afin de simuler une imprécision sur le modèle d'écoulement.

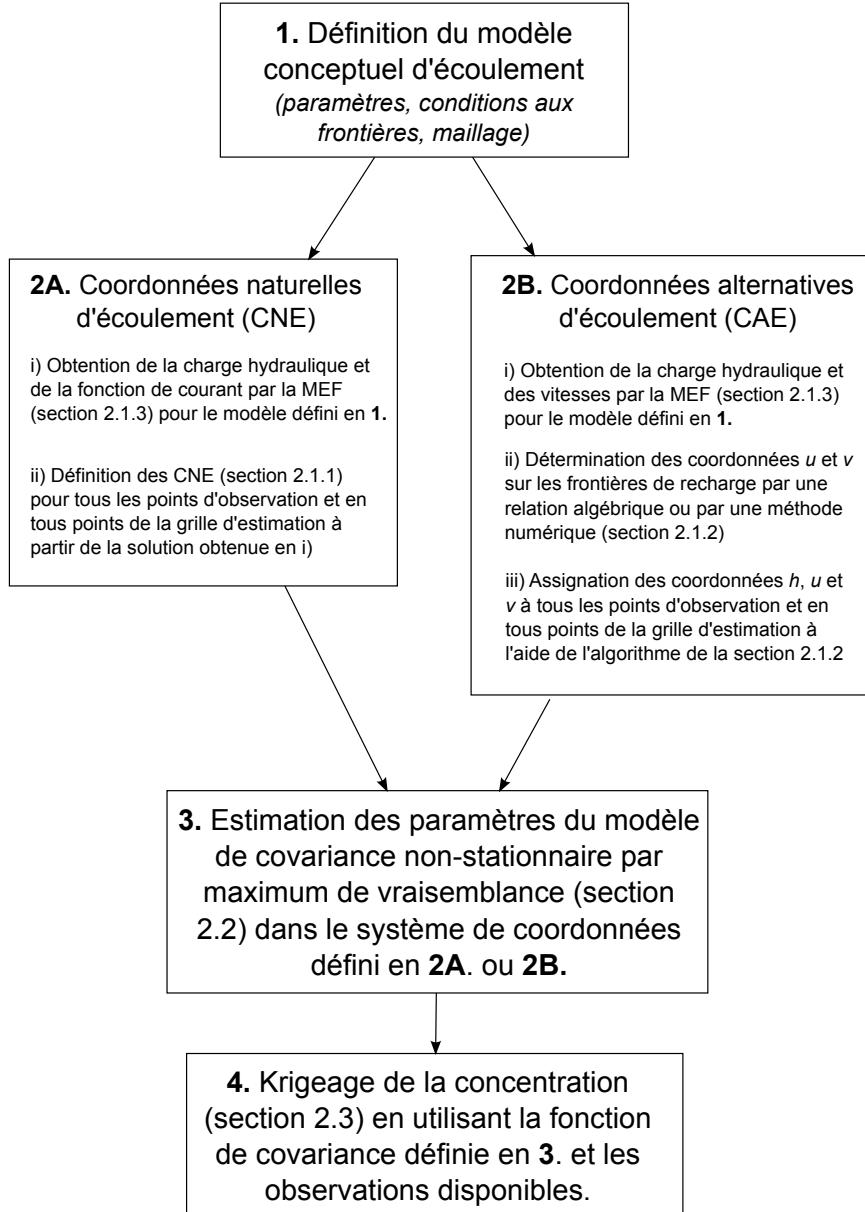


Figure 3.6 Schématisation de l'approche proposée.

CHAPITRE 4

ÉTUDES DE CAS AVEC MODÈLES D'ÉCOULEMENT CONNUS

Le présent chapitre vise à tester la méthode de krigage proposée sur des modèles synthétiques, en considérant que le réseau d'écoulement est connu. Autrement dit, les coordonnées d'écoulement sont calculées à l'aide du modèle d'écoulement de référence. Bien qu'une parfaite connaissance du réseau d'écoulement ne soit pas réaliste en pratique, on peut voir dans les exemples qui suivent un sorte de cas limite ou l'extrémité d'un spectre représentant différents niveaux d'information. Contrairement à un cas réel, l'utilisation de modèles synthétiques permet de quantifier le gain obtenu pour l'estimation de la masse de contaminant présente et du volume contaminé. D'autres statistiques mesurent plutôt la ressemblance entre la carte de concentration krigée et celle de référence (ϕ, r). Les résultats obtenus sont comparés à ceux de l'approche conventionnelle de krigage stationnaire en coordonnées cartésiennes. Il est important de noter que la connaissance exacte du réseau d'écoulement n'est pas une nécessité et qu'en pratique un modèle conceptuel du problème sera utilisé. Ceci est traité au chapitre suivant.

Dans ce qui suit, on présente deux exemples 2D représentant des coupes (exemples 1 et 3), où une grande partie de la trajectoire curviligne des panaches est due à la recharge. Dans le cas de l'exemple 1, des hétérogénéités entraînent des variations de la largeur du panache. À l'exemple 3, une structure moins perméable entraîne la division du panache en deux doigts distincts. L'exemple 2 présente un modèle 3D, où la direction de migration du panache varie latéralement et verticalement. Dans tout les cas, les modèles représentent des aquifères à nappe semi-captive (exemples 1 et 3) ou captive (exemple 2), en marge d'une ligne de partage des eaux. Les panaches sont produits par une source continue située sur une des frontières du modèle et le nombre d'observations disponibles se veut limité. Les exemples 1 et 2 sont des versions plus détaillées de ceux présentés dans Rivest et Marcotte (2012). En plus de servir à illustrer la performance de la méthode proposée, tous les exemples présentés ci-dessous visent des objectifs spécifiques énoncés au début de chaque section.

Un exemple simple tiré de Rivest *et al.* (2012b), où le modèle est composé de deux unités homogènes, est également présenté à l'Annexe A. Cet exemple utilise uniquement la transformation de coordonnée et illustre les avantages et les inconvénients de celle-ci, pour diverses configurations données. De plus, l'Annexe E détaille les paramètres géostatistiques utilisés

pour tous les exemples présentés dans ce chapitre.

4.1 Exemple 1. Cas 2D

Ce premier exemple 2D vise particulièrement à :

1. illustrer l'impact des paramétrisations non stationnaires et de la transformation de coordonnées sur la fonction de covariance ;
2. comparer l'utilisation des coordonnées naturelles d'écoulement (CNE) et les coordonnées alternatives (CAE_h) ;
3. tester l'utilisation d'une dérive linéaire décroissante pour tenir compte d'une moyenne non stationnaire ;
4. montrer l'effet de l'utilisation combinée de la transformation de coordonnées et des paramétrisations de la covariance sur la variance d'estimation.

4.1.1 Description du modèle de référence

La figure 4.1(a) présente le champ de conductivité hydraulique du modèle de référence, lequel comporte deux unités. L'Annexe B détaille la technique utilisée pour construire les champs de conductivité hydraulique utilisés ici et dans les exemples suivants. La méthode des éléments finis est utilisée afin d'obtenir le champ de charge hydraulique, la fonction de courant et le champ de vitesses, lesquels sont requis pour calculer les coordonnées d'écoulement. Le panache de référence est également simulé par MEF, pour un modèle de transport comprenant une source située en $7 \text{ m} \leq x \leq 52 \text{ m}$, sur la frontière $y = 50 \text{ m}$. On choisit comme référence une solution transitoire ($t \approx 29 \text{ ans}$)¹ et celle-ci est échantillonnée sur une grille régulière de 12 points (voir figure 4.1(b)), répartis sur quatre puits d'observation fictifs. Le tableau 4.1 résume les différents paramètres d'écoulement et de transport du modèle. On note que la valeur de dispersivité transversale utilisée ici est relativement élevée pour un système naturel et que cette valeur a été utilisée afin d'éviter l'utilisation d'un maillage très fin. Des valeurs similaires sont retrouvées dans certaines études stochastiques pour des modèles synthétiques de dimensions similaires, par exemple Nowak *et al.* (2010) et de Barros et Nowak (2010).

La figure 4.2(a) montre la fonction de courant (ψ) et les lignes de courant correspondant au modèle d'écoulement. L'approche combinant le traçage des lignes de courant et l'interpolation linéaire décrite au chapitre précédent est également utilisée pour obtenir la coordonnée u

1. On aurait également pu utiliser la solution en régime permanent.

présentée à la figure 4.2(b). Malgré que l'on observe localement des différences entre les deux, u reproduit les caractéristiques principales de ψ et semble être une approximation raisonnable pour cette dernière. En fait, la corrélation entre u et ψ est de 0.998. Il est à noter que les échelles des valeurs de ces deux variables sont très différentes ; ceci est pris en compte par les paramètres de portée lors du krigage. Pour l'exemple 1, la densité de lignes de courant utilisées pour l'obtention de u est de une ligne à tous les 10 m, sur les frontières $x = 0$ et $y = 50$. L'Annexe C montre que l'utilisation d'un espacement de 5 m des lignes de courant produit des résultats comparables à ceux avec l'espacement de 10 m.

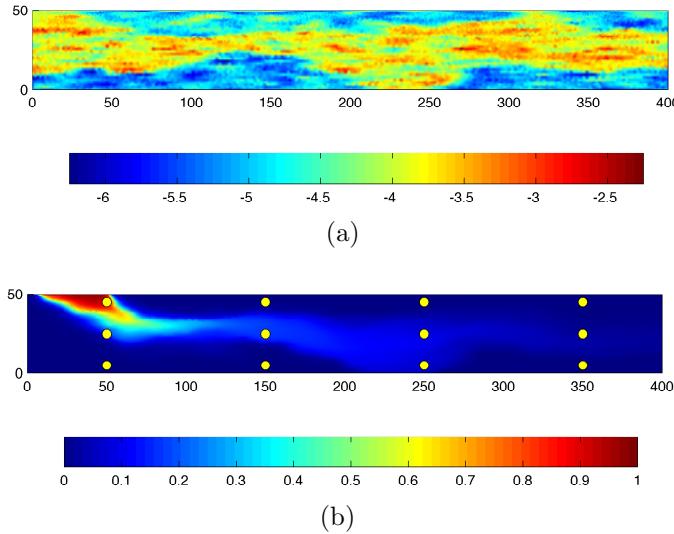


Figure 4.1 Exemple 1. Modèle de référence (a) Champ de conductivité hydraulique ($\log_{10}K$), (b) Concentrations de référence.

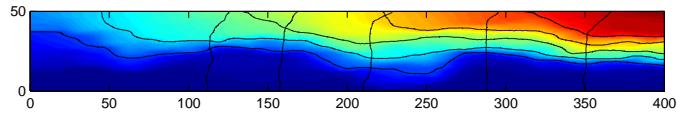
4.1.2 Aperçu des fonctions de covariance non stationnaires

Le tableau 4.2 résume les modèles de covariance candidats ayant été présentés au chapitre 3. Ces modèles ont des impacts différents sur la covariance de la concentration. Afin d'illustrer ceci on montre la fonction de covariance pour différentes paramétrisations, en trois points situés (50,50), (150,25) et (350,25). Les figures 4.3 et 4.4 illustrent les modèles de covariance NS1 et NS2 en coordonnées cartésiennes, où l'on utilise x comme coordonnée longitudinale pour modéliser la variation spatiale des paramètres. Pour le modèle NS1, le palier de la covariance est fixe, mais la distance pour laquelle les points sont corrélés augmente avec x . Pour le modèle NS2, le palier de la covariance diminue, mais la portée demeure constante².

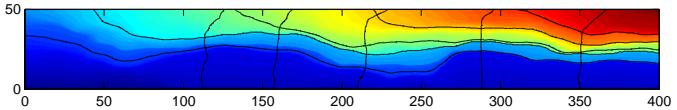
2. Ceci n'est pas visible sur la figure 4.4(c) en raison des valeurs atténuees de covariance lesquelles donnent l'impression que la distance de corrélation a diminué.

Tableau 4.1 Exemple 1. Paramètres du modèle synthétique

Paramètre	Valeur
Géométrie	
Dimensions	400 m \times 50 m
Taille max. éléments	2 m
Paramètres d'écoulement	
Moyenne géométrique K	5.4×10^{-5} m/s
Variance de $\ln(K)$	2.13
Gradient hydraulique moyen	0.00075
Conditions frontières	
h en $x = 0$	51 m
h en $x = 400$	50.7 m
Recharge ($y = 50$)	1.25×10^{-8} m/s
q_n en $y = 0$	nul
Paramètres de transport	
porosité	0.35
D_m	1.34×10^{-9} m ² /s
α_l, α_{tv}	5 m, 0.25 m



(a)



(b)

Figure 4.2 Exemple 1. Comparaison des coordonnées naturelles et alternatives d'écoulement. (a) Fonction de courant ψ , (b) Coordonnée alternative u . Les lignes de courant et les équipotentielles sont également représentées.

De son côté, la figure 4.5 montre que, sans égard à la fonction de covariance utilisée (stationnaire ou non), le passage aux coordonnées d'écoulement a lui-même un impact sur la

stationnarité. Cette figure présente le modèle S1 en coordonnées d'écoulement et montre ainsi que ce modèle de covariance, pourtant stationnaire dans l'espace des coordonnées d'écoulement, résulte en une covariance non stationnaire lorsque représenté dans l'espace cartésien. Ceci illustre comment la transformation de coordonnées permet à la fonction de covariance de se conformer à l'anisotropie variable associée au réseau d'écoulement de la figure 4.2. Cette anisotropie se manifeste autant par la prise en compte des directions d'écoulement, que des variations de taille des tubes de courant.

Tableau 4.2 Résumé des modèles candidats pour la paramétrisation du modèle de covariance non stationnaire. A : Anisotrope, S : Stationnaire, NS : Non stationnaire

Modèle	Portée	Variance
S1	A-S	S
NS1	A-NS	S
NS2	A-S	NS
NS3	A-NS	NS

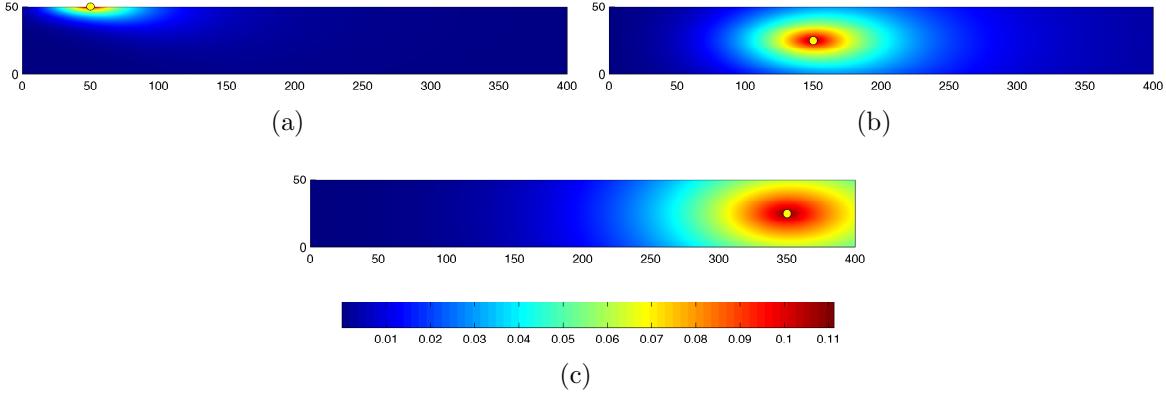


Figure 4.3 Fonction de covariance NS1 en coordonnées cartésiennes (CC/NS1) pour les points (a) (50,50), (b) (150,25) et (c) (350,25).

4.1.3 Sélection du modèle de covariance

Bien que tous les modèles candidats soient des simplifications du modèle sous-jacent (et inconnu), la question est de savoir lequel d'entre-eux est le mieux supporté par les données disponibles. Les tableaux 4.3 à 4.5 présentent les résultats du test de vraisemblance ($\alpha = 0.05$) et l'AIC_c pour les quatre paramétrisations candidates de la fonction de covariance, et ce pour les trois systèmes de coordonnées : coordonnées cartésiennes (CC), coordonnées naturelles

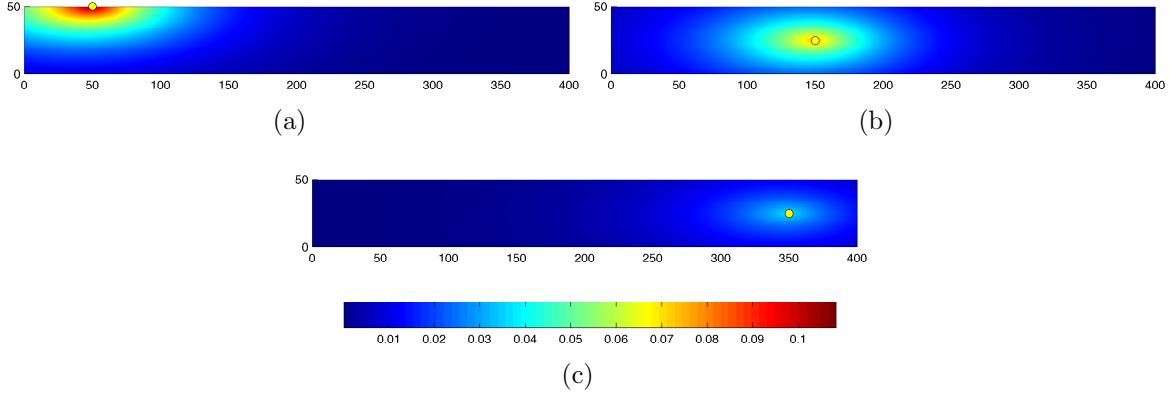


Figure 4.4 Fonction de covariance NS2 en coordonnées cartésiennes (CC/NS2) pour les points (a) (50,50), (b) (150,25) et (c) (350,25).

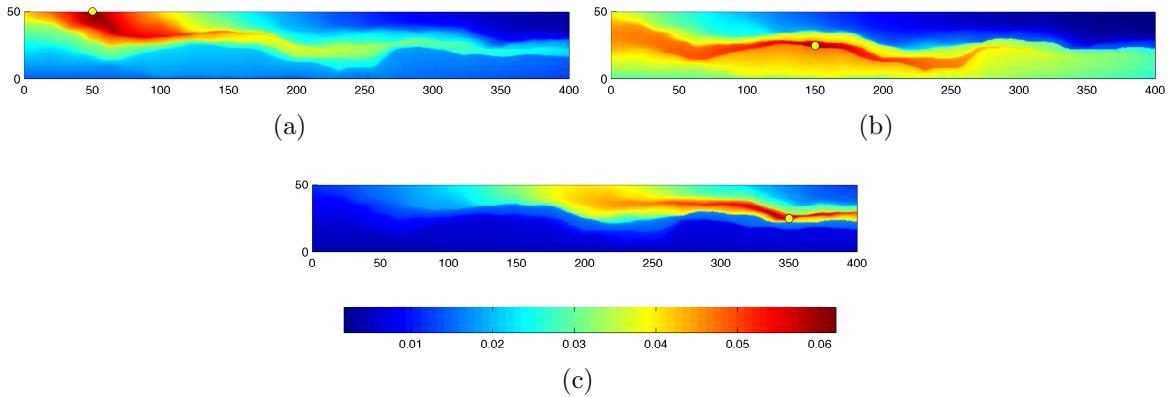


Figure 4.5 Fonction de covariance S1 en coordonnées alternatives d’écoulement (CAE_h/S1) pour les points (a) (50,50), (b) (150,25) et (c) (350,25).

d’écoulement (CNE) et coordonnées alternatives d’écoulement (CAE_h).

Tel qu’attendu, les modèles qui utilisent des coordonnées d’écoulement (tableaux 4.4 et 4.5) sont associés à des valeurs de vraisemblance supérieures à celles des modèles définis en coordonnées cartésiennes. Pour chaque système de coordonnées, la vraisemblance augmente avec la complexité du modèle et le modèle NS3 est celui ayant une vraisemblance supérieure. Dans tous les cas, ce gain en vraisemblance ne justifie pas l’emploi du modèle NS3 par rapport au modèle NS2. Ce résultat est également confirmé par les valeurs de l’AIC_c, pour lesquelles le modèle NS2 obtient les valeurs les plus basses.

Ces résultats sont encourageants car ils montrent d’une part, que les modèles en coor-

données d'écoulement ont des vraisemblances supérieures aux modèles définis en coordonnées cartésiennes, et d'autre part, que leur utilisation conjointe avec les modèles de covariance non stationnaires engendre également une augmentation significative de la vraisemblance. Ceci semble indiquer que le modèle de covariance combinant une transformation de coordonnée et une diminution de la variance (NS2) permet de mieux expliquer les données de concentration.

Tableau 4.3 Exemple 1. Maximum de vraisemblance (MV), *p-value* du test du rapport de vraisemblance ($\alpha = 0.05$) et AIC_c pour les modèles en **coordonnées cartésiennes**. k est le nombre de paramètres estimés pour chaque modèle.

Modèle	k	MV	H_0	<i>p-value</i>	AIC_c
CC/S1	3	0.94	-	-	9.31
CC/NS1	5	2.77	CC/S1	0.34	17.98
CC/NS2	4	143.51	CC/S1	0.0015	3.80
CC/NS3	6	448.10	CC/NS2	0.32	16.66

Tableau 4.4 Exemple 1. Maximum de vraisemblance (MV), *p-value* du test du rapport de vraisemblance ($\alpha = 0.05$) et AIC_c pour les modèles en **coordonnées naturelles d'écoulement (CNE)**. k est le nombre de paramètres estimés pour chaque modèle.

Modèle	k	MV	H_0	<i>p-value</i>	AIC_c
CNE/S1	3	1.23	-	-	9.56
CNE/NS1	5	4.19	S1	0.29	17.14
CNE/NS2	4	227.38	S1	0.0012	2.87
CNE/NS3	6	1578.80	NS2	0.14	14.07

Tableau 4.5 Exemple 1. Maximum de vraisemblance (MV), *p-value* du test du rapport de vraisemblance ($\alpha = 0.05$) et AIC_c pour les modèles en **coordonnées alternatives d'écoulement (CAE_h)**. k est le nombre de paramètres estimés pour chaque modèle.

Modèle	k	MV	H_0	<i>p-value</i>	AIC_c
CAE _h /S1	3	2.24	-	-	7.39
CAE _h /NS1	5	13.03	S1	0.17	14.87
CAE _h /NS2	4	219.00	S1	0.003	2.94
CAE _h /NS3	6	1914.83	NS2	0.11	13.69

4.1.4 Estimation de la concentration

Le tableau 4.6 présente les statistiques obtenues en comparant³ les valeurs de concentration de référence aux valeurs krigées, pour les différents modèles de covariance. On note d'abord qu'il n'y a pas de différences marquées dans les valeurs de RMSE. Peu importe le système de coordonnées considéré, les résultats montrent également que les modèles non stationnaires engendrent une amélioration de l'estimation, en particulier en ce qui a trait à l'erreur relative sur la masse et le volume (RAE_m , RAE_{vol}), ainsi que le ϕ de Pearson. L'amélioration est moins marquée lorsqu'on compare un même modèle de covariance entre les différents systèmes de coordonnées. Dans ce cas, les modèles en coordonnées d'écoulement produisent des résultats légèrement meilleurs que ceux en coordonnées cartésiennes. Par exemple, pour le modèle NS2 sélectionné précédemment, il n'y a pas une amélioration systématique de chaque statistique lorsqu'on passe aux coordonnées d'écoulement (CNE ou CAE_h). Toutefois, l'utilisation des coordonnées d'écoulement engendre une amélioration systématique des valeurs de ϕ ; lequel est un indicateur de la ressemblance entre les panaches krigés et de référence. Il indique notamment que les valeurs estimées délimitent mieux la zone occupée par le panache, pour les seuils indiqués.

