

Table des matières

Liste des symboles et abréviations	v
Introduction	1
I Contexte et méthodes développées	5
1 Approches usuelles et état de l'art	7
1.1 L'approche géostatistique pour caractériser les contaminations	7
1.2 Le cas particulier du cadre spatio-temporel	12
1.3 La modélisation à base physique	15
1.4 Combinaison de simulations numériques et de la géostatistique	18
1.5 Synthèse du chapitre	20
2 Mise en œuvre des méthodes géostatistiques développées	21
2.1 Le krigeage avec dérive externe	21
2.2 Les covariances et variogrammes numériques	22
2.3 Synthèse du chapitre	26
II Applications à un cas synthétique	27
3 Construction d'un cas synthétique en zone non saturée	29
3.1 Les paramètres hydrodynamiques de la zone non saturée	29
3.2 Influence de la variabilité spatiale des paramètres du modèle de Mualem-van Genuchten sur la migration d'un panache de tritium	30
3.3 Utilisation des résultats précédents pour construire un cas synthétique	38
3.4 Synthèse du chapitre	44
4 Application des méthodes combinant la géostatistique et les simulations numériques	45
4.1 Scénarios d'échantillonnage des activités en tritium	45
4.2 Exploitation des simulations de panaches de tritium	45
4.3 Estimations de l'activité en tritium	47
4.4 Quantification des incertitudes	59
4.5 Un cas multivariable	64
4.6 Synthèse du chapitre	70
5 Estimations spatio-temporelles et prévisions	71
5.1 Position du problème et ensemble de particules	71
5.2 Mise en œuvre des méthodes spatio-temporelles	72
5.3 Prévisions	74
5.4 Krigeage spatial ou krigeage spatio-temporel ?	77
5.5 Recherche de la localisation de la source	79
5.6 Synthèse du chapitre	81

III Applications à deux jeux de données réels acquis autour de la centrale de Tchernobyl	83
Présentation des sites	85
6 Application à une expérience de traçage	87
6.1 Présentation de l'expérience et des données	87
6.2 Simulations d'écoulement et de transport avec MELODIE	89
6.3 Estimations de l'activité en ^{36}Cl	91
6.4 Synthèse du chapitre	96
7 Application à un site d'enfouissement de déchets radioactifs	97
7.1 Présentation du site et des données	97
7.2 Simulations d'écoulement et de transport avec MELODIE	101
7.3 Combinaison du krigeage et de simulations d'écoulement et de transport : cadre spatial . .	107
7.4 Combinaison du krigeage et de simulations d'écoulement et de transport : cadre temporel .	114
7.5 Synthèse du chapitre	121
Conclusion	123
Annexes	128
A Vignette RGeostats	131
B Influence des portées des champs de paramètres de Mualem-van Genuchten	137
C Résultats complémentaires sur l'influence de la variabilité des paramètres	139
D Construction de références supplémentaires	141
E Résultats sur les références supplémentaires	143
F Résultats complémentaires de prévisions	153
G Analyses granulométriques à Prypiat Zaton et sur le site Epic	157
H Paramètres de Mualem-van Genuchten (zone de traçage)	159
I Résultats complémentaires sur les données de la T22	161
Références	165

Liste des symboles et abréviations

Abréviations

ANDRA : Agence Nationale pour la gestion des Déchets Radioactifs

ASN : Autorité de Sûreté Nucléaire

BRGM : Bureau de Recherches Géologiques et Minières

CEA : Commissariat à l'Énergie Atomique et aux Énergies Alternatives

IRSN : Institut de Radioprotection et de Sûreté Nucléaire

OCDE : Organisation de Coopération et de Développement Économiques

Symboles mathématiques

x point d'un espace à une ou plusieurs dimensions (\mathbb{R} , \mathbb{R}^2 ou \mathbb{R}^3)

$E[Z(x)]$ espérance de $Z(x)$, parfois notée $m(x)$

$\text{Var}[Z(x)]$ variance de $Z(x)$

$\text{Cov}[Z(x_1), Z(x_2)]$ covariance de $Z(x_1)$ et $Z(x_2)$

$\overline{Z(x)}$ moyenne empirique de $Z(x)$

$C(x, x')$ covariance de Z entre les points x et x'

$\gamma(x, x')$ variogramme de Z entre les points x et x'

$\mathcal{N}(\mu, \sigma^2)$ loi normale de moyenne μ et de variance σ^2

$\mathbb{1}_{\mathcal{A}}$ indicatrice de l'ensemble \mathcal{A} ; $\mathbb{1}_{\mathcal{A}}(x) = 1$ si $x \in \mathcal{A}$ et $\mathbb{1}_{\mathcal{A}}(x) = 0$ si $x \notin \mathcal{A}$

I_n matrice identité de taille $n \times n$

tM transposée de la matrice M

$\text{Tr}(M)$ trace de la matrice M

$||.||$ distance euclidienne

$\langle . \rangle$ produit scalaire usuel

Unités des grandeurs introduites

M : unité de masse dans le système international

L : unité de longueur dans le système international

T : unité de temps dans le système international

Introduction

La caractérisation des zones contaminées et ses enjeux

Les sites et sols pollués constituent une problématique importante dans les anciennes régions industrielles. En France, de nombreux outils sont développés dans l'objectif de les recenser, comme la base de données Basol [Site du Ministère de la Transition Écologique et Solidaire (a)]. Ces sites représentent un enjeu conséquent en terme d'aménagement du territoire : la reconquête de friches industrielles permet de réhabiliter des espaces, souvent en zone urbaine, et constitue une grande part du marché de la dépollution [Site du Ministère de la Transition Écologique et Solidaire (b)]. Selon les usages prévus des sites réhabilités (zone industrielle, logements, etc.), les exigences de dépollution et de surveillance varient.

Les sites contaminés par des substances radiologiques sont traités différemment des autres sites pollués. Les modalités de stockage des déchets radioactifs, et par conséquent les coûts associés, varient selon le type de déchet. Le problème des sites marqués par des substances radiologiques se pose surtout au moment de leur démantèlement. L'IRSN recense les installations nucléaires en cours de démantèlement en France : neufs réacteurs EDF, une usine de retraitement de combustible d'Orano et des installations du CEA en janvier 2020 [Site de l'IRSN (a)]. Les installations nucléaires ne sont pas la seule source de contamination radiologique : les anciennes mines d'uranium [Site de l'IRSN (b)] ou les sites industriels qui utilisaient du radium jusqu'aux années 1960 [Site de l'IRSN (c)] sont également des sources importantes de contamination radiologique.

L'Agence pour l'Énergie Nucléaire de l'OCDE rappelle l'importance de délimiter la contamination dans les sols et les nappes pendant le fonctionnement des installations, sans attendre leur arrêt [Nuclear Energy Agency, 2018]. Les difficultés à délimiter ces contaminations sont souvent soulignées [OECD et NEA, 2014], bien que des campagnes de caractérisation des sols et des nappes soient en général réalisées en amont (comme recommandé par l'IRSN et l'ASN [IRSN et ASN, 2011]). Ces difficultés rendent complexe la prévision des coûts associés à l'excavation et au stockage des terres contaminées.

Ainsi, la caractérisation des contaminations résultant d'activités industrielles dans les sols ou les nappes est un enjeu majeur de la réhabilitation de ces sites [Last *et al.*, 2004 ; Zhang *et al.*, 2010]. Ces enjeux sont à la fois sanitaires, environnementaux, économiques (coûts de traitement et stockage) et fonciers (reconquête de territoires). Il est donc important de fournir des caractérisations précises des étendues et niveaux de contamination. Cela permet de décider des méthodes de remédiation et de surveillance du site adaptées, en prenant en compte à la fois la contamination présente et l'usage futur du site, dans l'objectif de réduire les risques environnementaux, économiques et sociétaux [Schädler *et al.*, 2011 ; Chen *et al.*, 2019].

Objectifs et contexte scientifique : le projet Kri-Terres

Cette thèse fait partie du projet Kri-Terres^{1 2}, accompagné par l'ANDRA dans le cadre du programme "Nucléaire de demain" des Investissements d'Avenir. Ce projet vise à améliorer la caractérisation de la contamination radioactive autour des centrales à démanteler, via deux axes d'étude. Le premier axe, traité dans cette thèse, consiste à développer des méthodes de modélisation permettant d'estimer précisément les étendues et niveaux de contamination. Le second axe, non développé dans cette thèse, consiste à mettre au point des méthodes de mesure *in situ* des propriétés des sols, pour mieux caractériser la variabilité spatiale de ces propriétés à l'échelle d'un site afin d'améliorer la modélisation des écoulements et du transport de contaminants dans la zone non saturée [Houzé *et al.*, 2019].

L'objectif du projet est d'apporter des éléments de réflexion sur des questions opérationnelles de remédiation et surveillance des sites : méthodes les plus adaptées, emplacement optimal des piézomètres pour

1. Fiche de présentation du projet : <https://www.andra.fr/sites/default/files/2019-09/Fiche%20projet%20KRI-TERRES%20VF-FR.pdf>

2. Vidéo de présentation du projet : <https://www.youtube.com/watch?v=Qf52qL8IiDA>

la surveillance du site, paramètres de modélisation sur lesquels réaliser les études préliminaires les plus approfondies, ...

La caractérisation de la contamination dans les sols et les nappes autour d'installations industrielles ou nucléaires est complexe, pour plusieurs raisons :

- l'historique des activités (différents usages du site, substances utilisées, éventuels incidents, sources potentielles de contamination), notamment sur les petites installations industrielles utilisant des substances radioactives depuis le début du XX^{ème} siècle, n'est pas toujours connu précisément ;
- l'accès aux données peut être difficile (coûts des prélèvements et analyses, présence d'infrastructures empêchant l'échantillonnage, risques sanitaires lors de l'accès à des zones fortement contaminées) ;
- les niveaux de contamination sont fortement non stationnaires, les panaches peuvent présenter des formes complexes, et ces niveaux et formes évoluent au cours du temps ;
- la contamination peut être due à plusieurs éléments, avec des propriétés physico-chimiques contrastées, et par conséquent des évolutions spatiales et temporelles différentes ;
- la compréhension des phénomènes régissant la migration des contaminants est complexe, ainsi que leur modélisation (nombreux phénomènes physiques et chimiques, difficulté à estimer les paramètres de modélisation).

Ainsi, les méthodes de caractérisation des contaminations dans les sols et les nappes doivent tenir compte de ces difficultés et permettre de travailler dans différents cadres : modélisation spatiale (estimation des niveaux de contamination à une date fixée), modélisation spatio-temporelle (prévision de l'évolution de la contamination à plus ou moins long terme), modélisation multivariable (plusieurs contaminants). La démarche développée dans le projet Kri-Terres vise à combiner deux types d'approches de caractérisation, les approches probabilistes et les approches à base physique, afin de bénéficier des forces de chacune pour obtenir des estimations plus réalistes ou plus précises.

Dans les études de sites contaminés, les approches probabilistes sont encore majoritairement géostatistiques, bien que les méthodes d'apprentissage statistique se développent (Tableau 1). Les approches à base physique, ou approches phénoménologiques, sont constituées par des modélisations des phénomènes physiques à l'origine de la migration des contaminants, à différents niveaux de complexité (modèles conceptuels simples, modélisations mathématiques effectuées avec des codes de calcul, etc.).

TABLEAU 1 – Comparaison du nombre de résultats pour deux recherches différentes sur *ScienceDirect*. Recherche 1 : '(geostatistics OR kriging) AND (contamination OR pollution)' dans le titre, le résumé ou les mots-clés des publications. Recherche 2 : 'machine learning AND (contamination OR pollution)' dans le titre, le résumé ou les mots-clés des publications.

	Recherche 1	Recherche 2
Entre 2000 et 2010	151	5
Entre 2010 et 2020	495	288

Le principal avantage de la géostatistique est d'estimer les niveaux de contamination en respectant les mesures (interpolation exacte), par la modélisation de la structure spatiale de la variable étudiée. La grande variété des modèles géostatistiques et les nombreux outils opérationnels disponibles rendent l'estimation géostatistique facilement réalisable et adaptable. Cependant, ces estimations n'intègrent en général pas les processus physiques régissant les écoulements et le transport, ce qui peut conduire à des estimations qui ne sont pas réalistes.

Au contraire, les modèles à base physique reproduisent une partie des phénomènes physiques et chimiques identifiés dans les processus de migration des contaminants. Cela garantit le réalisme des panaches de contaminant simulés, avec une précision dépendante de la complexité de la modélisation. Cependant, pour réaliser des simulations reproduisant au mieux les mesures, il est nécessaire de caractériser précisément les phénomènes à modéliser et les paramètres associés, ce qui est en général une procédure longue, coûteuse, et mal posée (plusieurs jeux de paramètres donnant des résultats similaires). Enfin, si les paramètres inconnus sont trop nombreux, les phénomènes complexes, et l'origine de la contamination mal caractérisée, il n'est pas garanti que cette modélisation reflète convenablement la réalité.

Au regard des avantages et limites des approches classiques, il apparaît nécessaire de mettre au point des méthodes combinant l'estimation géostatistique et la modélisation d'écoulement et de transport, afin de fournir des estimations physiquement réalistes respectant les mesures. Des méthodes de ce type existent dans la littérature. Il s'agit par exemple d'utiliser la modélisation à base physique pour :

- modéliser la tendance du phénomène étudié (krigeage avec dérive externe [Rivest *et al.*, 2008 ; Varentsov *et al.*, 2019]) ;
- inférer les propriétés spatiales ou spatio-temporelles du phénomène étudié (covariances empiriques [Roth *et al.*, 1998 ; Yang *et al.*, 2018] ou filtre de Kalman d'ensemble [Evensen, 2006]).

Des travaux récents proposent également des méthodes d'apprentissage statistique à base physique [Raissi *et al.*, 2019 ; Tartakovsky *et al.*, 2019].

Dans la littérature, plusieurs exemples d'application concernent des estimations de charges hydrauliques [Roth *et al.*, 1998 ; Rivest *et al.*, 2008], et seul l'écoulement est modélisé. Lorsque le transport de contaminants est considéré, les applications sont effectuées sur des cas synthétiques à deux dimensions et en zone saturée [Shlomi et Michalak, 2007 ; Schwede et Cirpka, 2010 ; Yang *et al.*, 2018]. Dans cette thèse, les deux méthodes (krigeage avec dérive externe et calcul de covariances empiriques) sont comparées de façon détaillée et mises en œuvre dans des cadres plus variés (zone non saturée, modélisation multivariable, modélisation spatio-temporelle, cas réels complexes à deux et trois dimensions).

Organisation du manuscrit

L'objectif de cette thèse est de mettre en œuvre et valider des méthodes d'estimation combinant la géostatistique et les simulations à base physique, pour améliorer la caractérisation de la contamination dans les sols et les nappes autour d'installations industrielles.

La Partie I présente les approches classiques (estimation géostatistique et modélisation à base physique), ainsi qu'un état de l'art des méthodes de combinaison de la géostatistique et des simulations numériques (Chapitre 1).

Deux méthodes sont mises en œuvre et comparées dans ce travail. La première est un krigeage avec dérive externe, où la dérive est constituée par la moyenne d'un ensemble de simulations d'écoulement et de transport. La seconde est un krigeage ordinaire, pour lequel le modèle de variogramme usuellement ajusté sur les données est remplacé par un variogramme empirique, calculé à partir de quelques centaines de simulations d'écoulement et de transport. Le Chapitre 2 présente ces deux méthodes en détails ; l'accent est mis sur les éléments de mise en œuvre pratique.

Pour implémenter ces deux méthodes et les valider, un cas synthétique à deux dimensions est construit dans la Partie II. Ce cas synthétique simule la migration d'un panache de tritium dans la zone non saturée d'un aquifère sableux. Il est inspiré de l'environnement typique des installations nucléaires françaises, en zone alluviale, avec une zone non saturée de quelques mètres de profondeur, et est construit autour de l'incertitude sur la variabilité spatiale des champs de paramètres hydrodynamiques. Au Chapitre 3, une analyse de sensibilité montre que la variabilité spatiale de ces paramètres, que le second axe du projet Kri-Terres vise à caractériser, influence fortement la migration de contaminants dans la zone non saturée. Les résultats de cette analyse ont été publiés au cours de la thèse [Pannecoucke *et al.*, 2019].

Ce cas synthétique est utilisé pour mettre en œuvre les méthodes de combinaison proposées et pour valider leur intérêt pour étudier des contaminations en zone non saturée, lorsque le terme source ainsi que l'environnement hydrogéologique sont bien caractérisés. Des estimations d'activité en radionucléides sont réalisées dans plusieurs contextes. Au Chapitre 4, des estimations d'activité en tritium sont effectuées, dans un cadre spatial (*i.e.* à une date fixée) ; un cas multivariable (deux radionucléides) est également examiné. Une partie des résultats de ce chapitre a été publiée au cours de la thèse [Pannecoucke *et al.*, 2020a]. Au Chapitre 5, des estimations sont réalisées dans un cadre spatio-temporel, pour faire des prévisions de la migration de panaches de tritium.

Dans la Partie III, les méthodes sont appliquées à deux jeux de données réels acquis au voisinage de la centrale nucléaire de Tchernobyl.

Le premier jeu de données (Chapitre 6) est issu d'une expérience de traçage au chlore 36 (^{36}Cl), réalisée en 2000 pour étudier la migration d'un traceur dans la zone non saturée autour de la centrale de Tchernobyl. Ce jeu de données est dense dans l'espace (plus de 500 mesures dans un mètre cube), et les conditions dans lesquelles le traçage a été réalisé sont relativement bien connues, ce qui permet de valider les méthodes de combinaison sur un premier cas réel (en zone non saturée, avec des paramètres hydrodynamiques proches du cas synthétique). Il permet aussi de mettre en œuvre les méthodes sur un cas tridimensionnel.

Le second jeu de données (Chapitre 7) est constitué par des mesures d'activité en strontium 90 (^{90}Sr) à l'aval d'une tranchée d'enfouissement de déchets radioactifs, à proximité de la centrale de Tchernobyl, acquises entre 2000 et 2015. Ce jeu diffère du cas synthétique et du premier jeu de données réelles pour plusieurs raisons : il s'agit de mesures d'activité en zone saturée et la source de contamination en ^{90}Sr , ainsi que sa migration, sont moins bien caractérisées que précédemment. Les méthodes de combinaison sont implémentées dans un cadre spatial, pour quantifier la contamination dans la nappe à l'aval de la tranchée, ainsi que dans un cadre temporel, pour estimer la quantité de radionucléides ayant migré de la tranchée vers la nappe au cours du temps. Une partie des résultats de ce chapitre fait l'objet d'une publication en préparation [Pannecoucke *et al.*, 2020b].

Enfin, la conclusion récapitule les principaux résultats et propose de nouveaux champs d'application.

Première partie

Contexte et méthodes développées

Chapitre 1

Approches usuelles et état de l'art

Ce chapitre présente les approches classiquement utilisées pour caractériser les sites contaminés, ainsi que l'état de l'art sur lequel s'appuie ce travail.

1.1 L'approche géostatistique pour caractériser les contaminations

L'intérêt de la géostatistique est aujourd'hui reconnu par les professionnels du secteur des sites et sols pollués [INERIS *et al.*, 2005 ; UPDS, 2016]. La géostatistique a été développée dans les années 1950 par D. Krige [Krige, 1951], au cours de travaux sur des mines d'or en Afrique du Sud. Elle a ensuite été théorisée à partir des années 1960 par G. Matheron [Matheron, 1963, 1971, 1978]. Historiquement, son domaine d'application est la caractérisation de gisements miniers. Cependant, elle est appliquée à d'autres domaines des géosciences depuis les années 1970, notamment à des problèmes environnementaux [Delhomme, 1978]. Actuellement, la géostatistique est couramment utilisée dans d'autres domaines, par exemple pour la construction de méta-modèles remplaçant des codes de simulations numériques lorsque ceux-ci sont trop coûteux [Asher *et al.*, 2015 ; Kleijnen, 2015].

1.1.1 Variables régionalisées et fonctions aléatoires

La géostatistique s'appuie sur la théorie des variables régionalisées, qui caractérisent des phénomènes présentant des structures dans l'espace et/ou dans le temps. Par exemple, les altitudes, les concentrations en polluant ou les teneurs en minerai sont des variables régionalisées. La qualité de la représentation d'une variable régionalisée par un modèle déterministe dépend des connaissances que l'on a sur le phénomène étudié et des moyens qui peuvent être consacrés à son étude. Par ailleurs, une modélisation déterministe ne permet pas de prendre en compte l'incertitude résultant d'observations ponctuelles.

Par conséquent, la modélisation géostatistique, qui est probabiliste, est utilisée. En tout point x d'un domaine \mathcal{D} , la valeur de la variable régionalisée est considérée comme une réalisation particulière ω d'un mécanisme aléatoire : $z(x) = Z(x, \omega)$. Chaque fonction $Z(\cdot, \omega)$ est une réalisation de la fonction aléatoire $Z(x, \omega)$, plus simplement notée $Z(x)$ [Chilès et Delfiner, 2012]. À noter que lorsque \mathcal{D} est un espace à une dimension, une fonction aléatoire est également appelée processus stochastique. Lorsque la dimension de \mathcal{D} vaut au moins deux, on parle de champ aléatoire.

Une fonction aléatoire Z est représentée par sa loi spatiale, dont la variance et la covariance sont exprimées ainsi :

$$\text{Var}[Z(x)] = E[\{Z(x) - E[Z(x)]\}^2] \quad (1.1)$$

$$\text{Cov}[Z(x_1), Z(x_2)] = E[\{Z(x_1) - E[Z(x_1)]\}\{Z(x_2) - E[Z(x_2)]\}] \quad (1.2)$$

Une fonction aléatoire est dite strictement stationnaire lorsque sa loi spatiale est invariante par translation. La stationnarité stricte est souvent réduite à une stationnarité d'ordre deux, *i.e.* les deux premiers moments de la fonction aléatoire sont définis et invariants par translation pour tout x :

$$E[Z(x)] = m \quad (1.3)$$

$$\text{Var}[Z(x)] = \text{Cov}[Z(x), Z(x)] = C(0) \quad (1.4)$$

$$\text{Cov}[Z(x), Z(x+h)] = C(h) \quad \text{pour tout } h \quad (1.5)$$

Le cadre des fonctions aléatoires intrinsèques permet de relâcher certaines hypothèses de stationnarité. Une fonction aléatoire est intrinsèque d'ordre 0 lorsque les deux premiers moments de ses accroissements

sont stationnaires d'ordre deux :

$$E[Z(x+h) - Z(x)] = 0 \quad \text{pour tout } h \quad (1.6)$$

$$\frac{1}{2} \text{Var}[Z(x+h) - Z(x)] = \gamma(h) \quad \text{pour tout } h \quad (1.7)$$

Plus généralement, une fonction aléatoire intrinsèque d'ordre k est une fonction aléatoire dont les incréments d'ordre k sont stationnaires.

1.1.2 Fonction de covariance et variogramme

La fonction C définie dans l'Équation (1.5) est la fonction de covariance de Z . Il s'agit d'une fonction paire et de type positif :

$$\sum_{k=1}^K \sum_{\ell=1}^K \omega_k \omega_\ell C(x_k, x_\ell) \geq 0, \quad x_k \in \mathcal{D}, \quad \omega_k \in \mathbb{R} \quad \text{pour } k = 1 \dots K \quad (1.8)$$

L'une des conséquences de ces propriétés est que la fonction de covariance est bornée par $C(0)$.

La fonction γ définie dans l'Équation (1.7) est le variogramme de Z . Il s'agit d'une fonction paire, positive et de type négatif conditionnel :

$$-\sum_{k=1}^K \sum_{\ell=1}^K \omega_k \omega_\ell \gamma(x_k, x_\ell) \geq 0, \quad x_k \in \mathcal{D}, \quad \omega_k \in \mathbb{R} \quad \text{pour } k = 1 \dots K \quad \text{et avec} \quad \sum_{k=1}^K \omega_k = 0 \quad (1.9)$$

1.1.3 Le krigeage et ses variantes

Les méthodes d'estimation géostatistique les plus couramment utilisées sont le krigeage simple et le krigeage ordinaire (Complément 1.1), ainsi que les méthodes non stationnaires (krigeage universel, krigeage avec dérive externe, etc.) [Chilès et Delfiner, 2012].

Le krigeage ordinaire

La variable Z est estimée en un point x_0 , appelé cible, comme combinaison linéaire de N observations localisées aux points $x_n, n=1 \dots N$:

$$Z^*(x_0) = \sum_{n=1}^N \lambda_n Z(x_n) \quad (1.10)$$

L'erreur de krigeage s'écrit :

$$R(x_0) = Z(x_0) - Z^*(x_0) = Z(x_0) - \sum_{n=1}^N \lambda_n Z(x_n) \quad (1.11)$$

Dans le cas du krigeage ordinaire et sous l'hypothèse d'une stationnarité d'ordre deux, la moyenne de Z existe, est constante et finie, mais inconnue : $m(x) = m$. L'estimateur de krigeage étant sans biais, on cherche à annuler l'espérance de $R(x_0)$, quelle que soit la valeur de m :

$$E[R(x_0)] = 0 \quad \Leftrightarrow \quad \sum_{n=1}^N \lambda_n = 1 \quad (1.12)$$

L'Équation (1.12) donne la condition de non biais. La variance de $R(x_0)$ vaut :

$$\text{Var}[R(x_0)] = C(0) - 2 \sum_{n=1}^N \lambda_n C(x_n, x_0) + \sum_{n=1}^N \sum_{n'=1}^N \lambda_n \lambda_{n'} C(x_n, x_{n'}) \quad (1.13)$$

Le krigeage ordinaire étant un estimateur optimal, on cherche à minimiser cette variance sous la contrainte de l'Équation (1.12). Soit μ un multiplicateur de Lagrange. On minimise ϕ :

$$\phi = \text{Var}[R(x_0)] + 2\mu \left(\sum_{n=1}^N \lambda_n - 1 \right) \quad (1.14)$$

La dérivation partielle de ϕ par rapport au multiplicateur de Lagrange et par rapport à chaque λ_n donne :

$$\frac{\partial \phi}{\partial \mu} = 2 \left(\sum_{n=1}^N \lambda_n - 1 \right) \quad (1.15)$$

$$\forall n, n = 1 \dots N, \quad \frac{\partial \phi}{\partial \lambda_n} = -2C(x_n, x_0) + 2 \sum_{n'=1}^N \lambda_{n'} C(x_n, x_{n'}) + 2\mu \quad (1.16)$$

L'annulation de l'Équation (1.15) permet de retrouver la condition de non biais, et l'annulation de l'Équation (1.16) donne :

$$\forall n, n = 1 \dots N, \quad \sum_{n'=1}^N \lambda_{n'} C(x_n, x_{n'}) + \mu = C(x_n, x_0) \quad (1.17)$$

En notation matricielle, le système de krigeage ainsi obtenu s'écrit :

$$\begin{bmatrix} \mathbf{C} & \mathbf{1} \\ \mathbf{1}^t & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda} \\ \mu \end{bmatrix} = \begin{bmatrix} \mathbf{C}_0 \\ 1 \end{bmatrix} \quad (1.18)$$

où $\mathbf{C} = [C(x_n, x_{n'})]$ est la matrice des covariances entre observations, $\mathbf{1}$ est un vecteur de valeurs unitaires, $\boldsymbol{\Lambda} = [\lambda_n]$ est le vecteur des poids de krigeage et $\mathbf{C}_0 = [C(x_n, x_0)]$ est le vecteur des covariances entre les observations et x_0 . La résolution de ce système donne les poids $\lambda_{n,n=1\dots N}$ nécessaires pour calculer $Z^*(x_0)$ et $\text{Var}[R(x_0)]$.

À noter que pour travailler avec des variogrammes, il suffit de remplacer C par $-\gamma$ dans les équations précédentes.

Complément 1.1

De nombreux exemples d'application du krigeage à l'étude de sites pollués sont disponibles dans la littérature [Demougeot-Renard et de Fouquet, 2004 ; Saby *et al.*, 2006 ; Horta *et al.*, 2013 ; Liang *et al.*, 2018]. Le krigeage avec dérive externe est également couramment utilisé [Desbarats *et al.*, 2002 ; Romary *et al.*, 2014 ; Beauchamp *et al.*, 2018]. Beaucoup de travaux proposent des extensions des méthodes de krigeage existantes pour améliorer les estimations. Des exemples de ces travaux, en lien avec le sujet traité dans ce manuscrit, sont détaillés à la Section 1.4.

Le krigeage présente de nombreux avantages par rapport à d'autres méthodes d'interpolation : il s'agit d'un interpolateur sans biais et optimal qui est exact, aux erreurs de mesure près, et qui prend en compte la structure spatiale de la variable étudiée. Les variantes du krigeage permettent de traiter différents types de problèmes : des problèmes non stationnaires (krigeage universel, krigeage avec dérive externe), des problèmes multivariés (cokrigeage), des problèmes spatio-temporels. La modularité du voisinage de krigeage (voisinage unique, voisinage glissant) permet par ailleurs de ne conserver qu'une partie des observations, les plus proches de la cible, pour réaliser une estimation. Enfin, il s'agit d'une représentation probabiliste et l'estimation est accompagnée d'une quantification des incertitudes.

Cependant, la qualité de l'estimateur dépend de la modélisation de la structure spatiale du phénomène et des hypothèses sous-jacentes à cette modélisation. En particulier, l'estimateur de krigeage est peu précis si le nombre d'observations est trop faible ou si le phénomène étudié est fortement non stationnaire et gouverné par des processus complexes [Webster et Oliver, 2007 ; Wang *et al.*, 2017]. Par ailleurs, les données requises pour réaliser une estimation géostatistique sont souvent complexes à obtenir et coûteuses, notamment dans le cas d'études de contaminations (forages, analyses en laboratoire, zones non accessibles

à l'échantillonnage). Enfin, le krigeage ne prend pas en compte les processus physiques gouvernant le phénomène étudié, ce qui peut donner lieu à des estimations manquant de réalisme ou qui ne reproduisent pas correctement la variabilité du phénomène étudié à l'échelle macroscopique [Michael *et al.*, 2010 ; Yong *et al.*, 2019].

1.1.4 Traitement usuel de la non-stationnarité

Pour de nombreuses applications, l'hypothèse de stationnarité d'ordre deux est discutable, qu'il s'agisse d'une non-stationnarité de la moyenne ou de la covariance, et les caractéristiques non stationnaires du phénomène étudié doivent être prises en compte pour obtenir une estimation de bonne qualité. Par exemple, Reed *et al.* [2004] soulignent le caractère non stationnaire de la distribution spatiale de panaches de contaminant.

Différentes méthodes sont utilisées pour prendre en compte la non-stationnarité de la variable étudiée. Par exemple, si la non-stationnarité est liée à une tendance spatiale, le krigeage universel ou le krigeage avec dérive externe permettent de se ramener à l'étude de résidus stationnaires [Chilès et Delfiner, 2012], en introduisant la dichotomie suivante : $Z(x) = Y(x) + m(x)$, où $Y(x)$ est un résidu stationnaire d'espérance nulle et $m(x)$ une dérive déterministe. Cela permet de modéliser la non-stationnarité de la moyenne. Le cadre des fonctions aléatoires intrinsèques d'ordre k permet également de modéliser des phénomènes non stationnaires, par un ajustement indirect des structures [Chilès, 1977]. De nombreuses publications suggèrent l'utilisation de classes de modèles de covariances non stationnaires paramétriques, ou basés sur des convolutions, ou encore des déformations de l'espace de travail [Sampson et Guttorp, 1992 ; Rivest et Marcotte, 2012 ; Liang et Marcotte, 2016]. Une revue plus complète des méthodes de traitement des non-stationnarités est disponible dans la thèse de F. Fouedjio [Fouedjio, 2014]. Ces méthodes sont cependant propres à un type de non-stationnarité et leur application requiert une ou plusieurs hypothèses que la variable étudiée ne vérifie pas toujours.

Enfin, une autre méthode de modélisation des fonctions aléatoires non stationnaires consiste à travailler dans le cadre des équations aux dérivées partielles stochastiques : le champ aléatoire modélisé est considéré comme solution d'une équation aux dérivées partielles stochastique. La structure de covariance du champ aléatoire est contrôlée par les coefficients de l'équation et une approximation de la solution de l'équation est obtenue par éléments finis [Lindgren *et al.*, 2011 ; Carrizo Vergara, 2018 ; Pereira, 2019]. Cette méthode a initialement été développée dans le cadre des fonctions aléatoires stationnaires d'ordre deux, mais elle est désormais utilisée dans le cadre des fonctions aléatoires non stationnaires. À nouveau, cette méthode ne s'applique qu'à certains types de fonctions aléatoires et de covariances.

1.1.5 Les simulations géostatistiques

Les méthodes présentées précédemment ont tendance à conduire à des résultats lissés et sont peu adaptées à l'estimation d'une grandeur non linéaire par rapport à la variable d'intérêt. Par exemple, l'estimation par krigeage d'une distance à parcourir sur un profil escarpé, connaissant les altitudes en quelques points, est systématiquement sous-estimée. Les simulations géostatistiques permettent de contourner ce problème, en générant des variables régionalisées qui respectent les premiers moments (moyenne et covariance) du phénomène étudié. Ces simulations sont par exemple utilisées pour simuler des concentrations en contaminants [Boudreault *et al.*, 2016]. Il peut s'agir de simulations non conditionnelles ou de simulations conditionnelles, qui respectent alors les observations de la variable d'intérêt.