Les modèles de covariance non stationnaires permettent également d'améliorer l'estimation en coordonnées cartésiennes. Toutefois, si l'on compare les résultats des modèles CNE/NS2 et CAE_h /NS2 avec ceux du modèle de krigeage conventionnel CC/S1, l'amélioration est importante. Par exemple, avec le modèle CNE/NS2, on améliore l'estimation du volume de 50% pour le seuil 0.01 et de 15% pour le seuil 0.1, et celle de la masse d'un peu plus que 10 %. On note par ailleurs que la performance du modèle CC/S1 est bonifiée par l'utilisation de paramètres *ad hoc*; la différence (à l'avantage de CNE/NS2) aurait été encore plus marquée avec les paramètres estimés par maximum de vraisemblance.

Visuellement, la différence avec les modèles en coordonnées cartésiennes est appréciable. La figure 4.6 montre les cartes de concentration obtenues à l'aide du modèle stationnaire S1 et du modèle sélectionné NS2, pour les trois systèmes de coordonnées. Malgré les faibles différences observées précédemment pour certaines statistiques comparatives, les cartes de concentrations krigées montrent clairement l'amélioration engendrée par le recours à des coordonnées d'écoulement. Les figures 4.6(c), 4.6(e), 4.6(d) et 4.6(f) illustrent bien comment la transformation permet de tenir compte de l'anisotropie locale induite par la recharge et ainsi d'améliorer le réalisme des valeurs de concentration estimées. On remarque que les modèles

3. Sauf la RMSE, qui est obtenue par validation croisée et ne dépend pas de la connaissance du modèle de référence.

stationnaires CNE/S1 et $\text{CAE}_h/\text{S1}$ produisent des panaches qui s'étirent excessivement dans la direction de l'écoulement. Ceci n'est pas retrouvé sur les cartes produites avec le modèle de covariance NS2 (figures 4.6(d) et 4.6(f)) et démontre l'intérêt du modèle non stationnaire NS2.

De plus, la figure 4.6 permet d'apprécier la ressemblance entre les cartes issues des modèles en coordonnées naturelles et celles obtenues à l'aide des coordonnées approximatives CAE_h . Cette ressemblance semble indiquer que les CAE_h , basées sur le trajectographie, sont effectivement une bonne alternative aux CNE. En effet, les CNE sont uniquement applicables dans les cas où la fonction de courant peut être définie. Ainsi, elles ne sont pas applicables en 3D, ni aux modèles présentant des puits.

Finalement, bien que le modèle sélectionné (NS2) présente une amélioration par rapport aux modèles S1 et NS1, ce dernier n'apparaît pas clairement supérieur au modèle NS3 d'après les statistiques colligées. La figure 4.7 montre que les cartes estimées à l'aide de NS3 présentent un panache plus étroit près de la source, ce qui semble se refléter positivement sur l'erreur sur la masse et l'estimation du volume pour le seuil plus élevée (0.1). On remarque également une différence plus importante entre la carte obtenue avec les coordonnées naturelles et celle utilisant les coordonnées alternatives, pour ce modèle de covariance. Dans le cas présent, cette différence est attribuable aux paramètres estimés et non à la transformation elle-même, et pourrait être due à l'atteinte d'un minimum local pour $-\log(\mathcal{L})$.

4.1.5 Utilisation d'une dérive linéaire

Les résultats précédents ont été obtenus par krigeage ordinaire, lequel suppose une moyenne non conditionnelle constante (stationnaire). On présente ici les résultats obtenus par krigeage universel en utilisant la coordonnée longitudinale comme fonction de base. La moyenne est donc modélisée comme un plan décroissant dans le sens de l'écoulement. On se limite ici à montrer les effets de la dérive pour le krigeage en coordonnées naturelles d'écoulement (CNE).

L'utilisation de la dérive entraîne l'ajout de deux paramètres à chaque modèle : la constante et la pente associée à la coordonnée longitudinale (équation 3.62). Le tableau 4.7 montre que la dérive engendre une augmentation de la vraisemblance pour chacun des modèles (comparer avec le tableau 4.4). Tel que précédemment, c'est le modèle NS2 qui est sélectionné. Lorsqu'on compare la vraisemblance obtenue pour ce modèle avec celle du même modèle sans dérive, le LRT et l'AIC_c divergent. L'ajout des paramètres engendre une augmentation significative selon le LRT mais l'AIC_c n'est pas inférieur à celui du modèle sans dérive.

Tableau 4.6 Exemple 1. Comparaison des résultats pour les différentes normes d'évaluation. *En gris* : Les modèles retenus par la procédure de sélection pour chaque système de coordonnées. c_s : seuil de contamination.

Modèle	RMSE	RAE_m	RAE_{vol}		ϕ		r
			$c_s = 0.01$	$c_s = 0.1$	$c_s = 0.01$	$c_s = 0.1$	
CC/S1 ^a	0.2653	0.6174	0.6677	0.4372	0.07423	0.5778	0.6849
CC/NS1	0.2685	0.4237	0.6496	0.5048	0.1479	0.3796	0.6793
CC/NS2	0.2576	0.4247	0.5706	0.1732	0.3090	0.4486	0.7107
CC/NS3	0.2590	0.3766	0.5629	0.1157	0.3154	0.4655	0.7084
CNE/S1	0.2614	0.6331	0.5243	0.6335	0.3301	0.5761	0.7665
CNE/NS1	0.2510	0.4594	0.2658	0.3202	0.4732	0.6747	0.7716
CNE/NS2	0.2360	0.4715	0.1681	0.2891	0.6438	0.6564	0.7789
CNE/NS3	0.2520	0.01459	0.2309	0.09750	0.5185	0.6776	0.7298
CAE _h /S1	0.2621	0.8057	0.5545	0.5686	0.2996	0.5651	0.7359
CAE _h /NS1	0.2580	0.3241	0.4030	0.4240	0.3588	0.5822	0.7090
CAE _h /NS2	0.2363	0.5058	0.2967	0.3490	0.5425	0.6288	0.7832
CAE _h /NS3	0.2446	0.1976	0.1797	0.0169	0.6145	0.6889	0.7933

^a Des paramètres *ad hoc* ont été utilisés pour le modèle CC/S1 puisque l'estimation par maximum de vraisemblance a fourni des paramètres de portée excessivement petits. Ainsi, les paramètres utilisés pour ce modèle ne sont pas ceux associés au maximum de vraisemblance.

Le tableau 4.8 compare les résultats des statistiques pour les modèles CNE/NS2 avec et sans dérive. L'ajout de la dérive entraîne une diminution de la moitié de l'erreur sur la masse. Le volume est quant à lui moins bien estimé pour le modèle avec dérive. De même, les corrélations ϕ et r sont inférieures lorsque la dérive est utilisée. Ainsi, malgré l'augmentation de la vraisemblance, on ne note pas une amélioration marquée au niveau des autres statistiques.

La figure 4.8 montre que la dérive utilisée a un effet indésirable sur les estimés de concentration. Pour le modèle NS2, celle-ci a tendance à engendrer l'estimation de valeurs trop fortes loin des données, dans la partie amont du modèle (voir figure 4.8(b)). Ceci est plus marqué sur la figure 4.8(a) qui correspond à l'estimé fait avec le modèle S1. Inversement, le modèle NS3 semble atténuer ceci (figure 4.8(c)).

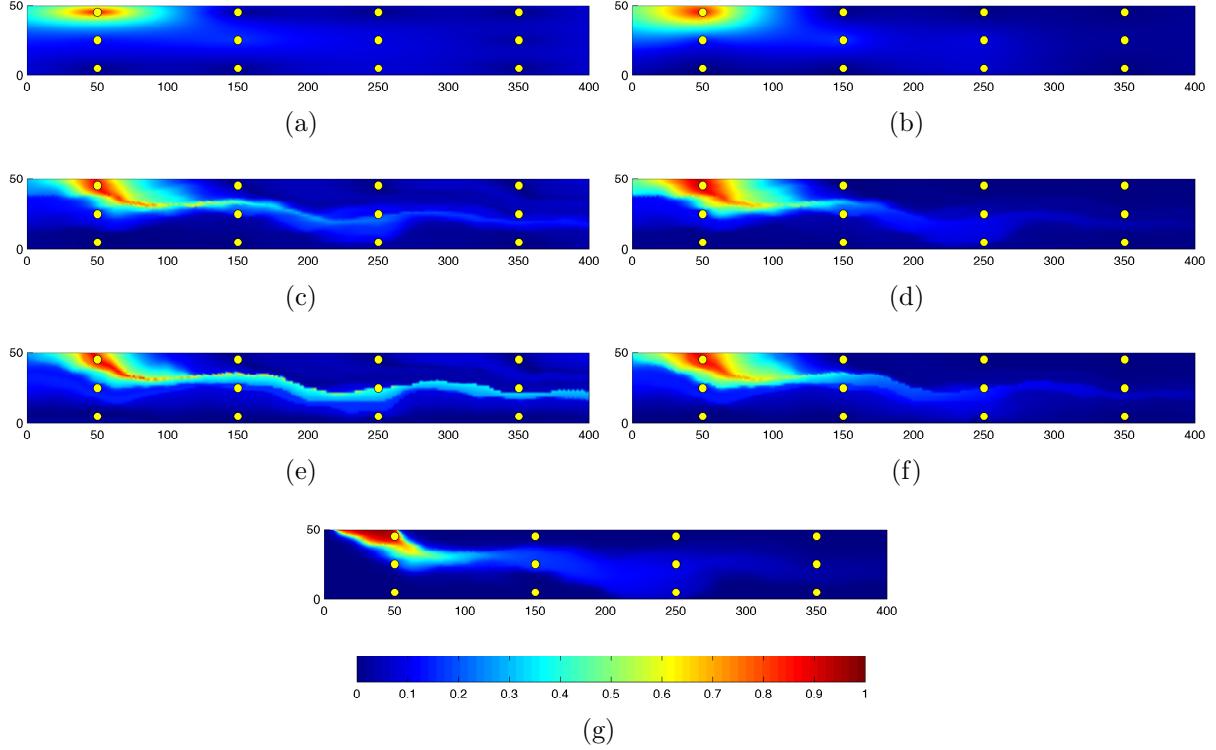


Figure 4.6 Cartes de concentration krigées et de référence pour l'exemple 1. (a) CC/S1, (b) CC/NS2, (c) CNE/S1, (d) CNE/NS2, (e) CAE_h/S1, (f) CAE_h/NS2 et (g) Concentrations de référence.

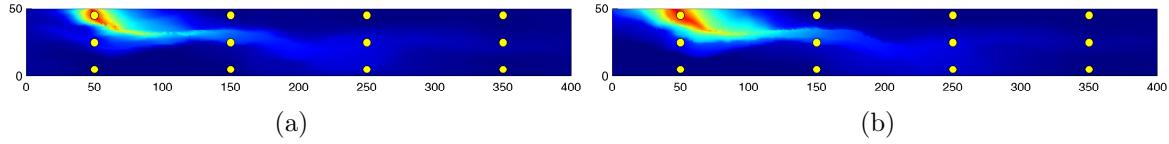


Figure 4.7 Exemple 1. Résultats du modèle NS3 en coordonnées d'écoulement. (a) CNE/NS3, (b) CAE_h/NS3.

Tableau 4.7 Exemple 1. Maximum de vraisemblance MV, *p-value* du test du rapport de vraisemblance ($\alpha = 0.05$) et AIC_c pour les modèles en **coordonnées naturelles d'écoulement (CNE) avec dérive**. k est le nombre de paramètres estimés pour chaque modèle.

Modèle	k	MV	H_0	<i>p-value</i>	AIC _c
CNE/NS2 sans dérive	4	227.38	-	-	2.8
CNE/S1	5	6.31	-	-	16.32
CNE/NS1	7	38.36	S1	0.17	34.71
CNE/NS2	6	75612.00	S1	1.47×10^{-5}	6.33
CNE/NS2	6	75612.00	NS2 sans dérive	0.003	6.33
CNE/NS3	8	148247.92	NS2	0.51	40.17

Tableau 4.8 Exemple 1. Comparaison des résultats pour le modèle CNE/NS2 avec et sans dérive

Modèle	RMSE	EAR_m	EAR_{vol}				ϕ	r
			$c_s = 0.01$	$c_s = 0.1$	$c_s = 0.01$	$c_s = 0.1$		
SANS dérive	0.2360	0.4715	0.1681	0.2891	0.6438	0.6564	0.7789	
AVEC dérive	0.2287	0.2261	0.2931	0.3372	0.3514	0.5396	0.7538	

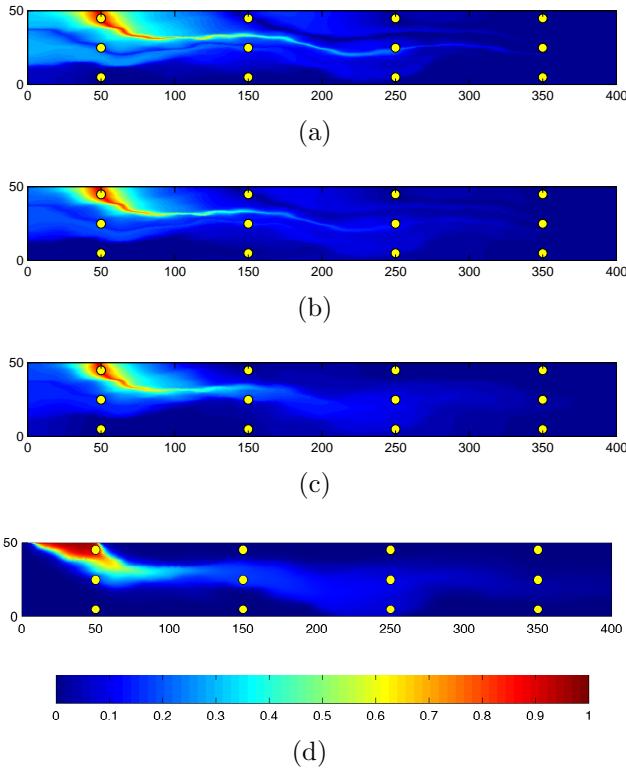


Figure 4.8 Cartes de concentration krigées pour l'exemple 1 avec dérive linéaire. (a) CNE/S1, (b) CNE/NS2, (c) CNE/NS3, (d) Concentrations de référence

4.1.6 Impacts sur la variance de krigeage

Pour conclure l'exemple 1, on s'intéresse à l'effet combiné de la transformation de coordonnée et des modèles de covariance non stationnaires sur la variance d'estimation, ou variance de krigeage. Comme pour la covariance (figures 4.3 à 4.5), les diverses paramétrisations proposées ont des conséquences différentes sur la variance d'estimation.

La figure 4.9 présente les cartes de l'écart-type d'estimation, σ_e , pour certains des estimés de la section 4.1.4. Pour le modèle CAE_h/S1 (figure 4.9(a)), de faibles valeurs de σ_e se concentrent le long des tubes de courant qui contiennent les observations. Inversement, des valeurs de σ_e élevées reflètent l'absence de données le long des lignes de courant sous-jacentes. Ceci est également retrouvé sur les cartes associées aux modèles NS2 et NS3 (figures 4.9(b) et 4.9(c)). Pour ces deux modèles, on observe aussi une diminution longitudinale de l'incertitude. Pour le modèle NS2, ceci est dû uniquement à la paramétrisation décroissante de la variance, alors que pour le modèle NS3 la décroissance de σ_e est davantage marquée, puisqu'elle est due aux effets combinés de l'augmentation des portées et de la diminution de la variance. Avec ces modèles, la variance d'estimation reflète donc une partie de l'incertitude liée à la position de la source. Les écart-types d'estimation obtenus avec les modèles en coordonnées cartésiennes sont également fournis à titre comparatif (figures 4.9(d) à 4.9(f)).

Il est à noter que pour une source de contamination continue, où les concentrations les plus élevées sont nécessairement retrouvées en amont dans le réseau d'écoulement, les modèles NS2 et NS3 ont pour effet de relier des valeurs élevées de variance d'estimation à des valeurs élevées de concentration, et inversement. Ainsi, la variance d'estimation hérite donc de la non-stationnarité modélisée par la transformation de coordonnées et de la fonction de covariance non stationnaire et σ_e^2 ne dépend pas uniquement de la localisation des données et du modèle de covariance (gaussien, exponentiel, etc).

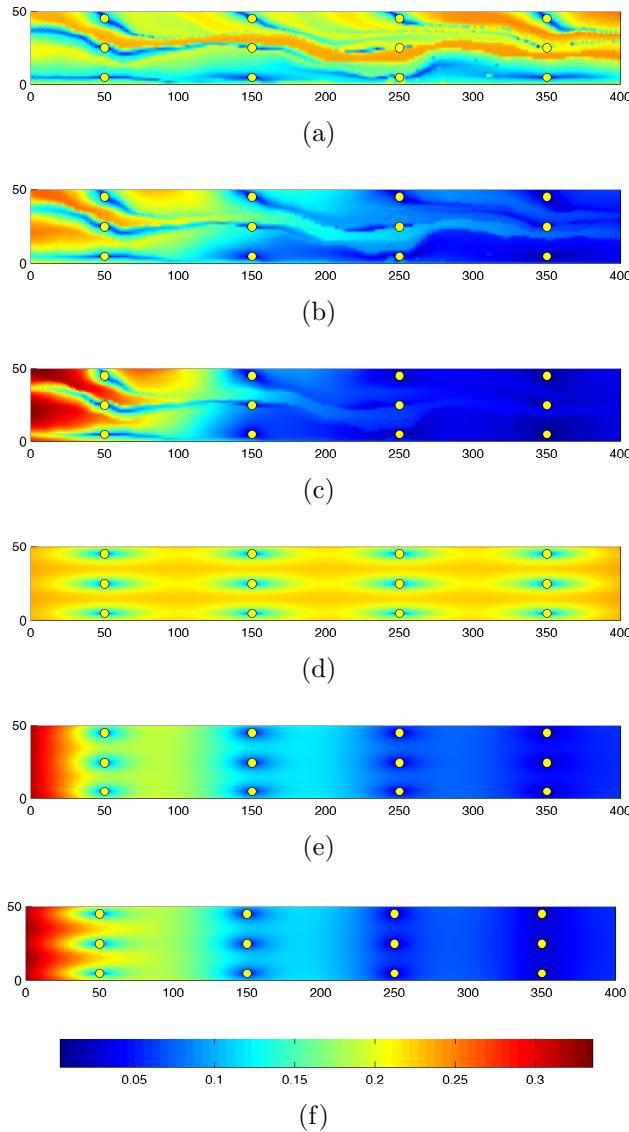


Figure 4.9 Écarts-types d'estimation associés à certains résultats de krigeage de l'exemple 1.
 (a) CAE_h/S1, (b) CAE_h/NS2, (c) CAE_h/NS3, (d) CC/S1, (e) CC/NS2, (f) CC/NS3.

4.2 Exemple 2. Cas 3D

Dans ce qui suit, on applique le krigeage non stationnaire à un cas synthétique 3D. L'exemple précédent a montré que les coordonnées d'écoulement approximatives permettent d'obtenir des résultats similaires à ceux obtenus en utilisant les coordonnées naturelles, h et ψ . Cet exemple a donc pour objectif spécifique de démontrer l'utilisation des coordonnées alternatives d'écoulement (CAE_h) en 3D.

4.2.1 Description du modèle de référence

La figure 4.10(a) présente le champ de conductivité hydraulique du modèle 3D. Un panache est généré par la libération constante d'un contaminant sur la frontière de gauche à partir d'une zone définie par $40 \text{ m} \leq y \leq 70 \text{ m}$ et $20 \text{ m} \leq z \leq 35 \text{ m}$ et celui-ci est échantillonné en 24 points (figure 4.10(b)). En ce qui concerne la densité d'échantillonnage et la taille du modèle, cet exemple se compare au *small test case* présenté dans les travaux de Reed *et al.* (2004). Les paramètres du modèle d'écoulement et de transport sont résumés au tableau 4.9. La figure 4.11 illustre les coordonnées alternatives d'écoulement en montrant un exemple de surfaces h , u et v .

Tableau 4.9 Exemple 2. Paramètres du modèle synthétique

Paramètre	Valeur
Géométrie	
Dimensions	$200 \text{ m} \times 100 \text{ m} \times 50 \text{ m}$
Taille max. éléments	7 m
Paramètres d'écoulement	
Moyenne géométrique K	$2.3 \times 10^{-6} \text{ m/s}$
Variance de $\ln(K)$	6.3
Gradient hydraulique moyen	0.0016
Conditions frontières	
q_n en $x = 0$	$1 \times 10^{-8} \text{ m/s}$
h en $x = 200$	50.7 m
q_n front. latérales	nul
Paramètres de transport	
Porosité	0.35
D_m	$1.34 \times 10^{-9} \text{ m}^2/\text{s}$
$\alpha_l, \alpha_{th}, \alpha_{tv}$	5 m, 0.25 m, 0.25 m

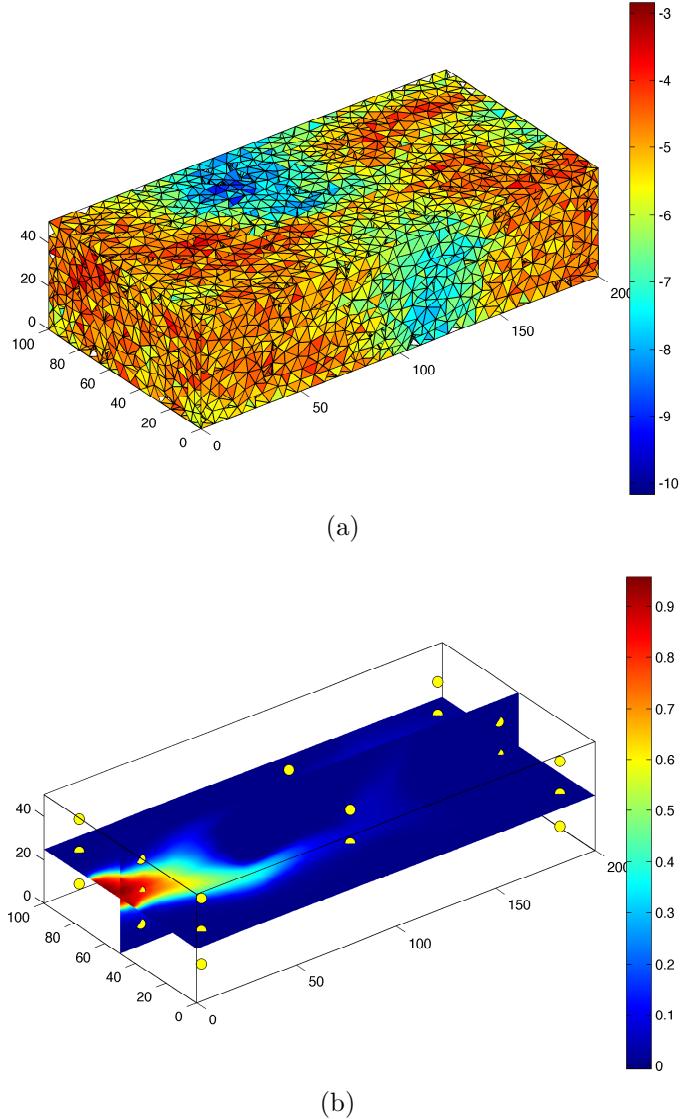


Figure 4.10 Exemple 2. Modèle de référence (a) Champ de conductivité hydraulique ($\log_{10}K$)
(b) Concentration de référence.

4.2.2 Sélection du modèle de covariance

Les tableaux 4.10 et 4.11 présentent les résultats du test de vraisemblance et de l’AIC_c pour les modèles en coordonnées cartésiennes et en CAE_h, respectivement. D’abord, le test de vraisemblance montre que les modèles avec des portées transversales différentes ne sont pas significatifs par rapport aux modèles plus parcimonieux où les portées transversales sont considérées égales (les modèles NSX*). Ce résultat est en conformité avec le modèle de transport de référence, lequel comporte des dispersivités transversales, α_{th} et α_{tv} , égales (voir tableau 4.9).

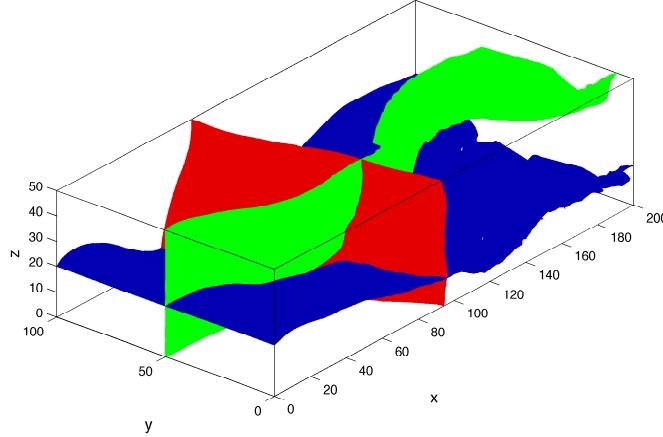


Figure 4.11 Exemple 2. Surfaces h (rouge), u (vert) et v (bleu). u et v sont obtenues par l'approche combinée de traçage des lignes de courant et d'interpolation linéaire présentée au chapitre 3. h est obtenu à partir de la solution numérique d'écoulement.

De plus, à l'instar des résultats de l'exemple 1, les valeurs du maximum de vraisemblance sont plus élevées pour les modèles définis en coordonnées d'écoulement. Toutefois, contrairement au cas 2D précédent, les paramètres ajoutés au modèle NS3*, par rapport au modèle NS2*, sont jugés significatifs par le LRT, et ce, autant en coordonnées cartésiennes qu'en CAE _{h} . Ce résultat est également confirmé par l'AIC _{c} , laquelle est la plus faible pour le modèle NS3*.

Tableau 4.10 Exemple 2. Maximum de vraisemblance (MV), *p-value* du test du rapport de vraisemblance ($\alpha = 0.05$) et AIC _{c} pour les modèles en **coordonnées cartésiennes**. k est le nombre de paramètres estimés pour chaque modèle.

Modèle	k	MV	H_0	<i>p-value</i>	AIC _{c}
CC/S1*	3	6.90×10^3	-	-	-10.48
CC/S1	4	6.90×10^3	S1*	0.98	-7.57
CC/NS1*	5	3.39×10^5	S1*	0.02	-12.14
CC/NS1	7	6.73×10^5	NS1*	0.50	-5.84
CC/NS2*	4	2.15×10^{10}	S1*	3×10^{-7}	-37.47
CC/NS2	5	6.52×10^{10}	NS2*	0.14	-36.47
CC/NS3*	6	4.45×10^{13}	NS2*	0.004	-45.91
CC/NS3	8	1.30×10^{14}	NS3*	0.34	-39.39

Tableau 4.11 Exemple 2. Maximum de vraisemblance maximum (MV), *p-value* du test du rapport de vraisemblance ($\alpha = 0.05$) et AIC_c pour les modèles en **coordonnées alternatives d’écoulement**. k est le nombre de paramètres estimés pour chaque modèle.

Model	k	ML	H_0	<i>p-value</i>	AIC_c
$CAE_h/S1^*$	3	1.3915×10^4	-	-	-11.88
$CAE_h/S1$	4	1.4169×10^4	S1*	0.85	-9.01
$CAE_h/NS1^*$	5	1.6909×10^6	S1*	0.008	-15.34
$CAE_h/NS1$	7	4.7368×10^6	NS1*	0.36	-9.74
$CAE_h/NS2^*$	4	8.5317×10^{10}	S1*	2.26×10^{-8}	-40.23
$CAE_h/NS2$	5	2.1510×10^{11}	NS2*	0.17	-38.86
$CAE_h/NS3^*$	6	8.0922×10^{13}	NS2*	0.001	-47.11
$CAE_h/NS3$	8	2.2044×10^{14}	NS3*	0.37	-40.45

4.2.3 Estimation de la concentration

Comme dans l'exemple précédent, l'utilisation des coordonnées d'écoulement n'entraîne pas une amélioration systématique de toutes les statistiques colligées, comme le montre le tableau 4.12. Toutefois, à l'exception du modèle S1, tous les modèles en coordonnées d'écoulement présentent des meilleurs résultats pour ϕ et r comparativement aux modèles en CC. Les modèles NS2 et NS3 en coordonnées d'écoulement améliorent l'estimation du volume contaminé pour les deux seuils considérés. Les valeurs de RMSE sont inférieures pour tous les modèles en coordonnées d'écoulement. Bien que le modèle NS3* ait été sélectionné comme le meilleur modèle en terme du maximum de vraisemblance, les résultats présentés au tableau 4.12 montrent que le modèle NS2* offre une meilleure précision.

Les figures 4.12 et 4.13 présentent les cartes obtenues avec les modèles $CAE_h/NS3^*$, $CAE_h/NS2^*$ et $CC/NS3^*$, pour une vue en plan ($z = 25$ m) et une coupe ($y = 50$ m). Ces cartes montrent que les panaches interpolés dépendent en grande partie d'une seule observation, située en (10, 50, 25). Les résultats pour les modèles $CAE_h/NS3^*$ et $CAE_h/NS2^*$ montrent que l'approche proposée tire mieux profit d'un échantillonnage limité de la concentration. On remarque aussi que la différence entre les cartes obtenues avec les modèles $CAE_h/NS2^*$ et $CAE_h/NS3^*$ réside principalement dans la portée transversale estimée pour chacun de ces modèles et que l'augmentation de celle-ci, dans le cas $CAE_h/NS3^*$, améliorera probablement les résultats. Le modèle $CAE_h/NS3^*$ produit néanmoins une carte plus satisfaisante que celle obtenue avec l'utilisation des coordonnées cartésiennes (figures 4.12(c) et 4.13(c)).

À titre indicatif, lorsqu'on utilise la portée transversale estimée pour le modèle NS2 pour

le modèle NS3, on obtient des résultats similaires pour ces deux modèles. Le modèle NS3 a alors une vraisemblance moindre que celle présentée au tableau 4.11, mais celle-ci est suffisamment élevée pour que le LRT la juge significative par rapport au modèle NS2. La valeur d'AIC_c obtenue est toutefois plus élevée que celle du modèle NS2. Dans un cas réel, où l'ajustement visuel de la portée n'est pas une option, on dispose néanmoins d'une alternative à l'estimation des paramètres : la validation croisée. Pour cet exemple, l'estimation des paramètres géostatistiques par validation croisée produit des RMSE semblables pour les modèles NS2 et NS3 (0.1469 et 0.1463, respectivement). Dans ce cas, les portées transversales pour ces deux modèles sont comparables et légèrement inférieures à celles ayant produit les résultats illustrés aux cartes 4.12(d) et 4.13(d).

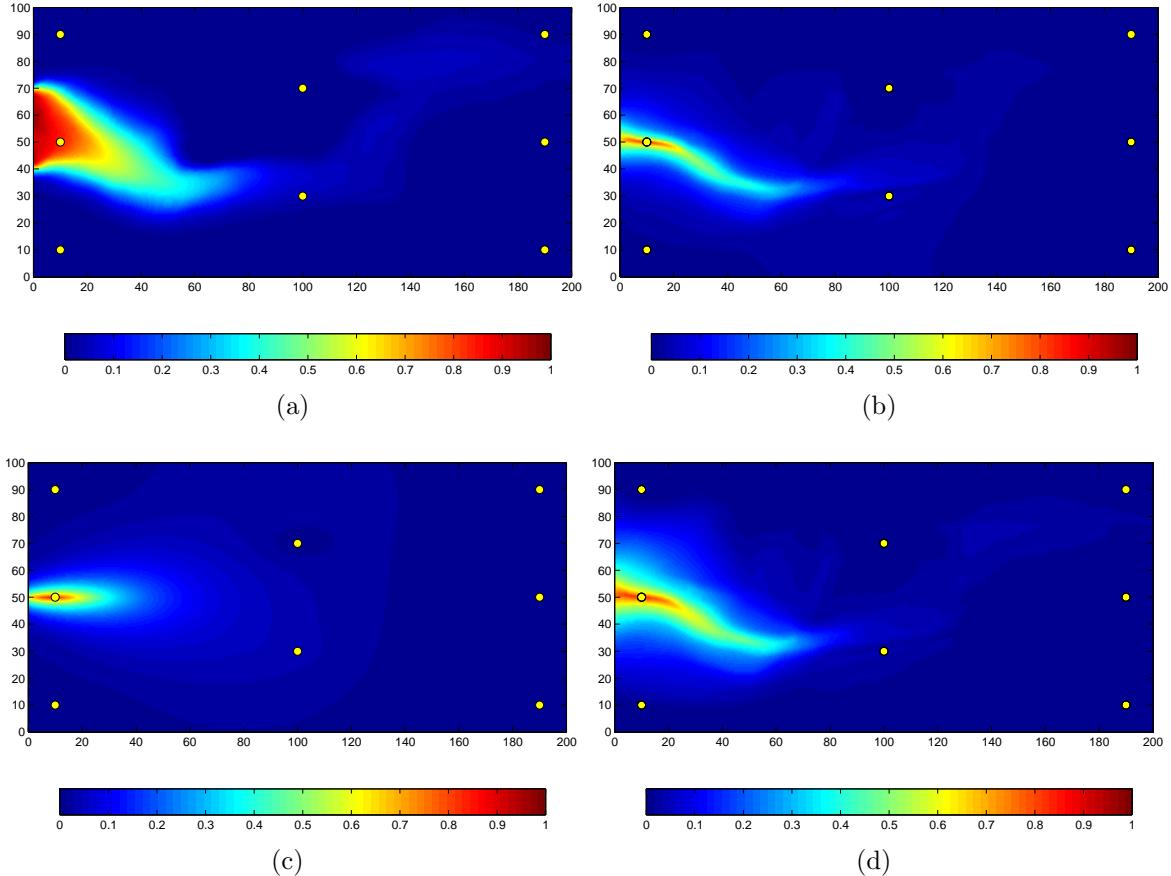


Figure 4.12 Exemple 2. Vue en plan ($z = 25$ m) des valeurs de concentration de référence et krigées. (a) Concentrations de référence. (b) CAE_h/NS3 - meilleur modèle selon le maximum de vraisemblance. (c) CC/NS3 - meilleur modèle en coordonnées cartésiennes et (d) CAE_h/NS2 - meilleur modèle selon les statistiques comparatives.

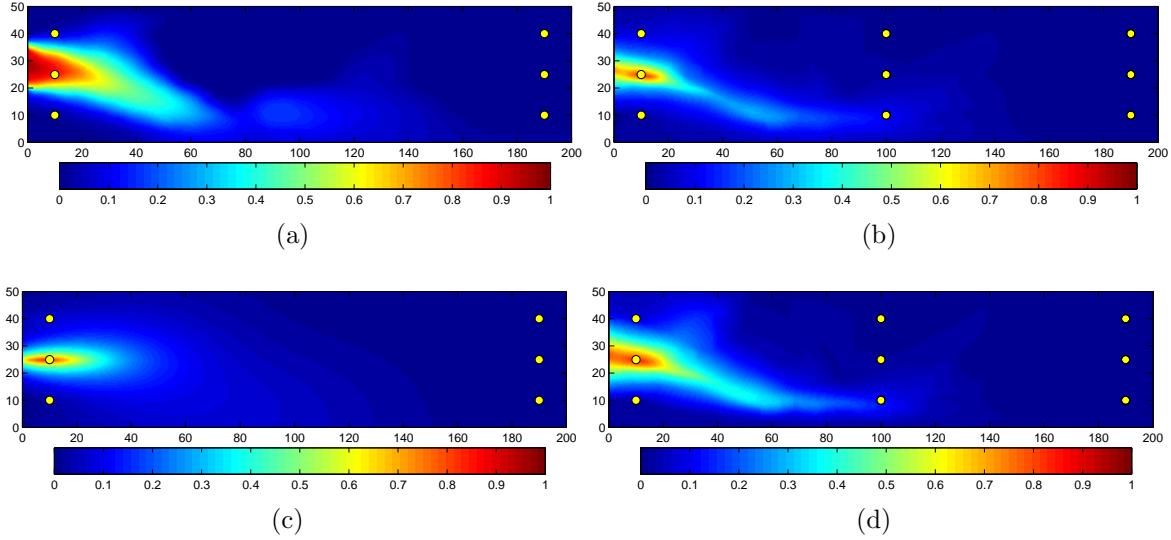


Figure 4.13 Exemple 2. Vue en coupe ($y = 50$ m) des valeurs de concentration de référence et krigées. (a) Concentration de référence. (b) $\text{CAE}_h/\text{NS3}$ - meilleur modèle selon le maximum de vraisemblance. (c) $\text{CC}/\text{NS3}$ - meilleur modèle en coordonnées cartésiennes et (d) $\text{CAE}_h/\text{NS2}$ - meilleur modèle selon les statistiques comparatives.

Tableau 4.12 Exemple 2. Comparaison des résultats pour les différentes normes d'évaluation. *En gris* : Les modèles retenus par la procédure de sélection pour chaque système de coordonnées. c_s : seuil de contamination.

Modèle	RMSE	RAE_m	RAE_{vol}		ϕ		r
			$c_s = 0.01$	$c_s = 0.1$	$c_s = 0.01$	$c_s = 0.1$	
CC/S1* ^a	0.1756	0.2994	1.5721	0.0549	0.1096	0.5273	0.7788
CC/NS1*	0.1746	0.1314	0.9409	0.1650	0.2948	0.5627	0.6579
CC/NS2*	0.1890	0.0256	0.6006	0.2211	0.3714	0.5087	0.8006
CC/NS3*	0.1706	0.3034	0.6191	0.7183	0.3881	0.4533	0.7423
CAE _h /S1*	0.1684	0.2607	1.5044	0.4196	0.1164	0.4175	0.5768
CAE _h /NS1*	0.1690	0.1304	0.9695	0.1881	0.3298	0.7970	0.7978
CAE _h /NS2*	0.1689	0.0209	0.3975	0.01353	0.5199	0.7337	0.8860
CAE _h /NS3*	0.1653	0.2326	0.5218	0.4108	0.4486	0.7335	0.8467

^a Des paramètres *ad hoc* ont été utilisés pour le modèle CC/S1* puisque l'estimation par vraisemblance maximum a fourni des paramètres de portée excessivement petits. Ainsi, les paramètres utilisés pour ce modèle ne sont pas ceux associés au maximum de vraisemblance.

4.3 Exemple 3. Utilisation du temps de parcours comme coordonnée longitudinale

En s'inspirant des coordonnées lagrangiennes, l'idée d'utiliser le temps comme coordonnée longitudinale a été évoquée au chapitre 3. Le temps permet éventuellement de tenir compte des chemins d'écoulement préférentiels dans la représentation spatiale des points, ce qui peut être déterminant pour l'interpolation lorsque le panache comporte une digitation.

Dans le présent exemple, on s'intéresse particulièrement à :

1. tester l'utilisation du temps comme coordonnée longitudinale et comparer les résultats avec ceux obtenus en utilisant la charge hydraulique ;
2. montrer la performance de l'approche d'interpolation dans un contexte où le panache se divise ;
3. illustrer la différence liée à la méthode d'estimation des paramètres (maximum de vraisemblance ou validation croisée) sur les cartes krigées.

4.3.1 Description du modèle de référence

Un modèle 2D, de même dimension que celui de l'exemple 1, est utilisé pour tester les coordonnées CAE_t . Le modèle est construit de manière à produire un panache qui se divise en deux « doigts » distincts pour tester l'efficacité de l'approche dans un contexte où le panache a une forme plus complexe. Le champ de conductivité hydraulique utilisé est présenté à la figure 4.14(a). La coordonnée u et le réseau d'écoulement correspondant au modèle sont présentés à la figure 4.14(b). Le panache de référence provient d'une source continue située en $20 \text{ m} \leq x \leq 70 \text{ m}$ sur la frontière $y = 50 \text{ m}$ (figure 4.15(c)). Les paramètres du modèle sont présentés au tableau 4.13.

4.3.2 Sélection du modèle de covariance

Comme pour les modèles précédents, on procède à l'estimation des paramètres du modèle pour les quatre modèles candidats et à l'identification du modèle le plus significatif. Afin d'alléger la présentation, on compare uniquement les modèles définis en coordonnées CAE_t et CAE_h . Les tableaux 4.14 et 4.15 présentent les résultats obtenus pour le test de vraisemblance et l'AIC_c. Le test de vraisemblance retient le modèle NS2 dans les deux cas. Par contre, dans le cas des coordonnées CAE_t , l'AIC_c du modèle NS2 est légèrement supérieur à celui du modèle S1. Étant donné les résultats obtenus aux exemples 1 et 2, où le modèle

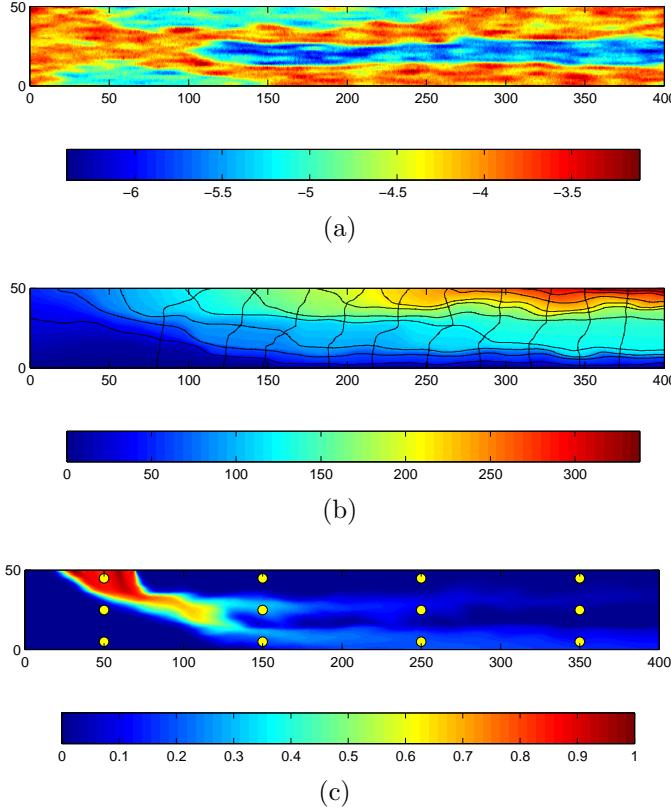


Figure 4.14 Exemple 3. (a) Champ de conductivité hydraulique ($\log_{10}K$), (b) Coordonnée u (divisée par u_{max}), (c) Panache de référence.

NS2 était systématiquement significatif par rapport au modèle S1, on considère que cette différence d'AIC_c est négligeable. On note toutefois que les valeurs de MV sont plus élevées pour les CAE_h.

4.3.3 Estimation de la concentration

Le tableau 4.16 compare les statistiques obtenues pour le krigage avec les modèles NS2 et NS3, dans les deux systèmes de coordonnées considérés ici. Les résultats montrent que les valeurs de concentration estimées en utilisant les CAE_h sont mieux corrélées avec le modèle de référence que celles estimées avec les CAE_t. Par contre, ces dernières engendrent un meilleur estimé du volume contaminé, et ce pour les deux seuils considérés⁴. On remarque également que le modèle NS2 n'est pas celui qui produit les meilleures statistiques comparatives. Ceci est confirmé visuellement sur les cartes de concentration obtenues, où l'on voit que celles produites par le modèle NS3 semblent plus réalistes que celles du modèle NS2 (figure 4.15).

4. Il est à noter que l'on considère ici un seuil plus élevé que dans les exemples précédents.

Tableau 4.13 Exemple 3. Paramètres du modèle synthétique

Paramètres	Valeur
Géométrie	
Dimensions	400 m \times 50 m
Taille max. éléments	1.25 m
Paramètres d'écoulement	
Moyenne géométrique K	3.13×10^{-5} m/s
Variance de ln(K)	2.2092
Gradient hydraulique moyen	0.000875
Conditions frontières	
h en $x = 0$	51 m
h en $x = 400$	50.65 m
q_n en $y = 50$	1×10^{-8} m/s
q_n en $y = 0$	nul
Paramètres de transport	
Porosité	0.35
D_m	1.34×10^{-9} m ² /s
α_l, α_{tv}	2 m, 0.075 m

Tableau 4.14 Exemple 3. Maximum de vraisemblance (MV), *p-value* du test du rapport de vraisemblance ($\alpha = 0.05$) et AIC_c pour les modèles en **coordonnées alternative CAE_t**. k est le nombre de paramètres estimés pour chaque modèle.

Modèle	k	MV	H_0	<i>p-value</i>	AIC _c
CAE _t /S1	3	3.26	-	-	6.64
CAE _t /NS1	5	6.90	CAE _t /S1	0.47	16.14
CAE _t /NS2	4	28.54	CAE _t /S1	0.037	7.01
CAE _t /NS3	6	34.60	CAE _t /NS2	0.83	21.71

Tableau 4.15 Exemple 3. Maximum de vraisemblance (MV), *p-value* du test du rapport de vraisemblance ($\alpha = 0.05$) et AIC_c pour les modèles en **coordonnées alternative CAE_h**. k est le nombre de paramètres estimés pour chaque modèle.

Modèle	k	MV	H_0	<i>p-value</i>	AIC _c
CAE _h /S1	3	2.68	-	-	7.03
CAE _h /NS1	5	33.20	CAE _h /S1	0.08	13.00
CAE _h /NS2	4	232.85	CAE _h /S1	0.0028	2.81
CAE _h /NS3	6	1477.91	CAE _h /NS2	0.16	14.20

Le résultat le plus intéressant est observé sur la figure 4.15(b). On y voit que la transformation spatiale utilisant le temps comme coordonnée longitudinale reproduit la séparation

entre les deux « doigts » du panache en permettant l'estimation de valeurs nulles (ou très faibles) dans la lentille de faible conductivité hydraulique, et ce avec un échantillonnage relativement faible de cette zone. Ceci n'est pas possible en utilisant la charge hydraulique comme coordonnée longitudinale d'écoulement, comme le montre les figures 4.15(c) et 4.15(d). Toutefois, l'examen de la partie située près de la source, sous le panache estimé sur la figure 4.15(b), illustre que ceci ne se traduit pas par une amélioration des statistiques en raison des valeurs trop élevées estimées dans cette zone.