Dans la suite de ce travail, nous nous intéressons en particulier aux simulations de champs gaussiens (Complément 1.2 et Figure 1.1), qui sont couramment utilisés pour représenter des paramètres de la subsurface. Par exemple, il est généralement admis que la conductivité hydraulique peut être représentée par une loi lognormale (transformée logarithmique d'un champ gaussien) [Freeze, 1975 ; Dagan, 1989 ; de Marsily *et al.*, 2005]. Différentes méthodes de simulation sont utilisées pour simuler des variables gaussiennes, comme la méthode des bandes tournantes ou les méthodes de décomposition matricielle [Alfaro, 1979 ; Boulanger, 1990 ; Lantuéjoul, 2002 ; Chilès et Delfiner, 2012] (Complément 1.3).

À noter que cette présentation des méthodes de simulation est centrée sur les méthodes utilisées dans ce travail et n'est pas exhaustive.

Vecteurs gaussiens et fonctions aléatoires gaussiennes

Un vecteur (Y_1, \dots, Y_n) est gaussien si toute combinaison linéaire de ses composantes est une variable gaussienne. Il est caractérisé par sa moyenne $\mathbf{m} = [m_{i,i=1\dots n}]$ et sa matrice de covariance $\Sigma = [C(Y_i, Y_j)_{i=1\dots n, j=1\dots n}]$.

Une fonction aléatoire stationnaire d'ordre deux $(Y(x), x \in \mathcal{D})$ est gaussienne si pour tout n , tout vecteur $(Y(x_1), \dots, Y(x_n))$ est gaussien. Cela implique en particulier que Y est bi-gaussienne, c'est-à-dire que toute paire $[Y(x), Y(x+h)]$ est gaussienne. Pour vérifier qu'une paire $[Y(x), Y(x+h)]$ est gaussienne, les deux propriétés suivantes peuvent être examinées :

1. le nuage de corrélation $[Y(x), Y(x+h)]$ est elliptique ;
2. la racine du variogramme est proportionnelle au madogramme γ_1 :

$$\text{pour tout } h, \quad \frac{\sqrt{\gamma(h)}}{\gamma_1(h)} = \sqrt{\pi} \quad \text{où} \quad \gamma_1(h) = \frac{1}{2} E[|Y(x+h) - Y(x)|] \quad (1.19)$$

La distribution spatiale d'une fonction aléatoire gaussienne stationnaire d'ordre deux est entièrement caractérisée par sa moyenne et sa matrice de covariance.

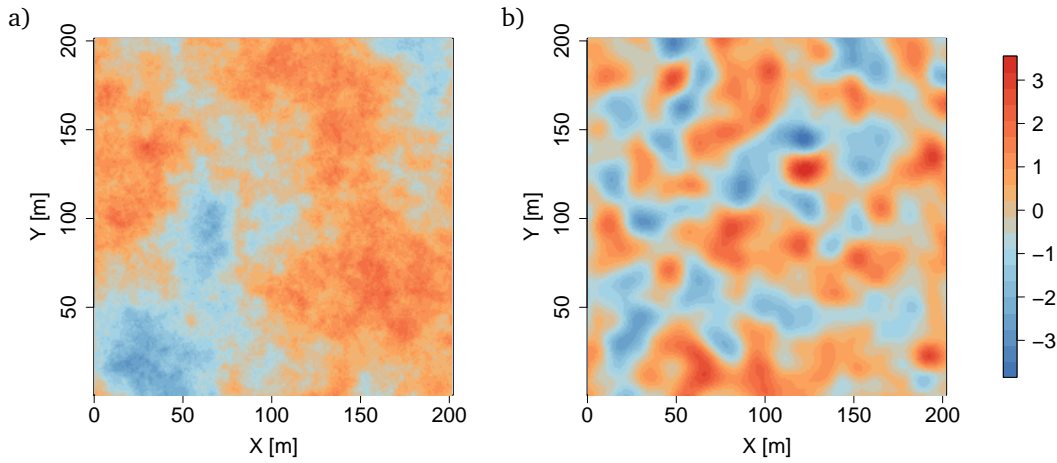


FIGURE 1.1 – Exemples de champs gaussiens simulés sur des carrés de 200 m de côté. a) Covariance sphérique de portée 50 m ; b) covariance gaussienne de portée 10 m.

Quelques méthodes de simulations de champs gaussiens**a) Bandes tournantes**

La méthode des bandes tournantes repose sur le lien entre fonction de covariance et mesure spectrale. Le principe de la méthode est de simuler des processus aléatoires ayant une structure donnée sur des droites, puis de projeter les contributions de chacun de ces processus aux points où l'on veut simuler une variable gaussienne. Dans le cas particulier d'un champ isotrope dans \mathbb{R}^3 de covariance C_3 , il existe un lien entre C_3 et la covariance unidimensionnelle C_1 du processus stochastique qui doit être généré sur chaque bande : $C_1(r) = \frac{d}{dr}(rC_3(r))$. L'algorithme des bandes tournantes est alors le suivant :

- (i) simuler n directions indépendantes et uniformes de \mathbb{R}^3 $(\theta_1, \dots, \theta_n)$;
- (ii) simuler des réalisations (z_1, \dots, z_n) de n processus stochastiques (Z_1, \dots, Z_n) de moyenne nulle et de covariance C_1 ;
- (iii) pour tout x de \mathbb{R}^3 , retourner $y(x) = \frac{1}{\sqrt{n}} \sum_{j=1}^n z_j(\langle {}^t x, \theta_j \rangle)$.

b) Décomposition matricielle

La méthode de décomposition matricielle permet de simuler un vecteur gaussien $Y = (Y_1, \dots, Y_n)$ de moyenne \mathbf{m} et de covariance Σ . Puisque Σ est définie positive, il existe une matrice \mathbf{A} de taille $n \times n$ telle que $\Sigma = \mathbf{A}(\mathbf{A}^t \mathbf{A})$ (obtenue par exemple par décomposition de Cholesky). L'algorithme de simulation est alors le suivant :

- (i) calculer \mathbf{A} ;
- (ii) simuler n variables gaussiennes indépendantes, centrées et normées $X = (X_1, \dots, X_n)$;
- (iii) retourner $Y = \mathbf{A}X + \mathbf{m}$.

Par combinaison linéaire, on a bien :

$$E[\mathbf{A}X + \mathbf{m}] = \mathbf{A}E[X] + \mathbf{m} = \mathbf{m} \quad \text{car } E[X] = 0 \quad (1.20)$$

$$\text{Var}[\mathbf{A}X + \mathbf{m}] = \mathbf{A}\Sigma_X(\mathbf{A}^t \mathbf{A}) = \Sigma \quad \text{car } \Sigma_X = \mathbf{I}_n \quad (1.21)$$

Quelle que soit la méthode de simulation employée, une réalisation peut être conditionnée par krigeage simple des résidus entre les données et la réalisation non conditionnée.

Les méthodes de simulation de variables gaussiennes peuvent également servir à simuler des variables à distribution non gaussienne. Cela requiert une étape de modélisation supplémentaire : l'anamorphose gaussienne, qui permet de passer de la distribution de la variable d'intérêt à une distribution gaussienne. Le principe de l'anamorphose est de trouver la fonction φ inversible telle que $Z = \varphi(Y)$ où Y est une variable gaussienne. Cette méthode pose problème lorsque la variable de départ présente une proportion non négligeable de valeurs identiques, comme un effet zéro.

Outre les hypothèses nécessaires sur la distribution des variables étudiées, la principale limite des simulations géostatistiques, à un instant donné, concerne la dimension des problèmes : certaines méthodes sont coûteuses en temps de calcul dès que les domaines de simulation dépassent quelques centaines de milliers de points. Cependant, cette limitation est rarement définitive, car l'efficacité des algorithmes de simulation est constamment améliorée.

Dans la suite de ce manuscrit, ces simulations sont appelées simulations géostatistiques ou réalisations d'une fonction aléatoire (qui seront à distinguer des simulations numériques ou simulations d'écoulement et de transport).

1.2 Le cas particulier du cadre spatio-temporel

Les méthodes géostatistiques mentionnées précédemment s'appliquent à des problèmes d'estimation dans l'espace, dans le temps, ou dans un cadre spatio-temporel. Quelques méthodes d'estimation spatio-temporelles, issues de la géostatistique ou de l'assimilation de données, et utilisées dans la suite de ce travail sont décrites ici. Il ne s'agit pas d'une liste exhaustive des méthodes existantes.

1.2.1 Le krigeage spatio-temporel

Le krigeage spatio-temporel consiste en une extension du krigeage à des données spatio-temporelles. Il est largement décrit dans la littérature, pour étudier différents phénomènes [Yang et al., 2015 ; Li et al., 2019 ; Hu et al., 2019 ; van Zoest et al., 2019], et de plus en plus utilisé (Figure 1.2).

Lorsque les observations sont réparties dans l'espace (ensemble \mathcal{S}) et dans le temps (ensemble \mathcal{T}), la valeur de Z en un point cible x_0 et à une date cible t_0 est estimée par combinaison linéaire des observations :

$$Z^*(x_0, t_0) = \sum_{x \in \mathcal{S}, t \in \mathcal{T}} \lambda_{x,t} Z(x, t) \quad (1.22)$$

À noter que les observations ne sont pas nécessairement sur une grille $(x, t) \in \mathcal{S} \times \mathcal{T}$. Dans cette section, cette hypothèse est faite pour simplifier les équations. La date cible peut être une date à laquelle

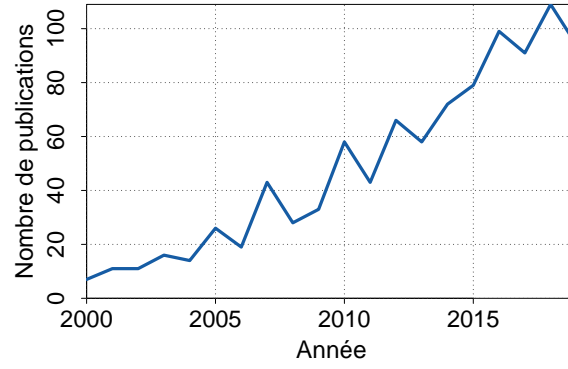


FIGURE 1.2 – Nombre de résultats sur *Web Of Science* pour la recherche 'space time kriging' ou 'spatio-temporal kriging' dans le titre, le résumé ou les mots-clés des publications entre 2000 et 2019.

aucune observation n'est disponible. Le krigeage est alors utilisé en extrapolation temporelle et permet de faire une prévision de l'évolution de Z . Différentes variantes du krigeage sont utilisées dans un cadre spatio-temporel : il s'agit la plupart du temps de krigeage ordinaire, mais également de krigeage avec dérive externe [Snepvangers *et al.*, 2003 ; Pearce *et al.*, 2009 ; Varouchakis *et al.*, 2019] ou de méthodes multivariées [Bourotte *et al.*, 2016 ; Salvaña et Genton, 2020].

Le krigeage spatio-temporel présente deux principales limitations : la dimension de la matrice de covariance $([Card(\mathcal{S}) \times Card(\mathcal{T})]^2)$ au plus dans le cas d'un krigeage simple) et l'ajustement d'un modèle autorisé de covariance spatio-temporelle. De nombreux travaux sur les modèles de covariance ont été effectués [Kyriakidis et Journel, 1999 ; Stein, 2005 ; Gneiting *et al.*, 2007 ; Huang et Sun, 2017]. Par exemple, Gneiting *et al.* [2007] distinguent les classes de modèles séparables et de modèles symétriques. Pour pouvoir utiliser ces modèles, la variable étudiée doit respecter de nombreuses hypothèses, parfois contraignantes et peu réalistes. Des modèles paramétriques permettant de relâcher certaines hypothèses sont proposés dans la littérature [Kolovos *et al.*, 2004 ; Varouchakis *et al.*, 2019], mais s'appliquent en général à des cas très particuliers. Par conséquent, le krigeage spatio-temporel est parfois considéré comme une méthode trop coûteuse et complexe à mettre en œuvre [McLean *et al.*, 2019].

Enfin, plusieurs variantes peuvent être envisagées quant aux conditions d'universalité résultant de la condition de non biais [Bogaert, 1996] :

$$E[Z(x_0, t_0) - Z^*(x_0, t_0)] = 0 \quad (1.23)$$

Trois exemples de variantes sont donnés dans les Équations (1.24), (1.25) et (1.26). La première est obtenue en supposant que la moyenne de Z est stationnaire dans le temps, alors que l'on considère que cette moyenne diffère pour chaque date dans la seconde. Enfin, la troisième variante est un exemple de combinaison des deux précédentes.

$$\sum_{x \in \mathcal{S}, t \in \mathcal{T}} \lambda_{x,t} = 1 \quad \text{si} \quad m(x, t) = m \quad (1.24)$$

$$\forall t_i \in \mathcal{T} \setminus \{t_0\} \quad \sum_{x \in \mathcal{S}} \lambda_{x,t_i} = 0 \quad \text{et} \quad \sum_{x \in \mathcal{S}} \lambda_{x,t_0} = 1 \quad \text{si} \quad m(x, t) = m(t) \quad (1.25)$$

$$\sum_{x \in \mathcal{S}, t \in \mathcal{T}, t \neq t_0} \lambda_{x,t} = 0 \quad \text{et} \quad \sum_{x \in \mathcal{S}} \lambda_{x,t_0} = 1 \quad \text{si} \quad m(x, t) = m \quad \text{et} \quad m(x, t_0) \neq m \quad (1.26)$$

1.2.2 Les méthodes d'assimilation de données

L'assimilation de données consiste à combiner des observations et les sorties d'un modèle pour obtenir une estimation de l'état du système étudié en utilisant les dernières informations disponibles. Deux principales familles de méthodes d'assimilation de données sont distinguées : les méthodes variationnelles, basées sur la minimisation d'une fonction de coût, et les méthodes stochastiques, dont les méthodes d'ensemble détaillées ci-après. Les méthodes d'ensemble utilisent des particules, qui correspondent à des réalisations du phénomène étudié, et consistent généralement en deux étapes :

1. la prévision (*forecast*), où l'ensemble des particules est propagé par le modèle non linéaire décrivant l'évolution du système (un code d'écoulement et de transport par exemple) ;
2. la correction ou l'analyse (*update* ou *analysis*), lors de laquelle l'ensemble ayant été propagé est modifié au regard des dernières observations acquises.

Ces deux étapes sont répétées séquentiellement à chaque fois que de nouvelles observations sont disponibles. Deux méthodes d'ensemble sont couramment utilisées en géosciences : le filtre de Kalman d'ensemble [Bertino *et al.*, 2002 ; Huang *et al.*, 2008 ; Li *et al.*, 2018] et le filtre particulaire [van Leeuwen, 2010 ; Farchi et Bocquet, 2018]. Ces deux méthodes sont comparables, car elles utilisent toutes les deux des connaissances *a priori* sur le système étudié, à travers des simulations numériques qui forment un ensemble de particules. Elles diffèrent cependant par leur étape d'analyse.

Le filtre de Kalman d'ensemble

Le principe du filtre de Kalman d'ensemble [Evensen, 2006 ; Katzfuss *et al.*, 2016] est de représenter l'état du système par un ensemble de M particules. Chacune de ces particules évolue suivant un modèle H et est corrigée lors de l'acquisition de nouvelles observations. Cette correction s'effectue en appliquant une dynamique de type filtre de Kalman à chaque particule [Kalman, 1960].

Ces deux étapes sont respectivement décrites dans les équations suivantes :

$$\textbf{Prévision : pour } m = 1 \dots M, \quad p_t^{(m)} = H(p_{t-1}^{(m)}) + q_t^{(m)}, \quad q_t^{(m)} \sim \mathcal{N}(0, \mathbf{Q}_t) \quad (1.27)$$

$$\textbf{Correction : pour } m = 1 \dots M, \quad \hat{p}_t^{(m)} = p_t^{(m)} + \mathbf{K}_t(y_t - \mathbf{B}_t p_t^{(m)} + r_t^{(m)}), \quad r_t^{(m)} \sim \mathcal{N}(0, \mathbf{R}_t) \quad (1.28)$$

où $p_t^{(m)}$ (respectivement $\hat{p}_t^{(m)}$) décrit la particule m à la date t avant (respectivement après) correction, q et r sont des erreurs de modèle et de mesure, \mathbf{K}_t est le gain de Kalman, y_t sont les observations à la date t et \mathbf{B}_t est l'opérateur (linéaire ici) permettant de passer de la particule aux observations sur cette particule. L'étape de correction est présentée de manière schématique sur la Figure 1.3.

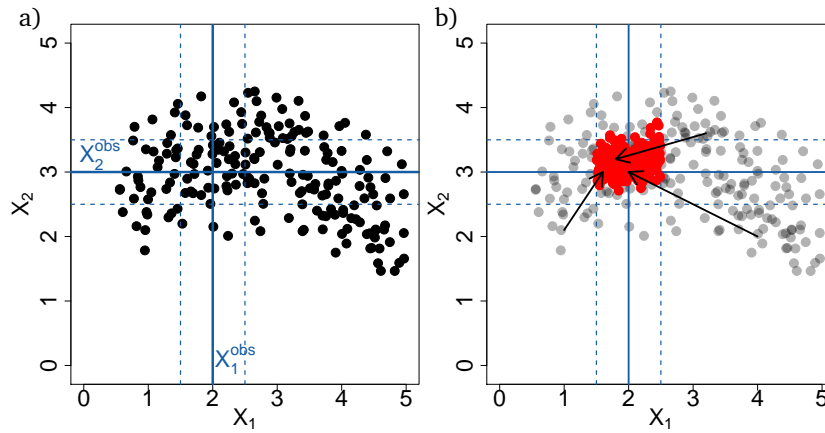


FIGURE 1.3 – Principe de l'étape de correction pour le filtre de Kalman d'ensemble sur un exemple à deux dimensions. Les coordonnées de l'observation sont en bleu (avec prise en compte d'une erreur de mesure). a) Ensemble avant correction ; b) ensemble après correction (les particules, représentées par les points rouges, ont été déplacées vers l'observation).

Le gain de Kalman \mathbf{K}_t de l'Équation (1.28) est en pratique remplacé par une approximation $\hat{\mathbf{K}}_t$ basée sur les éléments de l'ensemble :

$$\hat{\mathbf{K}} = \mathbf{C}^t \mathbf{B} (\mathbf{B} \mathbf{C}^t \mathbf{B} + \mathbf{R})^{-1} \quad (1.29)$$

où \mathbf{C} est la matrice de covariance empirique de la population et \mathbf{R} est la matrice de covariance des erreurs de mesure. Pour des raisons de lisibilité, l'indice t n'est pas repris dans cette équation.

Le filtre particulaire

L'étape de prévision du filtre particulaire est identique à celle du filtre de Kalman d'ensemble. En revanche, l'étape de correction du filtre particulaire consiste à affecter à chaque particule un poids calculé

selon une distance entre les particules et les observations. Ensuite, suivant leur poids, les particules sont éliminées ou conservées, voire dupliquées (étape de ré-échantillonnage) [Doucet *et al.*, 2001 ; Cappe *et al.*, 2007]. L'étape de correction du filtre particulaire est présentée de manière schématique sur la Figure 1.4. L'une des principales limites de cette méthode est la dégénérescence du système : la sélection/duplication successive de particules conduit à une chute du nombre de particules considérées.

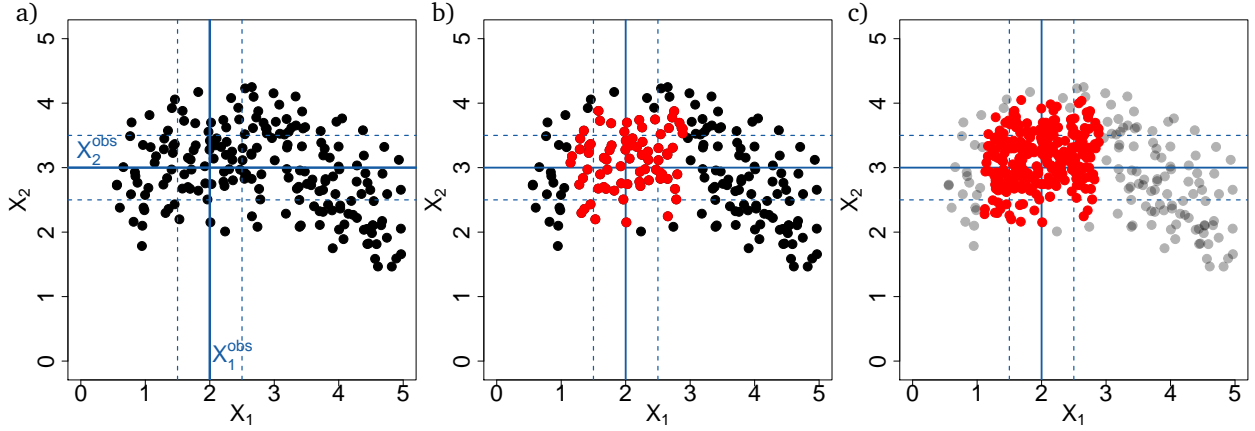


FIGURE 1.4 – Principe de l'étape de correction pour le filtre particulaire sur le même exemple à deux dimensions que pour le filtre de Kalman d'ensemble. a) Ensemble avant correction ; b) sélection des particules proches de l'observation en rouge ; c) ré-échantillonnage de la population sélectionnée.

Classiquement, les poids sont calculés par vraisemblance suivant l'Équation (1.30), ici dans le cas d'erreurs de mesures indépendantes et identiques pour toutes les simulations :

$$\omega_t^{(m)} = A \exp \left[-\frac{1}{2s} \|y_t - Bp_t^{(m)}\|^2 \right] \quad (1.30)$$

où A est un coefficient permettant de s'assurer que la somme des poids vaut 1 et s est la variance d'erreur de mesure.

1.3 La modélisation à base physique

Les méthodes géostatistiques présentées précédemment sont des approches qui peuvent être qualifiées de probabilistes, basées sur l'estimation des moments (moyenne, covariance) des phénomènes étudiés. D'autres approches de ce type, non présentées ici, sont par exemple les modèles d'apprentissage statistique orientés données.

À l'inverse, une autre classe d'approches consiste à représenter les processus (physiques, chimiques, biologiques, *etc.*) menant au phénomène observé. Il s'agit de modélisations basées sur une conceptualisation ou une résolution déterministe des équations régissant ces processus. Elles sont ici qualifiées de modélisations "à base physique", bien que cette terminologie soit parfois réservée aux modèles les plus complexes [Beven, 1991].

1.3.1 La diversité des modèles à base physique

Les modèles à base physique présentent des degrés de complexité très différents. Les modèles les plus simples sont les modèles dits conceptuels, pour lesquels les processus sont représentés par des fonctions mathématiques très simples. Par exemple, l'approche K_d revient à modéliser l'interaction complexe entre un soluté et la matrice rocheuse par un coefficient de partage. Un autre exemple assez courant en hydrologie est la représentation d'un sol par un réservoir, ou un ensemble de réservoirs, sur lequel des bilans sont effectués (par exemple le modèle GARDÉNIA [Thiéry, 2014]). Les modèles plus complexes sont les modèles dits mécanistes ou déterministes [Konikow et Mercer, 1988 ; Anderson *et al.*, 2015], qui sont basés sur une formalisation mathématique des processus observés. En pratique, ils sont implémentés dans des logiciels de calcul, qui résolvent par des méthodes numériques les équations régissant le phénomène étudié.

Quel que soit le niveau de complexité du modèle à base physique utilisé, il s'agit d'une représentation simplifiée de la réalité. Le niveau de simplification acceptable dépend de l'utilisation du modèle [Beven, 2019] : s'il est utilisé pour comprendre le fonctionnement d'un système, il faut que celui-ci soit assez détaillé. En revanche, s'il est utilisé dans un processus de décision, les simplifications peuvent être plus importantes, tant que les prévisions restent de qualité acceptable.

1.3.2 Modélisation de l'écoulement et du transport

Dans l'étude de la migration d'un contaminant, on cherche à modéliser les phénomènes d'écoulement et de transport qui décrivent d'une part le transfert de charges hydrauliques, et d'autre part la migration des solutés présents dans l'eau souterraine qui peuvent interagir avec l'eau et la matrice rocheuse. Les codes d'écoulement et de transport sont couramment utilisés pour quantifier les risques liés à une contamination dans les sols ou les nappes [Zhang *et al.*, 2010]. Ces outils permettent de simuler la migration d'un contaminant à partir d'hypothèses de modélisation (processus physiques et chimiques pris en compte, domaine de modélisation, maillage, *etc.*) et d'un certain nombre de paramètres d'entrée (conditions initiales, conditions aux limites, paramètres du milieu simulé, *etc.*). Ils sont en général un compromis entre des approches déterministes et des approches plus conceptuelles (approche K_d , modélisation de la zone non saturée par un réservoir, *etc.*). Ils sont aujourd'hui suffisamment performants pour prévoir de manière fiable la migration de panaches de contaminants [Fogg et Zhang, 2016], bien que la modélisation de certains processus soit encore trop conceptuelle pour des applications particulières [Stockmann *et al.*, 2017].

Nous utilisons dans ce travail le code MELODIE développé à l'IRSN [Amaziane *et al.*, 2008 ; IRSN, 2009 ; Amor *et al.*, 2015]. Les équations résolues par ce code sont détaillées dans le Complément 1.4.

MELODIE (Modèle d'Évaluation à Long terme des Déchets Irradiants Enterrés)

Le logiciel MELODIE permet de réaliser des études de sûreté sur les sites de stockage de déchets radioactifs. Il simule les écoulements souterrains et le transport de solutés dans des milieux poreux à saturation variable en deux ou trois dimensions. Les principaux phénomènes pris en compte pour modéliser la migration de radionucléides sont la convection, la dispersion, la diffusion, la solubilité des radionucléides, la rétention des radionucléides par les matériaux naturels ou artificiels ainsi que la décroissance et la filiation radioactives. Les équations aux dérivées partielles à résoudre sont discrétisées par une combinaison de volumes finis et d'éléments finis (développement de Galerkin pour les termes temporels et diffusifs et schéma de Godunov pour le terme convectif).

Écoulement en milieu à saturation variable

L'écoulement en milieu à saturation variable est régi par l'équation de Richards généralisée (donnée ici en l'absence d'apport) :

$$\frac{\partial \theta(h)}{\partial t} + S_s S(\theta) \frac{\partial h}{\partial t} = \operatorname{div} \left(\overline{\overline{K}}(h) \overrightarrow{\operatorname{grad} h} \right) \quad (1.31)$$

où t est le temps [T], θ [$L^3.L^{-3}$] la teneur en eau du milieu, $\overline{\overline{K}}$ [$L.T^{-1}$] le tenseur de conductivité hydraulique, h [L] la charge hydraulique, S_s [L^{-1}] le coefficient d'emmagasinement et S [-] la saturation en eau. La résolution de cette équation nécessite de connaître le lien entre charge h , teneur en eau θ et conductivité hydraulique relative $\overline{\overline{K}}$. Ces liens sont donnés par des modèles empiriques, comme par exemple le modèle de Mualem-van Genuchten, qui sera décrit au Chapitre 3.

Transport de radionucléides

L'équation globale du transport de radionucléides prenant en compte la dispersion, la diffusion, l'adsorption et la décroissance radioactive est la suivante :

$$\operatorname{div} \left[\left(\overline{\overline{D}} \overrightarrow{U} + \omega d \right) \overrightarrow{\operatorname{grad} C} - \overrightarrow{U} C \right] = \omega' R \frac{\partial C}{\partial t} + \omega' \lambda R C \quad (1.32)$$

où \vec{U} [L.T⁻¹] est la vitesse de Darcy, C [M.L⁻³] la concentration en radionucléide, \bar{D} [L] le tenseur de dispersivité, d [L².T⁻¹] le coefficient de diffusion moléculaire, ω [-] la porosité, ω' [-] la porosité efficace, λ [T⁻¹] la constante de décroissance radioactive et R [-] le coefficient de retard (valant 1 lorsque le radionucléide est un traceur de l'eau et n'interagit pas avec la matrice).

Dans le cas où la masse volumique de l'eau est constante, la vitesse de Darcy vaut :

$$\vec{U} = -\bar{K} \text{grad} h \quad (1.33)$$

On trouve dans la littérature de nombreuses études de risques basées sur l'utilisation de codes d'écoulement et de transport [Nichols *et al.*, 2005 ; Neukum et Azzam, 2009 ; Bugai *et al.*, 2012 ; Cadini *et al.*, 2016 ; Testoni *et al.*, 2017]. Malgré les progrès des codes et l'augmentation de la puissance des calculateurs, certaines applications requièrent des simulations pouvant durer de quelques heures à plusieurs jours. C'est par exemple le cas lorsque le milieu simulé est complexe, que le domaine de modélisation est étendu (trois dimensions) et que la simulation est réalisée sur plusieurs années à dizaines d'années.

1.3.3 La calibration des modèles et la modélisation inverse

Une autre difficulté majeure lors de l'utilisation des codes d'écoulement et de transport est leur calibration. En effet, après le choix des processus physico-chimiques à modéliser (réactivité chimique, décroissance radioactive, *etc.*) et la construction géométrique du modèle (domaine, maillage), il faut déterminer les paramètres associés à la modélisation des processus. Les grandes familles de paramètres d'entrée nécessaires à la génération de simulations d'écoulement et de transport sont décrites dans la liste ci-dessous, qui n'est pas exhaustive.

- Paramètres liés à la source : localisation, date et durée du rejet, type de contaminant.
- Paramètres liés à l'écoulement : charge et gradient hydraulique, quantité d'eau infiltrée et paramètres du sol gouvernant les écoulements (conductivité hydraulique et porosité par exemple).
- Paramètres liés au transport (modélisation des interactions entre les contaminants et le milieu) : coefficient de retard, coefficient de diffusion.

La calibration d'un modèle consiste à trouver le jeu de paramètres permettant de reproduire au mieux les observations, sachant que seule une partie des phénomènes physico-chimiques intervenant en réalité est reproduite. C'est un problème inverse souvent mal posé et sans unicité de la solution [Carrera *et al.*, 2005]. Beaucoup d'observations de l'environnement, de mesures *in situ* et d'analyses en laboratoire sont nécessaires pour construire puis valider les modèles d'écoulement et de transport.

Certaines méthodes de modélisation inverse caractérisent la source de la contamination [Shlomi et Michalak, 2007 ; Datta *et al.*, 2011 ; Ayvaz, 2016], et d'autres le milieu naturel [Huang *et al.*, 2008 ; Schwede et Cirpka, 2010]. Il est souvent nécessaire de réaliser des études géostatistiques en amont des simulations d'écoulement et de transport, pour caractériser la variabilité du milieu ou la source [Morakinyo et Mackay, 2006 ; Mackay et Morakinyo, 2006 ; Bennett *et al.*, 2019]. Par ailleurs, des méthodes géostatistiques peuvent intervenir pour déterminer les champs de paramètres du milieu. L'une d'entre elles est la méthode des points pilotes [RamaRao *et al.*, 1995 ; Doherty, 2003 ; Alcolea *et al.*, 2006], dont l'idée est d'assigner à un certain nombre de points du domaine de modélisation des valeurs de paramètres hydrodynamiques et d'interpoler entre ces points par krigeage. Les paramètres à retrouver par modélisation inverse sont les valeurs assignées aux points pilotes. Cependant, cette méthode n'est pas toujours très stable et ses performances dépendent du nombre et de la position des points pilotes. D'autres méthodes d'inversion géostatistique sont basées sur un couplage entre de l'information *a priori* sur la structure de covariance et des mesures [Schwede et Cirpka, 2009]. Cependant, ces méthodes sont assez complexes et coûteuses en temps de calcul. De nombreux travaux ont été réalisés pour les améliorer, et notamment réduire les coûts de calcul [Pau *et al.*, 2013 ; Lee et Kitanidis, 2014 ; Hörning *et al.*, 2019].

Dans la suite de ce manuscrit, les simulations d'écoulement et de transport pourront être appelées simulations numériques, à ne pas confondre avec les simulations géostatistiques ou réalisations d'une fonction aléatoire.

1.4 Combinaison de simulations numériques et de la géostatistique

1.4.1 Intérêt de la combinaison

Les Sections 1.1 et 1.3 ont mis en évidence les avantages et limites de deux méthodes classiquement utilisées pour quantifier les risques de contamination dans les sols et les nappes : les estimations ou simulations géostatistiques et les simulations numériques.

En résumé, l'estimation géostatistique peut manquer de précision lorsque les données ne sont pas suffisamment nombreuses pour ajuster de manière satisfaisante un modèle de variogramme. Il est aussi possible que les méthodes de traitement des non-stationnarités mentionnées précédemment ne soient pas applicables. Par ailleurs, les méthodes géostatistiques présentées sont basées sur des modélisations à caractère mathématique et ne prennent pas en compte les connaissances physiques sur le phénomène étudié, ce qui peut conduire à des estimations manquant de réalisme.

La simulation numérique permet quant à elle de reproduire une partie des caractéristiques physiques du phénomène étudié, selon les choix de modélisation de l'utilisateur. Elle demande cependant parfois beaucoup d'observations et d'analyses en laboratoire, ainsi que de temps de calcul, pour obtenir une calibration convenable, permettant de reproduire les observations de façon suffisamment précise.