Comme on l'a vu à l'exemple précédent, le modèle sélectionné par le LRT et l'AIC_c n'est pas toujours celui qui entraîne la meilleure estimation de la concentration. On examine ici les résultats obtenus en considérant la validation croisée pour l'estimation des paramètres. Ceci est fait uniquement pour les modèles en CAE_t. Dans le cas de la validation croisée, le critère minimisé est la RMSE. Le tableau 4.17 présente les valeurs obtenues pour les quatres modèles de covariance et montre que la RMSE est moins élevée pour le modèle NS2. Les paramètres estimés ne sont toutefois pas les mêmes que dans le cas du maximum de vraisemblance. Le tableau 4.18 montre que dans ce cas le modèle NS2 avec les paramètres estimés par VC présente une meilleure performance au niveau des statistiques comparatives. L'estimé du volume et la valeur de ϕ pour le seuil 0.2, ainsi que de la corrélation, sont améliorées par rapport aux résultats du krigage avec les paramètres estimés par maximum de vraisemblance. Cette amélioration est également visible sur la figure 4.16(a). Le résultat obtenu avec le modèle NS3 est également présenté à la figure 4.16(b) et montre peu de différence avec le modèle NS2. Ce résultat est semblable à celui obtenu pour l'exemple 2, où une vérification indiquait que les paramètres estimés par validation croisée produisaient des résultats équivalents pour les modèles NS2 et NS3.

Tableau 4.16 Exemple 3. Comparaison des résultats pour les différentes normes d'évaluation. *En gris* : Les modèles retenus par la procédure de sélection pour chaque système de coordonnées. c_s : seuil de contamination.

Modèle	RMSE	RAE _m	RAE _{vol}		ϕ		r
			$c_s = 0.1$	$c_s = 0.2$	$c_s = 0.1$	$c_s = 0.2$	
CAE _h /NS2	0.1903	0.1375	0.1491	0.4856	0.7006	0.7596	0.8751
CAE _h /NS3	0.1775	0.0058	0.1424	0.2233	0.7938	0.8340	0.9020
CAE _t /NS2	0.1988	0.1121	0.0360	0.3527	0.6166	0.6402	0.8107
CAE _t /NS3	0.2196	0.0243	0.0555	0.0819	0.4739	0.7856	0.8399

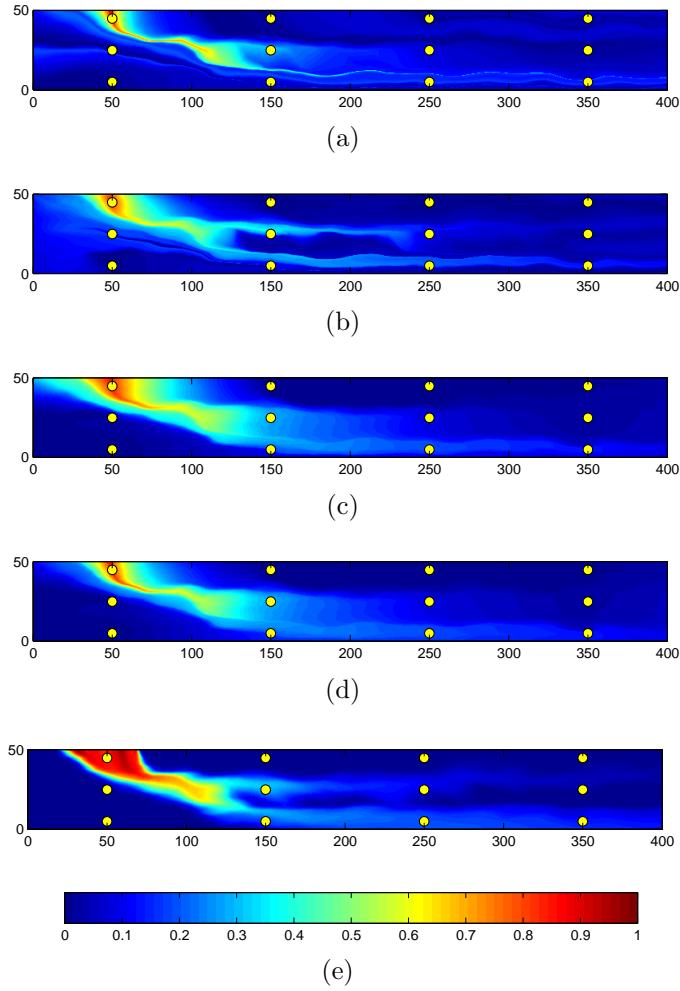


Figure 4.15 Exemple 3. Cartes de concentration krigées. (a) CAE_t/NS2, (b) CAE_t/NS3, (c) CAE_h/NS2, (d) CAE_h/NS3, (e) Concentrations de référence.

Tableau 4.17 Exemple 3. RMSE optimisées pour l'estimation des paramètres par validation croisée

Modèle	RMSE
CAE _t /S1	0.2070
CAE _t /NS1	0.2079
CAE _t /NS2	0.1745
CAE _t /NS3	0.1968

Tableau 4.18 Exemple 3. Comparaison des résultats pour les différentes normes d'évaluation pour le modèle CAE_t avec les paramètres estimés par **validation croisée**. c_s : seuil de contamination.

Modèle	RMSE	RAE_m	RAE_{vol}		ϕ		r
			$c_s = 0.1$	$c_s = 0.2$	$c_s = 0.1$	$c_s = 0.2$	
$\text{CAE}_t/\text{NS2}^{VC}$	0.1745	0.1069	0.0155	0.0401	0.5496	0.7370	0.8578
$\text{CAE}_t/\text{NS3}^{VC}$	0.1968	0.0795	0.0315	0.1592	0.5613	0.7559	0.8761

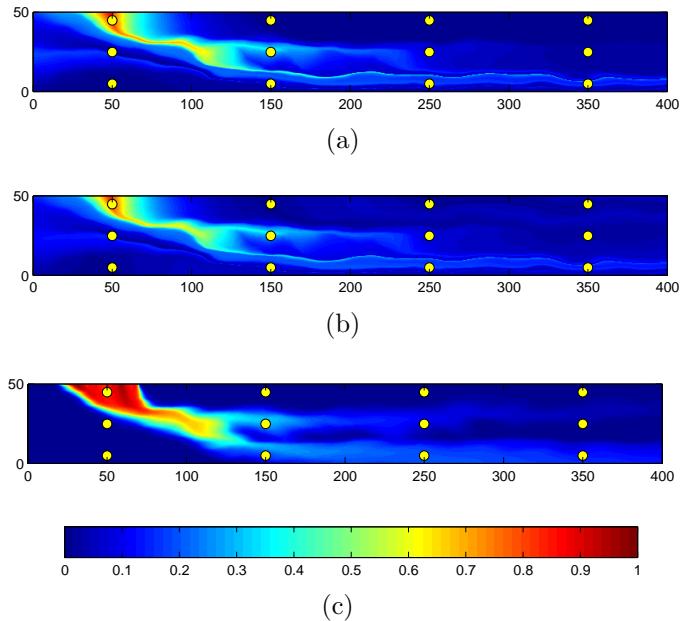


Figure 4.16 Exemple 3. Cartes de concentration krigées avec les paramètres obtenus par validation croisée. (a) $\text{CAE}_t/\text{NS2}$, (b) $\text{CAE}_t/\text{NS3}$, (c) Concentrations de référence.

4.4 Discussion

Dans cette section, on revient sur les principaux résultats en lien avec les éléments méthodologiques spécifiques à l'approche proposée, soit l'utilisation d'une transformation spatiale basée sur les coordonnées d'écoulement et la paramétrisation du modèle de covariance non stationnaire. On discute également de l'estimation des paramètres géostatistiques et ainsi que des gains obtenus sur les estimés de concentration.

4.4.1 Coordonnées d'écoulement

Les exemples précédents montrent bien l'intérêt d'utiliser une transformation basée sur les coordonnées d'écoulement pour le krigeage des valeurs de concentration dans l'eau souterraine. Avec un nombre limité d'observations, le recours à une transformation de coordonnées améliore considérablement le réalisme des cartes interpolées, en permettant que l'anisotropie propre au phénomène soit mieux prise en compte. Bien que l'écoulement ne soit jamais connu de façon aussi détaillée que dans les exemples présentés, les résultats donnent tout de même un aperçu des gains potentiels et de la capacité des coordonnées d'écoulement (autant CAE que CNE) à capturer l'anisotropie d'un panache en milieu hétérogène, tel qu'illustré par les fonctions de covariance montrées à la figure 4.5.

Un autre objectif visé était de tester la performance des coordonnées alternatives d'écoulement. Dans un premier temps, l'exemple 1 a permis de montrer que les CAE peuvent être une alternative viable aux coordonnées naturelles. Étant donnée la ressemblance entre u et ψ , montrée à la figure 4.2, les résultats obtenus avec ces deux systèmes de coordonnées sont très similaires. En 2D, u et ψ sont équivalents si le gradient de u est proportionnel au flux entrant. Autrement, et c'est le cas dans les exemples proposés, u est une version approximative de ψ . Pour cette raison, les résultats ne sont pas identiques, mais néanmoins similaires.

De plus, l'utilisation du temps de parcours comme coordonnée longitudinale a permis de reproduire les deux doigts du panache de l'exemple 3 sur la carte interpolée. Le temps de parcours est plus directement affecté par les contrastes de conductivité hydraulique retrouvés le long des lignes de courant. Toutefois, les distances redéfinies dans l'espace (t, u) peuvent également rapprocher certains points et contribuer à l'estimation de valeurs trop élevées où la concentration devrait être nulle. De même, les valeurs du temps de parcours sont sensibles aux éventuelles imprécisions numériques sur les flux hydrauliques estimés par MEF. De telles imprécisions pourraient expliquer les différences importantes entre les valeurs de vraisem-

blance en CAE_t et en CAE_h .

L'impact de la précision numérique sur les résultats n'a pas été vérifié. À toutes fins pratiques, on considère que les erreurs numériques associées aux algorithmes et aux méthodes utilisées pour obtenir les coordonnées d'écoulement (CNE ou CAE) sont négligeables face à l'incertitude sur le modèle d'écoulement lui-même (valeurs de conductivité hydraulique, connectivité des unités hydrogéologiques, conditions frontières). Pour un cas donné, ceci pourrait être vérifié en ayant recours à un grand nombre de réalisations du champ de conductivité hydraulique, de manière à quantifier une partie de l'incertitude sur la transformation de coordonnées. Il est envisageable que l'erreur numérique puisse avoir un impact important sur la précision des coordonnées d'écoulement, par exemple, lorsque l'emprise horizontale du modèle est beaucoup plus importante que son épaisseur (ratio $x : z$ très élevé).

4.4.2 Modèles de covariance non stationnaires

Tel qu'illustré aux figures 4.6(c) et 4.6(e), l'utilisation d'un modèle de covariance stationnaire en coordonnées d'écoulement ne permet pas de modéliser la diminution des valeurs de concentration en s'éloignant de la source. Les panaches krigés ont donc tendance à s'étirer exagérément en raison des paramètres estimés. Toutefois, la transformation en coordonnées d'écoulement simplifie grandement la paramétrisation du modèle de covariance non stationnaire utilisé en évitant d'avoir à modéliser les changements de direction d'anisotropie dans la matrice de covariance des noyaux gaussiens, Σ_i . Le modèle de covariance permet quant à lui faire varier la corrélation spatiale (modèles NS1 et NS3) et la variance (modèles NS2 et NS3) en fonction de la coordonnée longitudinale, et ainsi de modéliser approximativement les effets de la dispersion.

Les résultats montrent que dans chaque cas ce sont les paramétrisations non stationnaires incluant la diminution de la variance (NS2 et NS3) qui sont considérées significatives lors du LRT et qui génèrent le plus faible AIC_c . Ces paramétrisations sont également celles qui produisent les meilleurs résultats, autant quantitativement que visuellement. Ceci démontre bien que les paramétrisations proposées, même en étant relativement simples, ont un impact positif sur le krigeage de la concentration. De plus, les modèles non stationnaires améliorent également les résultats du krigeage en coordonnées cartésiennes, du moins au niveau des statistiques comparatives.

4.4.3 Estimation des paramètres par maximum de vraisemblance

Puisque les paramètres des modèles de covariance non stationnaires ne peuvent pas être estimés par variographie, l'approche est très dépendante de la procédure d'estimation des paramètres utilisée. Tel qu'il a été mentionné au chapitre 3, les paramètres estimés ne sont pas exempts de biais, en raison de l'écart par rapport à l'hypothèse gaussienne, inhérente aux données de concentration, et de l'incertitude sur la moyenne.

En dépit de cela, les paramètres estimés par maximum de vraisemblance ont produit des résultats acceptables, à l'exception de certains modèles en coordonnées cartésiennes où des paramètres *ad hoc* ont été employés. Pour le cas 3D, la procédure d'estimation des paramètres et de sélection des modèles s'est montrée efficace à rejeter les modèles complètement anisotropes (à trois portées différentes) dans un contexte où le panache de référence est issu d'un modèle de transport qui comporte des dispersivités transversales égales. Dans les exemples 1 et 2, les valeurs du maximum de vraisemblance sont systématiquement plus élevées pour les modèles en coordonnées d'écoulement (autant CAE_{*h*} que CNE) que pour ceux en coordonnées cartésiennes (l'exemple 3 ne comprend pas les résultats en coordonnées cartésiennes).

Toutefois, le modèle sélectionné sur la base du maximum de vraisemblance n'est pas nécessairement le modèle le plus performant en ce qui a trait à des statistiques telles que l'erreur sur la masse ou le volume. De plus, l'exemple 2 montre que les résultats obtenus sont relativement sensibles aux paramètres utilisées. L'exemple 3 semble indiquer que la validation croisée est plus robuste, possiblement parce que celle-ci est directement basée sur la capacité du modèle à reproduire les observations. Ainsi, les exemples montrent que l'ajustement automatique des paramètres peut être problématique.

4.4.4 Résultats du krigeage

Visuellement, l'approche de krigeage non stationnaire a permis d'améliorer le réalisme des cartes de concentration interpolées. Ces améliorations se reflètent avant tout dans les statistiques comparatives liées à la délimitation du panache, telles que le ϕ de Pearson et la corrélation r . L'amélioration est moins systématique dans le cas des statistiques associées à l'estimation de la masse et du volume de contaminant (RAE_{*m*}, RAE_{*vol*}). La statistique de validation croisée, RMSE, présente de très légères diminutions pour les modèles obtenus par l'approche de krigeage proposée.

Les fonctions de covariance non stationnaires, combinées à la transformation de coordonnées, ont également des implications appréciables sur la variance d'estimation. Pour un modèle de covariance stationnaire, il est fréquemment reproché à la variance de krigeage de tenir uniquement compte de la disposition géométrique des données. Ceci est toujours vrai avec le modèle non stationnaire dans l'espace des coordonnées d'écoulement. Par contre, le fait de paramétriser une diminution de la variance en fonction de l'écoulement produit une variance d'estimation qui tient compte du processus physique sous-jacent. Sans quelle soit équivalente aux variances obtenues par modélisation stochastique, la variance d'estimation obtenue avec les modèles non stationnaires s'en approche par ses nouvelles caractéristiques. Ces résultats sont intéressants, puisque le potentiel pour des résidus élevés est plus important près de la source (continue) que loin de celle-ci. De plus, la variance d'estimation représente bien le fait que l'on ne connaît pas la position exacte de la source. L'examen des résidus normalisés (non présenté ici) pour les modèles NS2 et NS3 de l'exemple 1 montre des valeurs importantes de $|\frac{c_R - c_*}{\sigma_e}|$ près de la source, suggérant que la variance de krigeage n'est pas suffisamment élevée pour être complètement représentative des erreurs dans l'estimation des valeurs élevées retrouvées dans ce secteur, et ce malgré les améliorations notées. Ailleurs sur le domaine, les résidus normalisés étaient compris entre ± 2 .

L'estimation est effectuée par krigeage ordinaire lequel suppose une moyenne (non conditionnelle) constante. Bien qu'une telle moyenne ne soit pas réaliste compte tenu du modèle physique, celle-ci a permis d'obtenir des résultats plus probants qu'en ayant recours au krigeage avec dérive externe avec une moyenne modélisée comme une décroissance linéaire. Un biais important a été observé avec cette dérive puisqu'elle entraîne une surestimation des concentrations dans la partie amont du domaine, de part et d'autre du panache. Ceci est dû au fait que la moyenne locale est plus élevée dans cette zone et que le krigeage tend à retrouver la valeur de la moyenne loin des données. De plus, une dérive décroissante peut éventuellement mener à l'estimation de valeurs de concentration négatives dans la partie aval du domaine, selon l'ajustement. Le problème de biais est plus marqué pour le modèle stationnaire et en présence de faibles valeurs de portées spatiales. Un biais similaire est également présent avec une moyenne constante, mais le problème est de moindre ampleur. Il demeure que ce choix entraîne localement une surestimation dans les zones non contaminées, ce qui pourrait expliquer les résultats parfois incohérents obtenus pour l'estimation de la masse et du volume. Une moyenne réaliste qui tiendrait compte du profil transversal du panache de contamination améliorerait probablement l'estimation. Par contre, la modélisation (paramétrisation) d'une telle moyenne nécessiterait vraisemblablement des informations ou des hypothèses concernant la position et la taille de la source. En cas de spécification d'une

moyenne non représentative, on risque toutefois des problèmes de biais similaires à ceux évoqués précédemment. Le krigeage ordinaire apparaît ainsi comme un choix prudent puisque, pour des points situés en dehors de la portée des observations, celui-ci tend à ramener les valeurs krigées vers la moyenne (constante) estimée à partir des données et du modèle de covariance.

Les paramètres qui semblent avoir le plus d'impact sur le panache estimé sont les paramètres de portées transversales, tel que le suggère la comparaison entre les résultats du krigeage pour les modèles NS2 et NS3 aux exemples 1 et 2. Dans les deux cas, la différence importante entre les panaches est essentiellement liée au fait que les portées transversales estimées pour le modèle NS3 sont plus faibles que celles estimées pour le modèle NS2.

On note qu'une transformation des données telle que celle utilisée par Kitanidis et Shen (1996), dont la transformation logarithmique apparaît comme un cas particulier, aurait pu être utilisée. Ce type de transformation permet d'éviter le krigeage de valeurs négatives et de tenir compte de l'asymétrie de la distribution de la concentration. Toutefois, ceci peut introduire une dépendance aux moments statistiques d'ordres supérieurs du logarithme plutôt que dans ceux de la variable originale, dans les valeurs retransformées. De plus, celles-ci peuvent être fortement biaisées si la variance d'estimation de la variable transformée n'est pas parfaitement estimée (David, 1988; Chilès et Delfiner, 1999). Cette option est également rejetée pour les raisons évoquées au chapitre précédent au sujet des incrément (et de l'estimation par maximum de vraisemblance restreint), à savoir qu'une transformation pourrait éventuellement contribuer à masquer la structure spatiale postulée dans le modèle de covariance non stationnaire.

Dans le cas des coordonnées alternatives, l'interpolation linéaire utilisée pour estimer u et v sur l'ensemble du domaine est également une source d'imprécision. Toutefois, les résultats présentés à l'Annexe C montrent que les résultats changent peu pour différentes densités de lignes de courant utilisées pour interpoler les coordonnées d'écoulement à l'intérieur du modèle. Finalement, le fait que la proportionnalité des coordonnées u et v avec le flux hydraulique puisse varier localement est également une source d'imprécision. Globalement, ces erreurs numériques sur les coordonnées se traduisent par des erreurs de positionnement des points dans l'espace, lesquelles peuvent être prise en compte par l'effet de pépite (variabilité à petite échelle).

Les coordonnées d'écoulement utilisées dans les exemples présentés dans ce chapitre ont

été calculées à partir du modèle d’écoulement de référence. En pratique, le modèle utilisé est plutôt un modèle conceptuel qui, selon les données disponibles et leur qualité, incorpore les éléments/structures déterminants de l’écoulement. Dans le chapitre suivant, on s’intéresse à l’impact de l’utilisation d’un modèle d’écoulement différent du modèle de référence sur l’estimation des paramètres et les résultats du krigage et on propose une méthodologie pour l’intégration des données hydrogéologiques disponibles.

CHAPITRE 5

EXEMPLE D'APPLICATION À UN MODÈLE D'ÉCOULEMENT PARTIELLEMENT CONNU

Dans ce chapitre, plutôt que d'utiliser directement le modèle d'écoulement de référence, les coordonnées d'écoulement sont calculées à partir de modèles conceptuels simplifiés, lesquels correspondent à des interprétations différentes de la géologie. Toutefois, ces modèles sont d'abord calés sur un nombre restreint de données de charge hydraulique. Ainsi, en plus de donner un aperçu de l'efficacité de l'approche dans un contexte où l'écoulement est partiellement connu, cet exemple propose une approche d'intégration des données, applicable dans des cas réels. Le modèle de référence et les modèles conceptuels sont repris de Rivest *et al.* (2012b), mais on utilise ici les fonctions de covariance non stationnaires et la procédure d'estimation des paramètres testées au chapitre précédent.

5.1 Description du modèle de référence et intégration des données

Le modèle de référence est semblable à celui utilisé à l'exemple 1 du chapitre précédent. Toutefois, il comporte quelques différences au niveau des paramètres d'écoulement et de transport. Ces différences sont résumées au tableau 5.1. De plus, la solution utilisée correspond à un temps plus élevé ($t = 48$ ans). Les figures 5.1(a) et 5.1(b) présentent respectivement le champ de conductivité hydraulique et le panache de référence. On dispose de 12 observations de concentration. À partir de ce modèle, on propose deux modèles conceptuels d'écoulement :

Scénario 1 Aucune information sur la stratigraphie n'est disponible et le modèle est composé d'une seule unité hydrogéologique homogène (valeur moyenne des conductivités hydrauliques de référence). Les conditions frontières utilisées exagèrent le gradient hydraulique et la recharge est considérée connue.

Scénario 2 On suppose que des forages ont permis l'interprétation proposée aux figures 5.1(c) et 5.1(d). Ce modèle comprend deux unités hydrostratigraphiques de conductivités hydrauliques $K_1 = 3.2 \times 10^{-4}$ m/s et $K_2 = 3.2 \times 10^{-5}$ m/s, lesquelles sont du même ordre de grandeur que les valeurs moyennes dans chaque unité du modèle de référence. Les conditions frontières utilisées sont les mêmes que pour le scénario 1.

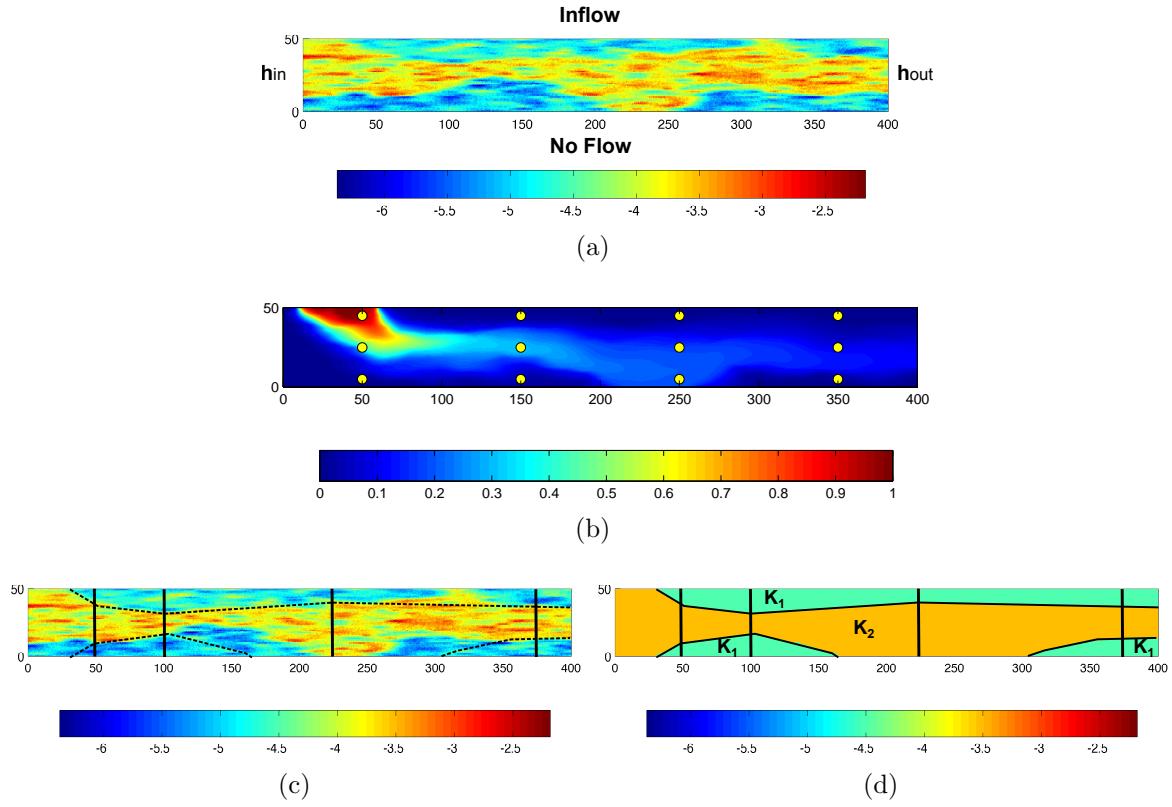


Figure 5.1 Modèle de référence et modèle simplifié du scénario 2. (a) Champ de conductivité hydraulique de référence ($\log_{10}K$), (b) Concentrations de référence, (c) Interprétation à partir de forages fictifs (illustrés par les lignes noires), (d) Modèle conceptuel du scénario 2.

Ces modèles conceptuels ne sont pas utilisés directement pour le calcul des coordonnées d’écoulement. On utilise plutôt une procédure d’intégration des données en deux étapes, dont le but est de caler les modèles d’écoulement sur des observations de charge hydraulique.

Dans un premier temps, les solutions de charge hydraulique correspondant aux modèles conceptuels décrits plus haut et des observations de charges hydrauliques prélevées sur le modèle de référence sont utilisées pour produire un estimé de la carte de charge en utilisant l’approche de krigeage avec dérive externe proposée par Rivest *et al.* (2008). Dans cette approche, la solution numérique de la charge hydraulique associée au modèle conceptuel est utilisée comme dérive externe (ou moyenne) dans le système de krigeage. Cette étape permet de produire un estimé de la charge hydraulique qui tienne compte à la fois des observations (de charges) et des caractéristiques du modèle conceptuel. Douze données de charge hydraulique sont utilisées pour ce krigeage ; celles-ci sont situées le long des « forages » représentés aux figures 5.1(c) et 5.1(d).

La deuxième étape consiste à utiliser le champ de charge krigé dans un algorithme d'inversion pour produire un estimé du champ de conductivité hydraulique calibré aux observations de charge. On obtient ensuite le réseau d'écoulement correspondant par simulation numérique. L'algorithme d'estimation successive des flux de Pasquier et Marcotte (2006) est utilisé à cette fin, tel qu'implémenté dans le programme *Hydrogeo_Inv* (Pasquier, 2005). Cette approche d'inversion, similaire à la méthode du modèle de comparaison de Ponzini et Lozej (1982), nécessite un estimé de la charge hydraulique en tout point du domaine et est donc particulièrement complémentaire avec l'approche de krigeage avec dérive externe de Rivest *et al.* (2008). De plus, l'implémentation de ces deux approches à l'aide du même logiciel d'éléments finis (COMSOL) facilite les calculs et permet l'utilisation d'un maillage commun pour le modèle conceptuel utilisé dans le krigeage et l'inversion. Dans l'algorithme d'estimation successive des flux, la conductivité hydraulique est obtenue de manière itérative par la résolution d'une version discrète du problème de Cauchy sur chaque élément fini du domaine, tel que décrit par l'équation suivante :

$$K_j^{i+1} = K_j^i \cdot \frac{\|\nabla h_j^i\|_2 + \beta^i}{\|\nabla \hat{h}_j\|_2 + \beta^i} \quad (5.1)$$

où i est le compteur itératif, j est l'indice de l'élément fini, \hat{h} est le champ de charge hydraulique estimé par l'approche de krigeage et β^i est un facteur d'amortissement décroissant avec l'itération i . Les conductivités hydrauliques sont modifiées jusqu'à ce que les charges calibrées reproduisent les charges estimées de manière satisfaisante. L'algorithme requiert également un champ de conductivité hydraulique initial, pour lequel on utilise les champs de K des modèles conceptuels décrits dans les scénarios 1 et 2. Lors de l'inversion, on suppose que toutes les frontières de type Neumann sont connues, alors que les frontières de type Dirichlet sont extraites de la carte de charge hydraulique krigée. L'ensemble de la procédure d'intégration des données est illustrée à la figure 5.2. Bien que l'étape d'inversion vienne alourdir quelque peu l'obtention des coordonnées d'écoulement, l'algorithme utilisé pour l'estimation successive des flux est très rapide comparativement à d'autres méthodes inverses (Pasquier et Marcotte, 2006).

En appliquant la procédure décrite ci-haut, on obtient donc les réseaux d'écoulement correspondant aux deux scénarios envisagés, calés à des observations de charge hydraulique provenant du modèle de référence. La figure 5.3 présente les réseaux d'écoulement du modèle de référence (figure 5.3(a)) et ceux obtenus suite au calage des modèles d'écoulement des

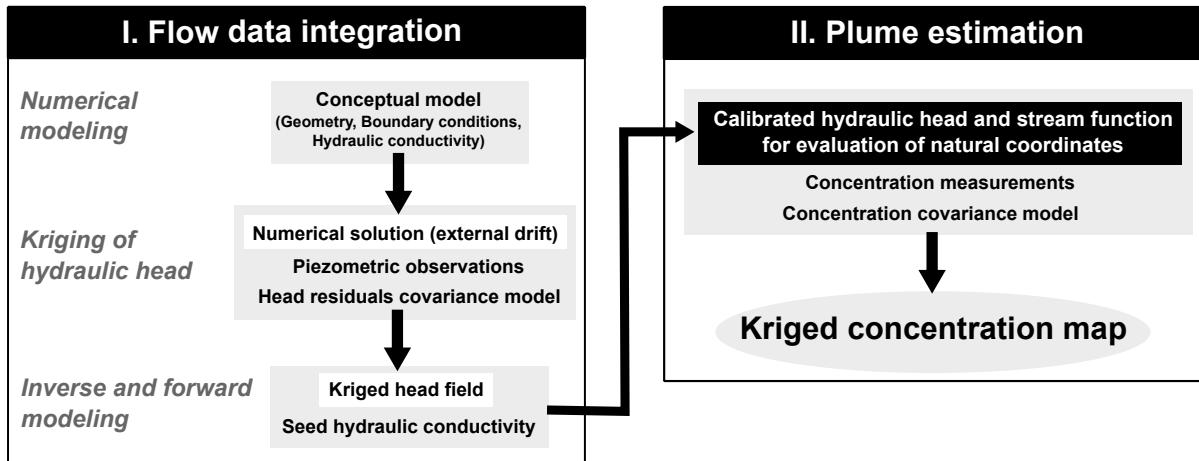


Figure 5.2 Schématisation de l'approche d'intégration des données (tiré de Rivest *et al.* (2012b)).

scénarios 1 et 2 (figures 5.3(b) et 5.3(c))¹. On y voit les différences entre les coordonnées naturelles estimées pour les scénarios 1 et 2, ainsi qu'avec les coordonnées naturelles de référence (figure 5.3(a)). Dans le cas du scénario 1 (figure 5.3(b)), la curvilinéarité observée dans les lignes de courant est principalement due à la recharge en provenance de la frontière du haut. Pour le scénario 2, l'écoulement est plus complexe et reproduit partiellement le rétrécissement des tubes de courant observé sur le modèle de référence (figure 5.3(c)).

5.2 Sélection du modèle de covariance

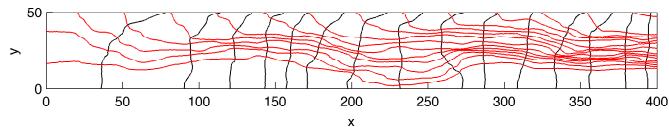
Comme au chapitre précédent, on utilise le maximum de vraisemblance pour l'estimation des paramètres géostatistiques et pour déterminer lequel des modèles de covariance candidats semble le mieux adapté aux données. Cette procédure est répétée pour les coordonnées d'écoulement associées aux deux scénarios proposés, de même que pour le modèle défini avec les coordonnées d'écoulement de référence (comme pour les exemples du chapitre 4). Afin d'alléger l'exemple, on se limite ici à la comparaison des résultats obtenus avec les coordonnées cartésiennes et avec le système de coordonnées naturelles d'écoulement (h et ψ).

Les valeurs du maximum de vraisemblance, la *p-value* du test de vraisemblance et les valeurs de l'AIC_c sont présentées aux tableaux 5.2 à 5.5. Comme pour l'exemple 1, le modèle jugé le plus significatif par le LRT et l'AIC_c est le modèle NS2 (portées fixes et variance décroissante) pour tous les cas considérés (coordonnées cartésiennes, scénarios 1 et 2, modèle

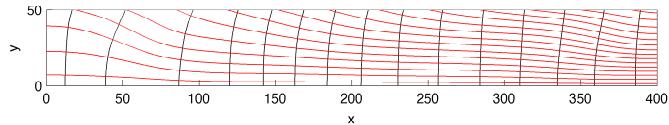
1. Il s'agit donc de réseaux d'écoulement basés sur des conductivités hydrauliques différentes de celles illustrées à la figure 5.1, selon le calage effectué par la méthode d'estimation successive des flux.

Tableau 5.1 Paramètres du modèle synthétique. *En gris* : les paramètres modifiés par rapport à ceux de l'exemple 1 du chapitre 4.

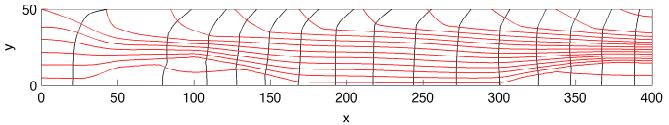
Paramètre	Valeur
Géométrie	
Dimensions	400 m × 50 m
Taille max. éléments	1.25 m
Paramètres d'écoulement	
Moyenne géométrique K	5.4×10^{-5} m/s
Variance de ln(K)	2.13
Gradient hydraulique moyen	0.00075
Conditions frontières	
h en $x = 0$	51 m
h en $x = 400$	50.7 m
q_n en $y = 50$	1.6×10^{-8} m/s
q_n en $y = 0$	nul
Paramètres de transport	
porosité	0.35
D_m	1.34×10^{-9} m ² /s
α_l, α_{tv}	5 m, 0.005 m



(a)



(b)



(c)

Figure 5.3 Réseaux d'écoulement associés aux différents modèles (a) Référence, (b) Scénario 1, (c) Scénario 2. Les réseaux d'écoulement illustrés en (b) et (c) ont été calés à 12 observations de charge hydraulique par l'approche d'estimation successive des flux et sont associés à des conductivités hydrauliques modifiées par rapport à celles des modèles conceptuels.

de référence). Ici encore, les paramètres de la plupart des modèles en coordonnées cartésiennes sont peu réalistes.

Tableau 5.2 Maximum de vraisemblance (MV), *p-value* du test du rapport de vraisemblance et AIC_c pour les modèles en **coordonnées cartésiennes**. *k* est le nombre de paramètres estimés pour chaque modèle.

Modèle	<i>k</i>	MV	<i>H</i> ₀	p-value	AICc
CC/S1	3	0.49	-	-	10.41
CC/NS1	5	1.70	S1	0.29	18.94
CC/NS2	4	12.08	S1	0.012	8.73
CC/NS3	6	21.64	NS2	0.56	22.65

Tableau 5.3 **Scénario 1.** Maximum de vraisemblance (MV), *p-value* du test du rapport de vraisemblance et AIC_c pour les modèles en coordonnées naturelles. *k* est le nombre de paramètres estimés pour chaque modèle.

Modèle	<i>k</i>	MV	<i>H</i> ₀	p-value	AICc
S1	3	0.54	-	-	10.24
NS1	5	2.63	S1	0.20	18.07
NS2	4	11.97	S1	0.013	8.75
NS3	6	35.57	NS2	0.34	21.66

Tableau 5.4 **Scénario 2.** Maximum de vraisemblance (MV), *p-value* du test du rapport de vraisemblance et AIC_c pour les modèles en coordonnées naturelles. *k* est le nombre de paramètres estimés pour chaque modèle.

Modèle	<i>k</i>	MV	<i>H</i> ₀	p-value	AICc
S1	3	0.52	-	-	10.31
NS1	5	3.10	S1	0.17	17.74
NS2	4	17.07	S1	0.01	8.04
NS3	6	51.70	NS2	0.33	20.91

5.3 Estimation de la concentration

Le tableau 5.6 compile les statistiques comparatives obtenues pour chaque scénario. On se limite ici aux résultats obtenus avec le modèle sélectionné NS2. Comme dans les exemples du chapitre 4, l'utilisation des modèles en coordonnées d'écoulement entraîne avant tout une amélioration du maximum de vraisemblance, de RMSE, de ϕ et de r , comparativement aux résultats du krigeage en coordonnées cartésiennes. Les statistiques liés à l'estimation de la

Tableau 5.5 Coordonnées issues du **modèle d’écoulement de référence**. Maximum de vraisemblance (MV), *p-value* du test du rapport de vraisemblance et AIC_c. *k* est le nombre de paramètres estimés pour chaque modèle.

Modèle	<i>k</i>	MV	H_0	p-value	AICc
S1	3	1.84	-	-	7.78
NS1	5	3.94	S1	0.47	17.26
NS2	4	50.73	S1	0.01	5.86
NS3	6	74.51	NS2	0.68	20.18

masse et du volume sont moins précises dans le cas des scénarios 1 et 2. Le krigage qui utilise les coordonnées issues du modèle de référence ne présente pas une performance vraiment supérieure à celles des scénarios approximatifs. Les résultats obtenus avec le modèle du scénario 1 semblent indiquer que la modélisation de la recharge provenant de la frontière $y = 50$ m est suffisante pour améliorer le krigage de la concentration. Ce résultat montre que l’utilisation de modèles d’écoulement simplifiés ou approximatifs permet tout de même des gains par rapport au krigage en coordonnées cartésiennes.

Tel qu’illustré à la figure 5.4, l’utilisation de coordonnées d’écoulement basées sur des modèles plus complexes (scénario 2 et modèle de référence) a toutefois un impact sur le réalisme des cartes de concentration krigées. En plus de reproduire l’anisotropie induite par la recharge provenant de la frontière supérieure, la figure 5.4(c) reproduit (plus ou moins précisément) le rétrécissement du panache observé en $x = 100$ m. La carte associée au scénario 2 (figure 5.4(c)) apparait comme une version plus lisse de celle obtenue en utilisant les coordonnées de référence (figure 5.4(d)).

Tableau 5.6 Statistiques comparatives entre les concentrations krigées et de référence. c_s : seuil de contamination.

Modèle	RMSE	RAE_m	RAE_{vol}		ϕ		r
			$c_s = 0.01$	$c_s = 0.1$	$c_s = 0.01$	$c_s = 0.1$	
CC/NS2	0.2580	0.1745	0.1649	0.1009	0.05690	0.5464	0.7698
CNE/NS2₁	0.2201	0.2830	0.1610	0.1085	0.1366	0.6263	0.8146
CNE/NS2₂	0.2257	0.2042	0.1336	0.05767	0.3305	0.5527	0.7876
CNE/NS2_{ref}	0.2129	0.1725	0.1400	0.1450	0.2631	0.5372	0.8067

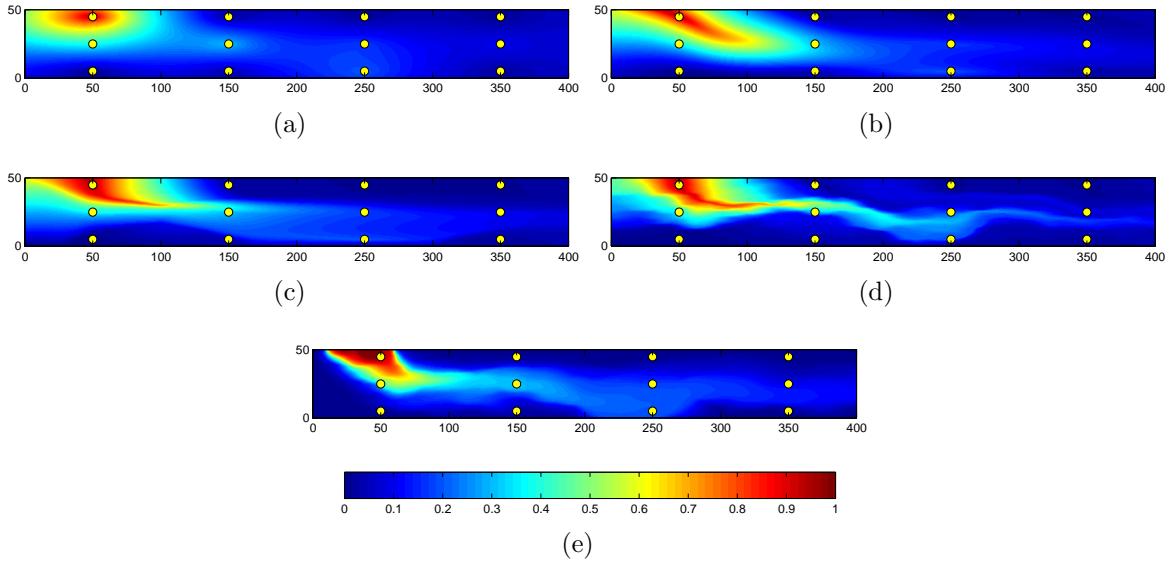


Figure 5.4 Cartes krigées et de référence. (a) CC/NS2, (b) CNE/NS2 pour le scénario 1, (c) CNE/NS2 pour le scénario 2, (d) CNE/NS2 en utilisant le réseau d'écoulement de référence, (e) Panache de référence.

5.4 Discussion

Le présent chapitre illustre la performance de l'approche dans un contexte où le modèle d'écoulement utilisé pour la transformation de coordonnées est partiellement connu. L'exemple utilisé montre qu'une transformation de coordonnées basée sur un modèle simplifié peut également contribuer à améliorer les résultats du krigage. Le modèle associé au scénario 1 (homogène) permet d'inclure l'anisotropie locale engendrée par la recharge, améliorant ainsi le résultat par rapport au krigage en coordonnées cartésiennes. Le scénario 2 comporte une interprétation simplifiée des structures présentes dans le modèle de référence, mais permet néanmoins d'introduire un rétrécissement dans la largeur du panache dû à la présence de structures moins perméables. Par contre, l'amélioration porte avant tout sur l'aspect visuel du panache. Comme aux exemples du chapitre précédent, ce sont les statistiques décrivant la capacité de délimitation du panache (i.e. ϕ) qui montrent une amélioration par rapport au résultat du krigage en coordonnées cartésiennes. Il en va autrement pour les statistiques liées à la quantité de contaminant. L'examen de la figure 5.4 montre que la taille de la zone de valeurs fortes a tendance à être surestimée sur les cartes employant la transformation de coordonnées, menant à la surestimation des masses et volumes présents. Ceci était également observé à l'exemple 1 du chapitre précédent, lequel était très similaire à celui présenté ici.

Bien entendu, les deux scénarios présentés ici comportaient une certaine ressemblance avec le modèle de référence. En pratique, si le modèle d'écoulement est loin de représenter la réalité, il est fort plausible qu'il n'aide pas à améliorer l'estimation, voire qu'il y nuise carrément. À titre d'exemple, la carte krigée en coordonnées cartésiennes peut être vue comme un cas où l'on aurait spécifié un modèle d'écoulement homogène et sans recharge, donc peu représentatif. Dans ce cas, la faible connectivité entre les observations situées en $x = 50$ m et en $x = 150$ m peut être vue comme une indication des lacunes du modèle conceptuel. L'estimation de faibles paramètres de portées longitudinales pourrait également être l'indicateur d'un manque de continuité entre les concentrations observées. Ceci a fréquemment été rencontré avec les modèles en coordonnées cartésiennes². De plus, lorsqu'on compare les paramètres de covariance estimés pour le modèle NS2 avec les différentes coordonnées d'écoulement (voir Annexe E), on note que la portée longitudinale augmente avec le niveau d'information (du scénario 1 au modèle de référence). Ceci semble aller de pair avec l'amélioration de la modélisation de la connectivité entre les données situées le long de mêmes tubes de courant.

Le calage à des données de charge hydraulique peut contribuer à améliorer la représentativité du modèle. Par contre, un réseau d'écoulement qui s'écarte fortement de la réalité produira éventuellement des fragments de panache déconnectés les uns des autres, indiquant que le modèle conceptuel devrait potentiellement être révisé. Dans ce cas, l'approche de krigage en coordonnées d'écoulement peut donc être vue comme un outil de diagnostic permettant d'évaluer la cohérence entre le modèle conceptuel d'écoulement et les données de concentration.

La qualité de la transformation dépend évidemment de la représentativité du modèle d'écoulement. Or, la validité du modèle d'écoulement est lié à la qualité et la quantité d'information utilisée et aux hypothèses retenues pour représenter le modèle (Oreskes *et al.*, 1994). En général, la représentativité du modèle peut donc être pondérée par la qualité des travaux ayant mené à l'élaboration de ce modèle. Or, la question de la représentativité du modèle d'écoulement mène à l'utilisation de modèles d'écoulement multiples dans le but d'incorporer l'incertitude sur la transformation de coordonnées dans l'analyse. Ceci peut être fait en ayant recours à des modèles conceptuels liés à des interprétations concurrentes (tels que les scénarios 1 et 2), de même qu'en utilisant des réalisations multiples du champ de conductivité hydraulique issues de simulations géostatistiques, multipoints, ou de toute autre approche appropriée. Cette alternative permet d'obtenir une série de modèles d'écoulement plausibles,

2. On a souvent eu recours à des paramètres de portée *ad hoc* au chapitre précédent en raison de l'estimation déficiente de ces paramètres en coordonnées cartésiennes.

de déterminer les systèmes de coordonnées d'écoulement correspondants et de procéder à plusieurs estimations de la carte de concentration. L'estimé final pourrait être obtenu en faisant la moyenne arithmétique des concentrations krigées. Dans le cas d'un nombre limité de modèles conceptuels, une approche permettant de combiner les résultats par *Bayesian Model Averaging* (Hoeting *et al.*, 1999) pourrait être envisagée, en s'inspirant des travaux de Troldborg *et al.* (2010).

CHAPITRE 6

CONCLUSION

La faible performance des outils d'interpolation face à l'estimation de la contamination dans l'eau souterraine rend nécessaire le développement d'approches qui permettent de tenir compte du modèle physique sous-jacent. Certaines approches récentes permettent d'utiliser un modèle d'écoulement et de transport lors de l'estimation des valeurs de concentration, mais celles-ci requièrent généralement des connaissances précises sur la source (position, taille, intensité) ou sur le modèle d'écoulement (conductivité hydraulique, conditions frontières). Le recours à l'approche de Monte Carlo exige des simulations répétées de l'écoulement et du transport, ce qui rend ce type de méthodes très coûteuses en temps de calcul et peut limiter leur applicabilité dans certains cas.