Ainsi, il semble intéressant de combiner les deux approches, pour compenser leurs limites respectives et, dans le cadre de ce travail, améliorer les estimations de concentrations en contaminants.

1.4.2 Méthodes existantes

Plusieurs méthodes combinant la géostatistique et les simulations numériques ont été développées, pour diverses applications.

Krigeage avec dérive externe

Le krigeage avec dérive externe permet d'intégrer de l'information issue de simulations numériques dans une estimation géostatistique. En effet, un ou plusieurs champs simulés peuvent être utilisés comme fonctions de dérive, donnant la tendance du phénomène étudié. Par exemple, Rivest *et al.* [2008] utilisent des charges simulées par un code éléments finis comme fonctions de dérive pour estimer des charges hydrauliques en régime permanent. Les estimations ainsi obtenues sont plus précises et plus réalistes que celles obtenues par krigeage ordinaire ou krigeage universel. Cette approche est également utilisée pour étudier d'autres phénomènes : pour estimer des températures de subsurface dans le cadre d'exploration géothermique [Rühaak *et al.*, 2014], pour quantifier des risques liés à une pollution de l'air à l'ozone [Wackernagel *et al.*, 2004], ou pour construire des cartes haute-résolution de températures [Varentsov *et al.*, 2019]. Les estimations obtenues par cette méthode sont plus réalistes et respectent les observations, tant que les simulations numériques donnent une tendance fiable du phénomène étudié et sont suffisamment corrélées aux observations.

Inférence de propriétés spatiales à partir de simulations numériques

Une autre classe de méthodes consiste à extraire certaines propriétés de la variable étudiée d'une ou de plusieurs simulations numériques. Par exemple, Michael *et al.* [2010] utilisent une simulation de dépôt sédimentaire basée sur des processus géologiques pour calibrer des simulations géostatistiques à base d'objet dans le but de simuler de manière réaliste l'hétérogénéité d'un aquifère. Varentsov *et al.* [2019] calculent un variogramme expérimental à partir d'un champ de température simulé et Yong *et al.* [2019] à partir d'une modélisation numérique de sédimentation. Dans ces deux exemples, le variogramme expérimental calculé sur une simulation permet d'avoir plus de points qu'un variogramme expérimental calculé sur les observations, ce qui facilite l'ajustement d'un modèle de variogramme stationnaire.

Dans d'autres cas, plusieurs simulations sont utilisées pour calculer des covariances empiriques. Une approche de calcul de covariances empiriques non stationnaires a été proposée par C. Roth [Roth, 1995 ; Roth *et al.*, 1998] pour estimer la transmissivité d'un aquifère (modélisation inverse). Les covariances non stationnaires sont calculées à partir de simulations numériques de charges hydrauliques obtenues pour différentes réalisations du modèle géostatistique de transmissivité. Cette approche est détaillée dans le Chapitre 2. On trouve dans la littérature d'autres exemples de calcul de covariances empiriques. Ainsi, Shlomi et Michalak [2007] utilisent des simulations d'écoulement et de transport pour estimer la distribution d'un panache de contaminant. Dans cet exemple, tous les paramètres d'entrée du modèle d'écoulement et de transport sont connus, sauf la source de contamination. Celle-ci est d'abord caractérisée par modélisation inverse, puis des simulations sont générées. La covariance de la distribution du panache est ensuite inférée à partir d'un ensemble de simulations. Elle est cohérente avec les équations physiques régissant le transport, et elle coïncide avec la covariance des observations puisque les simulations coïncident avec les observations. Schwede et Cirpka [2010] utilisent également des simulations d'écoulement et de transport pour estimer la distribution d'un panache de contaminant. Cette fois, seul le champ de conductivité hydraulique est inconnu. Une procédure de modélisation inverse permet de générer quelques milliers de champs de conductivité hydraulique, conditionnés et reproduisant les observations de concentration. Ces champs sont ensuite utilisés pour simuler des panaches de contaminant. La covariance de la distribution de la concentration est alors inférée à partir de l'ensemble de simulations.

Dans ces deux derniers exemples, la plupart des paramètres d'entrée sont supposés connus et le seul paramètre inconnu (la source dans le premier exemple, le champ de conductivité hydraulique dans le second) est inféré par modélisation inverse, sous certaines hypothèses. Les simulations générées sont donc cohérentes avec les observations.

Yang *et al.* [2018] calculent également des covariances empiriques à partir de sorties de simulations numériques, sous la dénomination de *physics-informed kriging*. Dans le cas synthétique présenté, seul le champ de conductivité hydraulique est inconnu. Celui-ci n'est pas inféré par modélisation inverse, mais sa structure est supposée parfaitement connue.

Enfin, Rivest et Marcotte [2012] et Rivest *et al.* [2012] utilisent des simulations d'écoulement et de transport en zone saturée pour transformer l'espace de travail et améliorer la prise en compte de l'anisotropie d'un panache de contaminant. Les simulations permettent de passer de l'espace cartésien à deux dimensions à l'espace {charge hydraulique, vitesse d'écoulement}, espace dit naturel qui permet de mieux reproduire l'anisotropie des panaches. Un krigeage ordinaire des concentrations est ensuite réalisé dans cet espace, avant de revenir aux coordonnées cartésiennes.

Combinaison de plusieurs types d'informations

La géostatistique est également utilisée pour coupler les informations provenant de plusieurs types de simulations numériques. Par exemple, Fetel et Caumon [2008] utilisent le krigeage avec dérive externe pour coupler les informations issues de simulations précises mais coûteuses (code dit haute-fidélité) et les informations de simulations moins précises mais rapides à obtenir (code dit basse-fidélité) pour caractériser les écoulements dans des réservoirs. Ce type d'approche est couramment utilisé pour construire des méta-modèles couplant codes haute-fidélité et codes basse-fidélité, quel que soit le phénomène étudié. Par exemple, Le Gratiet et Garnier [2014] construisent un méta-modèle pour un code d'hydrodynamique par cokrigeage de données issues d'un code haute-fidélité et d'un code basse-fidélité. Le cokrigeage multifidélité est également utilisé par Yang *et al.* [2019], qui considèrent que les données haute fidélité sont des mesures et que les données basse fidélité sont les sorties d'un code numérique.

Apprentissage statistique à base physique

L'apprentissage statistique à base physique (*physics-informed machine learning*) est une branche de l'apprentissage statistique développée depuis quelques années. Comme la géostatistique, les méthodes usuelles d'apprentissage statistique requièrent un grand nombre de données d'apprentissage pour être performantes, et ne garantissent pas la cohérence physique des résultats obtenus. Ainsi, l'apprentissage statistique à base physique vise les mêmes objectifs que la combinaison de la géostatistique et des simulations numériques : intégrer à un modèle statistique une information de nature physique, permettant de

compenser le manque de données et d'améliorer la cohérence physique des résultats.

L'idée générale de l'apprentissage statistique à base physique est de minimiser un écart entre les données d'entrée et de sortie, pénalisé par un terme lié aux équations régissant le phénomène étudié. Plus précisément, si l'on cherche à résoudre un problème général de la forme $f(x, t) = f[u(x, t), \lambda] = 0$, où f est un opérateur différentiel non linéaire paramétré par λ et $u(x, t)$ la variable d'état observée N fois aux points $(x_u^i, t_u^i)_{i=1\dots N}$, on construit un modèle en minimisant la somme de deux termes :

$$\phi = \frac{1}{N} \sum_{i=1}^N |u(x_u^i, t_u^i) - u^i|^2 + \frac{1}{N_f} \sum_{i=1}^{N_f} |f(x_f^i, t_f^i)|^2 \quad (1.34)$$

où u^i est la sortie du modèle aux points $(x_u^i, t_u^i)_{i=1\dots N}$ et $(x_f^i, t_f^i)_{i=1\dots N_f}$ est l'ensemble des points sur lesquels on cherche à imposer $f(x, t) = 0$. Par exemple, [Raissi et al. \[2019\]](#) construisent un réseau de neurones à base physique permettant de calibrer des modèles basés sur des équations aux dérivées partielles non linéaires de type Navier-Stokes. [Tartakovsky et al. \[2019\]](#) construisent un modèle d'apprentissage statistique (développement de Karhunen-Loève) appliqué à un problème de diffusion, pour créer un méta-modèle et rechercher les paramètres de l'équation.

1.5 Synthèse du chapitre

La caractérisation de la contamination dans les sols et les nappes est un enjeu important. Il est ainsi nécessaire de mettre au point des méthodes d'estimation fiables et robustes, tant dans un cadre spatial que dans un cadre spatio-temporel. L'estimation géostatistique fournit une première classe de méthodes couramment utilisées, mais qui peuvent s'avérer limitées lorsque le phénomène à modéliser est trop complexe par rapport au nombre de données disponibles. Les codes d'écoulement et de transport sont une autre classe de méthodes. Cependant, leur utilisation requiert souvent un long travail de calibration pour réussir à reproduire au mieux les données de concentration.

L'objectif de la combinaison de l'estimation géostatistique et des simulations numériques est d'utiliser une information physiquement réaliste dans une modélisation gouvernée par les mathématiques. Plusieurs exemples issus de la littérature ont été présentés dans ce chapitre. Ces exemples soulignent l'intérêt de la combinaison, pour diverses applications et sous certaines hypothèses. Deux méthodes parmi celles présentées sont adaptées et mises en œuvre dans ce travail. Elles sont détaillées dans le Chapitre 2.

Chapitre 2

Mise en œuvre des méthodes géostatistiques développées

Deux méthodes, parmi celles présentées dans le Chapitre 1 (Section 1.4), ont été développées dans ce travail : le krigeage avec dérive externe et le calcul de covariances ou de variogrammes à partir d'un ensemble de simulations d'écoulement et de transport. Quelques éléments de mise en œuvre de ces méthodes sont détaillés dans ce chapitre.

2.1 Le krigeage avec dérive externe

Le krigeage avec dérive externe permet de modéliser la non-stationnarité de la moyenne de la variable étudiée. L'information donnée par une ou plusieurs variables auxiliaires, qui peuvent être issues de simulations numériques, est prise en compte.

2.1.1 Principe de la méthode

Le krigeage avec dérive externe est basé sur la décomposition de $Z : Z(x) = Y(x) + m(x)$ où Y est un résidu stationnaire d'espérance nulle (Complément 2.1). La dérive est écrite sous la forme d'une combinaison linéaire de variables auxiliaires appelées fonctions de dérive :

$$m(x) = b_0 + \sum_{j=1}^J b_j W_j(x) \quad (2.1)$$

où les coefficients $b_{j,j=0\dots J}$ sont à déterminer. Si l'on travaille en voisinage glissant, ces coefficients peuvent varier d'un point cible à l'autre.

Le krigeage avec dérive externe

Comme en krigeage ordinaire, on cherche un estimateur de Z en un point x_0 de la forme :

$$Z^*(x_0) = \sum_{n=1}^N \lambda_n Z(x_n) \quad (2.2)$$

L'erreur d'estimation est donnée dans l'Équation (2.3) et son espérance dans l'Équation (2.4).

$$R(x_0) = Z(x_0) - Z^*(x_0) = Y(x_0) + b_0 + \sum_{j=1}^J b_j W_j(x_0) - \sum_{n=1}^N \lambda_n Z(x_n) \quad (2.3)$$

$$\begin{aligned} E[R(x_0)] &= b_0 + \sum_{j=1}^J b_j W_j(x_0) - \sum_{n=1}^N \lambda_n \left[b_0 + \sum_{j=1}^J b_j W_j(x_n) \right] \\ &= b_0 \left(1 - \sum_{n=1}^N \lambda_n \right) + \sum_{j=1}^J b_j \left[W_j(x_0) - \sum_{n=1}^N \lambda_n W_j(x_n) \right] \end{aligned} \quad (2.4)$$

L'annulation de cette erreur aboutit à la condition de non biais suivante :

$$\sum_{n=1}^N \lambda_n W_j(x_n) = W_j(x_0) \quad \text{pour tous } j = 0 \dots J, \text{ en posant } W_0 = 1 \quad (2.5)$$

La variance de R s'écrit :

$$\text{Var}[R(x_0)] = C(0) - 2 \sum_{n=1}^N \lambda_n C(x_n, x_0) + \sum_{n=1}^N \sum_{n'=1}^N \lambda_n \lambda_{n'} C(x_n, x_{n'}) \quad (2.6)$$

où C est la covariance de Y (égale à celle de Z). Pour minimiser cette variance sous les contraintes de l'Équation (2.5), on introduit $J + 1$ multiplicateurs de Lagrange. La dérivation partielle par rapport aux λ_n et aux multiplicateurs de Lagrange aboutit au système de krigeage suivant :

$$\begin{bmatrix} C & W \\ {}^t W & 0 \end{bmatrix} \begin{bmatrix} \Lambda \\ \mu \end{bmatrix} = \begin{bmatrix} C_0 \\ W_0 \end{bmatrix} \quad (2.7)$$

où $C = [C(x_n, x_{n'})]$ est la matrice de covariance des observations, $W = [W_j(x_n)]$ (dimension $N \times (J + 1)$) contient les W_j évalués aux points x_n , 0 est une matrice de 0, Λ est le vecteur des poids de krigeage, μ celui des multiplicateurs de Lagrange, $C_0 = [C(x_n, x_0)]$ est le vecteur des covariances entre les observations et le point x_0 et $W_0 = [W_j(x_0)]$ contient les W_j évalués en x_0 .

Complément 2.1

En pratique, l'inférence de la covariance ou du variogramme des résidus est complexe. En effet, puisque les coefficients de la dérive sont inconnus, les vrais résidus ne sont pas observés. Le variogramme calculé sur ces résidus est alors biaisé, et ce même si l'estimateur des coefficients de la dérive est optimal [Chilès et Delfiner, 2012 ; Li et al., 2015]. Dans le cadre de ce travail, l'inférence du variogramme des résidus est effectuée avec la procédure implémentée dans le package RGeostats [MINES ParisTech/ARMINES, 2020], qui consiste à calculer un variogramme ajusté sur les résidus empiriques de la régression linéaire. Il est possible d'utiliser une procédure itérative pour corriger le biais sur le variogramme expérimental, mais en pratique, cela ne modifie pas significativement les résultats d'estimation. D'autres méthodes existent, comme celles consistant à estimer simultanément par maximum de vraisemblance les coefficients de la dérive et les paramètres du modèle de covariance [Mardia et Marshall, 1984].

2.1.2 Choix de la dérive

Dans le cadre du krigeage avec dérive externe, les informations issues de simulations numériques sont prises en compte dans la dérive. Par exemple, Rivest et al. [2008] utilisent quatre fonctions de dérive, qui correspondent à quatre charges hydrauliques simulées avec des conditions aux limites simplifiées par rapport à la réalité. Varentsov et al. [2019] utilisent quant à eux les résultats d'une seule simulation, mais adaptée en fonction de certaines conditions au moment de l'estimation.

D'autres types de fonction de dérive sont envisagés dans ce travail : lorsque le nombre de paramètres d'entrée incertains est trop important pour obtenir une seule simulation représentative, ou un nombre restreint de cas simplifiés, la moyenne ou la médiane d'un ensemble de simulations sont utilisées. Ces alternatives seront discutées dans les applications sur cas synthétiques et cas réels des Parties II et III.

2.2 Les covariances et variogrammes numériques

Les modèles de covariance ou de variogramme classiquement utilisés dans les systèmes de krigeage peuvent être remplacés par des covariances ou variogrammes empiriques. Ceux-ci sont calculés à partir de plusieurs réalisations de la variable régionalisée obtenues à partir de simulations numériques.

2.2.1 Calcul et propriétés

On suppose que P réalisations de Z sont disponibles. Ces réalisations proviennent de l'application d'un opérateur non linéaire H à un jeu de paramètres d'entrée $\Theta : Z = H(\Theta)$, où H correspond ici à un code numérique d'écoulement et de transport (mais la méthode peut être étendue à un cadre plus général). La randomisation des paramètres Θ introduit une variabilité parmi les réalisations de Z .

La covariance empirique entre deux points x et x' d'un domaine \mathcal{D} est calculée comme suit :

$$C(x, x') = \frac{1}{P} \sum_{p=1}^P [Z_p(x) - \overline{Z(x)}] [Z_p(x') - \overline{Z(x')}] \quad (2.8)$$

où $Z_p(x)$ est la valeur obtenue au point x pour la p -ième réalisation et $\overline{Z(x)}$ est la moyenne empirique de $Z(x)$ sur les P réalisations. Il s'agit d'une covariance non stationnaire dépendant des points x et x' . Dans la suite de ce travail, on ne cherchera pas à moyenner cette covariance sur les couples (x, x') distants de la même distance. Cette covariance vérifie bien la condition de positivité (Complément 2.2). Le krigeage avec covariances numériques a été implémenté dans le package RGeostats [MINES ParisTech/ARMINES, 2020] (exemple à l'Annexe A).

Le variogramme empirique est calculé suivant le même principe :

$$\gamma(x, x') = \frac{1}{P} \sum_{p=1}^P \frac{1}{2} [Z_p(x) - Z_p(x')]^2 \quad (2.9)$$

Comme pour la covariance empirique, il s'agit d'un variogramme non stationnaire, dépendant des points x et x' et de type négatif conditionnel (Complément 2.2). Le lien entre covariance et variogramme empirique est donné par l'équation suivante :

$$\gamma(x, x') = \frac{1}{2} [s^2(x) + s^2(x') + (\overline{Z(x)} - \overline{Z(x')})^2] - C(x, x') \quad (2.10)$$

où $s^2(x)$ est la variance empirique des P réalisations au point x .

Propriété de la covariance empirique

Soient K réels $\omega_k, k=1\dots K$ et K points $x_k, k=1\dots K$.

$$\begin{aligned} \sum_{k=1}^K \sum_{\ell=1}^K \omega_k \omega_\ell C(x_k, x_\ell) &= \sum_{k=1}^K \sum_{\ell=1}^K \omega_k \omega_\ell \frac{1}{P} \sum_{p=1}^P [Z_p(x_k) - \overline{Z(x_k)}] [Z_p(x_\ell) - \overline{Z(x_\ell)}] \\ &= \frac{1}{P} \sum_{p=1}^P \left(\sum_{k=1}^K \omega_k [Z_p(x_k) - \overline{Z(x_k)}] \sum_{\ell=1}^K \omega_\ell [Z_p(x_\ell) - \overline{Z(x_\ell)}] \right) \\ &= \frac{1}{P} \sum_{p=1}^P \left(\sum_{k=1}^K \omega_k [Z_p(x_k) - \overline{Z(x_k)}] \right)^2 \end{aligned} \quad (2.11)$$

Donc cette covariance est bien définie positive.

Propriété du variogramme empirique

Soient K réels $\omega_k, k=1\dots K$ tels que $\sum_{k=1}^K \omega_k = 0$ et K points $x_k, k=1\dots K$.

$$\begin{aligned} - \sum_{k=1}^K \sum_{\ell=1}^K \omega_k \omega_\ell \gamma(x_k, x_\ell) &= - \sum_{k=1}^K \sum_{\ell=1}^K \omega_k \omega_\ell \frac{1}{P} \sum_{p=1}^P \frac{1}{2} [Z_p(x_k) - Z_p(x_\ell)]^2 \\ &= - \frac{1}{2P} \sum_{p=1}^P \left(\sum_{k=1}^K \sum_{\ell=1}^K \omega_k \omega_\ell [Z_p(x_k)^2 + Z_p(x_\ell)^2 - 2Z_p(x_k)Z_p(x_\ell)] \right) \\ &= - \frac{1}{P} \sum_{p=1}^P \left(\sum_{k=1}^K \omega_k \sum_{\ell=1}^K \omega_\ell Z_p(x_\ell)^2 - \sum_{k=1}^K \omega_k Z_p(x_k) \sum_{\ell=1}^K \omega_\ell Z_p(x_\ell) \right) \\ &= \frac{1}{P} \sum_{p=1}^P \left(\sum_{k=1}^K \omega_k Z_p(x_k) \right)^2 \quad \text{car} \quad \sum_{k=1}^K \omega_k = 0 \end{aligned} \quad (2.12)$$

Donc ce variogramme est bien de type négatif conditionnel.

Ces méthodes de calcul de covariances et variogrammes empiriques utilisent un ensemble de P réalisations, et non une unique réalisation comme c'est parfois le cas dans la littérature [Michael *et al.*, 2010 ; Yong *et al.*, 2019]. Dans la suite de ce travail, les covariances et variogrammes ainsi calculés sont appelés covariances numériques et variogrammes numériques. Les covariances et variogrammes numériques sont supposés reproduire précisément et de manière réaliste la structure spatiale de la variable étudiée, puisqu'ils sont calculés à partir de simulations à base physique. Chaque réalisation est générée à partir d'un jeu différent de paramètres d'entrée. Dans l'idéal, l'ensemble des réalisations rend compte des incertitudes sur les paramètres d'entrée et couvre tous les scénarios possibles.

Ces modèles peuvent être utilisés sans plus d'hypothèses sur la variable étudiée, comme sa stationnarité ou son isotropie, contrairement aux modèles ajustés sur des structures expérimentales. Cependant, pour que ces modèles reproduisent suffisamment bien la structure spatiale de la variable étudiée, il faut que les simulations numériques soient suffisamment proches de la réalité : un travail non négligeable de modélisation et de recherche d'informations sur les paramètres d'entrée est à mener en amont de l'estimation géostatistique. De plus, plusieurs dizaines à quelques centaines de simulations numériques sont nécessaires pour obtenir une estimation convenable des covariances et variogrammes.

Par ailleurs, il est intéressant de remarquer que la covariance numérique est la même covariance que celle utilisée dans l'étape de correction du filtre de Kalman d'ensemble : cette étape correspond en réalité à un krigeage simple avec des covariances numériques. Sous forme matricielle, l'équation de conditionnement par krigeage simple de la particule p s'écrit comme suit :

$$p_K = p + \Lambda(y - Bp + \mathcal{E}) \quad (2.13)$$

où p_K est la particule p après krigeage, Λ est le vecteur des poids de krigeage, y celui des observations, B est l'opérateur permettant de passer de p aux observations sur p et \mathcal{E} est l'erreur de mesure associée à y . Les poids Λ sont obtenus par résolution de l'équation suivante :

$$\Sigma_X(tB) - \Lambda\Sigma_Y = 0 \Leftrightarrow \Lambda = \Sigma_X(tB)\Sigma_Y^{-1} = \Sigma_X(tB)(B\Sigma_X(tB) + \Sigma_\epsilon)^{-1} \quad (2.14)$$

où Σ_X est la matrice de covariance du système, Σ_Y est la matrice de covariance aux points d'observations et Σ_ϵ la matrice de covariance des erreurs de mesure. On identifie dans l'Équation (2.14) les mêmes termes que dans l'Équation (1.29). En particulier, les matrices C de l'Équation (1.29) et Σ_X de l'Équation (2.14) sont identiques.

2.2.2 Krigeage ordinaire et stationnarité de la moyenne

Les covariances et variogrammes numériques peuvent être intégrés dans un système de krigeage ordinaire, où la moyenne de Z est inconnue mais supposée constante sur le domaine de travail. Cette hypothèse, parfois contraignante, peut être relâchée en prenant en compte la variabilité de la moyenne de Z . Si la moyenne de Z dépend de la localisation x , la condition de non biais devient :

$$\sum_{n=1}^N \lambda_n m(x_n) = m(x_0) \quad (2.15)$$

D'où le système de krigeage suivant :

$$\begin{bmatrix} C & M \\ {}^t M & 0 \end{bmatrix} \begin{bmatrix} \Lambda \\ \mu \end{bmatrix} = \begin{bmatrix} C_0 \\ M_0 \end{bmatrix} \quad (2.16)$$

où $M = [\overline{Z(x_n)}]$ est le vecteur contenant les moyennes empiriques de Z aux observations et $M_0 = \overline{Z(x_0)}$. Les différences entre les estimations réalisées avec une moyenne constante et une moyenne variable seront discutées dans les applications sur cas synthétiques et cas réels des Parties II et III.

2.2.3 Régularisation

Dans le cas où les covariances et variogrammes numériques conduisent à des systèmes de krigeage instables (si par exemple peu de simulations numériques ont été réalisées), ceux-ci peuvent être régularisés ou lissés, de plusieurs manières [Chilès et Delfiner, 2012].

Régularisation de Z

Soit v_x un voisinage du point x . La régularisée de Z sur v_x est calculée ainsi :

$$Z_v(x) = \frac{1}{|v_x|} \sum_{x_i \in v_x} Z(x_i) \quad (2.17)$$

où $|v_x|$ est la taille de v_x . Les covariances et variogrammes numériques des régularisées sont ensuite calculés directement sur les P réalisations de Z régularisées : $C_v(x, x') = C(Z_v(x), Z_v(x'))$ et $\gamma_v(x, x') = \gamma(Z_v(x), Z_v(x'))$. Cela modifie la variabilité à petite distance, car le support est différent.

Régularisation après calcul des covariances et variogrammes

Une autre approche, équivalente à la précédente, consiste à régulariser les covariances et variogrammes des réalisations de Z non modifiées.

Pour la covariance, on a :

$$C_v(x, x') = \frac{1}{|v_x|} \frac{1}{|v_{x'}|} \sum_{x_i \in v_x} \sum_{x_j \in v_{x'}} C(x_i, x_j) \quad (2.18)$$

Pour le variogramme, on a :

$$\begin{aligned} \gamma_v(x, x') &= \frac{1}{2} \text{Var}[Z_v(x) - Z_v(x')] = \frac{1}{2} C(Z_v(x) - Z_v(x'), Z_v(x) - Z_v(x')) \\ &= \frac{1}{2} C_v(x, x) + \frac{1}{2} C_v(x', x') - C_v(x, x') \\ &= \bar{\gamma}(x, x') - \frac{1}{2} \bar{\gamma}(x, x) - \frac{1}{2} \bar{\gamma}(x', x') \\ \text{avec } \bar{\gamma}(x, x') &= \frac{1}{|v_x|} \frac{1}{|v_{x'}|} \sum_{x_i \in v_x} \sum_{x_j \in v_{x'}} \gamma(x_i, x_j) \end{aligned} \quad (2.19)$$

Randomisation de l'origine de la grille

Une dernière approche consiste à randomiser l'origine X_0 de la grille des simulations numériques dans v_{X_0} . Contrairement à la régularisation, cette approche ne modifie pas la variance de Z , et revient à moyenner les covariances, en supposant que la randomisation est uniforme dans v_{X_0} :

$$C_{\text{random.}}(x, x') = \frac{1}{|v_{X_0}|} \sum_{\substack{x_i \in v_x, x_j \in v_{x'} \\ \text{avec } x_i - x_j = x - x'}} C(x_i, x_j) \quad (2.20)$$

2.2.4 Variantes spatio-temporelles et multivariées

Cas spatio-temporel

L'une des principales limites du krigeage spatio-temporel est la difficulté à trouver un modèle de covariance dont les hypothèses sous-jacentes ne soient pas trop limitantes et à l'ajuster. Les covariances spatio-temporelles numériques permettent de contourner cette difficulté. La covariance entre le point x à la date t et le point x' à la date t' est calculée sur un jeu de P réalisations obtenues par simulations numériques :

$$C(x, x', t, t') = \frac{1}{P} \sum_{p=1}^P [Z_p(x, t) - \overline{Z(x, t)}] [Z_p(x', t') - \overline{Z(x', t')}] \quad (2.21)$$

où $Z_p(x, t)$ est la valeur de Z au point x et à la date t pour la p -ième réalisation et $\overline{Z(x, t)}$ est la moyenne empirique des P réalisations au point x et à la date t . Ces covariances sont non stationnaires et non symétriques par rapport au temps : $C(x, x', t, t') \neq C(x, x', t', t)$.

Des variogrammes spatio-temporels numériques, non stationnaires et non symétriques, sont calculés ainsi :

$$\gamma(x, x', t, t') = \frac{1}{P} \sum_{p=1}^P \frac{1}{2} [Z_p(x, t) - Z_p(x', t')]^2 \quad (2.22)$$

Les performances du krigeage spatio-temporel avec covariances numériques sont comparées à celles des méthodes d'assimilation de données dans la Partie II.

Cas multivariable

Pour certaines applications, il est intéressant de travailler avec plusieurs variables et d'effectuer un cokrigeage. Au lieu d'ajuster des modèles sur les données pour modéliser les structures simples et croisées des variables étudiées, des structures empiriques croisées sont calculées à partir de P réalisations obtenues par simulations numériques :

$$C_{12}(x, x') = \frac{1}{P} \sum_{p=1}^P [Z_p^{(1)}(x) - \overline{Z^{(1)}(x)}] [Z_p^{(2)}(x') - \overline{Z^{(2)}(x')}] \quad (2.23)$$

$$\gamma_{12}(x, x') = \frac{1}{P} \sum_{p=1}^P \frac{1}{2} [Z_p^{(1)}(x) - Z_p^{(1)}(x')] [Z_p^{(2)}(x) - Z_p^{(2)}(x')] \quad (2.24)$$

où $C_{12}(x, x') = \text{Cov}[Z^{(1)}(x), Z^{(2)}(x')]$ est la covariance croisée entre les variables $Z^{(1)}$ et $Z^{(2)}$ aux points x et x' et $\gamma_{12}(x, x')$ est le variogramme croisé entre les variables $Z^{(1)}$ et $Z^{(2)}$ aux points x et x' .

2.2.5 Simulations géostatistiques

Les covariances numériques peuvent être utilisées pour simuler des fonctions aléatoires, éventuellement conditionnées. La méthode de simulation d'une fonction aléatoire est décomposée en plusieurs étapes.

Étape 1 : les P simulations numériques sont transformées en champs gaussiens par anamorphose inverse φ_x^{-1} . Dans le cadre de ce travail, on se propose de travailler avec des anamorphoses non stationnaires, puisque l'on a une distribution de Z en chaque point x grâce aux P simulations. Pour chaque point x du domaine de travail et chaque simulation, on a $Z_p(x) = \varphi_x(Y_p(x))$. Cette étape suppose une absence d'effet zéro.

Étape 2 : les covariances numériques $C(x, x')$ sont calculées à partir des $Y_{p,p=1\dots P}$. Soit $\Sigma = [C(x, x')]$ la matrice de covariance ainsi obtenue. La décomposition de Cholesky permet d'obtenir une matrice A telle que $\Sigma = A({}^t A)$.

Étape 3 : pour obtenir une nouvelle réalisation de Y , un vecteur gaussien X est généré, et le produit $Y_{sim} = AX$ est une nouvelle réalisation de Y . Cette étape suppose que les Y_p sont gaussiens.

Étape 4 : les anamorphoses non stationnaires permettent enfin de passer de Y_{sim} à Z_{sim} , nouvelle réalisation de Z dont la covariance est Σ .

L'applicabilité de cette méthode, notamment au regard des hypothèses sous-jacentes, est discutée dans la Partie II.

2.3 Synthèse du chapitre

Des éléments de mise en œuvre de deux méthodes combinant la géostatistique et les simulations numériques, le krigeage avec dérive externe et le krigeage avec covariances ou variogrammes numériques, ont été détaillés dans ce chapitre. Au cours de ce travail, ces méthodes sont mises en place sur un cas synthétique (Partie II) et sur deux jeux de données réels (Partie III).

Il est important de garder à l'esprit que les simulations générées pour implémenter ces méthodes ne sont pas réalisées dans l'objectif de calibrer un modèle numérique. On cherchera à trouver des jeux de paramètres reproduisant convenablement les observations, mais sans utiliser de méthodes inverses pour trouver le meilleur jeu de paramètres et sans générer des simulations d'écoulement et de transport conditionnées par les données de concentration.

Deuxième partie

Applications à un cas synthétique

Chapitre 3

Construction d'un cas synthétique en zone non saturée

Ce chapitre présente la construction d'un cas synthétique reproduisant la migration d'un contaminant en zone non saturée. La variabilité spatiale à grande échelle du milieu influence les écoulements souterrains et le transport de contaminants, tant en zone saturée [Knudby et Carrera, 2005 ; Renard et Allard, 2013] qu'en zone non saturée [Appels *et al.*, 2018]. L'intérêt est ici porté sur la variabilité à plus petite échelle (au sein d'un même faciès par exemple), qui influence également l'écoulement et le transport de contaminants, via les paramètres hydrodynamiques [Botros *et al.*, 2009].

La majorité des résultats de ce chapitre sont repris d'un article publié au cours de cette thèse [Pannecoucke *et al.*, 2019], présentant l'étude de l'influence de la variabilité spatiale des paramètres hydrodynamiques sur la migration d'un panache de contaminant.

3.1 Les paramètres hydrodynamiques de la zone non saturée

L'une des principales difficultés de la simulation numérique d'écoulement et de transport est la détermination des paramètres nécessaires à la résolution des systèmes d'équations (Chapitre 1, Complément 1.4), en particulier dans le cas d'une contamination en zone non saturée. En effet, les paramètres régissant l'écoulement sont complexes à caractériser mais influencent fortement la migration des contaminants, comme cela sera montré dans la Section 3.2.3.