La méthode de krigage développée dans cette thèse prend en compte le modèle d'écoulement par l'intermédiaire d'un modèle de covariance non stationnaire basé sur l'utilisation conjointe d'une transformation de coordonnées et une fonction de covariance comportant une paramétrisation spatialement variable. Dans un premier temps, la transformation en coordonnées d'écoulement permet de tenir compte de l'anisotropie locale induite par l'advection. En second lieu, le modèle de covariance non stationnaire peut être paramétrisé pour tenir compte de l'augmentation de la corrélation et de la diminution de la variance de la concentration avec la distance de transport, causées par la dispersion. Bien que la modélisation de l'écoulement soit requise, l'approche proposée ne nécessite pas la modélisation du transport de contaminant, ce qui réduit le niveau d'information (ou les hypothèses) requis, les efforts de modélisation et les temps de calcul. Un aspect important de l'approche proposée est qu'elle ne dépend pas d'une connaissance précise de la position de la source, ni de sa taille, ni de l'historique des émissions de contaminant (*release history*). De plus, l'applicabilité de la méthode n'est pas limitée aux contaminants non réactifs. Finalement, les paramètres du modèle de covariance sont estimés par maximum de vraisemblance ce qui permet de sélectionner par LRT, ou par AIC, le modèle le mieux adapté aux données parmi plusieurs paramétrisations candidates.

Les exemples numériques d'application montrent que les modèles proposés qui utilisent conjointement les coordonnées d'écoulement et les modèles de covariance non stationnaires expliquent mieux les données selon le maximum de vraisemblance. Les cartes de concentra-

tion obtenues apparaissent nettement plus réalistes que celles qui proviennent d'un krigage en coordonnées cartésiennes. Ceci est confirmé par l'amélioration des statistiques qui décrivent la capacité de classification des cartes krigées pour certains seuils de contamination. Les résultats obtenus pour l'estimation de la masse et du volume de contaminant présentent des améliorations moins systématiques lorsque comparés au krigage en coordonnées cartésiennes qui utilise les modèles de covariance non stationnaires. Toutefois, si l'on compare les résultats avec l'approche de krigage conventionnelle (coordonnées cartésiennes et modèle de covariance stationnaire), on note généralement une amélioration de l'ensemble des statistiques. Aucune des statistiques compilées n'exprime à elle seule la qualité de l'estimé. De bons estimés de la masse ou du volume contaminé ne veulent rien dire s'ils ne correspondent pas à des valeurs de concentration correctement localisées sur la carte. De même, une carte présentant une bonne corrélation ϕ , bien qu'elle indique la capacité à délimiter le panache, est peu utile en pratique si elle est associée à un estimé complètement erroné de la masse. Ainsi, il semble que les résultats issus de l'approche de krigage proposée montrent en général un bon compromis pour l'ensemble des statistiques colligées. Le faible nombre de données utilisées dans les exemples synthétiques présentés illustre la capacité de la méthode à faire une meilleure utilisation de l'information disponible, tel qu'illustré plus particulièrement par le cas 3D de l'exemple 2.

Face aux difficultés entourant la spécification des fonctions de courant en 3D, la présente thèse propose l'utilisation de coordonnées d'écoulement simplifiées (désignées coordonnées alternatives) basées sur les lignes de courant. Les résultats obtenus montrent qu'en dépit des approximations requises et des erreurs numériques sous-jacentes, ces coordonnées offrent une performance similaire aux coordonnées naturelles d'écoulement (en 2D) et permettent des gains face aux coordonnées cartésiennes en 3D. De plus, un exemple utilisant des coordonnées analogues aux coordonnées lagrangiennes (utilisation du temps comme coordonnée longitudinale) montre qu'il est possible de produire une interpolation plus réaliste d'un panache de contamination présentant des « doigts » distincts, sans recours à un échantillonnage exhaustif. De plus, bien que ceci ne soit pas présenté dans les exemples d'application, les coordonnées alternatives proposées, en permettant de s'affranchir de la fonction de courant, peuvent être utilisées pour des modèles d'écoulement comportant des puits ou des sources. En pratique, dans les cas où la fonction de courant est définie (2D sans puits, ni sources), on favorise l'utilisation des coordonnées naturelles (h, ψ) en raison de leur meilleure précision. Finalement, l'étude présentée au chapitre 5 montre que le recours à des réseaux d'écoulements simplifiés, équivalents à une certaine imprécision sur la transformation de coordonnées utilisée, n'engendre pas des différences importantes au niveau des statistiques compilées. Évi-

dement, ceci n'est pas vrai si le modèle d'écoulement est complètement erroné.

Dans les exemples d'application utilisés, les paramètres estimés par maximum de vraisemblance et les critères de sélection de modèles associés présentent un bilan mitigé. D'une part, les paramètres estimés pour les modèles non stationnaires produisent des résultats acceptables. Les modèles non stationnaires NS2 et NS3 sont systématiquement choisis comme optimaux. D'autre part, la procédure de sélection ne parvient pas à identifier le modèle (et les paramètres) produisant le meilleur estimé en ce qui a trait aux statistiques compilées et à l'aspect visuel des cartes obtenues. L'optimum de la vraisemblance ne garantit pas nécessairement une estimation optimale de la concentration. Bien que ceci n'ait pas été vérifié systématiquement, la validation croisée a semblé plus robuste que le maximum de vraisemblance pour l'estimation des paramètres et la sélection de modèle. L'absence d'une procédure fiable pour l'estimation des paramètres peut être problématique dans les applications réelles, puisqu'on ne pourra éventuellement pas identifier le modèle le plus performant. Une spécification de paramètres *ad hoc* peut être envisagée afin de permettre l'interpolation des données de concentration et la visualisation du panache, mais on veillera alors à interpréter le krigage comme un interpolateur déterministe.

Contribution originale

L'approche de krigage proposée se veut un compromis entre les approches d'interpolation conventionnelles et celles plus complexes basées sur l'inversion. Elle comporte les éléments novateurs suivants :

- une transformation de coordonnées basée sur le modèle physique de l'écoulement en milieu poreux ;
- l'utilisation conjointe de la classe de fonctions de covariance non stationnaires proposée par Higdon *et al.* (1999) et d'une transformation de coordonnées, en vue de faciliter la paramétrisation de la covariance ;
- des paramétrisations simplifiées du modèle de covariance, basées sur les études stochastique du transport de contaminant ;
- des coordonnées d'écoulement approximatives basées sur le trajectographie et une représentation en coordonnées locales des frontières de recharge du modèle d'écoulement.

On note également que l'approche développée peut s'avérer un outil intéressant pour évaluer la cohérence entre le modèle conceptuel d'écoulement et les données de concentration. Le manque de connectivité entre les observations qui présentent des valeurs élevées de

concentration peut éventuellement être perçu comme un indicateur de la déficience du modèle. L'incapacité des méthodes d'estimation des paramètres à ajuster un modèle ayant une certaine continuité spatiale, illustrée par des paramètres de portée longitudinale très faibles, peut également laisser présager une incohérence entre les données et le modèle d'écoulement utilisé pour la transformation de coordonnées. Toutefois, cette analyse demeure avant tout qualitative et doit faire appel au jugement et à l'expérience de l'hydrogéologue responsable de la modélisation. Étant donné qu'elle requiert une quantité d'information relativement faible, cette approche peut s'avérer un outil de visualisation intéressant lors des phases préliminaires d'une étude environnementale. Finalement, l'approche permet, par l'intermédiaire de variances d'estimation qui tiennent compte du modèle d'écoulement, de mieux cibler l'implantation de nouveaux forages ou d'un réseau de suivi.

Développements futurs

En ce qui a trait au modèle de covariance, d'autres paramétrisations pourraient éventuellement être envisagées. Par exemple, les modèles proposés supposent que la source est située complètement en amont du modèle, tel qu'exprimé par la variation des paramètres en fonction de $h_{max} - h$. En supposant davantage de connaissances sur la position de la source, il pourrait être possible d'élaborer des paramétrisations qui tiennent compte d'une position quelconque de la source dans le réseau d'écoulement. De plus, d'autres types de fonctions pourraient être envisagés pour modéliser l'évolution des portées et de la variance sur le domaine. On pense, par exemple, à l'utilisation de la racine carrée de $h_{max} - h$ pour les fonctions décrivant les portées (Cirpka *et al.*, 2011b). Cette fonction serait toutefois plus sensible à la localisation de h_{max} que la fonction linéaire utilisée. Les portées spatiales pourraient éventuellement être rattachées à la dispersion (Kitanidis, 1994).

Dans le même ordre d'idée, des connaissances liées à la position de la source permettrait éventuellement une modélisation plus réaliste de la moyenne non conditionnelle de la concentration et ainsi de minimiser le biais positif engendré par l'utilisation d'une moyenne constante (et positive) lors du krigeage ordinaire, laquelle n'est pas représentative en dehors du panache. La fonction de densité de probabilité gaussienne est parfois évoquée pour décrire la forme d'un panache et celle-ci pourrait servir de base à la définition d'une variable auxiliaire, laquelle pourrait alors être utilisée dans un krigeage avec dérive externe. Toutefois, en cas d'une mauvaise spécification d'une telle fonction, des biais importants sont à prévoir dans les concentrations estimées. Le recours à une modèle de dérive plus complexe ramène

la question de l'estimation des paramètres géostatistiques par maximum de vraisemblance restreint et les doutes quant la capacité de cette méthode à identifier le modèle de covariance proposé.

En ce qui concerne les coordonnées d'écoulement alternatives, la définition des coordonnées u et v devrait idéalement être reliée au flux traversant les frontières de recharge afin de se rapprocher de l'interprétation physique de la fonction de courant et ainsi, de tenir compte d'une distance plus plausible entre les lignes de courant et assurer que les tubes de courant représentent le même débit. Toutefois, ce problème semble loin d'être trivial en raison de l'indétermination associée à la définition des fonctions de courant en 3D. Des techniques similaires à celles utilisées pour la représentation 2D de surfaces tridimensionnelles pourraient être utilisées à cette fin, avec le flux hydraulique comme critère d'optimisation (voir Levy et Mallet (1998) et Mallet (2002)). Le recours à des éléments finis de type mixte hybride (Chavent et Robert, 1991) pourrait permettre d'améliorer la précision des vitesses et ainsi, le traçage des lignes de courant, tel que dans les travaux de Cirpka *et al.* (1999a).

L'établissement d'une procédure plus robuste pour l'estimation des paramètres et la sélection des modèles de covariance serait souhaitable. La validation croisée constitue une alternative et il n'a pas été vérifié systématiquement si elle permettait une sélection plus robuste des paramètres et du modèle. L'absence de tests statistiques pour évaluer le caractère significatif des gains liés à l'ajout de paramètres au modèle de covariance rend toutefois cette méthode moins attrayante. Comme il apparaît que les différences les plus marquées entre les panaches estimés sont dues aux paramètres de portée utilisés, on pourrait également envisager l'utilisation de paramètres *ad hoc* permettant de produire des panaches de tailles minimale, maximale et intermédiaire. Les analyses subséquentes pourraient alors tenir compte de cette étendue de possibilités.

La méthode peut être modifiée pour inclure l'incertitude sur les coordonnées d'écoulement en utilisant plusieurs réalisations du réseau d'écoulement. Ceci permettrait également de vérifier l'importance relative de l'erreur numérique par rapport à l'incertitude sur le modèle d'écoulement. L'incertitude liée à des modèles conceptuels multiples pourrait être intégrée à l'analyse par *Bayesian Model Averaging*. Ceci permettrait de tenir compte de conditions d'écoulement incertaines, telle que la recharge.

Il pourrait également être intéressant de tester de l'approche dans des contextes différents de ceux présentés aux chapitres quatre et cinq. On pourrait ainsi évaluer sa robustesse face

à des panaches de contamination produits par des conditions d'écoulement transitoires, ou face à des panaches issus d'une source instantanée ou de sources multiples. L'influence du nombre de données sur la performance de l'approche proposée par rapport à la méthode de krigeage conventionnelle pourrait aussi être explorée.

Globalement, les développements permettant l'amélioration de la précision des concentrations estimées sont primordiaux, car ils permettraient d'envisager sérieusement l'utilisation de l'approche développée pour l'estimation des flux de masses et ainsi, de quantifier l'atténuation naturelle sur un site.

RÉFÉRENCES

- ABOUFIRASSI, M. et MARINO, M. A. (1983). Kriging of water levels in the Souss aquifer, Morocco. *Mathematical Geology*, 15, 537–551.
- AHMED, S. et DE MARSILY, G. (1987). Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity. *Water Resources Research*, 23, 1717–1737.
- AKIMA, H. (1974). A method of bivariate interpolation and smooth surface fitting based on local procedures. *Commun. ACM*, 17, 18–20.
- ALMENDRAL, A., ABRAHAMSEN, P. et HAUGE, R. (2008). Multidimensional scaling and anisotropic covariance functions. *Proceedings of the Eight International Geostatistics Congress. Santiago, Chile*.
- ANDERES, E. B. et STEIN, M. (2011). Local likelihood estimation for nonstationary random fields. *Journal of Multivariate Analysis*, 102, 506–520.
- ANDERES, E. B. et STEIN, M. L. (2008). Estimating deformations of isotropic gaussian random fields on the plane. *Annals of Statistics*, 36, 710–714.
- ANDERSON, M. P. et WOESSNER, W. W. (2002). *Applied Groundwater Modeling : Flow and advective simulation*. Academic Press, New York.
- BARABAS, N., GOOVAERTS, P. et ADRIAENS, P. (2001). Geostatistical assessment and validation of uncertainty for three-dimensional dioxin data from sediments in an estuarine river. *Environmental Science and Technology*, 35, 3294–3301.
- BEAR, J. (1972). *Dynamics of Fluids in Porous Media*. American Elsevier.
- BHATTACHARYYA, B. K. (1969). Bicubic spline interpolation as a method for treatment of potential field data. *Geophysics*, 34, 402–423.
- BOCKELMANN, A., ZAMFIRESCU, D., PTAK, T., GRATHWOHL, P. et TEUTSCH, G. (2003). Quantification of mass fluxes and natural attenuation rates at an industrial site with limited monitoring network : a case study. *Journal of Contaminant Hydrology*, 60, 97–121.
- BOISVERT, J. B. et DEUTSCH, C. V. (2010). Programs for kriging and sequential gaussian simulation with locally varying anisotropy using non-euclidean distances. *Computers & Geosciences*, 41, 585–601.
- BOISVERT, J. B., MANCHUK, J. G. et DEUTSCH, C. V. (2009). Kriging in the presence of locally varying anisotropy using non-euclidean distances. *Mathematical Geosciences*, 41, 585–601.

- BRAMLETT, W. et BORDEN, R. C. (1990). Numerical generation of flow nets - the flows model. *Ground Water*, 28, 946–950.
- BROCHU, Y. et MARCOTTE, D. (2003). A simple approach to account for radial flow and boundary conditions when kriging hydraulic head fields for confined aquifers. *Mathematical Geology*, 35, 111–136.
- BURNHAM, K. et ANDERSON, D. (2002). *Model Selection and Multimodel Inference : A Practical Information-theoretic Approach*. Springer, New York.
- BUXTON, H. T. et MODICA, E. (1992). Patterns and rates of ground-water flow on long island, new york. *Ground Water*, 30, 857–866.
- CARONI, E. et FIOROTTO, V. (2005). Analysis of concentration as sampled in natural aquifers. *Transport in Porous Media*, 59, 19–45.
- CHAPRA, S. et CANALE, R. (2002). *Numerical Methods for Engineers*. McGraw-Hill, quatrième édition.
- CHAVENT, G. et ROBERT, J. (1991). A unified physical presentation of mixed, mixed-hybrid finite elements and standard finite difference approximations for the determination of velocities in waterflow problems. *Advances in Water Resources*, 14, 329–348.
- CHILÈS, J. P. (1977). *Géostatistique des phénomènes non stationnaires (dans le plan)*. Thèse de doctorat, Université de Nancy - I.
- CHILÈS, J. P. et DELFINER, P. (1999). *Geostatistics : Modeling Spatial Uncertainty*. Wiley-Interscience, New York, première édition.
- CHRISTAKOS, G., HRISTOPULOS, D. T. et BOGAERT, P. (2000). On the physical geometry concept at the basis of space/time geostatistical hydrology. *Advances in Water Resources*, 23, 799–810.
- CIRPKA, O., DE BARROS, F., CHIOGNA, G. et NOWAK, W. (2011a). Probability density function of steady-state concentration in two-dimensional heterogeneous porous media. *Water Resources Research*, 47, W11523.
- CIRPKA, O. A., DE BARROS, F. P. J., CHIOGNA, G., ROLLE, M. et NOWAK, W. (2011b). Stochastic flux related analysis of transverse mixing in two dimensional heterogeneous porous media. *Water Resources Research*, 47, W06515.
- CIRPKA, O. A., FRIND, E. O. et HELMIG, R. (1999a). Numerical methods for reactive transport on rectangular and streamline-oriented grids. *Advances in Water Resources*, 22, 711–728.
- CIRPKA, O. A., FRIND, E. O. et HELMIG, R. (1999b). Streamline-oriented grid generation for transport modelling in two-dimensional domains including wells. *Advances in Water Resources*, 22, 697–710.

- CLERC, M. et MALLAT, S. (2003). Estimating deformations of stationary processes. *The Annals of Statistics*, 31, 1772–1821.
- CORDES, C. et KINZELBACH, W. (1992). Continuous groundwater velocity fields and path lines in linear, bilinear, and trilinear finite elements. *Water Resources Research*, 28, 2903–2911.
- CRESSIE, N. (1993). *Statistics for Spatial Data, revised ed.* Wiley, New York.
- CURRIERO, F. C. (1996). *The use of non-euclidean distances in geostatistics.* Thèse de doctorat, Kansas State University.
- CURRIERO, F. C. (2006). On the use of non-euclidean distance measures in geostatistics. *Mathematical Geology*, 38, 907–926.
- DAGAN, G. (1982). Stochastic modeling of groundwater flow by unconditional and conditional probabilities. 2. The solute transport. *Water Resources Research*, 18, 835–848.
- DAGAN, G. (1984). Solute transport in heterogeneous porous formations. *Journal of Fluid Mechanics*, 145, 151–177.
- DAGBERT, M., DAVID, M., CROZEL, D. et DESBARATS, A. (1984). Computing variograms in folded strata-controlled deposits. G. Vergy, M. David, A. Journel et A. Maréchal, éditeurs, *Geostatistics for Natural Resources Characterization*. D. Reidel, 71–89.
- DAMIAN, D., SAMPSON, P. et GUTTORP, P. (2001). Bayesian estimation of semi-parametric non-stationary spatial covariance structure. *Environmetrics*, 12, 161–178.
- DAVENPORT, E. C. et EL-SANHURRY, N. A. (1991). Phi/phimax : review and synthesis. *Educational and Psychological Measurements*, 51, 821–828.
- DAVID, M. (1988). *Handbook of Applied Advanced Geostatistical Ore Reserve Estimation.* Elsevier, Amsterdam.
- DE BARROS, F. et NOWAK, W. (2010). On the link between contaminant source release conditions and plume prediction uncertainty. *Journal of Contaminant Hydrology*, 116, 24–34.
- DELHOMME, J. P. (1978). Kriging in the hydrosciences. *Advances in Water Resources*, 1, 251–266.
- DENTZ, M. (2012). Concentration statistics for transport in heterogeneous media due to stochastic fluctuations of the center of mass velocity. *Advances in Water Resources*, 36, 11–22.
- DEUTSCH, C. et WANG, L. (1996). Hierarchical object-based stochastic modeling of fluvial reservoirs. *Mathematical Geology*, 28, 857–879.
- DEUTSCH, C. V. (2002). *Geostatistical Reservoir Modeling.* Oxford University Press.

- DOMENICO, P. A. et SCHWARTZ, F. W. (1998). *Physical and chemical hydrogeology*. John Wiley & Sons, New York ; Toronto, seconde édition.
- DOWD, P. et PARDO-IGUZQUIZA, E. (2012). Parametric inference of spatial covariance. P. Abrahamsen, R. Hauge et O. Kolbjornsen, éditeurs, *Geostatistics Oslo 2012*.
- DUCHON, J. (1977). Splines minimizing rotation-invariant semi-norms in sobolev spaces. W. Schempp et K. Zeller, éditeurs, *Constructive Theory of Functions of Several Variables*, Springer Berlin / Heidelberg, vol. 571 de *Lecture Notes in Mathematics*. 85–100. 10.1007/BFb0086566.
- FETTER, C. W. (1993). *Contaminant Hydrogeology*. Macmillan Publishing Company.
- FEYEN, L. et CAERS, J. (2006). Quantifying geological uncertainty for flow and transport modeling in multi-modal heterogeneous formations. *Advances in Water Resources*, 29, 912–929.
- FIOROTTO, V. et CARONI, E. (2002). Solute concentration statistics in heterogeneous aquifers for finite peclet values. *Transport in porous media*, 48, 331–351.
- FITTS, C. (2002). *Groundwater science*. Academic Press, San Diego, CA.
- FOGG, G. E. et SENGER, R. K. (1985). Automatic generation of flow nets with conventional ground-water modeling algorithms. *Ground Water*, 23, 336–344.
- FRIND, E. O. (1982). The principal direction technique : a new approach to groundwater contaminant transport modeling. *Finite Elements in Water Resources. Proceedings of the 4th International Conference, Hanover, Germany, June 1982*.
- FRIND, E. O. et MATANGA, G. B. (1985). The dual formulation of flow for contaminant transport modeling 1. review of theory and accuracy aspects. *Water Resources Research*, 21, 159–169.
- FUENTES, M. (2001). A high frequency kriging approach for non-stationary environmental processes. *Environmetrics*, 12, 469–438.
- GAMBOLATI, G. et VOLPI, G. (1979a). A conceptual deterministic analysis of the kriging technique in hydrology. *Water Resources Research*, 15, 625–629.
- GAMBOLATI, G. et VOLPI, G. (1979b). Groundwater contour mapping in Venice by stochastic interpolators 1. Theory. *Water Resources Research*, 12, 281–290.
- GELHAR, L. W. (1993). *Stochastic subsurface hydrology*. Prentice-Hall, Englewood Cliffs, New Jersey.
- GELHAR, L. W. et AXNESS, C. L. (1983). Three-dimensional stochastic analysis of macrodispersion. *Water Resources Research*, 19, 161–180.