3.1.1 Les paramètres du modèle de Mualem-van Genuchten

La résolution de l'équation de Richards (Équation (1.31)) régissant l'écoulement en zone non saturée requiert la connaissance du lien entre charge (h), conductivité hydraulique (K) et teneur en eau (θ). Ces liens sont donnés par des relations empiriques. Dans ce travail, le modèle de Mualem-van Genuchten est utilisé [Mualem, 1976 ; van Genuchten, 1980]. Il est décrit par les équations suivantes :

$$\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{(1 + |\alpha h|^n)^m} \quad \text{avec} \quad m = 1 - \frac{1}{n} \quad (3.1)$$

$$\overline{K(h)} = \overline{K_s} S^{\frac{1}{2}} \left[1 - \left(1 - S^{\frac{1}{m}} \right)^2 \right] \quad \text{avec} \quad S = \frac{\theta(h) - \theta_r}{\theta_s - \theta_r} \quad (3.2)$$

où θ_r [$L^3.L^{-3}$] et θ_s [$L^3.L^{-3}$] sont respectivement la teneur en eau résiduelle et à saturation, α [L^{-1}] est un paramètre lié à l'inverse de la pression d'entrée d'air, n [-] est un paramètre de taille de pore et $\overline{K_s}$ [$L.T^{-1}$] est le tenseur de conductivité hydraulique à saturation. Les paramètres θ_r , θ_s , K_s , α et n sont les paramètres du modèle de Mualem-van Genuchten. Ils varient en fonction du type de sol considéré. Les Équations (3.1) et (3.2) décrivent les courbes donnant la teneur en eau et la conductivité hydraulique en fonction de la charge (Figure 3.1, pour trois types de sol décrits dans le Tableau 3.1).

TABEAU 3.1 – Valeurs moyennes des paramètres du modèle de Mualem-van Genuchten pour trois types de sols de la base de données ROSETTA [Site de l'USDA].

Type de sol	θ_r [$m^3.m^{-3}$]	θ_s [$m^3.m^{-3}$]	K_s [$m.s^{-1}$]	α [m^{-1}]	n [-]
Limon silto-argileux (<i>silty clay loam</i>)	0,090	0,482	$1,3.10^{-6}$	0,84	1,52
Sable limoneux (<i>sandy loam</i>)	0,039	0,387	$4,4.10^{-6}$	2,67	1,45
Sable (<i>sand</i>)	0,053	0,375	$7,4.10^{-5}$	3,52	3,17

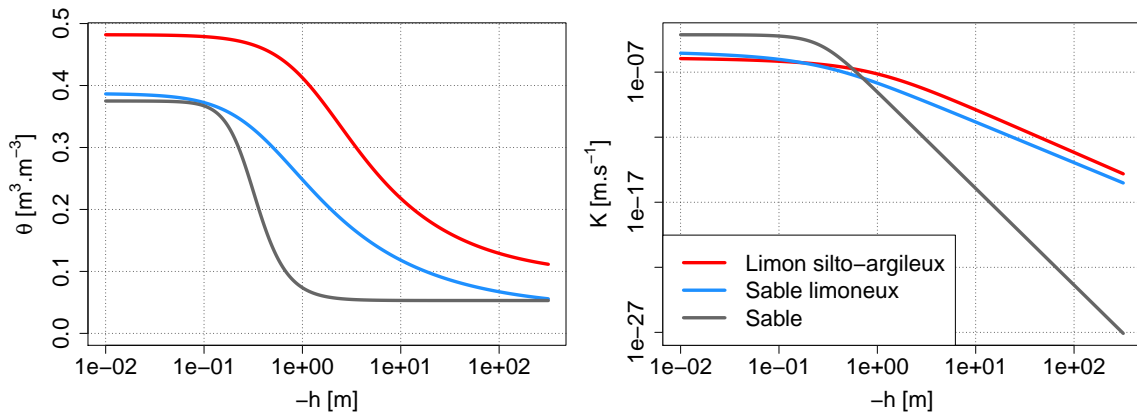


FIGURE 3.1 – Teneur en eau (θ) et conductivité hydraulique relative (K) en fonction de la charge (h) pour trois types de sol (Tableau 3.1).

3.1.2 Les fonctions de pédo-transfert

La caractérisation des paramètres hydrodynamiques nécessite des expérimentations *in situ* ou en laboratoire complexes [Schaap *et al.*, 2004 ; Le Coz *et al.*, 2013 ; Léger *et al.*, 2014]. Cependant, les paramètres du modèle de Mualem-van Genuchten sont liés à la texture du sol (répartition par taille des particules le constituant, qui est relativement facile à mesurer par analyse granulométrique) par des fonctions empiriques appelées fonctions de pédo-transfert [Wösten *et al.*, 1999 ; Tóth *et al.*, 2015 ; Zhang et Schaap, 2017, 2019]. Par exemple, la fonction de pédo-transfert Rosetta3 [Schaap *et al.*, 2001 ; Zhang et Schaap, 2017], basée sur un réseau de neurones, permet d'estimer les paramètres de Mualem-van Genuchten à partir des proportions de sable, de limon et d'argile de l'échantillon considéré. Cette fonction de pédo-transfert utilise un grand nombre d'échantillons couvrant des types de sols très différents [Patil et Singh, 2016 ; van Looy *et al.*, 2017]. Elle a été réévaluée récemment et un module Python est disponible [Zhang et Schaap, 2017], ce qui la rend accessible.

3.1.3 La variabilité des paramètres de Mualem-van Genuchten à l'échelle d'une parcelle

La base de données WoSIS [Site de l'IRSIC] contient les données d'environ 96 000 profils de sols, dont 20% sont certifiées et standardisées [Batjes *et al.*, 2017]. Les proportions de sable, de limon et d'argile sont disponibles pour la plupart des échantillons. Ces propriétés texturales sont utilisées pour estimer les paramètres de Mualem-van Genuchten et leur variabilité à l'échelle de la parcelle (constituée ici par un cercle d'un kilomètre de diamètre contenant les données d'au moins dix profils) à l'aide de la fonction de pédo-transfert Rosetta3. Quarante-neuf sites sont ainsi identifiés, et pour chacun d'entre eux, les moyennes et variances de K_s , α et n sont calculées (Figure 3.2), ce qui permet d'obtenir un ordre de grandeur de la variance des champs de paramètres en fonction de leur moyenne. Les variances des logarithmes des paramètres varient d'environ un ordre de grandeur parmi les sites considérés. Les variances augmentent avec la moyenne pour les paramètres K_s et n , et diminuent pour le paramètre α .

Les coefficients de corrélation entre les paires des logarithmes des paramètres ($\log K_s$ - $\log \alpha$, $\log K_s$ - $\log n$, $\log \alpha$ - $\log n$) sont également calculés (Figure 3.3). Ils dépendent de la valeur moyenne des paramètres. Pour les faibles valeurs de K_s (moins de $2 \cdot 10^{-6}$ à $3 \cdot 10^{-6}$ m.s⁻¹), ces coefficients sont très étalés. Pour des valeurs de K_s plus élevées, les coefficients de corrélation tendent à se stabiliser.

3.2 Influence de la variabilité spatiale des paramètres du modèle de Mualem-van Genuchten sur la migration d'un panache de tritium

La calibration des modèles d'écoulement et de transport est d'autant plus difficile que les paramètres de Mualem-van Genuchten sont complexes à caractériser. Ainsi, leur variabilité spatiale au sein d'une même formation est rarement prise en compte dans les études de migration de contaminants [Bugai *et al.*, 2012 ; Testoni *et al.*, 2017 ; Kwicklis *et al.*, 2019]. On s'intéresse ici à l'influence de la variabilité spatiale des

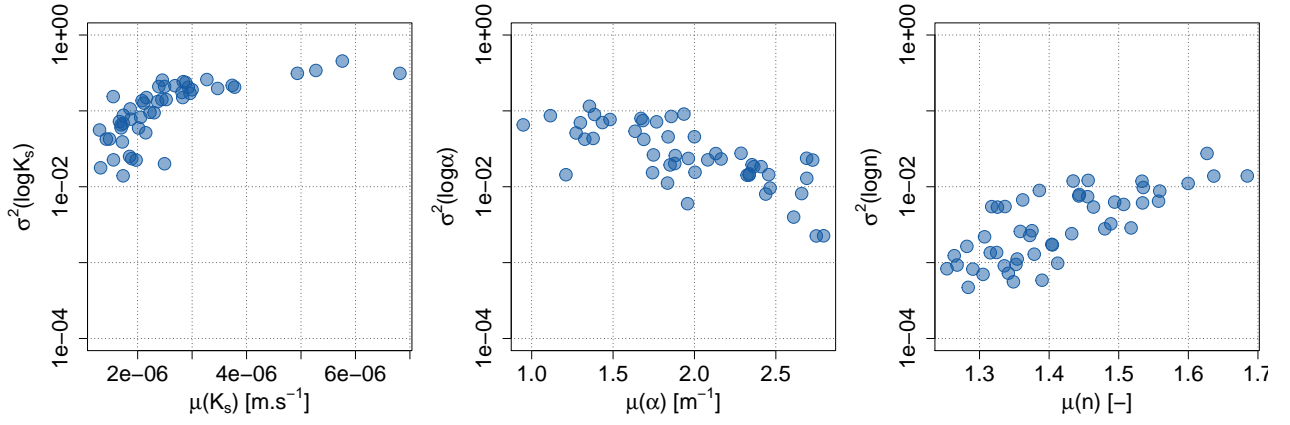


FIGURE 3.2 – Variances des logarithmes de K_s , α et n en fonction de leur moyenne. Chaque point représente un site identifié.

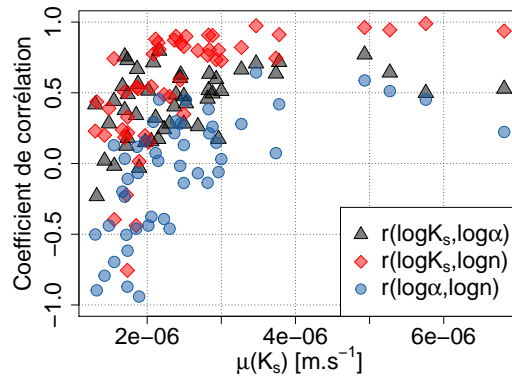


FIGURE 3.3 – Coefficients de corrélation entre les logarithmes des paramètres K_s , α et n en fonction de la moyenne de K_s pour chaque site identifié.

paramètres K_s , α et n sur la migration d'un panache de tritium. Les paramètres de teneur en eau θ_r et θ_s peuvent être mesurés plus facilement [Jadoon *et al.*, 2012] et sont donc des paramètres d'entrée moins critiques. Par ailleurs, des études de sensibilité préliminaires réalisées dans des conditions de climat tempéré montrent que la variabilité spatiale de θ_r et θ_s n'a pas d'influence significative sur la migration d'un contaminant.

3.2.1 Construction d'un cas d'étude

Le code MELODIE est utilisé pour simuler la migration de panaches de tritium, à partir de champs de paramètres de Mualem-van Genuchten variables.

Simulations géostatistiques de champs de paramètres de Mualem-van Genuchten

Dans de nombreux travaux, les paramètres K_s , α et n sont représentés par des champs aléatoires lognormaux [Botros *et al.*, 2009 ; Paleologos *et al.*, 2015 ; Tan *et al.*, 2017]. La distribution d'un champ aléatoire lognormal est caractérisée par la moyenne et la variance de la transformée logarithmique, et sa variabilité spatiale par la covariance ou le variogramme de la transformée logarithmique. Le variogramme choisi pour modéliser la variabilité spatiale du logarithme des paramètres K_s , α et n est le modèle sphérique, couramment ajusté sur des données de paramètres de sol [Herbst *et al.*, 2006 ; Botros *et al.*, 2009 ; Bevington *et al.*, 2016] :

$$\gamma(h) = \begin{cases} \frac{3}{2} \frac{|h|}{a} - \frac{1}{2} \left(\frac{|h|}{a} \right)^3, & \text{si } |h| \leq a \\ 1, & \text{si } |h| > a \end{cases} \quad (3.3)$$

où a est la portée du variogramme. Afin de prendre en compte l'anisotropie des structures géologiques, une anisotropie géométrique avec différentes portées dans les directions horizontales (a_x) et verticales (a_z) est introduite. Le milieu homogène équivalent, au sein duquel les paramètres ne varient pas dans l'espace, est un sable limoneux (Tableau 3.1, ligne 2). Des exemples de champs ainsi simulés sont donnés dans la Figure 3.4.

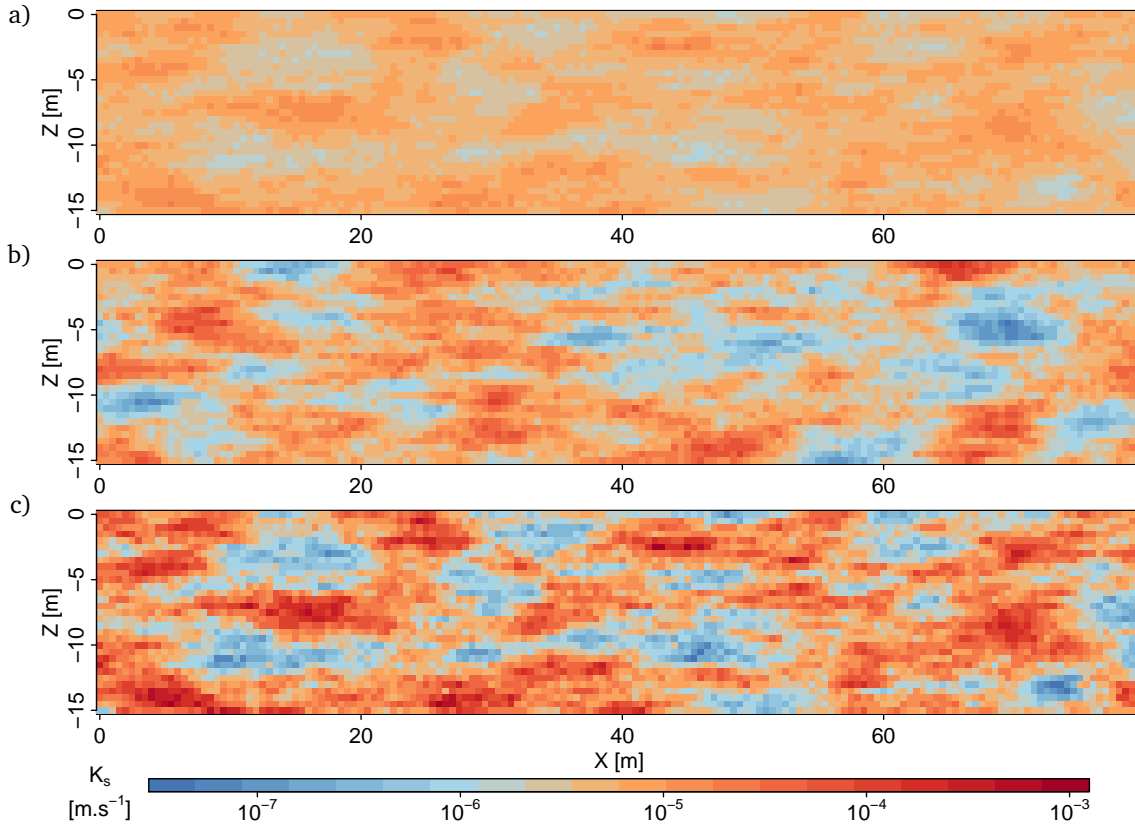


FIGURE 3.4 – Exemples de champs du paramètre K_s (échelle logarithmique) simulés sur des coupes verticales. La portée horizontale vaut 10 m, la portée verticale 3 m et la moyenne de K_s vaut $4,4 \cdot 10^{-6} \text{ m.s}^{-1}$. La variance change d'une simulation à l'autre. a) $\sigma^2(\log K_s) = 0,05$; b) $\sigma^2(\log K_s) = 0,3$; c) $\sigma^2(\log K_s) = 0,7$.

Autres paramètres et modélisation avec MELODIE

Les champs de paramètres de Mualem-van Genuchten ne constituent qu'une partie des paramètres d'entrée nécessaires pour simuler la migration d'un contaminant avec le code MELODIE.

Le domaine de modélisation considéré est une coupe verticale, parallèle au sens de l'écoulement. Le choix d'une modélisation à deux dimensions est d'abord lié au temps et à la capacité de calcul disponibles. Par ailleurs, on se place dans le cas d'un écoulement dont la direction est constante, mais avec un niveau piézométrique pouvant varier assez nettement au cours du temps. En effet, la plupart des installations nucléaires françaises sont implantées dans le lit majeur de cours d'eau importants, en milieu alluvial, et les conditions aux limites choisies sont en cohérence avec ce type de milieu. Le domaine de modélisation mesure 100 m de large par 15 m de profondeur, et est discrétisé en triangles de 0,5 m de côté. Les conditions aux limites sont décrites sur le schéma de la Figure 3.5. Les paramètres de transport sont donnés dans le Tableau 3.2. Une source ponctuelle de tritium est imposée au milieu de la surface supérieure du domaine de modélisation. La migration du panache de tritium est ensuite simulée pendant six ans (Figure 3.6). Les résultats de simulations correspondent à des activités volumiques, exprimées en Becquerels par mètres cubes d'eau (Complément 3.1).

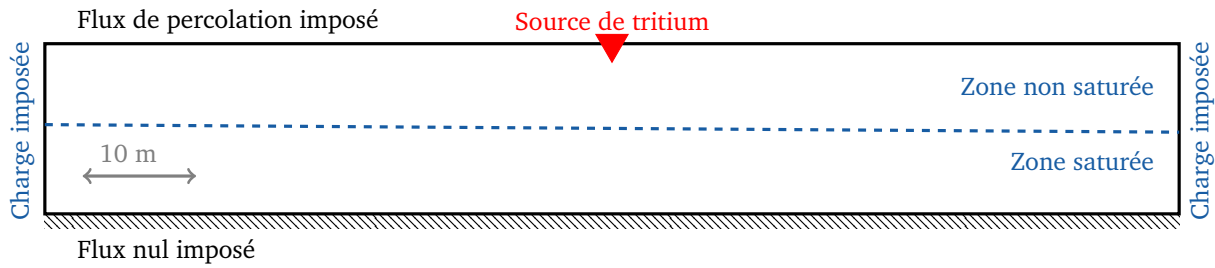


FIGURE 3.5 – Schéma du domaine de modélisation (100 m × 15 m). La charge imposée à gauche est égale à 7,5 m. Le gradient latéral vaut 0,004 m.m⁻¹.

TABLEAU 3.2 – Paramètres de transport (Équation (1.32)), dans le cas du tritium, considéré comme un traceur parfait. ω est la porosité totale, D_x et D_z sont les composantes horizontales et verticales du tenseur de dispersivité, d est le coefficient de diffusion moléculaire, λ est la constante de désintégration du tritium et R est le coefficient de retard.

ω [-]	D_x [m]	D_z [m]	d [m ² .s ⁻¹]	λ [s ⁻¹]	R [-]
0,38	0,05	0,005	$1,2 \cdot 10^{-10}$	$1,83 \cdot 10^{-9}$	1

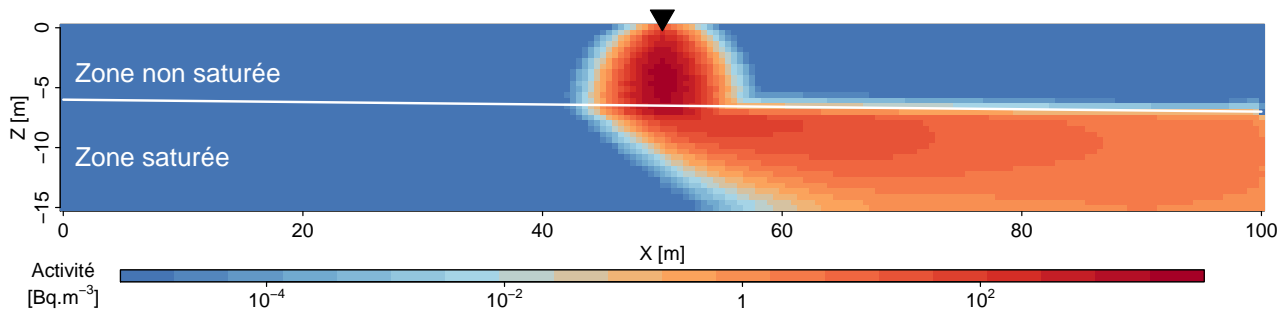


FIGURE 3.6 – Panache de tritium simulé dans un milieu homogène de type sable limoneux (Tableau 3.1, ligne 2), six ans après le déversement. Le triangle noir indique la localisation de la source de tritium.

Activité et décroissance radioactive

L'activité d'une source radioactive est la vitesse de désintégration de l'élément la constituant. Son unité dans le système international est le Becquerel (Bq), qui correspond à une désintégration par seconde. L'activité $A(t)$ est le nombre de désintégrations en une seconde :

$$A(t) = -\frac{dN}{dt} = -\frac{d(N_0 e^{-\lambda t})}{dt} = N_0 \lambda e^{-\lambda t} \quad (3.4)$$

où N_0 [-] est le nombre d'atomes initial et λ [T⁻¹] la constante de désintégration, liée à la demi-vie $t_{1/2}$ par la relation $\lambda = \frac{\ln 2}{t_{1/2}}$.

Complément 3.1

3.2.2 Outils d'analyse de sensibilité

Pour étudier l'influence de la variabilité spatiale des paramètres de Mualem-van Genuchten, des séries de 100 simulations d'écoulement et de transport sont générées. On considère d'abord un seul paramètre de Mualem-van Genuchten variable et les deux autres constants sur le domaine, puis les trois paramètres variant de manière corrélée. Chaque série de 100 simulations est générée en utilisant les valeurs de variances et portées données dans le Tableau 3.3 (Séries 1. pour K_s variable, Séries 2. pour α variable, Séries 3. pour n variable et Séries 4. pour les trois paramètres variables). Les moyennes sont celles du Tableau 3.1 (ligne 2). Les champs de paramètres de Mualem-van Genuchten sont les seuls paramètres variant d'une simulation

d'écoulement et de transport à l'autre. Ainsi, 41 séries de 100 simulations sont générées, permettant de couvrir les gammes de variance de la Figure 3.2.

Les analyses de sensibilité se concentrent sur la partie des panaches simulés située dans la zone non saturée. Pour chaque série de 100 panaches, des critères moyens sont calculés afin de quantifier l'influence des variances et des portées des champs de paramètres de Mualem-van Genuchten sur la migration du tritium. Trois critères sont pris en compte :

1. la surface moyenne des panaches d'une série (nombre de mailles dans la zone non saturée dont l'activité volumique est supérieure à 1 Bq.m^{-3}) ainsi que l'écart-type de ces surfaces ;
2. la moyenne des rapports d'élongation des ellipses équivalentes aux panaches d'une série (Complément 3.2) ;
3. la distance moyenne entre les centres de gravité des panaches d'une série et le centre de gravité du panache simulé dans le milieu homogène équivalent (panache représenté sur la Figure 3.6).

TABLEAU 3.3 – Paramètres utilisés pour les séries de 100 simulations géostatistiques de paramètres de Mualem-van Genuchten. $\sigma^2 = 0$ correspond à un champ constant.

		$\sigma^2(\log K_s)$	$\sigma^2(\log \alpha)$	$\sigma^2(\log n)$	a_x [m]	a_z [m]	$r(\log K_s, \log \alpha)$ et $r(\log K_s, \log n)$
Série 1.1	▲	0,05	0	0	10	3	-
Série 1.2	▲	0,1	0	0	10	3	-
Série 1.3	▲	0,2	0	0	5	3	-
Série 1.4	▲	0,2	0	0	10	3	-
Série 1.5	▲	0,2	0	0	15	3	-
Série 1.6	▲	0,2	0	0	20	3	-
Série 1.7	▲	0,2	0	0	10	1	-
Série 1.8	▲	0,2	0	0	10	5	-
Série 1.9	▲	0,3	0	0	10	3	-
Série 1.10	▲	0,5	0	0	10	3	-
Série 1.11	▲	0,7	0	0	10	3	-
Série 1.12	▲	1	0	0	10	3	-
Série 2.1	◆	0	0,01	0	10	3	-
Série 2.2	◆	0	0,03	0	10	3	-
Série 2.3	◆	0	0,04	0	5	3	-
Série 2.4	◆	0	0,04	0	10	3	-
Série 2.5	◆	0	0,04	0	15	3	-
Série 2.6	◆	0	0,04	0	20	3	-
Série 2.7	◆	0	0,04	0	10	1	-
Série 2.8	◆	0	0,04	0	10	5	-
Série 2.9	◆	0	0,05	0	10	3	-
Série 2.10	◆	0	0,06	0	10	3	-
Série 2.11	◆	0	0,08	0	10	3	-
Série 2.12	◆	0	0,1	0	10	3	-
Série 2.13	◆	0	0,15	0	10	3	-
Série 2.14	◆	0	0,20	0	10	3	-
Série 3.1	●	0	0	0,0001	10	3	-
Série 3.2	●	0	0	0,0005	10	3	-
Série 3.3	●	0	0	0,001	5	3	-
Série 3.4	●	0	0	0,001	10	3	-
Série 3.5	●	0	0	0,001	15	3	-
Série 3.6	●	0	0	0,001	20	3	-
Série 3.7	●	0	0	0,001	10	1	-
Série 3.8	●	0	0	0,001	10	5	-
Série 3.9	●	0	0	0,002	10	3	-
Série 3.10	●	0	0	0,003	10	3	-
Série 3.11	●	0	0	0,004	10	3	-
Série 3.12	●	0	0	0,005	10	3	-
Série 4.1	▲	0,2	0,04	0,001	10	3	0,8
Série 4.2	▲	0,4	0,08	0,002	10	3	0,8
Série 4.3	▲	0,6	0,12	0,003	10	3	0,8

Inertie et ellipse équivalente à un panache

a) Centre de gravité

Les coordonnées (x_G, z_G) du centre de gravité d'un panache sont définies par :

$$x_G = \frac{1}{A} \sum_{n=1}^N a_n x_n \quad \text{et} \quad z_G = \frac{1}{A} \sum_{n=1}^N a_n z_n \quad (3.5)$$

où dans un cas bidimensionnel un point est décrit par ses coordonnées (x_n, z_n) et sa valeur d'activité a_n , N est le nombre de points du panache et $A = \sum_{n=1}^N a_n$ est l'activité totale du panache.

b) Matrice d'inertie

La matrice d'inertie G d'un panache de centre de gravité (x_G, z_G) est donnée par :

$$G = \frac{1}{A} \begin{bmatrix} \sum_{n=1}^N a_n (z_n - z_G)^2 & - \sum_{n=1}^N a_n (x_n - x_G)(z_n - z_G) \\ - \sum_{n=1}^N a_n (x_n - x_G)(z_n - z_G) & \sum_{n=1}^N a_n (x_n - x_G)^2 \end{bmatrix} \quad (3.6)$$

Cette matrice décrit la répartition de l'activité d'un panache autour de son centre de gravité. Ses vecteurs propres donnent les directions principales d'inertie (les deux directions autour desquelles l'activité est équitablement répartie).

c) Ellipse équivalente

L'ellipse équivalente à un panache est définie comme l'ellipse de même centre de gravité et de même matrice d'inertie. Dans son repère principal, la matrice d'inertie G_{ell} d'une ellipse de demi-grand axe a et de demi-petit axe b vaut :

$$G_{ell} = \begin{bmatrix} \frac{\pi b^3 a}{4} & 0 \\ 0 & \frac{\pi a^3 b}{4} \end{bmatrix} \quad (3.7)$$

a et b sont alors donnés par :

$$a = \left(\frac{16G_2^3}{\pi^2 G_1} \right)^{\frac{1}{8}} \quad \text{et} \quad b = \left(\frac{16G_1^3}{\pi^2 G_2} \right)^{\frac{1}{8}} \quad (3.8)$$

où G_1 et G_2 sont les valeurs propres de la matrice G . L'ellipse équivalente à un panache est donc l'ellipse de centre (x_G, z_G) , orientée suivant les vecteurs propres de la matrice G et de demi-grand axe et demi-petit axe donnés par l'Équation (3.8). Un exemple de panache avec son ellipse équivalente est donné sur la Figure 3.7.

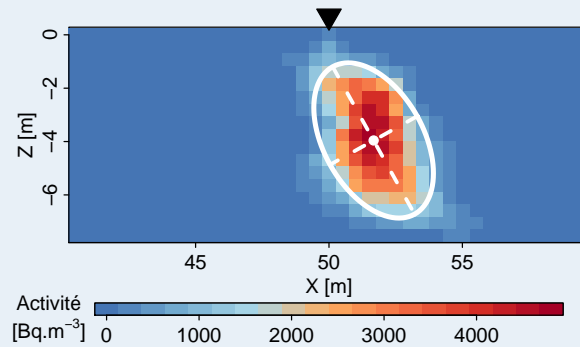


FIGURE 3.7 – Exemple de panache de tritium avec son ellipse équivalente, centrée sur le centre de gravité du panache. Le triangle noir marque la localisation de la source de tritium.

3.2.3 Résultats

Les résultats sont présentés pour le dernier pas de temps des simulations (six ans après le déversement).

Influence de la variance des champs de paramètres

Pour étudier l'influence de la variance des champs de paramètres, les portées sont fixées à 10 m dans la direction horizontale et 3 m dans la direction verticale, ce qui correspond aux séries 1.1, 1.2, 1.4 et 1.9 à 1.12 pour K_s , 2.1, 2.2, 2.4 et 2.9 à 2.14 pour α , 3.1, 3.2, 3.4 et 3.9 à 3.12 pour n (un paramètre variable à la fois). Quel que soit le paramètre considéré (K_s , α ou n), l'augmentation de la variance des champs induit une augmentation des surfaces des panaches simulés ainsi qu'une augmentation de la variabilité (écart-type) des surfaces des panaches (Figure 3.8). Par ailleurs, les relations entre les variances des champs et les surfaces des panaches semblent similaires pour les trois paramètres lorsque l'on considère les rapports suivants : $\sigma^2(\log K_s)/\sigma^2(\log n) = 200$ et $\sigma^2(\log K_s)/\sigma^2(\log \alpha) = 5$. Lorsque l'on considère les panaches simulés à partir de champs de paramètres corrélés (avec des rapports de 200 entre $\sigma^2(\log K_s)$ et $\sigma^2(\log n)$ d'une part et de 5 entre $\sigma^2(\log K_s)$ et $\sigma^2(\log \alpha)$ d'autre part) les moyennes et écart-types des surfaces sont proches des séries pour lesquelles un seul paramètre varie, à même variance. À noter que d'autres valeurs de coefficient de corrélation ont été testées et conduisent aux mêmes résultats.

Le rapport moyen entre les longueurs de petit axe (b) et de grand axe (a) des ellipses équivalentes aux panaches donne une indication sur la forme des panaches : plus il est faible, moins les panaches sont circulaires. L'augmentation de la variance des champs de paramètres conduit à une diminution du rapport b/a moyen, donc à des panaches plus allongés (Figure 3.9).

Enfin, lorsque les variances des champs de paramètres augmentent, les panaches sont plus dispersés dans l'espace (Figure 3.10). Cette information est résumée par la distance moyenne entre les centres de gravité des panaches d'une série et le centre du panache simulé dans le milieu homogène équivalent (Figure 3.11). Quel que soit le paramètre considéré, l'augmentation de la variance des champs conduit à une augmentation de la distance moyenne entre les centres de gravité, dans les deux directions. La tendance est cependant plus marquée dans la direction horizontale que dans la direction verticale. Les panaches simulés dans des milieux où les paramètres varient de manière corrélée présentent les mêmes caractéristiques que ceux simulés dans les milieux où un seul paramètre varie.

Influence des portées des champs de paramètres

Les mêmes indicateurs ont été calculés pour les séries où seules les portées varient. Les résultats détaillés sont donnés dans l'Annexe B. En résumé, les valeurs des portées ne semblent pas avoir d'influence sur les panaches de tritium simulés (au moins pour les plages de portées considérées), quel que soit le paramètre de Mualem-van Genuchten.

Résultats complémentaires

Quelques simulations complémentaires ont été réalisées avec un autre radionucléide et dans un autre milieu. Les résultats sont présentés à l'Annexe C. Dans ces deux cas, la variabilité spatiale des paramètres de Mualem-van Genuchten a également un impact sur les panaches simulés ; les conclusions sont les mêmes que pour le tritium dans un sable limoneux.

3.2.4 Discussion et conclusion

L'étude réalisée quantifie l'influence de la variabilité spatiale à l'échelle de la parcelle des champs de paramètres K_s , α et n sur le transport de tritium dans la zone non saturée. Si l'on considère les variances estimées à l'échelle de la parcelle avec la base de donnée WoSIS, le paramètre n est celui qui semble le plus critique (*i.e.* qui conduit aux panaches les plus étendus en moyenne, avec le plus d'incertitudes sur leur taille et localisation). En effet, les variances moyennes des paramètres à l'échelle de la parcelle sont : $\sigma^2(\log K_s)=0,14$, $\sigma^2(\log \alpha)=0,036$ et $\sigma^2(\log n)=0,005$. Par rapport au panache simulé dans le milieu homogène équivalent, les panaches simulés dans des milieux avec ces variances conduisent à :

- des augmentations respectives de 25%, 20% et 65% des surfaces des panaches ;

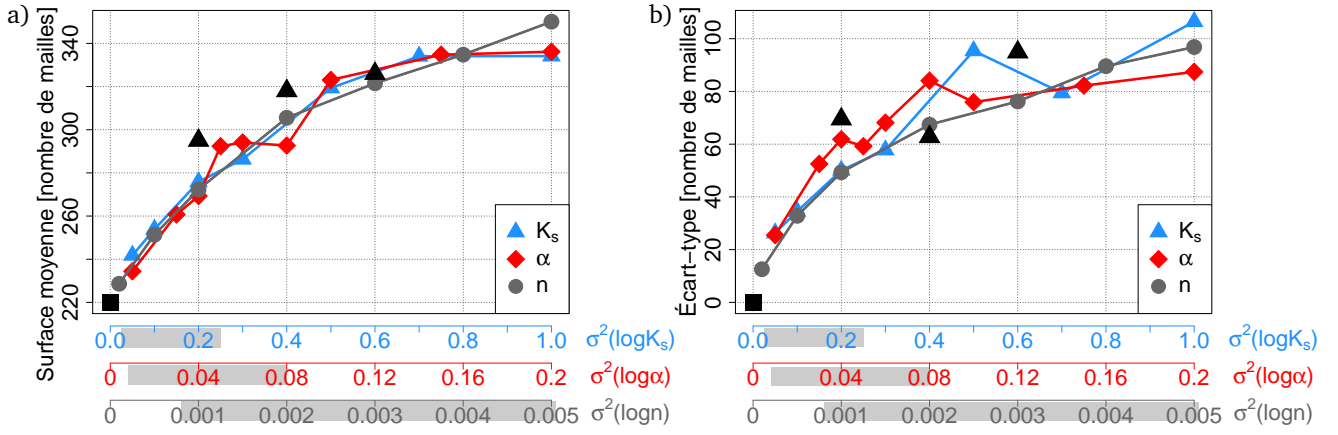


FIGURE 3.8 – a) Moyenne et b) écart-type des surfaces de chaque série de 100 panaches en fonction de la variance des champs de paramètres. Le carré noir correspond au panache simulé dans le milieu homogène équivalent. Les trois triangles correspondent aux séries avec champs de paramètres corrélés (séries 4.1 à 4.3). Les rectangles grisés sur les axes correspondent aux plages de variances à l'échelle de la parcelle trouvées avec la base de données WoSIS.

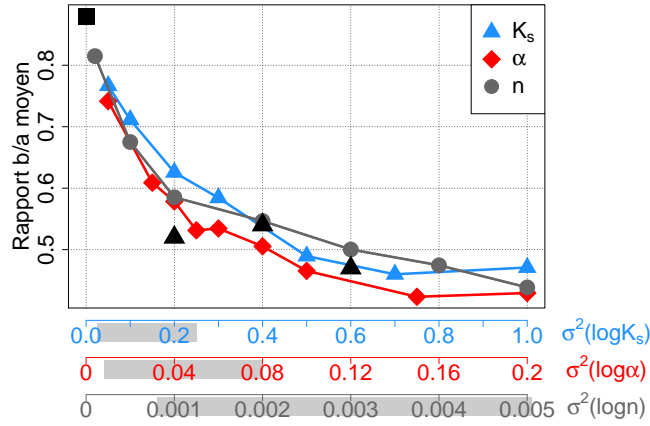


FIGURE 3.9 – Rapport moyen entre le petit axe b et le grand axe a des ellipses équivalentes aux panaches.

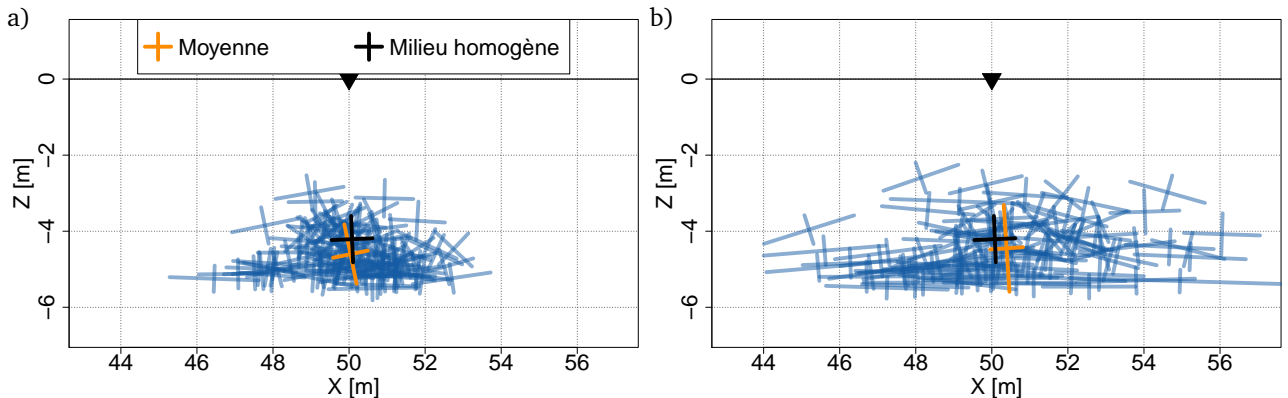


FIGURE 3.10 – Ensemble des positions des centres de gravité des panaches de deux séries de simulations. Les longueurs des branches des croix sont proportionnelles aux axes de l'ellipse équivalente à chaque panache et les croix sont orientées comme l'ellipse équivalente. Le triangle noir indique la localisation de la source de tritium. a) Série 2.2 ($\sigma^2(\log \alpha) = 0,03$) ; b) série 2.13 ($\sigma^2(\log \alpha) = 0,15$).

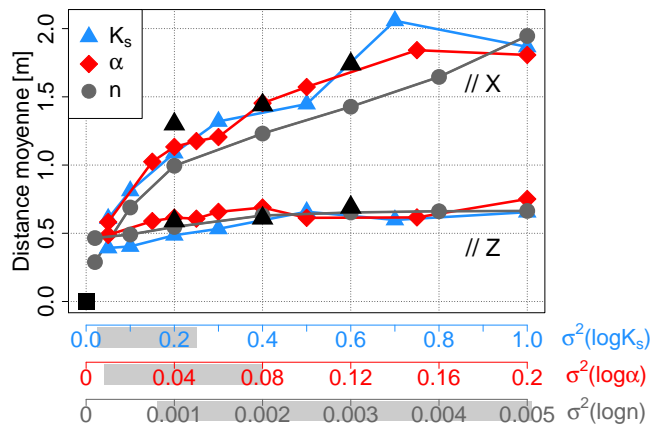


FIGURE 3.11 – Distance moyenne entre les centres de gravité des panaches d’une série et le centre de gravité du panache simulé dans le milieu homogène équivalent (représenté par la croix noire sur la Figure 3.10), en distinguant les distances dans la direction horizontale (X) et dans la direction verticale (Z).

- des diminutions respectives de 20%, 30% et 50% de la circularité des panaches ;
- des distances moyennes dans la direction horizontale de 0,8 m, 1 m et 1,8 m avec le centre de gravité du panache simulé dans le milieu homogène.

Ces résultats sont cohérents avec d’autres études soulignant la plus grande influence du paramètre n sur l’écoulement [Lu et Zhang, 2002 ; Tan et al., 2017].

Par ailleurs, les valeurs de portées utilisées ne permettent pas de faire apparaître une éventuelle influence des portées des champs de paramètres sur les panaches simulés. Ces valeurs sont du même ordre de grandeur que la dimension des panaches simulés (quelques mètres), ce qui peut expliquer cette influence peu marquée. Cependant, ces valeurs de portées sont cohérentes avec la littérature [Russo et al., 1997 ; Botros et al., 2009]. Pour des plus grandes valeurs de portée (de quelques centaines de mètres à quelques kilomètres), l’influence de la portée horizontale sur l’écoulement en zone non saturée pourrait être plus significative [Tan et al., 2017].

Finalement, ce travail souligne qu’une caractérisation insuffisante de la variabilité spatiale des paramètres de Mualem-van Genuchten (incertitude sur les variances des champs) induit une forte incertitude sur les panaches simulés. En effet, toutes les simulations sont réalisées avec les mêmes conditions initiales et conditions aux limites ; les seuls changements d’une simulation à l’autre concernent les champs de paramètres de Mualem-van Genuchten.

3.3 Utilisation des résultats précédents pour construire un cas synthétique

Les résultats de la Section 3.2.3 montrent que la variabilité spatiale à petite échelle des paramètres du modèle de Mualem-van Genuchten suffit à générer par simulation d’écoulement et de transport des panaches de contaminant significativement différents les uns des autres. Cette variabilité est exploitée pour construire un cas synthétique permettant de mettre en œuvre les méthodes d’estimation présentées au Chapitre 2.

Le point de départ de la construction de ce cas synthétique est la texture du sol (proportions de sable, de limon et d’argile) plutôt que directement les paramètres de Mualem-van Genuchten, pour des raisons opérationnelles. En effet, la texture du sol est un paramètre relativement simple à mesurer contrairement aux paramètres de Mualem-van Genuchten. Cette section reprend la construction du cas synthétique détaillée dans un article publié au cours de la thèse [Pannecoucke et al., 2020a].

3.3.1 Simulations de champs de texture et d’un panache de référence

La construction du cas synthétique consiste à générer :

1. un panache dit de référence, c’est-à-dire qui sera échantillonné et dont on cherchera à estimer l’activité avec les méthodes présentées au Chapitre 2 ;

2. un jeu de plusieurs centaines de simulations de panaches, permettant de calculer des covariances ou variogrammes numériques et des fonctions de dérive.

Champs de texture

Puisque la plupart des installations nucléaires sont situées en bordure d'un cours d'eau, on se place dans un milieu sableux dont les propriétés moyennes sont proches de celles du milieu homogène utilisé dans l'étude de sensibilité de la Section 3.2. Les proportions de sable, de limon et d'argile sont supposées suivre une distribution normale et la variabilité spatiale de ces proportions est modélisée par un variogramme exponentiel [Reza *et al.*, 2015 ; Usowicz et Lipiec, 2017 ; Taye *et al.*, 2018] :

$$\gamma(h) = 1 - \exp\left(-\frac{3|h|}{a}\right) \quad (3.9)$$

Une anisotropie entre les directions horizontales et verticales est introduite. Un triplet de champs aléatoires représentant les proportions de sable, de limon et d'argile est simulé suivant la méthode décrite dans le Complément 3.3, avec des portées de 20 m dans la direction horizontale et 5 m dans la direction verticale, et des moyennes et écart-types donnés dans le Tableau 3.4. Un exemple de triplet ainsi généré est donné sur la Figure 3.12. Ce triplet sert de référence.

TABLEAU 3.4 – Moyennes et écart-types des distributions des proportions de sable, de limon et d'argile.

	Sable	Limon	Argile
Moyenne [%]	75	12.5	12.5
Écart-type [%]	10	6	6

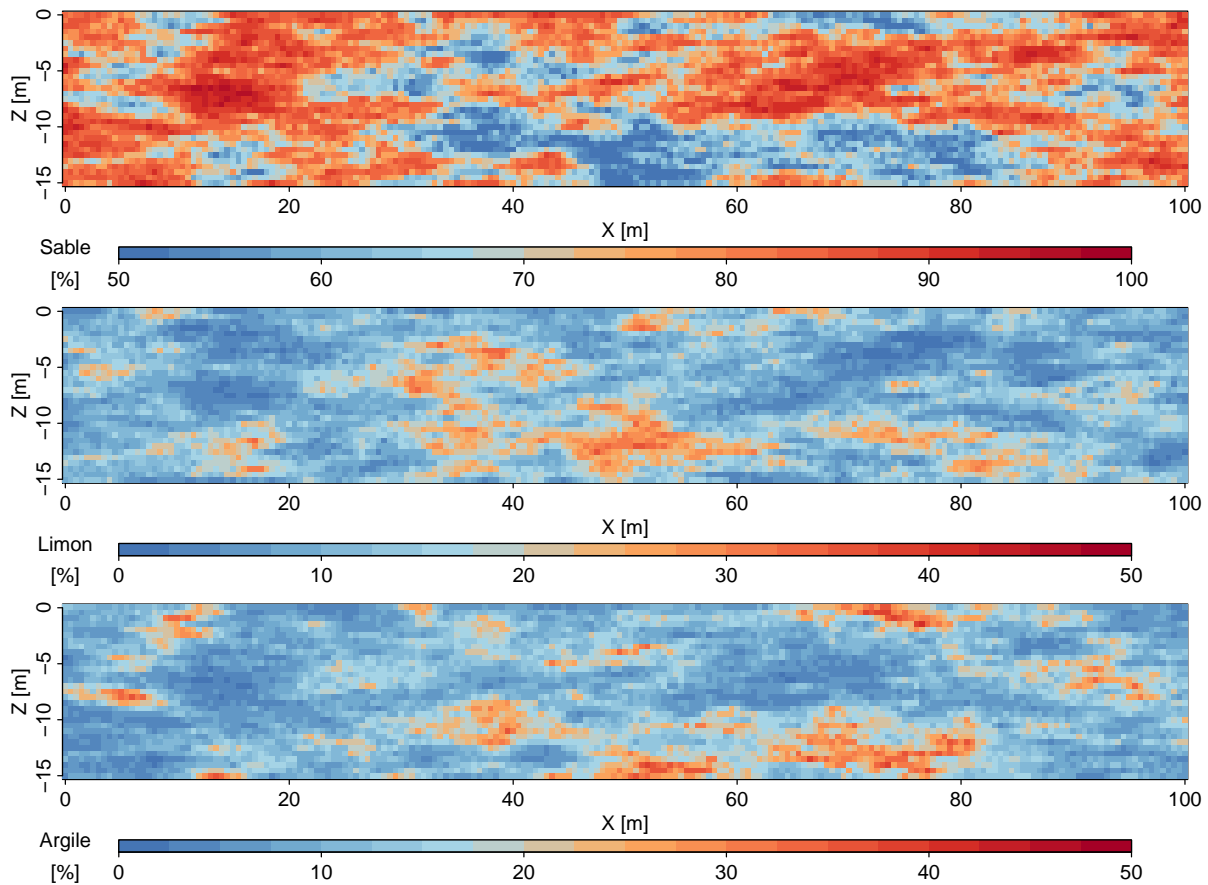


FIGURE 3.12 – Champs de proportions de sable, de limon et d'argile.

Simulations de champs de textures

L'objectif est de simuler un triplet de champs aléatoires Y_1, Y_2, Y_3 tels que pour tout $x \in \mathcal{D}$, $Y_1(x) + Y_2(x) + Y_3(x) = 1$. Pour cela, on génère d'abord un champ gaussien P_{sable} représentant la proportion de sable. Pour répartir le complémentaire de P_{sable} entre le limon et l'argile, un second champ aléatoire \tilde{P} de mêmes portées que P_{sable} est généré. On pose ensuite :

$$P_{limon} = (1 - P_{sable})\tilde{P} \quad (3.10)$$

$$P_{argile} = (1 - P_{sable})(1 - \tilde{P}) \quad (3.11)$$

On s'assure ainsi que $P_{sable}(x) + P_{limon}(x) + P_{argile}(x) = 1$ pour tout $x \in \mathcal{D}$. Les gaussiennes sont éventuellement tronquées pour conserver des proportions positives. Cette méthode de composition de gaussiennes est assez simpliste, mais le champ qui apporte le plus de variabilité et qui présente la plus grande valeur moyenne est le champ de proportion de sable, simulé sans composition. Des méthodes plus complexes existent, notamment celles basées sur les log-ratios [Aitchison, 1982 ; Tolosana-Delgado *et al.*, 2019].

Paramètres du modèle de Mualem-van Genuchten

Les champs de texture ainsi générés ne peuvent pas être directement utilisés par le code MELODIE. Ceux-ci doivent au préalable être convertis en champs de paramètres de Mualem-van Genuchten. Pour cela, la fonction de pédo-transfert Rosetta3 est utilisée, en considérant les paramètres moyens de la distribution obtenue, c'est-à-dire qu'aucune variabilité n'est ajoutée lors du passage des textures aux paramètres. On obtient alors les champs de paramètres K_s, α, n, θ_r et θ_s donnés sur la Figure 3.13.

Simulations avec MELODIE

Les autres paramètres d'entrée nécessaires pour simuler la migration de tritium avec le code MELODIE sont identiques à ceux utilisés pour l'analyse de sensibilité de la Section 3.2. On obtient finalement le panache de tritium de la Figure 3.14, qui constitue le panache de référence.

Pour tester la robustesse des méthodes d'estimation, et s'assurer de la reproductibilité des résultats, ce cas synthétique est reproduit sur deux autres références. La construction de ces deux références, ainsi que tous les résultats s'y rapportant, sont donnés en annexe (Annexe D pour la construction des références).

3.3.2 Échantillonnage et génération d'un ensemble de champs de paramètres de Mualem-van Genuchten**Échantillonnage**

Le panache de tritium simulé précédemment ainsi que les champs de textures à partir desquels il a été généré ne sont supposés connus que partiellement à partir d'ici. Deux types d'informations sont extraits de cette référence :

- la texture est échantillonnée sur huit sondages traversant la zone non saturée (7 m de profondeur). Ces sondages donnent les proportions précises de sable, de limon et d'argile avec une résolution verticale de 0,5 m (Figure 3.15). Le reste des champs (hors sondages), ainsi que leurs structures spatiales, sont inconnus ;
- le panache de tritium est également échantillonné. Les détails sur cet échantillonnage sont donnés au Chapitre 4.

Dans un cas réel, la texture est échantillonnée dans des forages, et les activités en tritium sont échantillonnées dans des piézomètres.

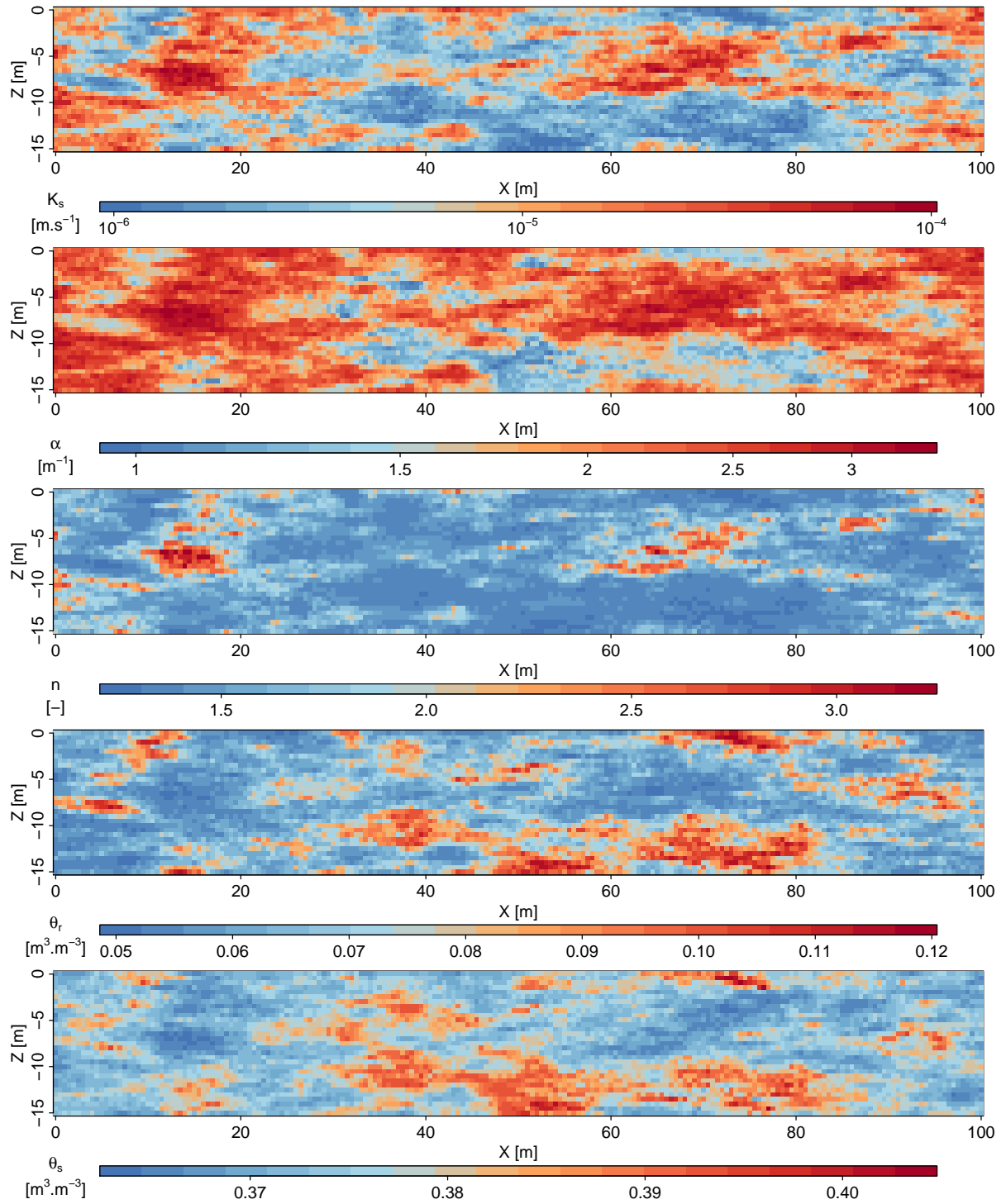


FIGURE 3.13 – Champs de paramètres de Mualem-van Genuchten. De haut en bas : K_s (en échelle logarithmique), α (en échelle logarithmique), n , θ_r et θ_s .

Génération d'un ensemble de champs

Le modèle utilisé pour générer les champs de proportions de sable, de limon et d'argile (modèle de variogramme, portées, etc.) est supposé inconnu. L'échantillonnage des textures permet de générer des champs de paramètres de Mualem-van Genuchten. Deux approches sont mises en œuvre. Elles sont résumées sur la Figure 3.16.

Approche 1 : À partir de l'échantillonnage des textures, supposé sans erreurs de mesures, un variogramme expérimental est construit (pour la proportion de sable et son complémentaire). Ensuite, un modèle de variogramme stationnaire est ajusté, avec des incertitudes (Complément 3.4). Des simulations

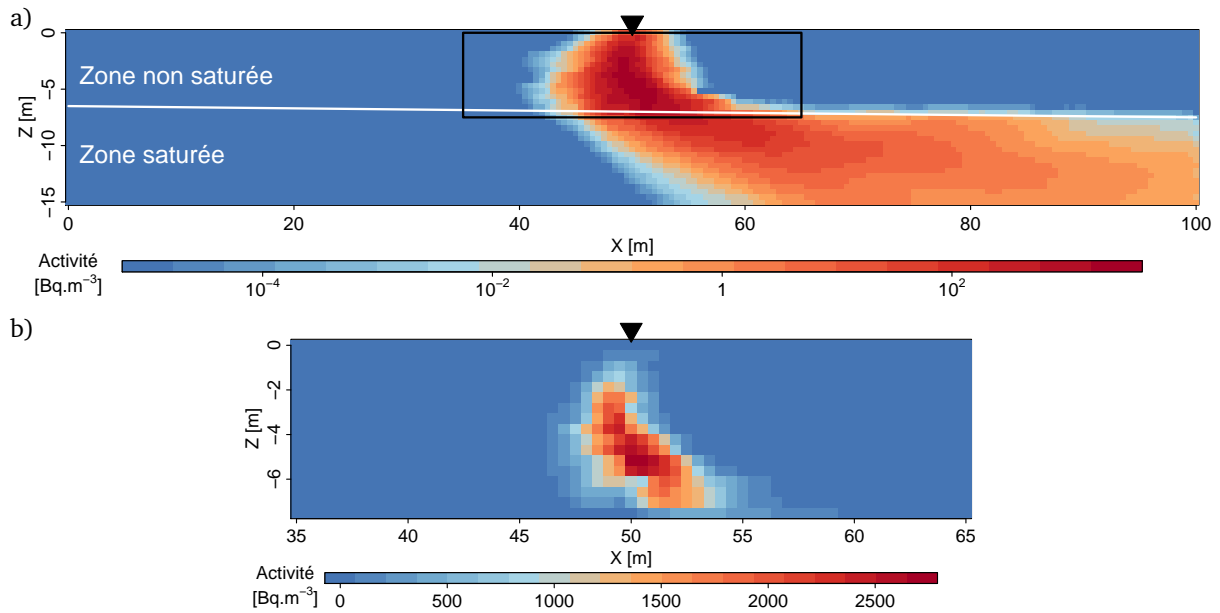


FIGURE 3.14 – Panache de tritium simulé avec MELODIE, cinq ans après le rejet. a) Panache sur le domaine complet, en échelle logarithmique ; b) zoom dans la zone non saturée (rectangle noir sur a)).

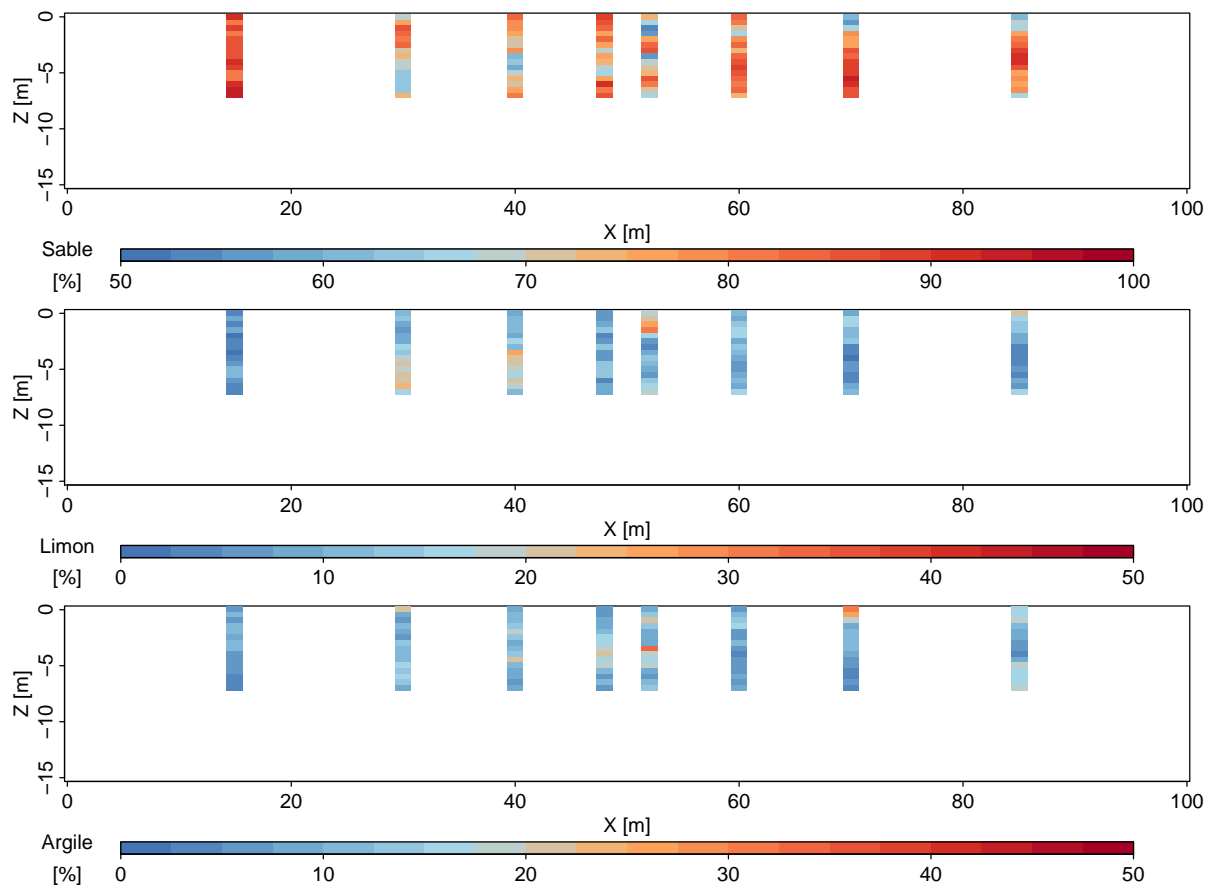


FIGURE 3.15 – Échantillonnage des champs de proportions de sable, de limon et d'argile en huit sondages traversant la zone non saturée. Les sondages sont élargis pour une meilleure lisibilité.

conditionnelles de champs de proportion de sable et de son complémentaire sont générées, en tirant un modèle de variogramme dans la gamme d'incertitudes. Les trois champs de proportions obtenus sont convertis en champs de paramètres de Mualem-van Genuchten par la fonction de pédotransfert Rosetta3. Mille quintuplets de champs de paramètres de Mualem-van Genuchten sont ainsi générés.

Approche 2 : L'échantillonnage des textures est directement converti en paramètres de Mualem-van Genuchten par la fonction de pédotransfert Rosetta3. Ensuite, pour chaque paramètre de Mualem-van Genuchten, des simulations conditionnées par les données de texture converties sont générées, en imposant une distribution lognormale (pour K_s , α et n) ou normale (pour θ_r et θ_s) de moyennes et variances égales aux moyennes et variances empiriques des données de chaque paramètre. Les paramètres des variogrammes sont tirés dans la gamme d'incertitude (Complément 3.4). Mille quintuplets de champs de paramètres de Mualem-van Genuchten sont ainsi générés.

Prise en compte d'incertitudes dans la modélisation du variogramme

Pour prendre en compte les incertitudes dans l'ajustement d'un modèle de variogramme sur les données de texture ou de paramètres de Mualem-van Genuchten, les modèles de variogramme sont randomisés :

- le palier est tiré dans une distribution normale centrée sur le palier du modèle ajusté sur le variogramme expérimental C_0 avec une incertitude de 20% ($3\sigma = 0,2C_0$) ;
- la portée verticale est tirée dans une distribution normale centrée sur la portée verticale du modèle ajusté sur le variogramme expérimental avec une incertitude de 20% ;
- la portée horizontale est tirée dans une distribution triangulaire de mode égal à la portée horizontale du modèle ajusté sur le variogramme expérimental et de valeurs minimales et maximales respectivement égales à 0,5 et 10 fois cette portée expérimentale. Cette distribution conduit volontairement à une incertitude plus forte sur la portée horizontale que sur la portée verticale, puisque l'inférence de la portée horizontale est moins précise que celle de la portée verticale au vu du schéma d'échantillonnage ;
- le comportement du variogramme à l'origine est choisi aléatoirement entre trois possibilités : un modèle cubique sans effet de pépité, un modèle exponentiel sans effet de pépité ou un modèle exponentiel avec effet de pépité (entre 0% et 5% du palier total).

Complément 3.4

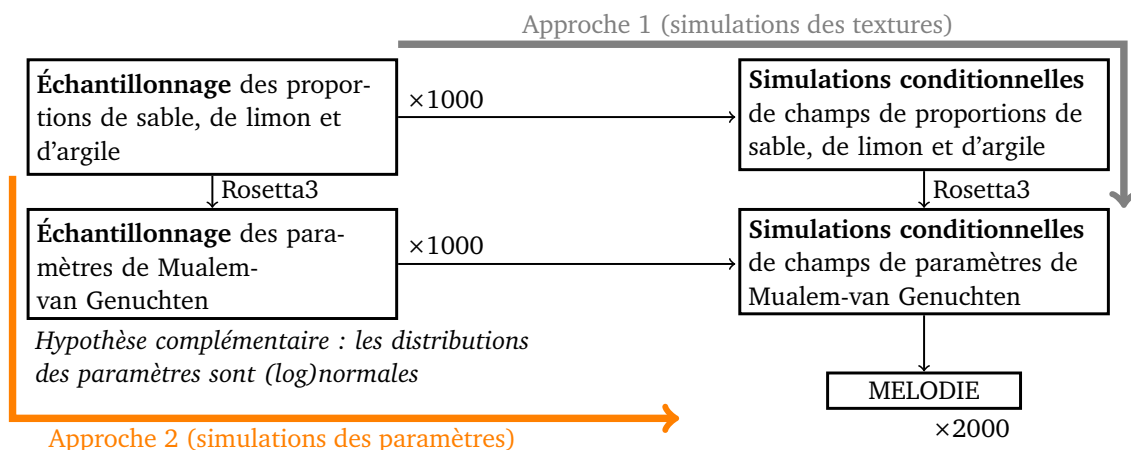


FIGURE 3.16 – Schéma récapitulatif des deux approches de simulations conditionnelles de champs de paramètres de Mualem-van Genuchten.

3.3.3 Simulation des panaches de tritium

À partir des 2000 quintuplets de champs de paramètres de Mualem-van Genuchten générés, 2000 panaches de tritium sont simulés avec le code MELODIE. Tous les autres paramètres des simulations sont

identiques à ceux de la référence. Les 2000 panaches obtenus seront utilisés au Chapitre 4.

Les deux approches pour générer des champs de paramètres de Mualem-van Genuchten sont comparées par l'analyse des panaches résultants, en terme de surface et de localisation. Les résultats sont présentés sur la Figure 3.17 : les panaches générés à partir de l'approche 2 sont en moyenne plus étendus que ceux générés à partir de l'approche 1 (259 mailles en moyenne pour l'approche 2 contre 213 pour l'approche 1), avec plus de variabilité sur cette surface (écart-type de 51 pour l'approche 2 contre 25 pour l'approche 1). Ils sont également plus éloignés les uns des autres.