- GRAHAM, W. D. et MCLAUGHLIN, D. (1989a). Stochastic analysis of nonstationary subsurface solute transport. 1. Unconditional moments. *Water Resources Research*, 25, 215–232.
- GRAHAM, W. D. et MCLAUGHLIN, D. (1989b). Stochastic analysis of nonstationary subsurface solute transport. 2. Conditional moments. *Water Resources Research*, 25, 2331–2355.
- GRAHAM, W. D. et MCLAUGHLIN, D. (1991). A stochastic model of solute transport in groundwater : application to the Borden, Ontario, tracer test. *Water Resources Research*, 27, 1345–1359.
- GUYON, X. et PERRIN, O. (2000). Identification of space deformation using linear and superficial quadratic deformation. *Statistics & Probability Letters*, 47, 307–316.
- HAITJEMA, H. M. et KELSON, V. A. (1996). Using the stream function for flow governed by poisson's equation. *Journal of Hydrology*, 187, 367.
- HIGDON, D., SWALL, J. et KERN, J. (1999). Non-stationary spatial modeling. J. Bernardo, J. O. Berger, A. P. Dawid et A. Smith, éditeurs, *Bayesian Statistics 6*, Oxford University Press. 761–768.
- HOETING, J., DAVIS, R., MERTON, A. et THOMPSON, S. (2006). Model selection for geostatistical models. *Ecological Applications*, 16, 87–98.
- HOETING, J. A., MADIGAN, D. AND RAFTERY, A. et VOLINSKY, C. (1999). Bayesian model averaging : A tutorial. *Statistical Science*, 14, 382–401.
- ISTOK, J. D., SMYTH, J. D. et FLINT, A. L. (1993). Multivariate geostatistical analysis of ground-water contamination : a case study. *Ground Water*, 31, 63–74.
- JONES, N. L., DAVIS, R. J. et SABBAH, W. (2003). A comparison of three-dimensional interpolation techniques for plume characterization. *Ground Water*, 41, 411–419.
- JUN, M. et STEIN, M. (2008). Nonstationary covariance models for global data. *The Annals of Applied Statistics*, 2, 1271 ?1289.
- KAPOOR, V. et GELHAR, L. W. (1994). Transport in 3-dimensionally heterogeneous aquifers .1. Dynamics of concentration fluctuations. *Water Resources Research*, 30, 1775–1788.
- KAPOOR, V. et KITANIDIS, P. (1998). Concentration fluctuations and dilution in aquifers. *Water Resources Research*, 34, 1181–1193.
- KENDALL, M. et STUART, A. (1977). *The Advanced Theory of Statistics*. Macmillan.
- KIM, H. J. et CAI, L. (1993). Robustness of the likelihood ratio test for a change in simple linear regression. *Journal of the American Statistical Association*, 88, 864–871.

- KITANIDIS, P. (1987). Parametric estimation of covariances of regionalized variables. *Water Resources Bulletin*, 23, 557–567.
- KITANIDIS, P. (1994). The concept of the dilution index. *Water Resources Research*, 30, 2011–2026.
- KITANIDIS, P. (1995). Quasi-linear geostatistical theory for inversing. *Water resources research*, 31, 2411–2420.
- KITANIDIS, P. et SHEN, K.-F. (1996). Geostatistical interpolation of chemical concentration. *Advances in Water Resources*, 19, 369–378.
- KLEIBER, W. et NYCHKA, D. (2012). Nonstationary modeling for multivariate spatial processes. *Journal of Multivariate Analysis*, 112, 76–91.
- KOHAVI, R. et PROVOST, F. (1998). Glossary of terms. *Machine Learning*, 30, 271–274.
- KUBAT, M., HOLTE, R. C. et MATWIN, S. (1998). Machine learning for the detection of oil spills in satellite radar images. *Machine Learning*, 30, 195–215.
- LEGLEITER, C. J. et KYRIAKIDIS, P. C. (2006). Forward and inverse transformation between cartesian and channel-fitted coordinate systems for meandering rivers. *Mathematical Geology*, 38, 927–958.
- LEVY, B. et MALLET, J.-L. (1998). Non-distorted texture mapping for sheared triangulated meshes. *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics*. New York, NY, USA, 343 – 52.
- LOLAND, A. et HOST, G. (2003). Spatial covariance modelling in a complex coastal domain by multidimensional scaling. *Environmetrics*, 14, 307–321.
- MALLET, J. (2002). *Geomodeling*. Oxford University Press.
- MARCOTTE, D. (1995). Generalized cross-validation for covariance model selection. *Mathematical Geology*, 27, 659–672.
- MARCOTTE, D. et DAVID, M. (1988). Trend surface analysis as a special case for IRF-k kriging. *Mathematical Geology*, 20, 821–824.
- MARDIA, K., KENT, J. et BIBBY, J. (1979). *Multivariate Analysis*. Academic Press.
- MATANGA, G. B. (1993). Stream functions in three-dimensional groundwater flow. *Water Resources Research*, 29, 3125–3133.
- MATEU, J., FERNÉNDEZ-AVILÉS, G. et MONTERO, J. (2011). On a class of non-stationary, compactly supported spatial covariance functions. *Stochastic Environmental Research and Risk Assessment*, 1–13. 10.1007/s00477-011-0510-8.
- MATHERON, G. (1971). La théorie des fonctions aléatoires intinsèques généralisées. Rapport technique, École des mines de Paris, Centre de morphologie mathématique. N-252.

- MATHERON, G. (1981). Down-to-earth statistics : Solutions looking for geological problems, chapter splines and kriging : their formal equivalence.
- MONESTIEZ, P., SAMPSON, P. et GUTTORP, P. (1993). Modeling of heterogeneous spatial correlation structure by spatial deformation. *Cahiers de Géostatistique, Fascicule 3*, 35–46.
- NARASIMHAN, T. N. (2008). Laplace equation and Faraday's lines of force. *Water Resources Research*, 44, W09412.
- NELDER, J. et MEAD, R. (1965). A simplex method for function minimization. *Computer Journal*, 7, 308–313.
- NOWAK, W., DE BARROS, F. P. et RUBIN, Y. (2010). Bayesian geostatistical design : Task-driven optimal site investigation when the geostatistical model is uncertain. *Water Resources Research*, 46, W03535.
- NYCHKA, D., W. C. et ROYLE, K. (2002). Multiresolution models for nonstationary spatial covariance functions. *Stat. Model.*, 2, 315–332.
- ORESKES, N., SHRADER-FRECHETTE, K. et BELITZ, K. (1994). Verification, validation and confirmation of numerical models in the earth sciences. *Science*, 263, 641.
- PACIOREK, C. (2003). *Nonstationary Gaussian Processes for Regression and Spatial Modelling*. Thèse de doctorat, Carnegie Mellon University.
- PACIOREK, C. et SCHERVISH, M. J. (2006). Spatial modeling using a new class of nonstationary covariance functions. *Environmetrics*, 17, 483–506.
- PARADIGMTM (2009). *GOCAD[®] 2009.1 User Guide. Part IV Foundation Modeling*.
- PASQUIER, P. (2005). *Résolution du problème inverse en hydrogéologie par une estimation successive des flux*. Thèse de doctorat, École Polytechnique de Montréal.
- PASQUIER, P. et MARCOTTE, D. (2006). Steady- and transient-state inversion in hydrogeology by successive flux estimation. *Advances in Water Resources*, 29, 1934–1952.
- PEPPER, D. et HEINRICH, J. C. (2006). *The Finite Element Method : Basic Concepts and Applications*. Taylor & Francis.
- PERRIN, O. et MEIRING, W. (1999). Identifiability for non-stationary spatial structure. *Journal of Applied Probability*, 36, 1244–1250.
- PERRIN, O. et SENOUSSI, R. (2000). Reducing non-stationary random fields to stationarity and isotropy using a space deformation. *Statistics & Probability Letters*, 48, 23–32.
- PINDER, F. et GRAY, W. (1977). *Finite Element Simulation in Surface and Subsurface Hydrology*. Academic Press. 295 p.

- PONZINI, G. et LOZEJ, A. (1982). Identification of aquifer transmissivities : The comparison model method. *Water Resources Research*, 18, 597–622.
- REED, P. M., ELLSWORTH, T. R. et MINSKER, B. S. (2004). Spatial interpolation methods for nonstationary plume data. *Ground Water*, 42, 190–202.
- RIVEST, M. et MARCOTTE, D. (2012). Kriging groundwater solute concentrations using flow coordinates and nonstationary covariance functions. *Journal of Hydrology*, 472-473, 238–253.
- RIVEST, M., MARCOTTE, D. et PASQUIER, P. (2008). Hydraulic head field estimation using kriging with an external drift : a way to consider conceptual model information. *Journal of Hydrology*, 361, 349–361.
- RIVEST, M., MARCOTTE, D. et PASQUIER, P. (2012a). Interpolation of concentration measurements by kriging using flow coordinates. P. Abrahamsen, R. Hauge et O. Kolbjornsen, éditeurs, *Geostatistics Oslo 2012*, Springer Netherlands, vol. 17. 519–530.
- RIVEST, M., MARCOTTE, D. et PASQUIER, P. (2012b). Sparse data integration for the interpolation of concentration measurements using kriging in natural coordinates. *Journal of Hydrology*, 416, 72–82.
- RUBIN, Y. (1990). Stochastic modeling of macrodispersion in heterogeneous porous media. *Water Resources Research*, 26, 133–141.
- RUBIN, Y. (1991). The spatial and temporal moments of tracer concentration in disordered porous media. *Water Resources Research*, 27, 2845–2854.
- SAITO, H. et GOOVAERTS, P. (2001). Accounting for source location and transport direction into geostatistical prediction of contaminants. *Environmental Science and Technology*, 35, 4823–4829.
- SAMPSON, P. D. et GUTTORP, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87, 108–119.
- SCHMIDT, A. et O'HAGAN, A. (2003). Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 65, 743–758.
- SCHMIDT, A. M. (2001). *Bayesian spatial interpolation of environmental monitoring stations*. Thèse de doctorat, University of Sheffield.
- SCHWARZ, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.
- SCHWEDE, R., CIRPKA, O., NOWAK, W. et NEUWEILER, I. (2008). Impact of sampling volume on the probability density function of steady state concentration. *Water Resources Research*, 44, W12433.

- SCHWEDE, R. L. et CIRPKA, O. A. (2010). Interpolation of steady-state concentration data by inverse modeling. *Ground Water*, 48, 569–579.
- SEMPRINI, L., KITANIDIS, P. K., KAMPBELL, D. et WILSON, J. (1995). Anaerobic transformation of chlorinated aliphatic hydrocarbons in a sand aquifer based on spatial chemical distributions. *Water Resources Research*, 31, 1051–1062.
- SHEPARD, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 23rd ACM National Conference*. ACM, New York, NY, USA, 517–524.
- SHLOMI, S. et MICHALAK, A. M. (2007). A geostatistical framework for incorporating transport information in estimating the distribution of a groundwater contaminant plume. *Water Resources Research*, 43, W03412.
- SIBSON, R. (1981). A brief description of natural neighbor interpolation. V. Barnett, éditeur, *Interpreting Multivariate Data*, John Wiley. pp. 21–36.
- SOARES, A. (1990). Geostatistical estimation of orebody geometry : Morphological kriging. *Mathematical Geology*, 22, 787–802.
- SRIVASTAVA, R. et BRUSSEAU, M. (1995). Darcy velocity computations in the finite element method for multidimensional randomly heterogeneous porous media. *Advances in Water Resources*, 18, 191–201.
- STATISTIQUE CANADA (2003). Human activity and the environment : Annual statistics 2003. Rapport technique.
- STEIN, M. (2005). Nonstationary spatial covariance functions. Rapport technique CISES Technical Report 21, University of Chicago.
- STEIN, M. L. (1999). *Interpolation of Spatial Data : Some Theory for Kriging*. Springer-Verlag, New York.
- STEWARD, D. (1998). Stream surfaces in two-dimensional and three-dimensional divergence-free flows. *Water Resources Research*, 34, 1345–1350.
- SWALL, J. L. (1999). *Non-Stationary spatial modeling using a process convolution approach*. Thèse de doctorat, Duke University.
- TE STROET, C. B. M. et SNEPVANGERS, J. J. J. C. (2005). Mapping curvilinear structures with local anisotropy kriging. *Mathematical Geology*, 36, 635–649.
- THOMPSON, J. F., WARSI, Z. U. A. et MASTIN, C. W. (1985). *Numerical Grid Generation : Foundations and Applications*. North-Holland.
- TONKIN, M. et LARSON, S. (2002). Kriging water levels with a regional-linear and point logarithmic drift. *Ground Water*, 33, 185–193.

- TROLDORG, M., NOWAK, W., TUXEN, N., BJERG, P., HELMIG, R. et BINNING, P. J. (2010). Uncertainty evaluation of mass discharge estimates from a contaminated site using a fully bayesian framework. *Water Resources Research*, 46, W12552.
- VANDERBORGHT, J. (2001). Concentration variance and spatial covariance in second-order stationary heterogeneous conductivity fields. *Water Resources Research*, 37, 1893–1912.
- VANDERBORGHT, J., KASTEEL, R. et VEREECKEN, H. (2006). Stochastic continuum transport equations for field-scale solute transport : Overview of theoretical and experimental results. *Vadose Zone Journal*, 5, 184–203.
- VOMVORIS, E. et GELHAR, L. (1986). Stochastic prediction of dispersive contaminant transport. Rapport technique CR-811135-01-2, U.S. Environmental Protection Agency, Washington D.C.
- VOMVORIS, E. G. et GELHAR, L. W. (1990). Stochastic analysis of the concentration variability in a three-dimensional heterogeneous aquifer. *Water Resources Research*, 26, 2591–2602.
- WARRENS, M. J. (2008). On association coefficients for 2x2 tables and properties that do not depend on the marginal distributions. *Psychometrika*, 73, 777–789.
- WATSON, D. (1992). *Contouring : A Guide to the Analysis and Display of Spatial Data, with programs on diskette*. Oxford ; Pergamon Press, New York.
- XU, W. (1996). Conditional curvilinear stochastic simulation using pixel-based algorithms. *Mathematical Geology*, 28, 937–949.
- YUAN, K. H. et BENTLER, P. M. (1999). On normal theory and associated test statistics in covariance structure analysis under two classes of nonnormal distributions. *Statistica Sinica*, 9, 831–853.

ANNEXE A

ANNEXE A : Exemple simple avec un nombre de données variable

On présente ici une étude de cas préliminaire retrouvée dans Rivest *et al.* (2012b). Cet exemple simple démontre la capacité de la transformation de coordonnée à tenir compte d'une anisotropie variable et d'exploiter plus efficacement un nombre limité d'observations. Il illustre aussi certains inconvénients liés à la transformation de coordonnées.

Un domaine de $100 \text{ m} \times 100 \text{ m}$ comporte deux zones de conductivité hydraulique différentes, illustrées à la figure A.1(a). Le réseau d'écoulement est représenté à la figure A.1(b). On simule le déversement d'un contaminant sur la frontière $x = 0 \text{ m}$, lequel produit le panache montré à la figure A.1(c).

L'écoulement est considéré parfaitement connu et les paramètres utilisés pour les interpolation ont été estimés par maximum de vraisemblance. Trois scénarios d'échantillonnage sont utilisés :

- 25 données localisées aléatoirement ;
- 18 données sur une grille régulière ;
- 8 données sur une grille régulière.

La figure A.2 montre les résultats du krigeage en coordonnées cartésiennes pour les trois scénarios présentés précédemment. On y voit que ce type de krigeage ne parvient pas à interpoler les valeurs dans la direction pertinente au phénomène. Le modèle ajusté produit un étalement artificiel des valeurs de concentrations (figure A.2(a)) ou un panache fragmenté (figure A.2(b)). L'anisotropie utilisée est parallèle aux axes, mais on note que l'utilisation d'un angle d'anisotropie n'a pas permis d'obtenir des résultats vraiment plus acceptables.

La figure A.3 montre quant à elle que les résultats sont considérablement améliorés par le recours à la transformation en coordonnées naturelles d'écoulement puisqu'elle permet d'étendre l'influence des observations disponibles dans la direction d'écoulement. Cet exemple montre que malgré le peu d'observations échantillonnant des valeurs élevées de concentration, il est possible d'avoir une carte réaliste, du moins en ce qui concerne la forme du panache estimé. Les erreurs absolues moyennes obtenues pour les six krigeages sont présentés au ta-

bleau A.1 et montrent que l'utilisation de la transformation permet d'améliorer la précision des cartes krigées.

En dépit de l'amélioration obtenue par la transformation de coordonnées, certains éléments problématiques sont retrouvés sur les cartes krigées de la figure A.3. Pour le scénario à huit données (figure A.3(c)), bien que la forme du panache soit relativement bien estimée considérant qu'une seule donnée échantillonne le panache, on remarque que la faible portée dans la direction transversale ne permet pas d'étendre latéralement les valeurs nulles de concentration. Loin des observations, le krigeage ordinaire retourne la moyenne des observations et entraîne l'estimation de valeurs positives dans une zone où les données de concentration environnantes sont pourtant nulles. Ce biais est dû à la moyenne constante utilisée, laquelle ne représente pas adéquatement le phénomène. L'augmentation de la portée, en augmentant la zone d'influence des données, pourrait améliorer la reproduction de valeurs nulles mais entraînerait également un panache trop large.

L'utilisation des coordonnées naturelles d'écoulement peut également donner lieu à des artefacts d'interpolation, tel que sur la figure A.3(b). Dans ce cas, la donnée située en (65,50) produit un étalement artificiel du panache dans la zone imperméable. Ceci est dû à la proximité de ces points dans l'espace transformé $h - \psi$.

L'exemple présenté illustre donc le potentiel de la transformation en coordonnées d'écoulement, de même que d'éventuels points faibles.

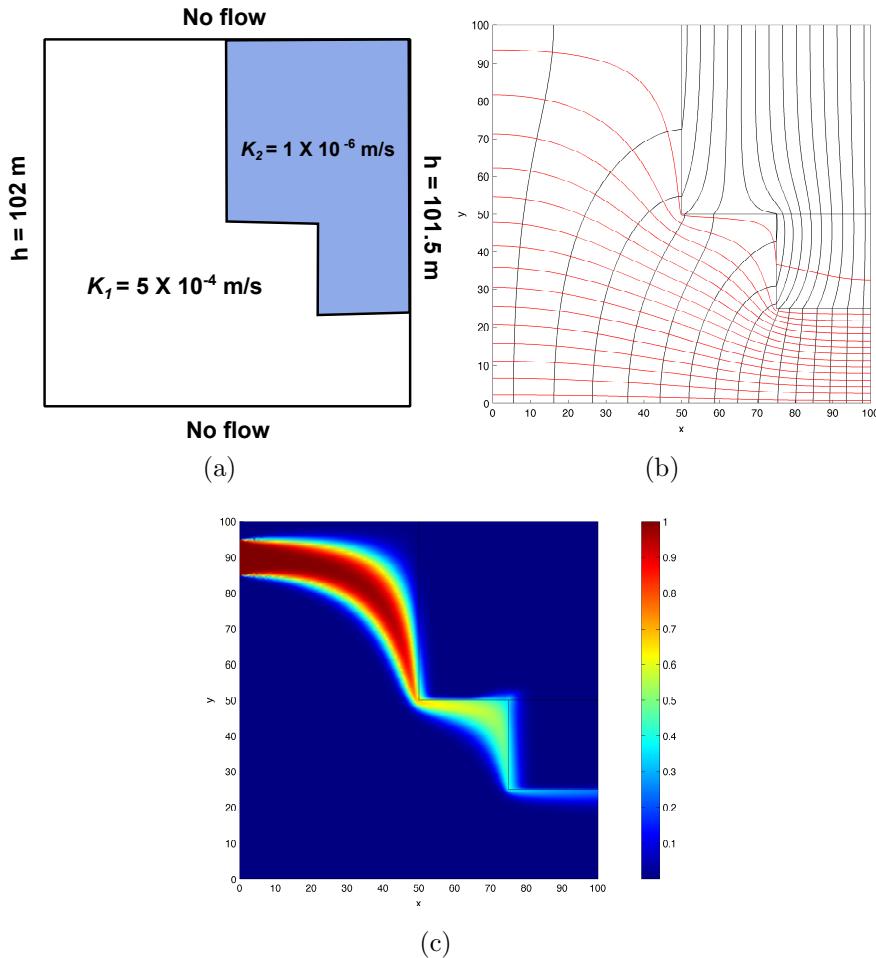


Figure A.1 Modèle de référence. (a) Modèle et conditions frontières, (b) Réseau d'écoulement de référence avec les équipotentielles en noir et les lignes de courant en rouge, (c) Concentration de référence

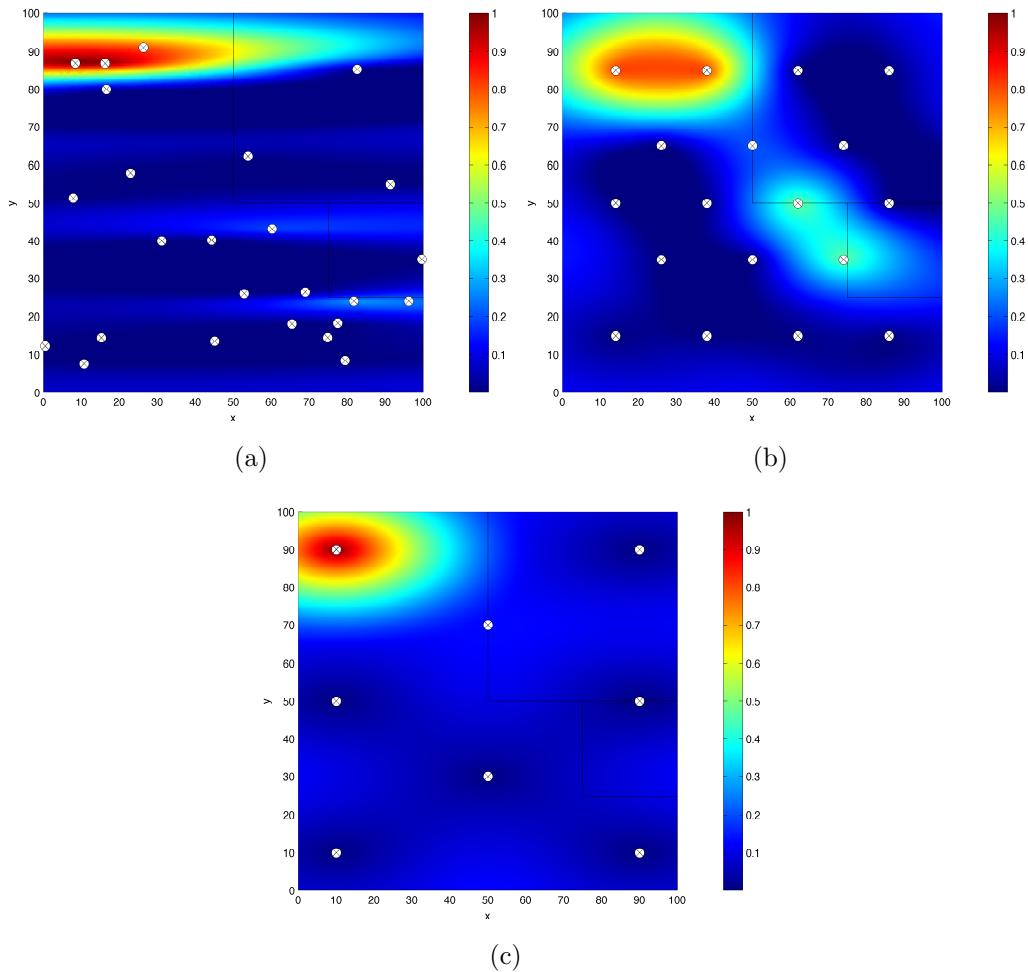


Figure A.2 Résultats du krigage en coordonnées cartésiennes avec (a) 25 données, (b) 18 données, et (c) 8 données.

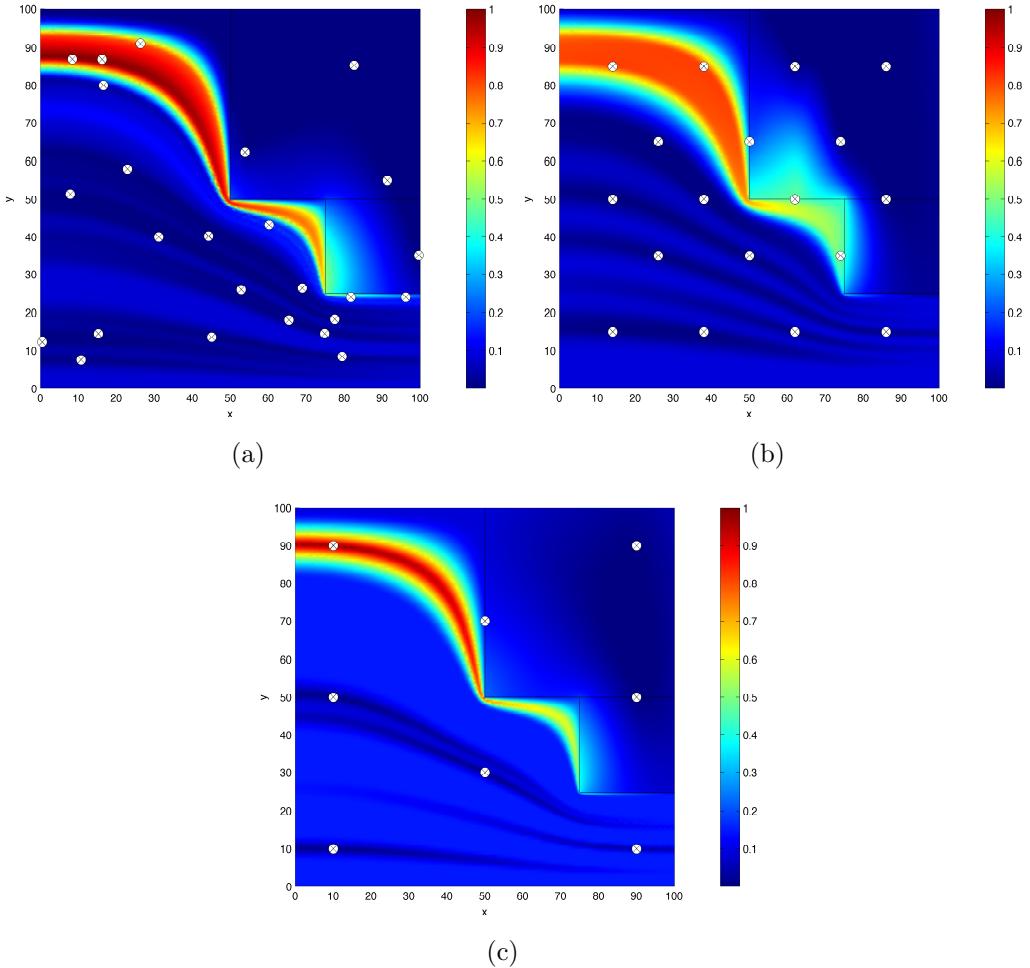


Figure A.3 Résultats du krigeage en coordonnées naturelles d'écoulement (a) 25 données, (b) 18 données, et (c) 8 données.

Tableau A.1 Erreur absolue moyenne (MAE) pour le krigeage en coordonnées cartésiennes (CCK) et le krigeage en coordonnées naturelles d'écoulement (NCK).

	MAE ₂₅	MAE ₁₈	MAE ₈
CCK	0.1099	0.1177	0.1263
NCK _R	0.0460	0.0721	0.1047

ANNEXE B

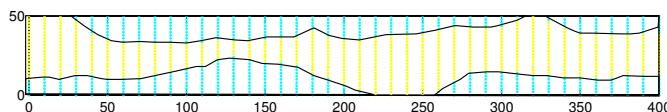
ANNEXE B : Construction des champs de conductivité hydraulique hétérogènes

La figure B.1 illustre la méthode utilisée pour construire les champs de conductivité hydraulique *de référence* utilisés dans les exemples des chapitres 4 et 5.

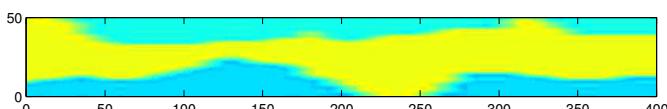
1. Définition des structures



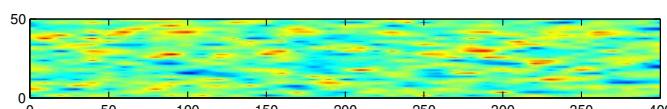
2. Assignation d'une grille dense de valeurs de K



3. Krigeage de K



4. Krigeage d'un ΔK de faible amplitude, aléatoire et structuré



5. Addition de champs définis en 3. et 4.

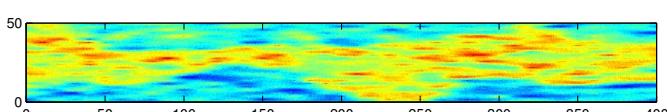


Figure B.1 Schématisation de la méthode utilisée pour construire les champs de conductivité hydraulique. On note que l'échelle de valeurs est différente pour les figures 4 et 5.

ANNEXE C

ANNEXE C : Évaluation sommaire de la sensibilité des résultats au nombre de lignes de courant utilisées pour obtenir les coordonnées u et v

Comme on l'a vu au chapitre 3, les coordonnées d'écoulement alternatives transversales (u et v) sont obtenues à l'aide d'une approche basée sur le traçage des lignes de courant. Une interpolation linéaire est utilisée pour étendre les valeurs obtenues à partir d'une « grille » de lignes de courant à l'ensemble des points d'estimation et d'observation requis pour le krigage.

On s'intéresse ici à l'influence de l'espacement des lignes de courant utilisées sur les résultats de l'exemple 1 du chapitre 4. Le tableau C.1 montre les résultats obtenus pour le modèle CAE/NS2 pour trois densités de lignes de courant correspondant à des lignes espacées de $\Delta = 5$ m, $\Delta = 10$ m et $\Delta = 25$ m. Il est à noter que le cas avec $\Delta = 10$ m est celui présenté au chapitre 4.

Au tableau C.1, on remarque que les statistiques associées au seuil 0.01 sont très similaires pour chaque densité de ligne de courant. Ceci est moins vrai pour celles associées au seuil 0.1. Ainsi, la précision des coordonnées semble avoir un impact sur la reproduction des valeurs de concentration plus élevées. La figure C.1 montre globalement une bonne ressemblance entre les cartes obtenues, en particulier celles où la coordonnée u a été calculée avec $\Delta = 10$ m et $\Delta = 5$ m. On note cependant que la carte krigée avec une coordonnées u moins précise ($\Delta = 25$ m) présente un panache légèrement moins large, lequel est toutefois entouré d'un halo diffus de valeurs non-nulles qui est moins prononcé sur les autres cartes.

Malgré ces différences, les résultats montrent tous une amélioration par rapport au krigage en coordonnées cartésiennes (voir tableau 4.6 au chapitre 4).

Tableau C.1 Exemple 1. Résultats obtenus avec le modèle CAE_{*h*}/NS2, pour différents espacements des lignes de courant utilisées lors du calcul de la coordonnées *u*.

<i>n</i> lignes	RMSE	RAE _{<i>m</i>}	RAE _{<i>vol</i>}		ϕ		<i>r</i>	VM
			$c_s = 0.01$	$c_s = 0.1$	$c_s = 0.01$	$c_s = 0.1$		
$\Delta = 5$ m (91 lignes)	0.2353	0.4333	0.2972	0.4319	0.5566	0.5986	0.7881	283.4805
$\Delta = 10$ m (46 lignes)	0.2363	0.5058	0.2967	0.3490	0.5425	0.6288	0.7504	219.0056
$\Delta = 25$ m (19 lignes)	0.2359	0.5094	0.2946	0.6180	0.5257	0.5917	0.8050	224.8098

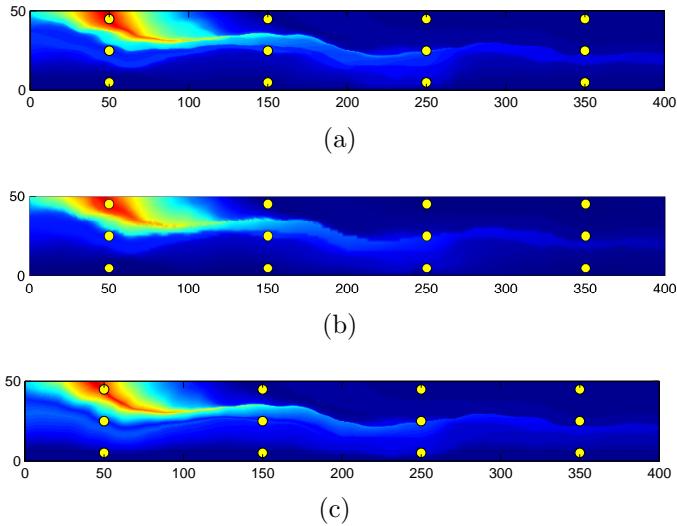


Figure C.1 Cartes de concentration krigées en coordonnées alternatives d'écoulement (modèle CAE_{*h*}/NS2) pour différentes densités de lignes de courant. (a) $\Delta = 5$ m, (b) $\Delta = 10$ m, (c) $\Delta = 25$ m.

ANNEXE D

ANNEXE D : Résultats pour l'exemple 1 avec une configuration de données différente

On présente ici les résultats obtenus pour l'exemple 1 du chapitre 4 en utilisant une configuration différente des observations. On utilise toujours douze observations, toutefois celles-ci sont réparties sur trois puits au lieu de quatre. Cette nouvelle configuration est illustrée à la figure D.1.

Comme pour le cas à quatre puits, le modèle NS2 est celui retenu, et ce pour les deux systèmes de coordonnées (LRT et AIC_c non-présentés). Le tableau D.1 montre les résultats obtenus pour tous les modèles, en coordonnées cartésiennes et en coordonnées naturelles d'écoulement. Les modèles NS2 et NS3 en coordonnées d'écoulement montrent une amélioration par rapport aux modèles en coordonnées cartésiennes pour l'ensemble des statistiques, exceptée l'erreur sur la masse. Ainsi, malgré certaines différences avec les résultats obtenus pour la configuration à quatre puits, la modification de la configuration des données n'altère pas les conclusions formulées pour l'exemple 1.

Tableau D.1 Résultats pour le modèle utilisé à l'exemple 1 avec une autre configuration des données. c_s : seuil de contamination.

Modèle	RMSE	RAE_m	RAE_{vol}		ϕ		r
			$c_s = 0.01$	$c_s = 0.1$	$c_s = 0.01$	$c_s = 0.1$	
CC/S1	0.2685	0.3602	0.6739	0.9358	-	0.2032	0.2567
CC/NS1	0.2757	0.06977	0.4484	0.2101	0.4190	0.4973	0.7143
CC/NS2	0.2780	0.0594	0.4551	0.3246	0.4702	0.3681	0.7018
CC/NS3	0.2610	0.2152	0.5055	0.4205	0.4049	0.3281	0.6666
CNE/S1	0.2581	0.5500	0.3909	0.3269	0.3326	0.4735	0.7383
CNE/NS1	0.2600	0.1512	0.2236	0.06630	0.4468	0.6132	0.8319
CNE/NS2	0.2321	0.3694	0.00808	0.1549	0.5586	0.5760	0.8039
CNE/NS3	0.2377	0.2999	0.2388	0.04348	0.5669	0.6250	0.8389

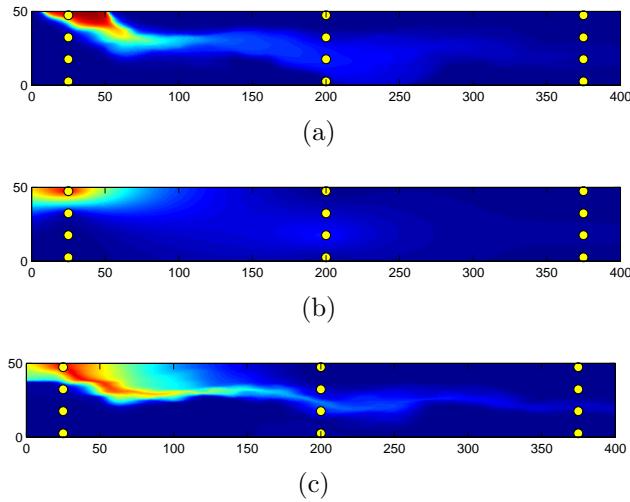


Figure D.1 Résultats pour l'exemple 1 avec une configuration différente des données (a) Concentration de référence, (b) CC/NS2, (c) CNE/NS2

ANNEXE E

ANNEXE E : Compilation des paramètres de covariance utilisés dans les exemples des chapitres 4 et 5

Les paramètres présentés aux tableaux suivants ont été introduits précédemment au chapitre 3. Les paramètres de portée a_{p0} et b_p sont définis à l'équation 3.43 et les paramètres de variance σ_0^2 et d sont définis à l'équation 3.45. Au tableau E.4, les paramètres associés à la dérive externe m_0 et m_1 sont présentés à l'équation 3.62.

Lors de l'estimation des paramètres (par EMV ou par validation croisée), ceux-ci sont contraints de manière à être positifs. Des limites supérieures *ad hoc* sont également imposées : 1) les portées estimées dans chaque direction sont limitées aux valeurs maximales de chaque coordonnée, 2) la variance maximale est limitée au double de la variance des observations.

Paramètres de l'exemple 1

Tableau E.1 Paramètres du modèle de covariance non-stationnaire : Exemple 1 - coordonnées cartésiennes. Coordonnées maximales : $x_{max} = 400$ m, $y_{max} = 50$ m

Modèle	a_{x0}	b_x	a_{y0}	b_y	σ_0^2	d
S1	3.08	0	8.45	0	0.059	0
NS1	17.08	0.23	0.20	0.12	0.056	0
NS2	61.86	0	21.99	0	0.13	0.0091
NS3	68.28	0.0063	9.23	0.10	0.13	0.0094

Tableau E.2 Paramètres du modèle de covariance non-stationnaire : Exemple 1 - coordonnées naturelles d'écoulement. Coordonnées maximales : $\Delta h_{max} = 0.3$ m, $\psi_{max} = 6.658 \times 10^{-6}$ m/s

Modèle	$a_{\Delta h0}$	$b_{\Delta h}$	$a_{\psi0}$	b_{ψ}	σ_0^2	d
S1	0.12	0	4.48×10^{-7}	0	0.062	0
NS1	0.30	0.0096	4.83×10^{-7}	1.19×10^{-5}	0.10	0
NS2	0.20	0	1.06×10^{-6}	0	0.13	12.21
NS3	0.29	0	2.01×10^{-8}	2.22×10^{-5}	0.10	9.97

Tableau E.3 Paramètres du modèle de covariance non-stationnaire : Exemple 1 - coordonnées alternatives d'écoulement. Coordonnées maximales : $\Delta h_{max} = 0.3$ m, $u_{max} = 200$ m

Modèle	$a_{\Delta h0}$	$b_{\Delta h}$	a_{u0}	b_u	σ_0^2	d
S1	0.30	0	11.51	0	0.062	0
NS1	0.30	0.013	0	641.28	0.062	0
NS2	0.22	0	25.23	0	0.13	11.74
NS3	0.30	0.0074	7.25	642.26	0.13	12.20

Tableau E.4 Paramètres du modèle de covariance non-stationnaire : Exemple 1 - coordonnées naturelles d'écoulement (avec dérive linéaire). Coordonnées maximales : $\Delta h_{max} = 0.3$ m, $\psi_{max} = 6.658 \times 10^{-6}$ m/s

Modèle	$a_{\Delta h0}$	$b_{\Delta h}$	$a_{\psi0}$	b_ψ	σ_0^2	d	m_0	m_1
S1	0.12	0	1.90×10^{-7}	0	0.046	0	0.31	1.35
NS1	0.30	1.30×10^{-8}	1.63×10^{-8}	2.22×10^{-5}	0.052	0	0.30	1.10
NS2	0.14	0	2.27×10^{-7}	0	0.07	15.35	0.17	0.67
NS3	0.27	0.10	8.28×10^{-8}	1.23×10^{-5}	0.081	15.35	0.16	0.59

Paramètres de l'exemple 2

Tableau E.5 Paramètres du modèle de covariance non-stationnaire : Exemple 2 - coordonnées cartésiennes). Coordonnées maximales : $x_{max} = 200$, $y_{max} = 100$ m, $z_{max} = 50$ m

Modèle	a_{x0}	b_x	a_{y0}	b_y	a_{z0}	b_z	σ_0^2	d
S1*	2.82	0	4.19	0	0	0	0.028	0
S1	1.19	0	1.19	0	4.23	0	0.028	0
NS1*	195.10	1.17×10^{-9}	1.87	0.24	0.03	0	0	0
NS1	198.55	0.0072	5.22	0.47	3.85	0.231	0.031	0
NS2*	38.88	0	18.03	0	0.00	0	0.059	0.022
NS2	34.08	0	36.89	0	15.89	0	0.059	0.023
NS3*	51.81	0.0024	2.13	0.24	0	0	0.057	0.023
NS3	139.92	1.59×10^{-5}	2.27	0.49	1.31	0.243	0.055	0.023

Tableau E.6 Paramètres du modèle de covariance non-stationnaire : Exemple 2 - coordonnées alternatives d'écoulement). Coordonnées maximales : $\Delta h_{max} = 0.3$ m, $u_{max} = 100$ m, $v_{max} = 50$ m

Modèle	$a_{\Delta h0}$	$b_{\Delta h}$	a_{u0}	b_u	a_{v0}	b_v	σ_0^2	d
S1*	0.30	0	4.87	0	0	0	0.028	0
S1	0.081	0	12.59	0	6.89	0	0.029	0
NS1*	0.24	0.22	1.90	166	0	0	0.028	0
NS1	0.20	0.23	0	333.00	0.21	166	0.027	0
NS2*	0.12	0	14.78	0	0	0	0.059	14.84
NS2	0.14	0	28.54	0	13.57	0	0.059	15.17
NS3*	0.095	0	1.95	162.78	0	0	0.057	16.00
NS3	0.096	0	0	323.90	0	151.55	0.053	15.84

Tableau E.7 Paramètres du modèle de covariance non-stationnaire (Validation-croisée) : Exemple 2 - coordonnées alternatives d'écoulement). Coordonnées maximales : $\Delta h_{max} = 0.3$ m, $u_{max} = 100$ m, $v_{max} = 50$ m

Modèle	$a_{\Delta h0}$	$b_{\Delta h}$	a_{u0}	b_u	a_{v0}	b_v	σ_0^2	d
S1*	0.3	0	5.01	0	0	0	0.036	0
NS1*	0.3	0	4.92	105.09	0	0	0.037	0
NS2*	0.3	0	9.40	0	0	0	0.0068	16.00
NS3*	0.28	0.058	8.44	31.60	0	0	0.021	16.00

Paramètres de l'exemple 3

Tableau E.8 Paramètres du modèle de covariance non-stationnaire : Exemple 3 - coordonnées naturelles d'écoulement CAE_{*h*}. Coordonnées maximales : $\Delta h_{max} = 0.3$ m, $u_{max} = 338$ m

Modèle	$a_{\Delta h0}$	$b_{\Delta h}$	a_{u0}	b_u	σ_0^2	d
S1	0.3	0	6.51	0	0.05	0
NS1	0.3	0	1.95	960.40	0.052	0
NS2	0.20	0	53.66	0	0.12	10.24
NS3	0.3	0	12.87	643.41	0.10	9.52

Tableau E.9 Paramètres du modèle de covariance non-stationnaire : Exemple 3 - coordonnées naturelles d'écoulement CAE_{*t*}. Coordonnées maximales : $\frac{t_{max}}{1 \times 10^7} = 6346$ s/1 $\times 10^7$, $u_{max} = 338$ m

Modèle	a_{t0}	b_t	a_{u0}	b_u	σ_0^2	d
S1	801.78	0	40.44	0	0.063	0
NS1	374.89	0.94	5.33	0.052	0.064	0
NS2	6345.81	0	9.31	0	0.081	0.00073
NS3	1496.83	0	11.80	0.051	0.12	0.00073

Tableau E.10 Paramètres du modèle de covariance non-stationnaire (Validation-croisée) : Exemple 3 - coordonnées naturelles d'écoulement CAE_{*t*}. Coordonnées maximales : $\frac{t_{max}}{1 \times 10^7} = 6346$ s/1 $\times 10^7$, $u_{max} = 338$ m

Modèle	a_{t0}	b_t	a_{u0}	b_u	σ_0^2	d
S1	6345.81	0	102.01	0	0.12	0
NS1	4214.41	0.34	82.14	0	3.24×10^{-8}	0
NS2	6014.86	0	26.73	0	0.12	0.00073
NS3	1313.09	0.79	15.59	0	0.042	0.00073

Paramètres pour le chapitre 5

Tableau E.11 Paramètres du modèle de covariance non-stationnaire : Exemple chapitre 5 - coordonnées cartésiennes. Coordonnées maximales : $x_{max} = 400$ m, $y_{max} = 50$ m

Modèle	a_{x0}	b_x	a_{y0}	b_y	σ_0^2	d
S1	3.82	0	17.48	0	0.071	0
NS1	0.003	0.99	8.07	0.10	0.081	0
NS2	57.66	0	19.63	0	0.15	0.0071
NS3	23.34	0.94	2.88	0.12	0.15	0.0061

Tableau E.12 Paramètres du modèle de covariance non-stationnaire : Exemple chapitre 5 - coordonnées naturelles d'écoulement pour le scénario 1. Coordonnées maximales : $\Delta h_{max} = 0.3$ m, $\psi_{max} = 7.948 \times 10^{-6}$ m/s

Modèle	$a_{\Delta h0}$	$b_{\Delta h}$	$a_{\psi0}$	b_ψ	σ_0^2	d
S1	0.06	0	5.57×10^{-7}	0	0.067	0
NS1	0.29	0.0297	0	2.60×10^{-5}	0.087	0
NS2	0.09	0	1.03×10^{-6}	0	0.15	8.99
NS3	0.24	0	0	2.42×10^{-5}	0.15	7.97

Tableau E.13 Paramètres du modèle de covariance non-stationnaire : Exemple chapitre 5 - coordonnées naturelles d'écoulement pour le scénario 2. Coordonnées maximales : $\Delta h_{max} = 0.3$ m, $\psi_{max} = 1.043 \times 10^{-5}$ m/s

Modèle	$a_{\Delta h0}$	$b_{\Delta h}$	$a_{\psi0}$	b_ψ	σ_0^2	d
S1	0.09	0	6.81×10^{-7}	0	0.068	0
NS1	0.30	5.39×10^{-9}	0	3.46×10^{-5}	0.074	0
NS2	0.17	0	1.51×10^{-6}	0	0.15	9.43
NS3	0.30	7.16×10^{-4}	0	2.89×10^{-5}	0.15	8.62

Tableau E.14 Paramètres du modèle de covariance non-stationnaire : Exemple chapitre 5 - coordonnées naturelles d'écoulement de référence. Coordonnées maximales : $\Delta h_{max} = 0.3$ m, $\psi_{max} = 7.397 \times 10^{-6}$ m/s

Modèle	$a_{\Delta h0}$	$b_{\Delta h}$	$a_{\psi0}$	b_ψ	σ_0^2	d
S1	0.30	0	5.70×10^{-7}	0	0.067	0
NS1	0.30	2.65×10^{-4}	3.40×10^{-7}	2.42×10^{-5}	0.078	0
NS2	0.30	0	9.49×10^{-7}	0	0.15	10.09
NS3	0.29	4.94×10^{-6}	7.00×10^{-7}	1.04×10^{-5}	0.15	9.87