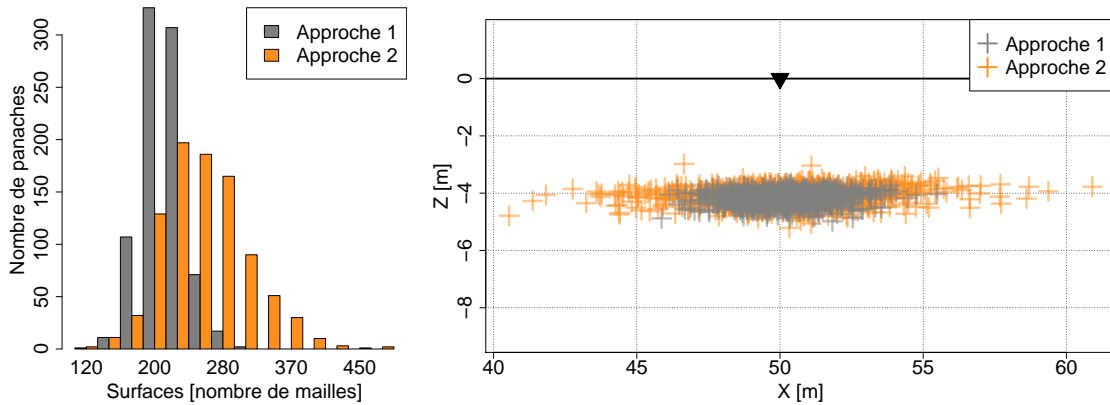


FIGURE 3.17 – Comparaison des panaches selon l'approche utilisée pour générer les champs de paramètres de Mualem-van Genuchten.

3.4 Synthèse du chapitre

Un premier cas synthétique a été construit pour étudier l'influence de la variabilité spatiale des paramètres de Mualem-van Genuchten sur la migration d'un panache de tritium. Ce cas synthétique est construit à partir de simulations de champs de paramètres K_s , α et n .

Des analyses de sensibilité ont été menées sur les caractéristiques moyennes de séries de simulations de 100 panaches. Celles-ci montrent que lorsque les variances des champs de paramètres augmentent, les panaches sont en moyenne plus étendus, plus allongés et plus distants les uns des autres. L'introduction de champs de paramètres corrélés ne change pas les caractéristiques moyennes des panaches simulés. Cette étude permet de souligner l'incertitude obtenue au sein d'un ensemble de panaches de tritium simulés lorsque seule la variabilité spatiale des paramètres de Mualem-van Genuchten n'est pas parfaitement caractérisée.

La variabilité obtenue au sein des panaches de tritium simulés apparaît suffisante pour construire un cas synthétique sur lequel mettre en œuvre les méthodes d'estimation décrites au Chapitre 2. Pour se rapprocher des conditions opérationnelles, le point de départ devient un triplet de champs de proportions de sable, de limon et d'argile. Une référence ainsi qu'un jeu de 2000 panaches de tritium sont générés à partir de ce cas synthétique. Le jeu de 2000 panaches est construit suivant deux approches : la première consiste à générer des champs de texture et à les convertir en paramètres de Mualem-van Genuchten en utilisant la fonction de pédo-transfert Rosetta3 et la seconde consiste à générer directement des champs de paramètres de Mualem-van Genuchten à partir d'observations de texture converties. Cette seconde approche apporte une plus grande variabilité dans les champs de paramètres obtenus (car elle permet de sortir des valeurs moyennes proposées par Rosetta3) et par conséquent dans les panaches simulés (Figure 3.17).

La variabilité obtenue au sein du jeu de 2000 panaches est comparable à celle du cas sur lequel les analyses de sensibilité ont été réalisées. Ainsi, cette variabilité apparaît réaliste, et les panaches simulés reflètent bien l'incertitude du modélisateur sur les paramètres du modèle de Mualem-van Genuchten.

Deux références supplémentaires sont également construites. Les champs de texture à l'origine de ces références supplémentaires coïncident avec ceux de la référence principale aux huit sondages d'échantillonnage des textures : cela permet d'utiliser le même jeu de 2000 simulations de panaches de tritium pour réaliser les estimations.

Chapitre 4

Application des méthodes combinant la géostatistique et les simulations numériques

Ce chapitre présente l'application des méthodes détaillées au Chapitre 2, sur le cas synthétique construit à la fin du Chapitre 3. Une partie de ces résultats ont été publiés au cours de la thèse [Pannecoucke *et al.*, 2020a].

4.1 Scénarios d'échantillonnage des activités en tritium

Le panache de référence généré au Chapitre 3 (Figure 3.14) est échantillonné pour obtenir des observations d'activité volumique en tritium, supposées sans erreur de mesure, avec une résolution verticale de 0,5 m. Trois scénarios d'échantillonnage sont distingués (Figure 4.1), avec des sondages répartis sur une zone de 20 m de large.

Scénario S1 : sept sondages traversant l'intégralité de la zone non saturée (7 m de profondeur).

Scénario S2 : quatre sondages traversant l'intégralité de la zone non saturée.

Scénario S3 : sept sondages traversant seulement une partie de la zone non saturée (4 m de profondeur), aux mêmes emplacements que ceux du scénario S1.

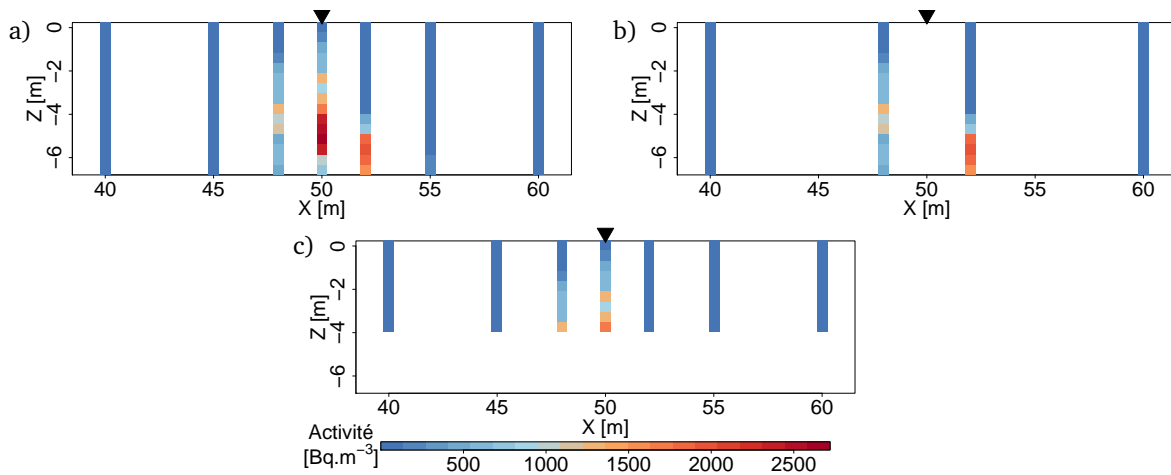


FIGURE 4.1 – Observations d'activité volumique en tritium cinq ans après le déversement. a) Scénario S1 ; b) scénario S2 ; c) scénario S3. Le triangle noir indique la localisation de la source de tritium.

4.2 Exploitation des simulations de panaches de tritium

Sur les 2000 simulations effectuées, environ 1700 sont exploitables, les autres n'ayant pas convergé.

4.2.1 Statistiques de l'ensemble

Quelques exemples de panaches simulés sont donnés sur la Figure 4.2. Ces exemples soulignent la variabilité de l'ensemble (en terme d'intensité, de forme, de localisation, *etc.*). Par ailleurs, plusieurs caractéristiques globales sont extraites de l'ensemble de panaches simulés, comme la moyenne empirique (Figure 4.3a), la médiane (Figure 4.3b) et l'écart-type (Figure 4.3c) (statistiques calculées pour chaque point du domaine). Un panache central est également calculé (Figure 4.3d) : il s'agit du panache minimisant la distance euclidienne avec tous les autres panaches dans l'espace ($x_G, z_G, b/a, \theta$). Dans cet espace,

un panache est décrit par les coordonnées de son centre de gravité (x_G, z_G), le rapport entre le demi-petit axe et le demi-grand axe (b/a) de son ellipse équivalente et l'angle θ entre la direction horizontale et la direction du grand axe de son ellipse équivalente. Toutes les grandeurs sont normalisées entre 0 et 1.

Le panache central est plus représentatif des panaches simulés que les panaches moyens ou médians. En effet, celui-ci présente des activités plus élevées et une forme moins régulière, ce qui est le cas de beaucoup des panaches de l'ensemble.

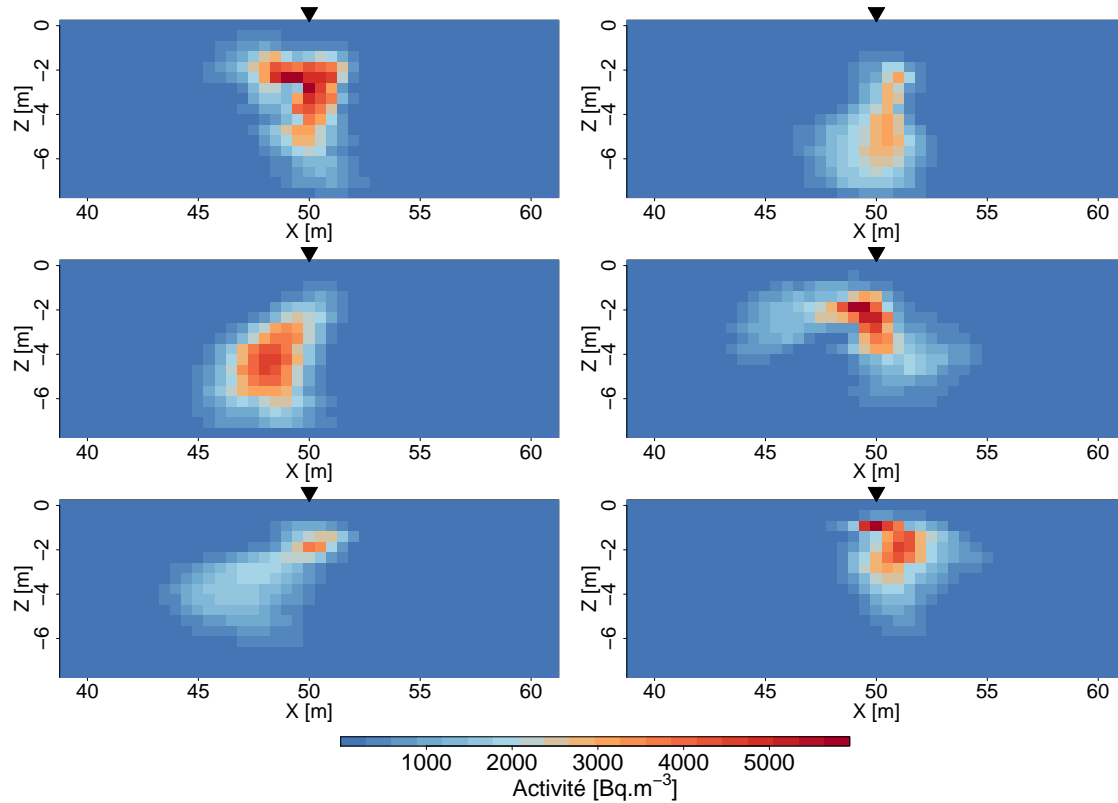


FIGURE 4.2 – Exemple de six panaches simulés tirés aléatoirement dans l'ensemble.

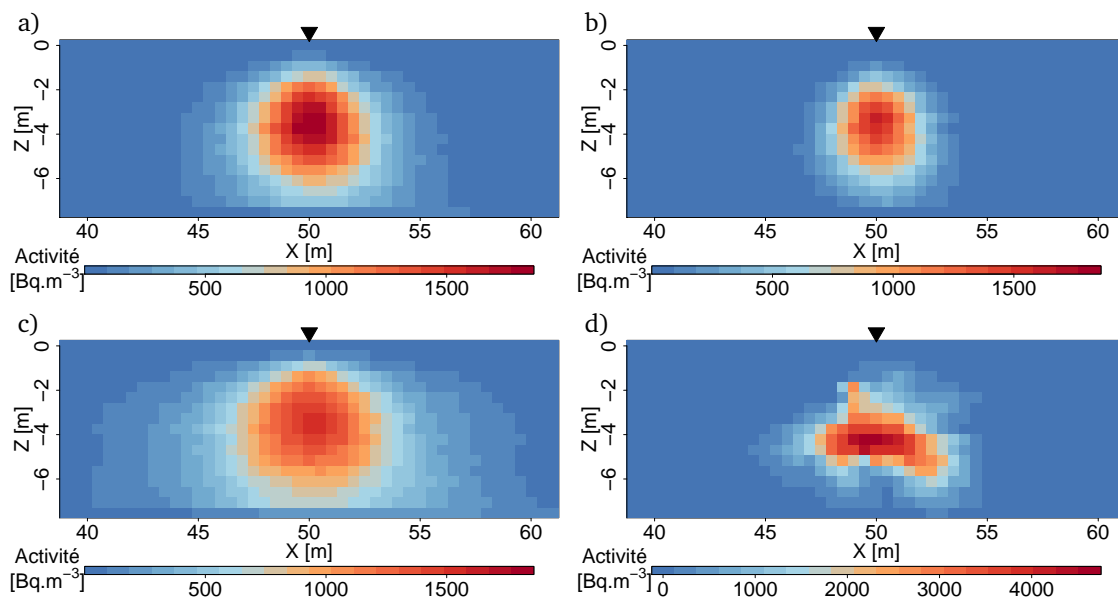


FIGURE 4.3 – Caractéristiques de l'ensemble de panaches : a) moyenne ; b) médiane ; c) écart-type ; d) panache central. L'échelle de couleur est la même pour les cartes a) à c).

4.2.2 Variogrammes numériques

Des variogrammes numériques sont calculés à partir de l'ensemble de panaches. Ceux-ci sont fortement non stationnaires (Figure 4.4). L'allure des cartes est très dépendante du point par rapport auquel sont tracées les valeurs de variogramme : sur la Figure 4.4a, le point de référence est hors de la zone du panache moyen, et la carte obtenue reproduit la localisation du panache moyen. À l'inverse, sur la Figure 4.4b, le point de référence est dans la zone du panache moyen. La carte obtenue reproduit également la localisation du panache, mais avec des valeurs faibles de variogramme au milieu du panache, et des valeurs fortes à l'extérieur. Enfin, sur les Figures 4.4c et 4.4d, les points de référence sont en bordure du panache moyen (en général en bordure des panaches simulés), ce qui se traduit sur les cartes par un découpage en deux zones distinctes, l'une de faibles valeurs de variogrammes autour du point de référence, l'autre de fortes valeurs dans la zone centrale du panache.

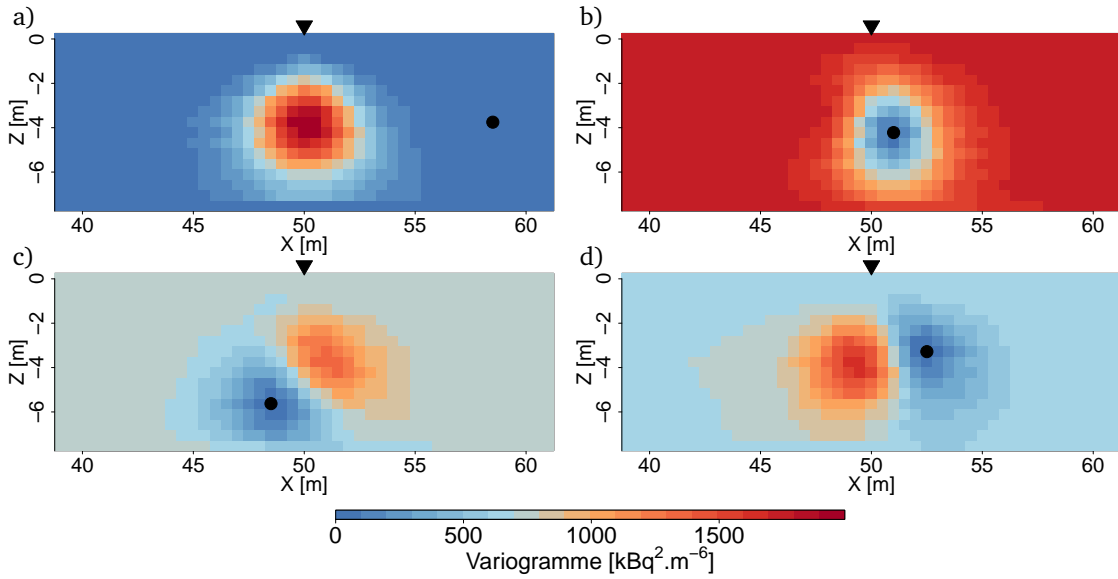


FIGURE 4.4 – Cartes de variogrammes $\gamma(x, x')$ pour quatre points x fixés, représentés par le point noir sur chaque carte, et x' parcourant le reste du domaine.

4.3 Estimations de l'activité en tritium

Cette section compare les estimations obtenues par krigeage avec variogrammes numériques, par krigeage avec dérive externe et par krigeage ordinaire avec un modèle de variogramme ajusté sur les données. Quelques variantes des méthodes d'estimation sont également présentées.

4.3.1 Indicateurs de la qualité de l'estimation

Plusieurs indicateurs sont calculés sur l'ensemble du domaine étudié, sans validation croisée puisque la référence est connue. À noter que les résultats obtenus par validation croisée aboutissent aux mêmes conclusions.

1. Les cartes d'estimation, d'erreur d'estimation (différence point par point entre la référence et l'estimation) et d'écart-type théorique d'erreur d'estimation sont tracées. Pour le krigeage ordinaire et le krigeage avec dérive externe, les cartes d'écart-type d'erreur de krigeage sont corrigées par un effet proportionnel [Donati et de Fouquet, 2018] qui permet de tenir compte des zones de fortes ou faibles valeurs d'activité. Une relation linéaire est recherchée entre la moyenne des activités et l'écart-type des activités sur chaque sondage. L'écart-type d'erreur de krigeage (σ_K) est ensuite corrigé :

$$\sigma_K^{\text{corr}}(x_0) = A\sigma_K(x_0)[a + bZ^*(x_0)] \quad (4.1)$$

où a et b sont les coefficients de la régression linéaire et A un coefficient de normalisation. Cette étape de modélisation supplémentaire n'est pas nécessaire pour le krigeage avec variogrammes numériques, car ces variogrammes prennent directement en compte la variabilité locale des activités dans la zone contaminée.

2. Les nuages de corrélation entre les valeurs d'activité de la référence et les activités estimées sont tracés, pour caractériser précisément les erreurs de chaque méthode. Des espérances conditionnelles empiriques sont également calculées (moyennes par classe des valeurs observées et estimées), afin de quantifier le biais conditionnel de l'estimateur.
3. Les erreurs sont quantifiées par l'erreur absolue moyenne (EAM), la racine de l'erreur quadratique moyenne (REQM) et l'erreur relative moyenne (ERM) :

$$\text{EAM} = \frac{1}{N_D} \sum_{i=1}^{N_D} |Z^{\text{ref}}(x_i) - Z^*(x_i)| \quad (4.2)$$

$$\text{REQM} = \sqrt{\frac{1}{N_D} \sum_{i=1}^{N_D} [Z^{\text{ref}}(x_i) - Z^*(x_i)]^2} \quad (4.3)$$

$$\text{ERM} = \frac{1}{N_D} \sum_{i=1}^{N_D} \frac{Z^{\text{ref}}(x_i) - Z^*(x_i)}{\max(1, Z^{\text{ref}}(x_i))} \quad (4.4)$$

où N_D est le nombre de mailles du domaine (sans les observations), $Z^{\text{ref}}(x_i)$ est l'activité de la référence au point x_i et $Z^*(x_i)$ est l'activité estimée au point x_i . Si la valeur de $Z^{\text{ref}}(x_i)$ est inférieure à 1 Bq.m^{-3} , le dénominateur de l'erreur relative est fixé à 1 Bq.m^{-3} , pour éviter des valeurs d'erreur relative très élevées lorsque $Z^{\text{ref}}(x_i)$ est proche de 0.

4. Les courbes de sélectivité permettent d'évaluer la capacité de l'estimateur à reproduire la distribution des valeurs d'activité [Chilès et Delfiner, 2012]. Ces courbes sont paramétrées par le seuil de contamination z . Pour chaque valeur de z , deux quantités sont calculées :
 - sur l'axe des abscisses, le pourcentage de mailles telles que $Z^*(x_i) \geq z$ ("mineral"), défini par

$$100 \times \frac{1}{N_D} \sum_{i=1}^{N_D} \mathbb{1}_{Z^*(x_i) \geq z} \quad (4.5)$$

- sur l'axe des ordonnées, le pourcentage de l'activité totale contenue par les mailles contaminées ("métal"), défini par

$$100 \times \frac{\sum_{i=1}^{N_D} Z^*(x_i) \mathbb{1}_{Z^*(x_i) \geq z}}{\sum_{i=1}^{N_D} Z^*(x_i)} \quad (4.6)$$

5. Les proportions de faux-négatifs et de faux-positifs sont également calculées, pour différents seuils de contamination z . Ces indicateurs permettent de quantifier les risques de laisser de la contamination en place (faux-négatifs) ou de surestimer l'étendue de la contamination et donc des coûts associés (faux-positifs). La proportion de faux-négatifs est définie comme le nombre de mailles telles que $Z^*(x_i) < z$ et $Z^{\text{ref}}(x_i) \geq z$ divisé par la surface contaminée sur la référence (le nombre de mailles telles que $Z^{\text{ref}}(x_i) \geq z$). Au contraire, la proportion de faux-positifs est définie comme le nombre de mailles telles que $Z^*(x_i) \geq z$ et $Z^{\text{ref}}(x_i) < z$ divisé par la surface contaminée sur la référence.

4.3.2 Résultats avec toutes les simulations

Dans cette section, toutes les simulations sont utilisées pour calculer les variogrammes numériques et la fonction de dérive (moyenne empirique du jeu de simulations), sans distinction selon l'approche de génération des champs de paramètres (Chapitre 3, Section 3.3.3). Le voisinage considéré est un voisinage unique. Les résultats détaillés pour les références supplémentaires (construites dans l'Annexe D) sont présentés dans l'Annexe E.

Scénario S1 (sept sondages traversant toute la zone non saturée)

Les cartes d'estimation et d'erreur d'estimation sont semblables pour les trois méthodes (Figure 4.5). Les écart-types d'erreur sont plus faibles pour le krigeage avec variogrammes numériques que pour les deux autres méthodes, même lorsque l'effet proportionnel est pris en compte. De manière quantitative, les erreurs moyennes les plus faibles sont obtenues par krigeage avec variogrammes numériques, puis par krigeage avec dérive externe et enfin par krigeage ordinaire (Tableau 4.1). Cela est confirmé par le nuage de la Figure 4.6a, quelles que soient les valeurs de la référence. Les espérances conditionnelles empiriques montrent que les estimations sont sans biais conditionnel (Figure 4.6a). Les courbes de sélectivité montrent que les trois méthodes reproduisent correctement la distribution de l'activité (Figure 4.6b). La proportion de faux-négatifs est légèrement plus élevée pour le krigeage avec variogrammes numériques que pour les autres méthodes, pour des seuils de contamination inférieurs à 500 Bq.m⁻³ (Figure 4.6c). Pour des seuils plus élevés, le krigeage avec variogrammes numériques donne des proportions de faux-négatifs plus faibles, car il permet de mieux reproduire les fortes valeurs d'activité. La proportion de faux-positifs est plus faible pour le krigeage avec variogrammes numériques que pour les autres méthodes, sauf pour des seuils de contamination supérieurs à 1200 Bq.m⁻³ (Figure 4.6d).

Scénario S2 (quatre sondages traversant toute la zone non saturée)

Les cartes d'estimation et d'erreur d'estimation sont différentes pour les trois méthodes (Figure 4.7) : la forme du panache estimé par krigeage ordinaire suggère un panache avec deux zones de forte activité, ce qui n'est pas le cas en réalité. Les formes des panaches estimés par krigeage avec dérive externe et par krigeage avec variogrammes numériques sont cohérentes, mais les activités estimées par krigeage avec variogrammes numériques sont plus fortes que par krigeage avec dérive externe. Les écart-types d'erreur sont plus faibles pour le krigeage avec variogrammes numériques que pour les deux autres méthodes. Les erreurs moyennes les plus faibles sont obtenues par krigeage avec variogrammes numériques, puis par krigeage avec dérive externe et enfin par krigeage ordinaire (Tableau 4.2). En particulier, le nuage de la Figure 4.8a et les espérances conditionnelles soulignent la tendance du krigeage ordinaire et du krigeage avec dérive externe à sous-estimer les fortes valeurs d'activité, alors que le krigeage avec variogrammes numériques a tendance à les surestimer légèrement. Ainsi, même si la zone centrale du panache, présentant de fortes valeurs d'activité, n'est pas échantillonnée, le krigeage avec variogrammes numériques reproduit ces fortes valeurs. Les courbes de sélectivité montrent que le krigeage ordinaire ne reproduit pas aussi bien la distribution d'activité que les deux autres méthodes (Figure 4.8b). La proportion de faux-négatifs est légèrement plus élevée pour le krigeage avec variogrammes numériques que pour le krigeage avec dérive externe, pour des seuils de contamination inférieurs à 500 Bq.m⁻³ (Figure 4.8c). Au delà, cette proportion est bien plus faible pour le krigeage avec variogrammes numériques que pour les deux autres méthodes. La proportion de faux-positifs est plus faible pour le krigeage avec variogrammes numériques que pour les autres méthodes, sauf pour des seuils de contamination supérieurs à 1000 Bq.m⁻³ (Figure 4.8d).

Scénario S3 (sept sondages peu profonds)

Les cartes d'estimation et d'erreur d'estimation sont semblables pour les trois méthodes (Figure 4.9) : aucune des méthodes ne reproduit correctement la forme et l'intensité du panache de référence. Les erreurs moyennes les plus faibles sont obtenues par krigeage avec variogrammes numériques, puis par krigeage avec dérive externe et enfin par krigeage ordinaire (Tableau 4.3). Le nuage de la Figure 4.10a ainsi que les espérances conditionnelles soulignent que toutes les méthodes sous-estiment les activités, surtout au-delà de 1000 Bq.m⁻³. Les courbes de sélectivité montrent que le krigeage ordinaire reproduit moins bien la distribution d'activité que les deux autres méthodes (Figure 4.10b). La proportion de faux-négatifs est plus élevée pour le krigeage avec variogrammes numériques pour des seuils de contamination inférieurs à 500 Bq.m⁻³ (Figure 4.10c). La proportion de faux-positifs est nettement plus faible pour le krigeage avec variogrammes numériques et le krigeage avec dérive externe que pour le krigeage ordinaire pour des seuils de contamination inférieurs à 200 Bq.m⁻³, similaire ensuite (Figure 4.10d). Pour ce scénario, les fortes valeurs d'activité se situent sous la zone échantillonnée, et aucune méthode ne parvient à reproduire ces fortes valeurs.

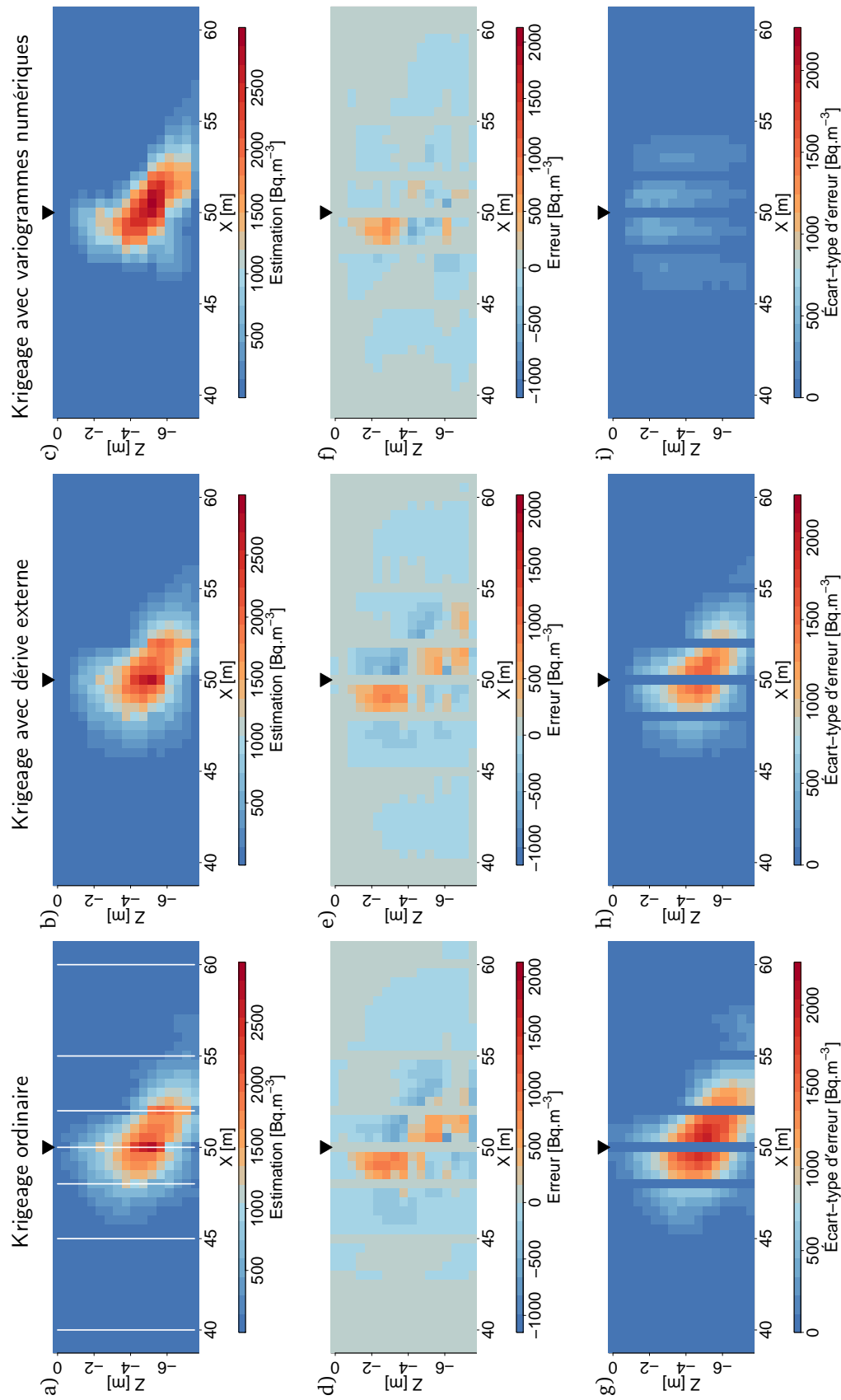


FIGURE 4.5 – Cartes d'estimation (a, b et c), d'erreur d'estimation (d, e et f) et d'écart-type d'erreur d'estimation (g, h et i) obtenues avec les trois méthodes pour le scénario S1 (rappelé sur la première carte). La même échelle de couleur est utilisée sur chaque ligne. Les erreurs négatives correspondent à des sous-estimations et les erreurs positives à des surestimations.

TABLEAU 4.1 – Erreurs moyennes pour les trois méthodes pour le scénario S1.

	KO	KDE	KVN
EAM [Bq.m^{-3}]	83	76	40
REQM [Bq.m^{-3}]	176	164	99
ERM [-]	-7.4	-5.1	-3

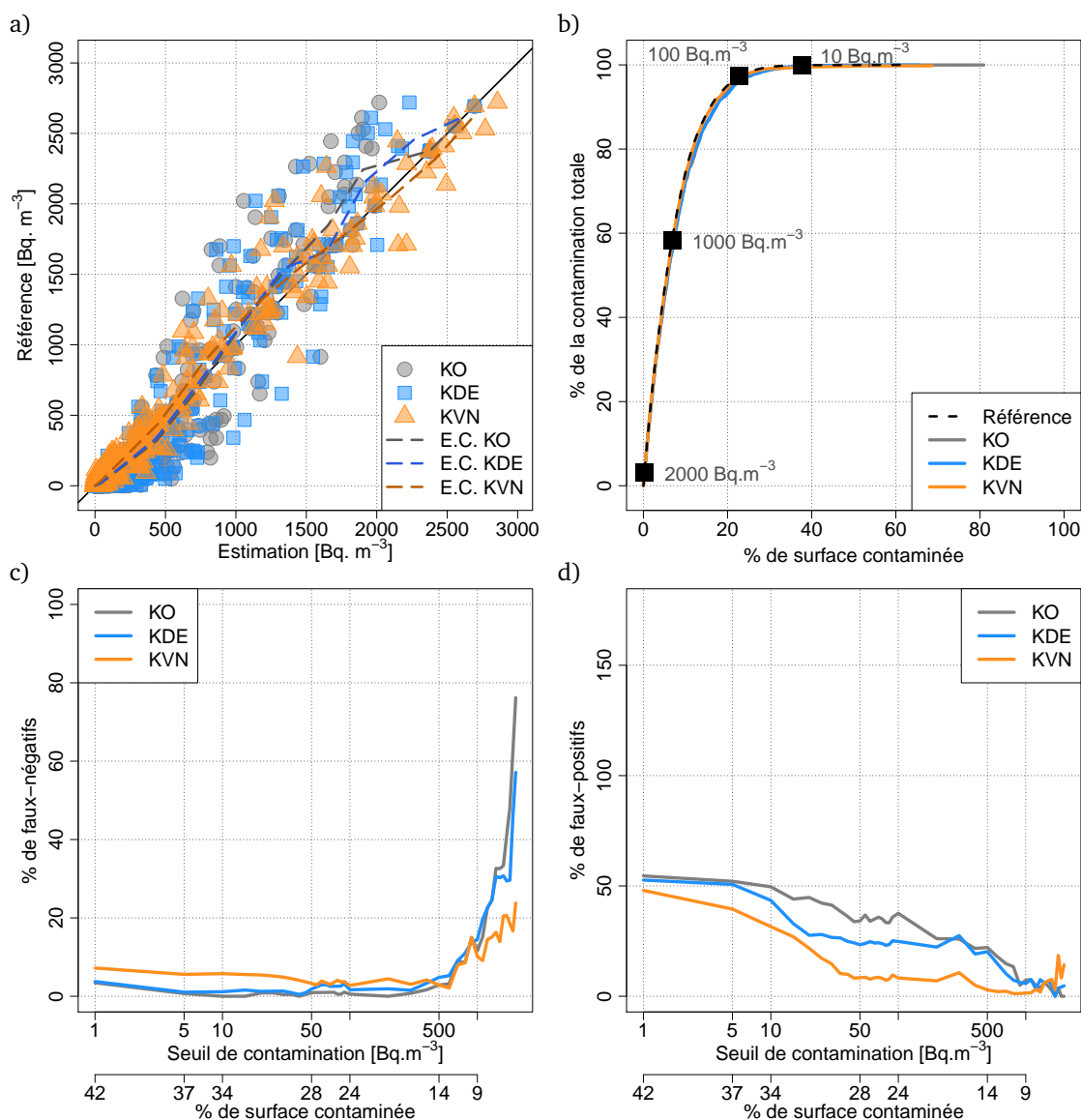


FIGURE 4.6 – Indicateurs de performance pour le scénario S1 : a) comparaison entre les valeurs d'activité de la référence et les activités estimées ; b) courbes de sélectivité ; c) proportion de faux-négatifs par rapport à la surface réellement contaminée ; d) proportion de faux-positifs par rapport à la surface réellement contaminée. Sur la figure a), les traits en pointillés donnent l'espérance conditionnelle empirique pour chaque méthode. Sur la figure b), les carrés noirs correspondent à quelques points de la courbe de sélectivité de la référence. Sur les figures c) et d), les axes des abscisses secondaires donnent la surface contaminée (en % du domaine de modélisation) de la référence en fonction du seuil de contamination. Les échelles des axes des ordonnées sont par ailleurs choisies pour être identiques sur les Figures 4.6, 4.8 et 4.10.

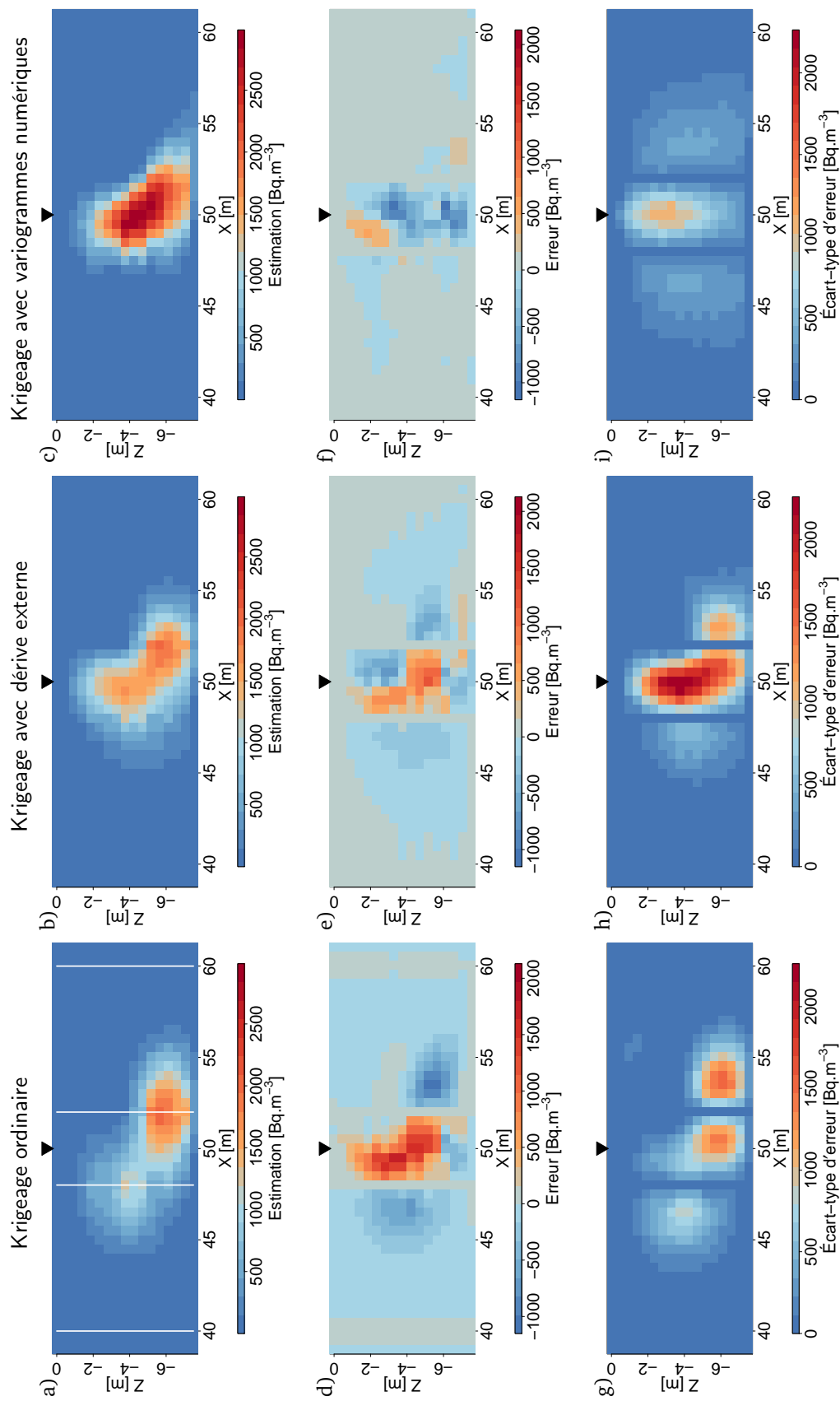


FIGURE 4.7 – Cartes d'estimation, d'erreur d'estimation et d'écart-type d'erreur obtenues avec les trois méthodes pour le scénario S2. Les échelles de couleur sont les mêmes que sur la Figure 4.5.

TABLEAU 4.2 – Erreurs moyennes pour les trois méthodes pour le scénario S2.

	KO	KDE	KVN
EAM [Bq.m^{-3}]	169	98	58
REQM [Bq.m^{-3}]	337	203	162
ERM [-]	-26.1	-9.4	-0.5

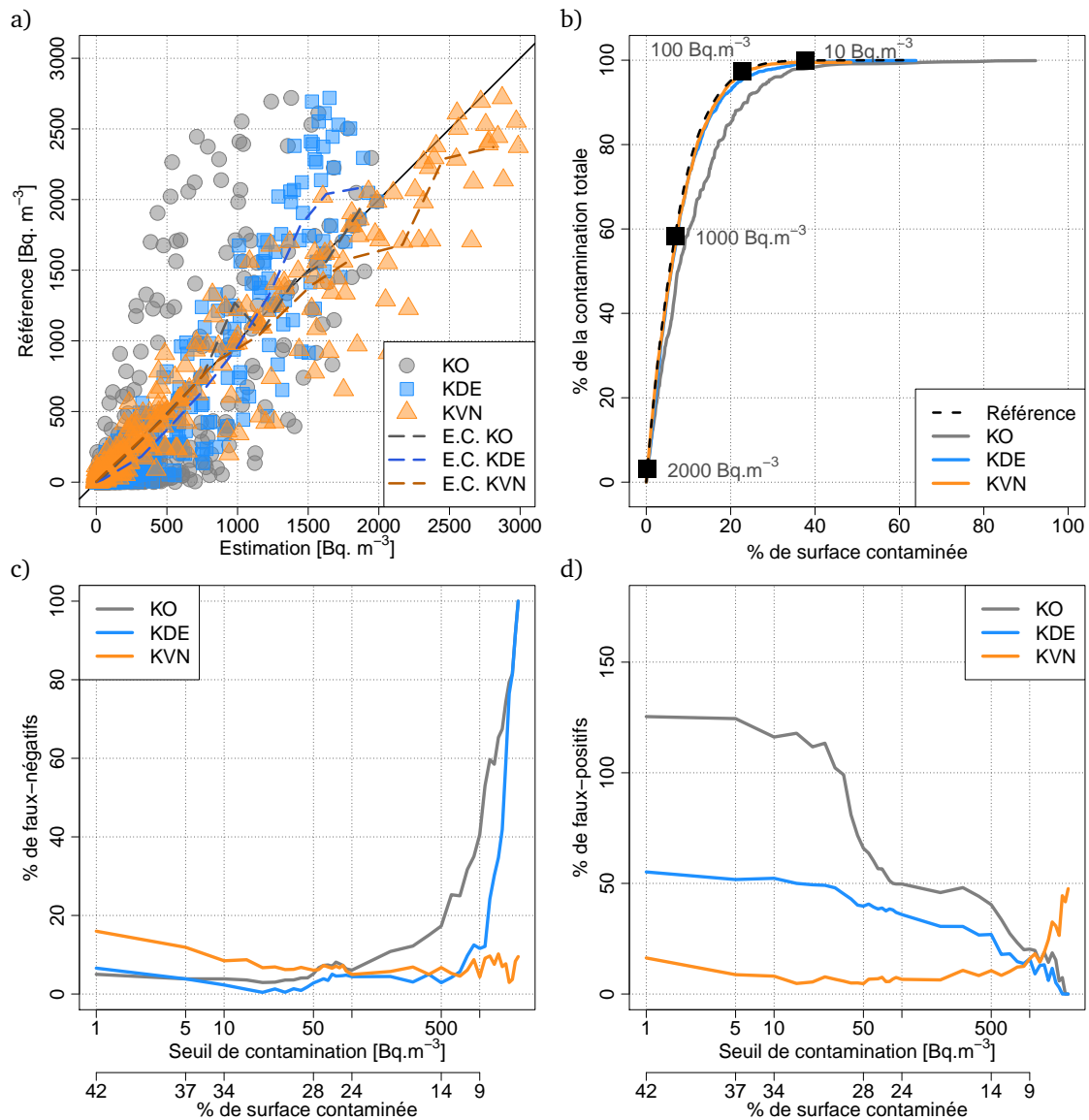


FIGURE 4.8 – Indicateurs de performance pour le scénario S2 : a) comparaison entre les valeurs d'activité de la référence et les activités estimées ; b) courbes de sélectivité ; c) proportion de faux-négatifs par rapport à la surface réellement contaminée ; d) proportion de faux-positifs par rapport à la surface réellement contaminée.

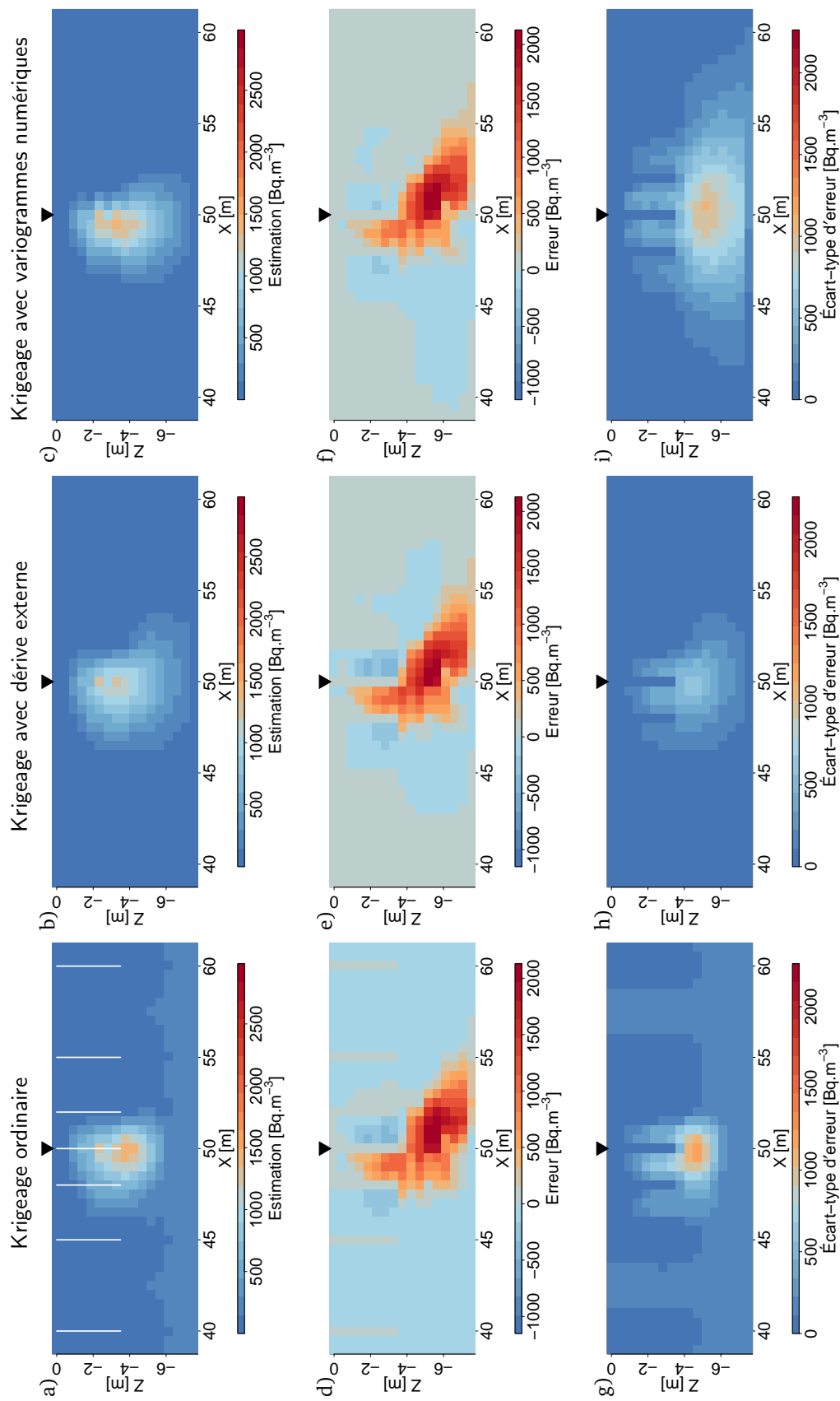


FIGURE 4.9 – Cartes d'estimation, d'erreur d'estimation et d'écart-type d'erreur obtenues avec les trois méthodes pour le scénario S3. Les échelles de couleur sont les mêmes que sur la Figure 4.5.

TABLEAU 4.3 – Erreurs moyennes pour les trois méthodes pour le scénario S3.

	KO	KDE	KVN
EAM [Bq.m^{-3}]	213	154	145
REQM [Bq.m^{-3}]	412	394	391
ERM [-]	-55.2	-2.9	-1.2

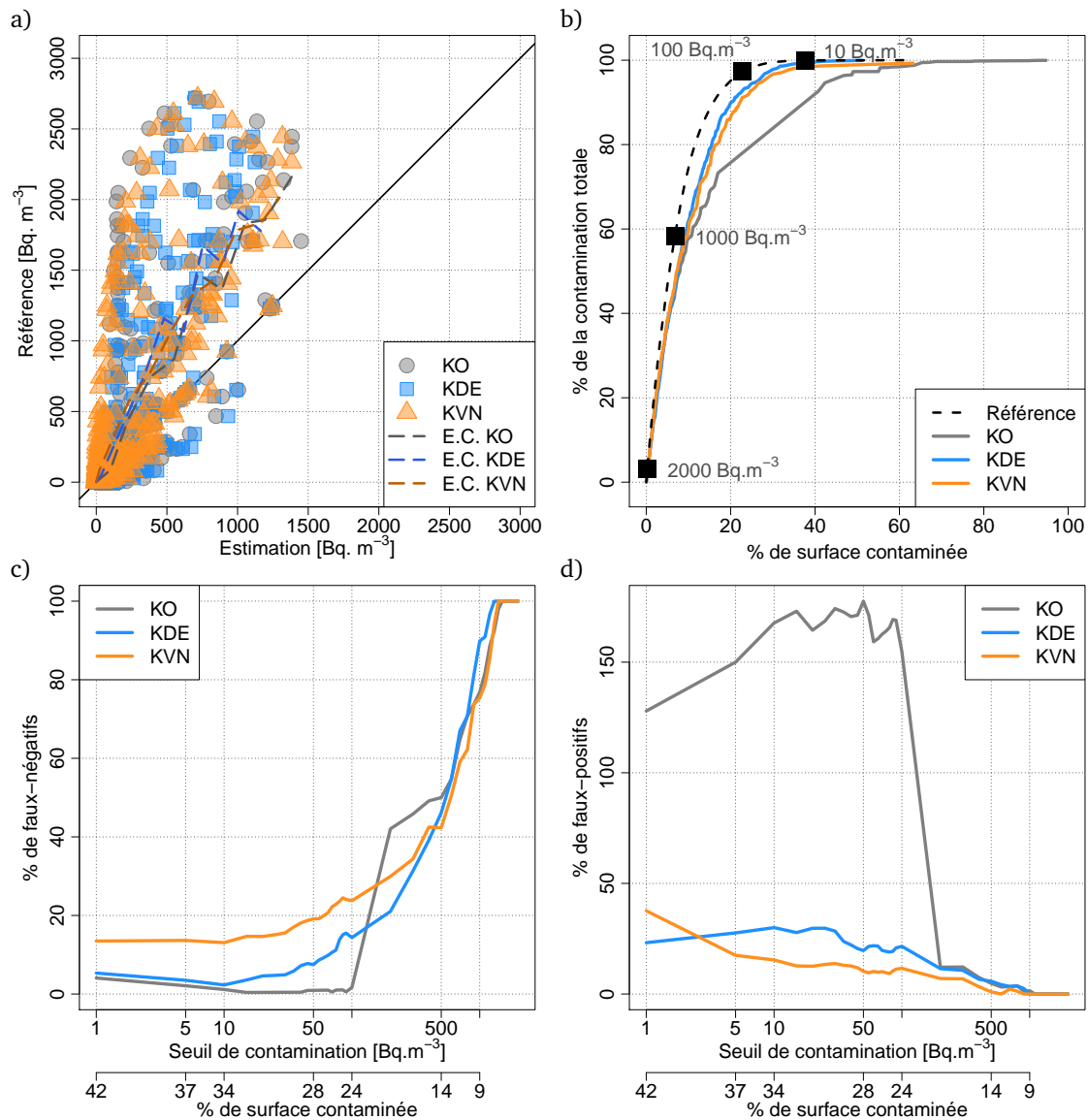


FIGURE 4.10 – Indicateurs de performance pour le scénario S3 : a) comparaison entre les valeurs d'activité de la référence et les activités estimées ; b) courbes de sélectivité ; c) proportion de faux-négatifs par rapport à la surface réellement contaminée ; d) proportion de faux-positifs par rapport à la surface réellement contaminée.

4.3.3 Influence de la méthode de construction des champs de paramètres

Au Chapitre 3 (Section 3.3.3), deux approches ont été utilisées pour générer les champs de paramètres de Mualem-van Genuchten à partir desquels sont simulés les panaches de tritium. Les estimations par krigeage avec variogrammes numériques ont été réalisées en utilisant dans un premier temps uniquement les 813 simulations obtenues avec l'approche 1 (simulations de champs de textures puis conversion en champs de paramètres), puis dans un second temps uniquement les 883 simulations obtenues avec l'approche 2 (conversion des données de texture et simulations de champs de paramètres). Les erreurs moyennes pour les trois scénarios d'échantillonnage sont données dans le Tableau 4.4.

TABLEAU 4.4 – Erreurs moyennes obtenues par krigeage avec variogrammes numériques en utilisant uniquement les simulations générées par l'approche 1 (App. 1) puis uniquement celles générées par l'approche 2 (App. 2).

	(a) Scénario S1		(b) Scénario S2		(c) Scénario S3	
	App. 1	App. 2	App. 1	App. 2	App. 1	App. 2
EAM [Bq.m ⁻³]	42	44	79	56	131	160
REQM [Bq.m ⁻³]	114	104	232	135	352	425
ERM [-]	-0.7	-4.9	-0.3	-2.4	-1.4	-1.7

Pour le scénario S1, les erreurs sont presque identiques pour les deux approches. Pour les scénarios S2 et S3, les erreurs sont légèrement différentes, mais selon le scénario, ce n'est pas la même approche qui donne les erreurs les plus faibles. Cela ne permet pas de conclure sur un groupe de simulations à utiliser en priorité. La même étude a été menée sur les références supplémentaires (Annexe E). Pour l'une des deux références supplémentaires, les résultats sont nettement dégradés lorsque seules les simulations obtenues à partir de l'approche 1 sont utilisées. Cette différence est due à la plus faible variabilité des simulations obtenues avec l'approche 1 par rapport à celles obtenues avec l'approche 2 (Figure 3.17), ce qui souligne qu'il est important d'utiliser un jeu de simulations suffisamment variable.

4.3.4 Régularisation

La régularisation, présentée au Chapitre 2 (Section 2.2.3), a pour objectif d'améliorer le conditionnement de la matrice de krigeage. Elle est mise en place en utilisant dans un premier temps toutes les simulations. Le voisinage de régularisation est un carré de trois mailles de côté centré sur le point d'intérêt. Le krigeage avec variogrammes numériques régularisés conduit à des erreurs nettement plus élevées que le krigeage avec variogrammes numériques non régularisés (Tableau 4.5).

TABLEAU 4.5 – Erreurs moyennes obtenues par krigeage avec variogrammes numériques sans régulariser (Non rég.) et en régularisant (Rég.) les variogrammes.

	(a) Scénario S1		(b) Scénario S2		(c) Scénario S3	
	Non rég.	Rég.	Non rég.	Rég.	Non rég.	Rég.
EAM [Bq.m ⁻³]	40	161	58	429	145	211
REQM [Bq.m ⁻³]	99	345	162	842	391	445
ERM [-]	-3	-19.7	-0.5	-83.3	-1.2	-16.9

L'intérêt de la régularisation a également été étudié en ne considérant qu'une centaine de simulations pour calculer les variogrammes numériques. Dans ce cas, la régularisation ne permet pas non plus d'améliorer les estimations. Enfin, la randomisation de l'origine de la grille des simulations a été mise en œuvre, ce qui ne permet pas non plus d'améliorer les résultats. Cela montre que les méthodes de régularisation utilisées ne suffisent pas à rendre les matrices de covariances ou de variogrammes plus robustes. D'autres méthodes sont mises en œuvre dans ce travail (voir Chapitre 7).

4.3.5 Covariances ou variogrammes numériques ?

Les estimations de la Section 4.3.2 ont été obtenues avec des variogrammes numériques. Les mêmes estimations sont ici réalisées avec des covariances numériques, qui prennent en compte la moyenne des simulations. Pour les scénarios S1 et S2, les résultats sont presque identiques avec les covariances ou avec les variogrammes numériques (Tableaux 4.6a et 4.6b). Pour le scénario S3, les erreurs sont plus faibles avec les variogrammes numériques (Tableau 4.6c).

TABLEAU 4.6 – Erreurs moyennes obtenues par krigeage avec variogrammes numériques et par krigeage avec covariances numériques.

(a) Scénario S1			(b) Scénario S2		(c) Scénario S3	
	KVN	KCN	KVN	KCN	KVN	KCN
EAM [Bq.m ⁻³]	40	40	58	56	145	174
REQM [Bq.m ⁻³]	99	99	162	151	391	464
ERM [-]	-3	-3.3	-0.5	0	-1.2	-0.1

4.3.6 Influence des conditions de non biais

Il est possible de travailler avec une moyenne variable sur le domaine (Chapitre 2, Section 2.2.2), ce qui conduit à une modification des conditions de non biais du système de krigeage. Cependant, cette modification des conditions de non biais n'affecte que très peu les résultats d'estimation (Tableau 4.7). Ainsi, il n'est pas nécessaire de relâcher l'hypothèse de la stationnarité de la moyenne : cela conduit à un modèle plus complexe que celui avec une moyenne constante, sans améliorer les résultats.

TABLEAU 4.7 – Erreurs moyennes obtenues par krigeage avec variogrammes numériques avec moyenne constante (KVN1) et avec moyenne variable (KVN2).

(a) Scénario S1			(b) Scénario S2		(c) Scénario S3	
	KVN1	KVN2	KVN1	KVN2	KVN1	KVN2
EAM [Bq.m ⁻³]	40	40	58	60	145	139
REQM [Bq.m ⁻³]	99	99	162	171	391	371
ERM [-]	-3	-3	-0.5	-0.7	-1.2	-2.2

4.3.7 Choix de la dérive

Les résultats présentés pour le krigeage avec dérive externe ont été obtenus en utilisant comme fonction de dérive la moyenne du jeu de simulations. D'autres fonctions de dérive ont été testées :

1. la médiane de l'ensemble (Figure 4.3b) ;
2. le panache central de l'ensemble (Figure 4.3d) ;
3. les trois panaches les plus proches de la référence (ceux qui minimisent les écarts aux observations).

Les résultats montrent que l'utilisation du panache central en dérive externe donne systématiquement de moins bons résultats que les autres fonctions de dérive (quel que soit le scénario d'échantillonnage et pour la référence principale comme pour les références supplémentaires). Parmi les trois autres fonctions de dérive (moyenne, médiane, panaches proches), aucune n'apparaît nettement meilleure que les autres. Les résultats dépendent de la référence et des scénarios d'échantillonnage et il faut adapter la fonction de dérive au cas d'étude.

Une autre approche, basée sur les méthodes de réduction de scénarios [Scheidt et Caers, 2009], consisterait à identifier un faible nombre de panaches représentatifs de la variabilité de l'ensemble et à utiliser ces panaches comme variables auxiliaires. Dans l'objectif de minimiser le recours à des codes d'écoulement et de transport, la réduction de scénarios pourrait également être effectuée sur les champs de paramètres d'entrée. Ainsi, seul un nombre minimal de panaches serait généré.

4.3.8 Discussion

Intérêt de la combinaison des simulations numériques et du krigeage

Les estimations réalisées sur ce cas synthétique mettent en évidence l'intérêt de combiner les simulations d'écoulement et de transport et le krigeage : quel que soit le scénario d'échantillonnage, les méthodes combinant les deux approches donnent de meilleures estimations que le krigeage ordinaire. L'intérêt de cette combinaison est d'autant plus marqué que le nombre de sondages diminue : pour les scénarios S1 (119 observations sur sept sondages) et S3 (56 observations sur sept sondages), l'erreur absolue moyenne est divisée par deux entre le krigeage ordinaire et le krigeage avec variogrammes numériques. Pour le scénario S2 (68 observations sur quatre sondages), elle est divisée par trois. Les indicateurs quantitatifs, mais également l'allure des panaches, montrent que le krigeage avec variogrammes numériques est plus performant que le krigeage avec dérive externe. Plusieurs variantes ont été mises en œuvre (la méthode de génération des champs de paramètres, la régularisation des variogrammes numériques, le remplacement des variogrammes par des covariances, la modification des conditions de non biais et le changement des fonctions de dérive), mais aucune ne modifie fortement les estimations.

Dans cet exemple, le modèle de variogramme utilisé en krigeage ordinaire est ajusté sur un variogramme expérimental obtenu à partir des observations. Comme proposé dans certaines publications, ce modèle pourrait être ajusté sur un variogramme expérimental calculé sur une unique simulation [Varentsov et al., 2019 ; Yong et al., 2019]. En tirant une réalisation dans le jeu de simulations et en ajustant un modèle de variogramme à partir de cette réalisation, les estimations ne sont pas améliorées. Cette approche est par ailleurs limitée, car il semble arbitraire de tirer aléatoirement une unique réalisation et de s'en servir comme seul moyen de calcul du variogramme. En effet, la Figure 4.2 souligne que les panaches sont très différents les uns des autres, et qu'une seule réalisation ne suffit pas à représenter l'ensemble. La même approche a été mise en place en utilisant la moyenne des simulations plutôt qu'une seule réalisation. Les résultats sont légèrement améliorés, mais restent moins bons que ceux obtenus par krigeage avec variogrammes numériques.

Enfin, les méthodes de combinaison améliorent la caractérisation de la contamination par rapport aux simulations seules. Bien que les seuls paramètres inconnus du modèle d'écoulement et de transport soient les champs de paramètres de Mualem-van Genuchten, les simulations sont éloignées de la référence. En effet, l'erreur absolue moyenne entre les simulations du jeu et la référence vaut au minimum 96 Bq.m^{-3} (soit plus que l'erreur absolue moyenne obtenue par krigeage avec variogrammes numériques pour les scénarios S1 et S2) et en moyenne 254 Bq.m^{-3} (soit plus que les erreurs obtenues pour tous les scénarios d'échantillonnage).

Influence du nombre de simulations

Les résultats présentés dans la Section 4.3.2 ont été obtenus en utilisant toutes les simulations disponibles pour calculer les variogrammes numériques et la fonction de dérive. Les mêmes estimations ont été réalisées avec moins de simulations (entre 100 et 1000), afin de quantifier la robustesse des méthodes par rapport au nombre de simulations utilisées. Vingt tirages de simulations ont été effectués à nombre de simulations fixé. Les erreurs moyennes obtenues par krigeage avec dérive externe dépendent peu du nombre de simulations et sont peu différentes d'un tirage à l'autre (Figure 4.11). En revanche, pour le krigeage avec variogrammes numériques, les erreurs moyennes diminuent lorsque le nombre de simulations augmente. Pour les faibles nombres de simulations (moins de quelques centaines), les estimations dépendent fortement du tirage. Malgré cette incertitude, les erreurs obtenues par krigeage avec variogrammes numériques restent plus faibles que celles obtenues par krigeage ordinaire. Pour le scénario S3, les erreurs obtenues par krigeage avec variogrammes numériques dépendent moins du nombre de simulations que pour les scénarios S1 et S2.

En résumé, lorsque peu de simulations sont disponibles (moins de 200 ici), le krigeage avec dérive externe est plus fiable et plus robuste que le krigeage avec variogrammes numériques. En revanche, si un plus grand nombre de simulations est disponible, le krigeage avec variogrammes numériques donne des erreurs plus faibles. Ce résultat n'est pas surprenant : le krigeage avec dérive externe est basé sur l'estimation d'un moment statistique d'ordre un (moyenne), alors que le krigeage avec variogrammes numériques est basé sur l'estimation d'un moment d'ordre deux (variogrammes).

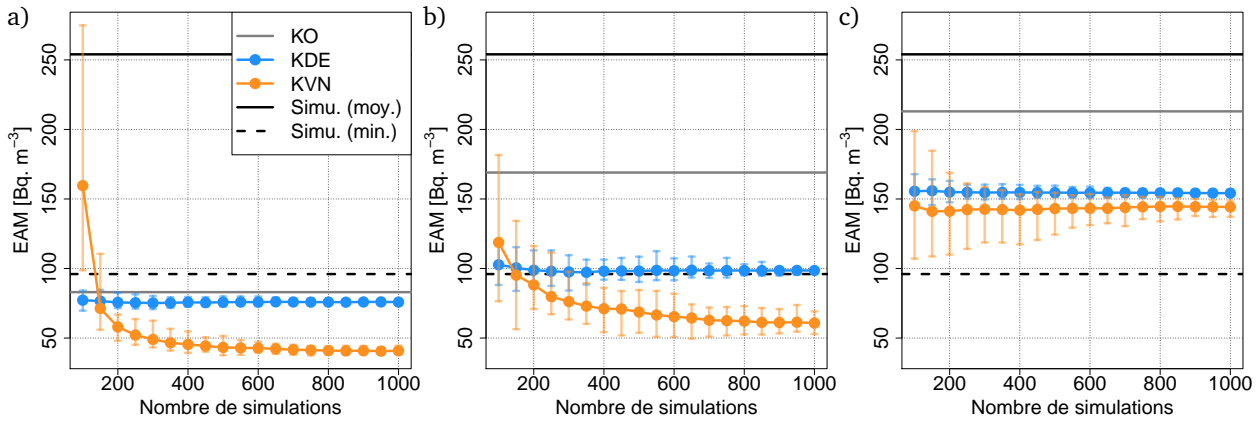


FIGURE 4.11 – Influence du nombre de simulations sur les erreurs absolues moyennes obtenues par krigeage avec variogrammes numériques et par krigeage avec dérive externe pour les scénarios S1 (a), S2 (b) et S3 (c). Les points représentent les erreurs moyennes obtenues sur 20 tirages de simulations, et les barres d'erreurs donnent les erreurs minimales et maximales obtenues sur les 20 tirages. Les erreurs obtenues par krigeage ordinaire et sur les simulations (en moyenne ou au minimum) sont également rappelées.

4.4 Quantification des incertitudes

L'objectif opérationnel d'une étude géostatistique de site contaminé est souvent d'estimer la quantité et la localisation des terres devant être dépolluées ou excavées, ainsi que les incertitudes associées. Trois méthodes sont comparées pour quantifier les risques liés aux incertitudes d'estimation.

4.4.1 Approche classique basée sur l'écart-type théorique d'erreur de krigeage

Dans le cas d'estimations par krigeage, la distinction entre zone contaminée et zone saine est rarement effectuée en se basant uniquement sur le résultat de l'estimation : l'écart-type théorique d'erreur de krigeage (σ_K) est pris en compte pour distinguer une zone très probablement contaminée, une zone très probablement saine, et une zone intermédiaire sur laquelle il est difficile de se prononcer [Bobbia *et al.*, 2008 ; de Fouquet *et al.*, 2011]. Cette méthode est fondée sur l'hypothèse de résidus R gaussiens et indépendants de l'estimation :

$$Z = Z^* + \sigma_K R \quad \text{avec} \quad R \sim \mathcal{N}(0, 1) \quad (4.7)$$

où Z^* est l'estimation de Z . Dans ce cas, des intervalles de confiance sont construits en utilisant les quantiles de la loi normale (voir Complément 4.1).

Intervalles de confiance

Sous l'hypothèse d'un résidu gaussien et indépendant de Z^* , un intervalle de confiance est construit comme suit :

$$Z^* + \sigma_K q_\alpha \leq Z < Z^* + \sigma_K q_{1-\alpha} \quad \text{au risque } 2\alpha \text{ près} \quad (4.8)$$

où q_α est le quantile de la loi normale associé à α . Il est également possible de calculer des intervalles de confiance non symétriques. Soient α et β les risques respectivement associés à la borne inférieure et à la borne supérieure de l'intervalle de confiance :

$$Z_{inf} = Z^* + \sigma_K q_\alpha \quad \text{et} \quad Z_{sup} = Z^* + \sigma_K q_{1-\beta} \quad (4.9)$$

Les trois configurations possibles sont schématisées sur la Figure 4.12.

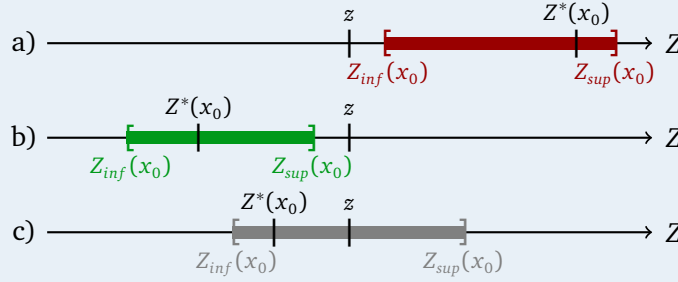


FIGURE 4.12 – Configurations possibles au point x_0 par rapport au seuil de contamination z : a) contaminé au risque α près ; b) sain au risque β près ; c) zone intermédiaire incertaine.

4.4.2 Indicatrice de dépassement de seuil

Dans le cas où la loi du résidu de l'estimation ne vérifie pas les hypothèses requises, une autre approche est proposée. Elle repose sur le krigage de l'indicatrice de dépassement de seuil $\mathbb{1}_{Z \geq z}$, qui est une approximation de l'espérance conditionnelle. Dans ce travail, l'indicatrice de dépassement de seuil est estimée par cokrigage avec des covariances simples et croisées numériques et en utilisant comme variable auxiliaire les valeurs d'activité. Des exemples d'indicatrices ainsi estimées sont donnés sur la Figure 4.13. Ces résultats sont satisfaisants, puisque l'on retrouve bien les fortes valeurs d'indicatrice à l'intérieur du contour.

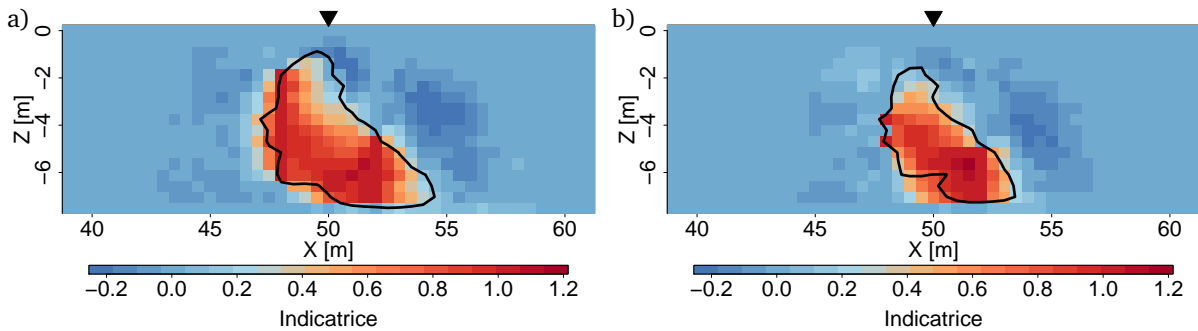


FIGURE 4.13 – Cokrigage de l'indicatrice pour les seuils 500 Bq.m^{-3} (a) et 1000 Bq.m^{-3} (b). Le trait noir indique l'enveloppe du panache de référence pour ces seuils. Ces cartes sont obtenues en considérant le scénario d'échantillonnage S2.

En utilisant l'estimation de l'indicatrice de dépassement d'un seuil z , les trois zones précédemment décrite sont définies comme suit :

- x_0 est dans la zone très probablement contaminée si $Z^*(x_0) \geq z$ et $\mathbb{1}_{Z \geq z}^*(x_0) \geq p_1$;
- x_0 est dans la zone très probablement saine si $Z^*(x_0) < z$ et $\mathbb{1}_{Z \geq z}^*(x_0) < p_2$;
- x_0 est dans la zone intermédiaire si aucune des deux conditions précédentes n'est remplie.

L'information sur l'indicatrice permet de conforter (cas 1 et 2) ou de nuancer (cas 3) le résultat de l'estimation. Les valeurs les plus pertinentes de p_1 et p_2 sont celles permettant d'avoir peu de risques sur la zone très probablement contaminée (risque économique de faux-positif), et encore moins sur la zone très probablement saine (risque sanitaire de faux-négatif). Lorsque p_1 augmente, la surface de la zone intermédiaire augmente et celle de la zone très probablement contaminée diminue. Le risque de trouver de la contamination dans la zone intermédiaire augmente et le risque de faux-positif diminue. Inversement, quand p_1 est plus faible, il y a moins de risque de retrouver de la contamination dans la zone intermédiaire. D'un point de vue sanitaire, il est donc préférable d'avoir une valeur de p_1 plutôt faible. Quand p_2 diminue, la surface de la zone très probablement saine diminue et celle de la zone intermédiaire augmente. Le risque de faux-négatif diminue. Ainsi, d'un point de vue sanitaire, il est également préférable d'avoir une valeur de p_2 plutôt faible.

4.4.3 Simulations géostatistiques

Des réalisations de fonctions aléatoires peuvent être générées à partir d'une matrice de covariance donnée (Chapitre 2, Section 2.2.5). Ces réalisations permettent de prendre en compte les incertitudes, en travaillant sur la distribution obtenue sur un grand nombre de réalisations. Plusieurs hypothèses sous-jacentes sont nécessaires pour générer ces réalisations, et bien que la validité d'une partie de ces hypothèses soit discutable sur le cas synthétique étudié (Complément 4.2), des simulations conditionnelles ont été réalisées (le conditionnement est effectué par krigeage simple des résidus). Les panaches simulés semblent réalistes (Figure 4.14). Pour tracer des cartes de probabilité de dépassement de seuil, 200 simulations conditionnelles sont générées. Les résultats sont intéressants, puisque les fortes probabilités sont situées à l'intérieur des contours (Figure 4.15).

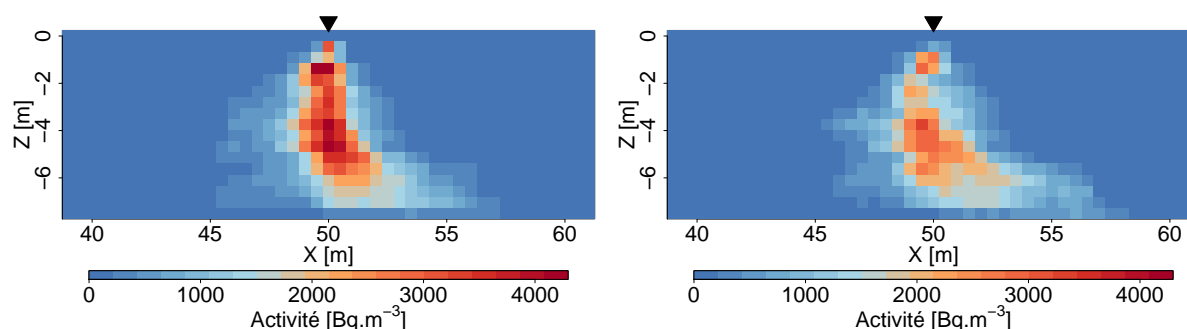


FIGURE 4.14 – Simulations conditionnées par les données de la référence pour le scénario S2.

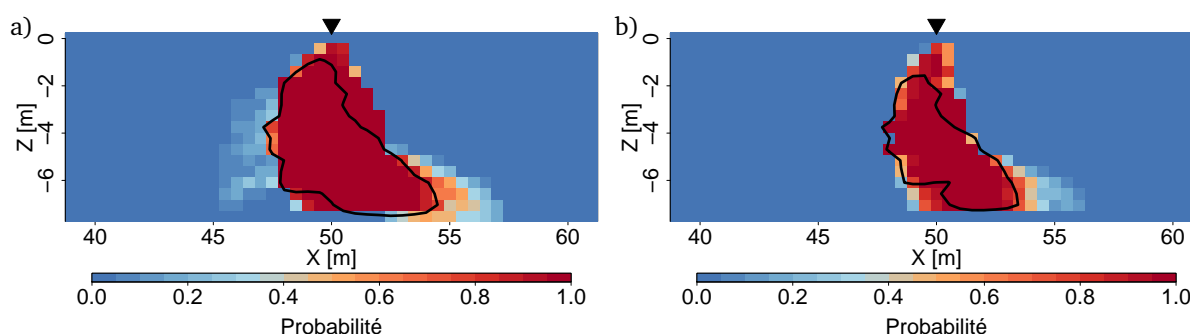


FIGURE 4.15 – Cartes des probabilités de dépassement des seuils 500 Bq.m^{-3} (a) et 1000 Bq.m^{-3} (b) obtenues à partir de 200 simulations conditionnées par les données du scénario S2.

Hypothèses sous-jacentes

La méthode de simulation a été détaillée au Chapitre 2 (Section 2.2.5). La première étape consiste à transformer par anamorphose la référence pour obtenir une distribution gaussienne. Cette étape est discutable dans le cas présent, puisque les simulations de panaches présentent un effet zéro important : sur une partie du domaine de modélisation, l'activité est nulle pour toutes les réalisations ou presque. Après anamorphose, il faut vérifier que les vecteurs obtenus sont gaussiens (*a minima* bi-gaussiens). Deux éléments sont étudiés pour confirmer ou infirmer cette hypothèse (Chapitre 1, Complément 1.2).

a) Nuages de corrélation

Deux exemples de nuages de corrélation $[Y(x_1), Y(x_2)]$, où Y est la transformée gaussienne de l'activité, sont présentés sur la Figure 4.16, pour deux couples de points pris dans la zone où l'effet zéro est faible. Ces deux nuages ne sont pas elliptiques, ce qui tend à infirmer l'hypothèse que les vecteurs obtenus sont bi-gaussiens.

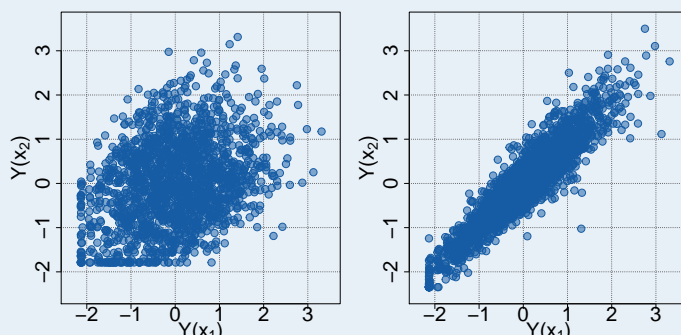


FIGURE 4.16 – Nuages de corrélation $[Y(x_1), Y(x_2)]$ pour deux couples de points.

b) Variogramme et madogramme

L'autre élément à vérifier est le rapport entre la racine du variogramme et le madogramme. Ce rapport est tracé pour tous les couples de points présentant un effet zéro faible en fonction de la distance entre les deux points (Figure 4.17). Ce rapport n'est pas constant, ce qui tend également à infirmer l'hypothèse que les vecteurs obtenus sont bi-gaussiens.

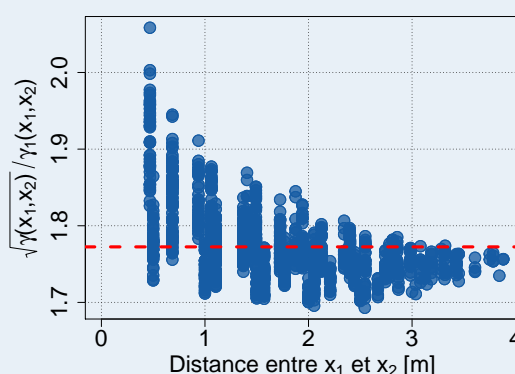


FIGURE 4.17 – Rapport entre la racine du variogramme et le madogramme en fonction de la distance entre les points. Le trait rouge indique la constante $\sqrt{\pi}$, à laquelle le rapport est égal dans le cas gaussien.

Complément 4.2

4.4.4 Comparaison des méthodes et discussion

Chacune des trois méthodes proposées présente des avantages et limites, qui sont discutés sur le cas synthétique étudié. L'approche classique basée sur la construction d'intervalles de confiance présente l'avantage d'être une méthode conventionnelle et reconnue. Cependant, dans le cas présent, les hypothèses sur la loi des résidus obtenus par krigeage avec variogrammes numériques ne sont pas vérifiées. En effet, sous cette hypothèse, près de la moitié des points du domaine d'étude présentent une borne inférieure d'intervalle de confiance négative, parfois assez largement. Par ailleurs, en choisissant des risques raisonnables, beaucoup de points de la référence ne sont pas contenus dans les intervalles de confiance calculés.

Le principal inconvénient de la méthode de cokrigeage d'indicateur est qu'elle repose sur des paramètres p_1 et p_2 arbitraires, sur lesquels aucune étude n'a été menée. Ces paramètres dépendent des risques que l'on peut se permettre de prendre, d'un point de vue sanitaire (selon la substance considérée) et d'un point de vue économique (méthode de dépollution). En revanche, cette méthode ne requiert pas d'hypo-

thèses sur les résidus de krigeage. De plus, puisque les variogrammes simples et croisés de l'activité et des indicatrices sont calculés sur des simulations numériques, ils sont cohérents entre eux.

Le tracé de cartes de probabilité de dépassement de seuil à partir d'un grand nombre de simulations géostatistiques permet d'avoir la distribution en tout point, si la méthode de simulation utilisée est robuste. À nouveau, il faut se fixer des probabilités au-dessus et en-deçà desquelles les zones contaminées et saines se trouvent. La méthode de simulation proposée ici repose sur un grand nombre d'hypothèses qui ne sont pas toujours vérifiées. Il serait intéressant d'améliorer la méthode pour la rendre plus robuste.

Finalement, des cartes de délimitation en zone contaminée, zone saine et zone intermédiaire sont tracées pour les trois méthodes à partir des données du scénario S2. Les paramètres propres à chaque méthode sont arbitraires, mais ont été choisis de manière à ce que les méthodes soient comparables, dans la mesure du possible. Les risques α et β du calcul d'intervalles de confiance valent respectivement 0,05 et 0,1 (ce qui revient à accepter un risque plus élevé sur les faux-positifs, *i.e.* le risque économique). Les paramètres p_1 et p_2 de l'indicateur de dépassement de seuil sont fixés à 0,5 et 0,3. Pour les simulations, les zones contaminées et saines sont situées au-dessus de la probabilité 0,9 et en-deçà de la probabilité 0,3. Les cartes obtenues avec les trois méthodes sont assez semblables (Figure 4.18). La zone contaminée apparaît cependant plus étendue pour la méthode de simulations, et la zone intermédiaire est plus étendue avec la méthode des intervalles de confiance.

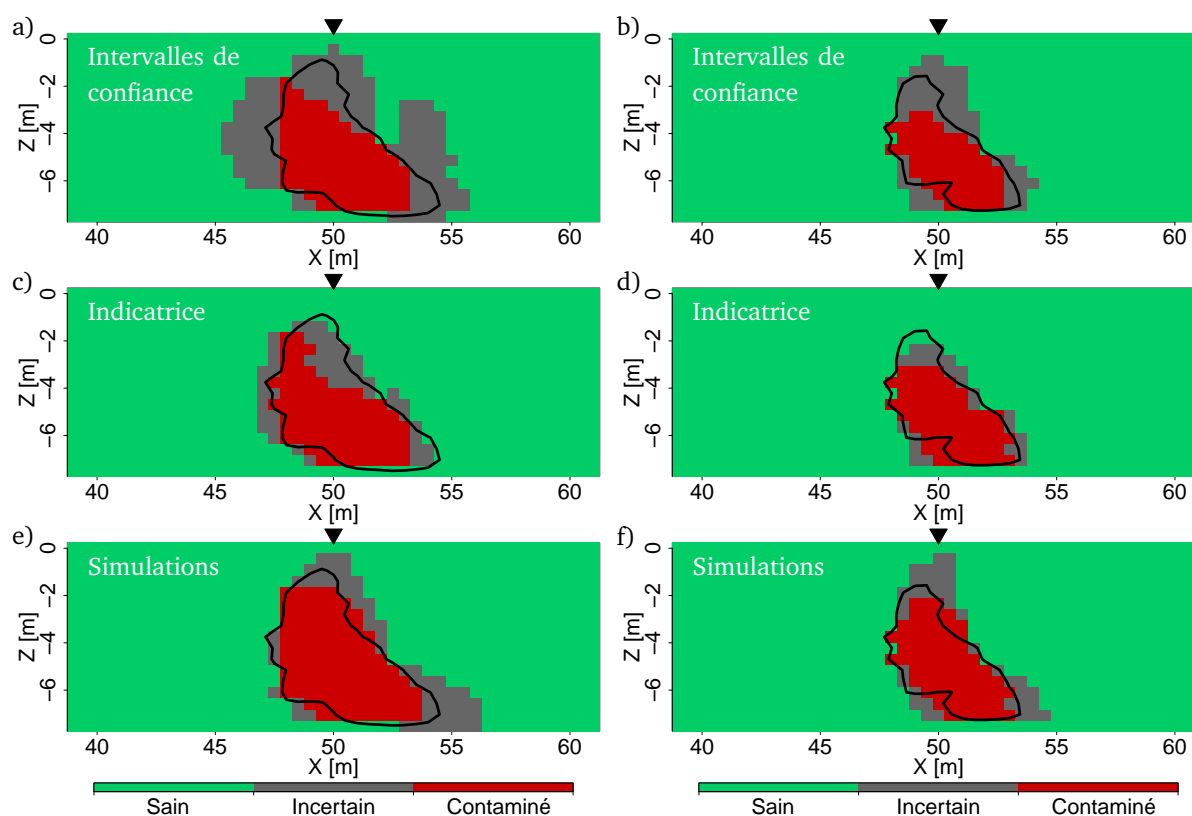


FIGURE 4.18 – Délimitation des trois zones pour les seuils 500 Bq.m⁻³ (a, c, e) et 1000 Bq.m⁻³ (b, d, f).

Les erreurs associées sont données dans le Tableau 4.8. Pour le seuil 500 Bq.m⁻³, très peu de mailles sont mal classées (faux-positifs et faux-négatifs), quelle que soit la méthode. Pour la méthode des intervalles de confiance, une très large partie de la zone intermédiaire est en réalité saine. C'est également le cas, dans une moindre mesure, pour les simulations. La zone intermédiaire est répartie de manière plus équilibrée entre mailles contaminées et mailles saines pour la méthode de l'indicateur. Pour le seuil 1000 Bq.m⁻³, la proportion de faux-positifs augmente, quelle que soit la méthode. À nouveau, une large partie de la zone intermédiaire est en réalité saine pour la méthode des intervalles de confiance, et dans une moindre mesure pour les simulations.

Ainsi, les approches usuelles (intervalles de confiance et simulations) apparaissent plus conservatives et moins risquées d'un point de vue sanitaire : le nombre de faux-négatifs est nul, et seule une faible partie de

la zone intermédiaire est en réalité contaminée. L'approche proposée dans ce travail, basée sur le krigeage d'indicatrices de dépassement de seuil, apparaît légèrement plus risquée (3% de faux négatifs contre 0% pour le seuil 500 Bq.m⁻³ et 4% pour le seuil 1000 Bq.m⁻³), mais conduit à une proportion nettement plus faible de la zone intermédiaire en réalité saine. Cela est intéressant d'un point de vue économique et logistique, notamment dans le cas d'une contamination radiologique : le volume de terre stocké dans des sites de stockage de déchets radiologiques, mais en réalité non contaminé, est plus faible que pour les autres approches.

Finalement, les trois méthodes sont assez comparables, et se distinguent surtout par leur mise en œuvre pratique et les hypothèses qu'elles requièrent. Par ailleurs, il est important de noter que dans cette étude, le support considéré pour la décontamination est ponctuel (maillage des simulations réalisées avec MELODIE). En réalité, le support dépend de la méthode de dépollution utilisée, et peut atteindre quelques mètres cubes.

TABLEAU 4.8 – Pourcentage de mailles mal classées avec les trois méthodes (Inter.=intervalles de confiance ; Indic.=indicatrices de dépassement de seuil ; Simu.=simulations) par rapport à la surface contaminée sur la référence (soit 104 mailles pour le seuil 500 Bq.m⁻³ et 69 pour le seuil 1000 Bq.m⁻³).

	(a) Seuil 500 Bq.m ⁻³			(b) Seuil 1000 Bq.m ⁻³		
	Inter.	Indic.	Simu.	Inter.	Indic.	Simu.
Faux-positifs	4	4	8	6	9	13
Faux-négatifs	0	3	0	0	4	0
Zone intermédiaire saine	97	24	43	51	20	42
Zone intermédiaire contaminée	23	23	10	28	14	13

4.5 Un cas multivariable

Le même cas synthétique est utilisé pour mettre en œuvre le cokrigeage avec covariances numériques (Chapitre 2, Section 2.2.4). Dans le cas du tritium, considéré comme un traceur parfait, le coefficient de retard R vaut 1. Ce coefficient de retard R donne le rapport entre la vitesse moyenne du fluide et la vitesse moyenne du soluté. Le transport de deux autres radionucléides a été simulé : le premier avec un coefficient R égal à 3, le second avec un coefficient R égal à 5. Ces deux autres radionucléides sont des éléments fictifs : les coefficients de retard considérés, associés à la constante de désintégration du tritium, ne correspondent pas à des éléments existants.

4.5.1 Références et jeux de simulations

Références

Les mêmes champs de paramètres de Mualem-van Genuchten et les mêmes conditions initiales et conditions aux limites ont été utilisées pour générer deux nouvelles références, l'une avec $R=3$ et l'autre avec $R=5$ (Figure 4.19). La date considérée ici est six ans après le déversement de radionucléide (contre cinq précédemment), afin d'avoir des panaches plus étendus pour les radionucléides retardés.

Jeux de simulations

Cinq cents quintuplets de champs de paramètres de Mualem-van Genuchten ont été tirés parmi les 2000 disponibles et utilisés en entrée du code MELODIE pour générer des panaches avec les coefficients de retard $R=3$ et $R=5$. La répartition des surfaces des panaches simulés est présentée sur la Figure 4.20 et comparée à la répartition des surfaces des panaches de tritium obtenus avec les mêmes champs de paramètres de Mualem-van Genuchten. Les moyennes des surfaces des panaches passent de 364 mailles pour $R=1$, à 190 pour $R=3$ et à 114 pour $R=5$. Les écart-types passent de 67 mailles pour $R=1$, à 48 pour $R=3$ et à 39 pour $R=5$. Ainsi, comme attendu, lorsque le coefficient de retard augmente, les surfaces des panaches diminuent

en moyenne, et l'incertitude sur ces surfaces diminue également. Cependant, l'incertitude relative (rapport entre écart-type et moyenne) augmente quand le coefficient de retard augmente.

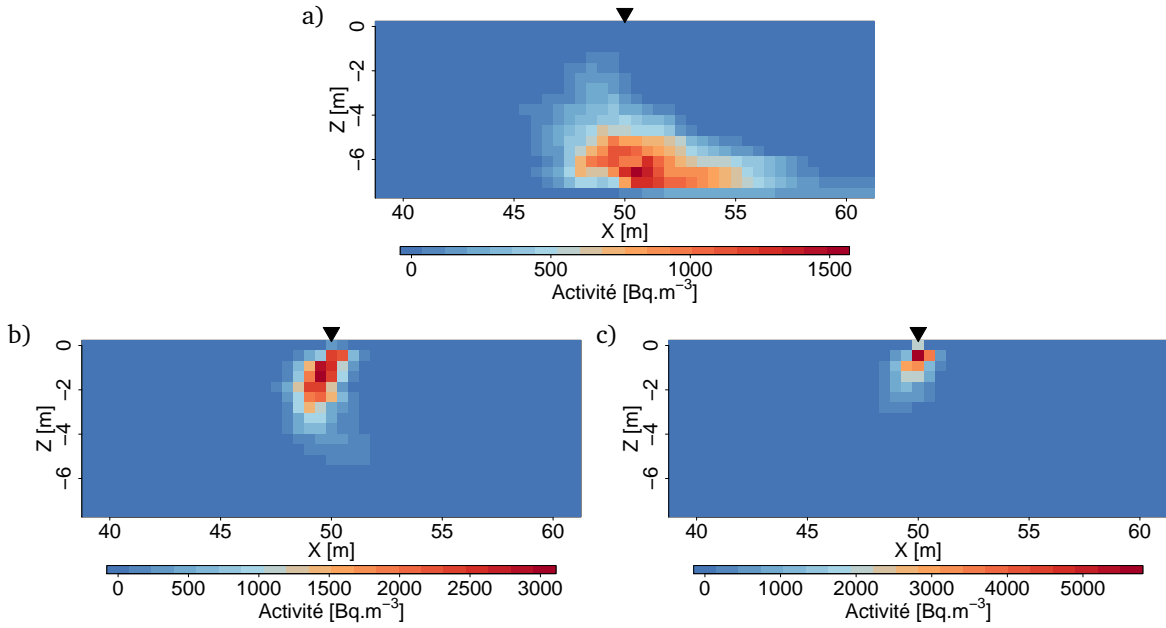


FIGURE 4.19 – Panaches de référence six ans après le déversement. a) $R=1$ (tritium) ; b) $R=3$; c) $R=5$. Les échelles de couleur sont différentes sur les trois cartes.

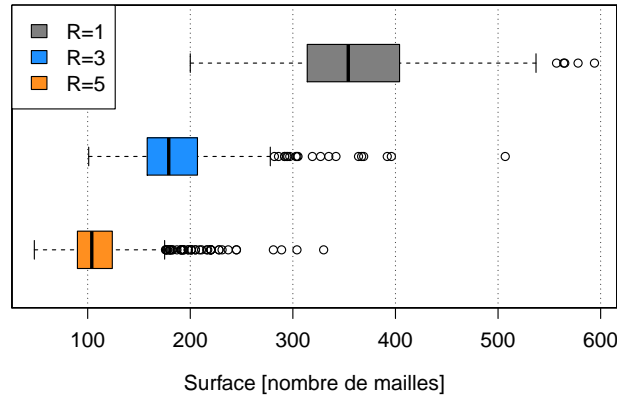


FIGURE 4.20 – Boîtes de dispersion (premier quartile, médiane, troisième quartile) représentant les surfaces des séries de 500 panaches simulés.

4.5.2 Covariances croisées numériques

Les covariances croisées numériques C_{ij} sont non stationnaires et vérifient les relations suivantes :

$$C_{ij}(x, x') \neq C_{ij}(x', x) \quad (4.10)$$

$$C_{ij}(x, x') \neq C_{ji}(x, x') \quad (4.11)$$

$$C_{ij}(x, x') = C_{ji}(x', x) \quad (4.12)$$

Ces relations sont visibles sur la Figure 4.21, sur laquelle sont représentées les valeurs de $C_{ij}(x, x')$ pour 100 points tirés aléatoirement dans le domaine de modélisation, en fonction de la distance algébrique entre x et x' . Les deux nuages de points sont symétriques par rapport à l'axe des abscisses.

Ces covariances croisées numériques, ainsi que les covariances simples numériques, sont utilisées pour estimer les valeurs d'activité par cokrigage. À noter que l'on travaille ici avec des covariances numériques

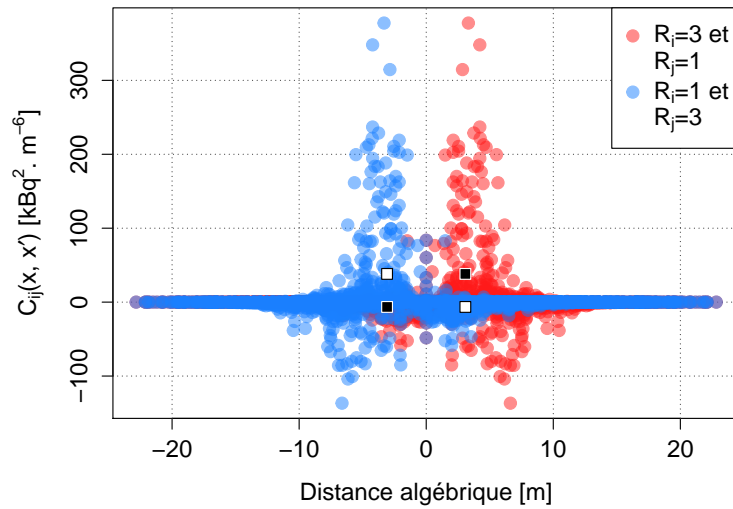


FIGURE 4.21 – Covariances croisées numériques entre l'activité du radionucléide avec $R=3$ et celui avec $R=1$ (notées C_{31} et C_{13}). Les deux carrés noirs représentent $C_{31}(x, x')$ et $C_{31}(x', x)$ pour deux points x et x' particuliers et les deux carrés blancs représentent $C_{13}(x, x')$ et $C_{13}(x', x)$ pour ces mêmes points.

plutôt que des variogrammes numériques, car les covariances sont plus usuelles et contiennent plus d'informations que les variogrammes dans le cas multivarié [Chilès et Delfiner, 2012].

4.5.3 Cokrigage

Scénario

Les activités des références sont supposées connues suivant le scénario d'échantillonnage S1 (sept sondages traversant l'intégralité de la zone non saturée). Le scénario S2 n'est pas utilisé car les panaches sont trop peu étendus par rapport à l'espacement entre les sondages pour ce scénario.

La variable représentant l'activité volumique du radionucléide tel que $R=i$ (pour i valant 1, 3 ou 5) est notée Z_i . Les estimations sont réalisées par cokrigage suivant plusieurs configurations :

- estimation de Z_1 en utilisant Z_3 puis Z_5 comme variable auxiliaire, en voisinage unique ;
- estimation de Z_3 en utilisant Z_1 puis Z_5 comme variable auxiliaire, en voisinage glissant (cercle d'un rayon de 5 m) ;
- estimation de Z_5 en utilisant Z_1 puis Z_3 comme variable auxiliaire, en voisinage glissant (cercle d'un rayon de 5 m).

Deux méthodes sont comparées : un cokrigage ordinaire avec des modèles de covariances simples et croisées ajustés sur les données et un cokrigage ordinaire avec des covariances simples et croisées numériques. Les performances des méthodes sont évaluées par le tracé des cartes d'estimation, d'erreur d'estimation et d'écart-type théorique d'erreur d'estimation (corrigé par un effet proportionnel pour le cokrigage ordinaire), ainsi que par le calcul d'erreurs moyennes (EAM, REQm, ERM).

Cokrigage de Z_1

Les cartes d'estimation et d'erreur d'estimation obtenues par cokrigage de Z_1 avec Z_3 sont semblables pour les deux méthodes (Figure 4.22). En revanche, les écart-types d'erreur d'estimation sont nettement plus élevés pour le cokrigage ordinaire que pour le cokrigage avec covariances numériques. Les erreurs moyennes (Tableau 4.9) sont plus faibles pour le cokrigage avec covariances numériques. Lorsque la variable auxiliaire est Z_5 , les erreurs moyennes augmentent légèrement : cela semble cohérent, puisque la variable auxiliaire est moins proche de la variable principale dans ce cas.

À noter par ailleurs que les erreurs moyennes sont légèrement plus faibles en utilisant Z_3 comme variable auxiliaire, et 500 simulations, que pour les estimations sans variable auxiliaire, et avec plus de 1500

simulations : l'erreur absolue moyenne est divisée par deux entre le krigage avec variogrammes numériques monovariante et le cokrigage avec covariances numériques. Cela montre l'intérêt du cokrigage par rapport au krigage.

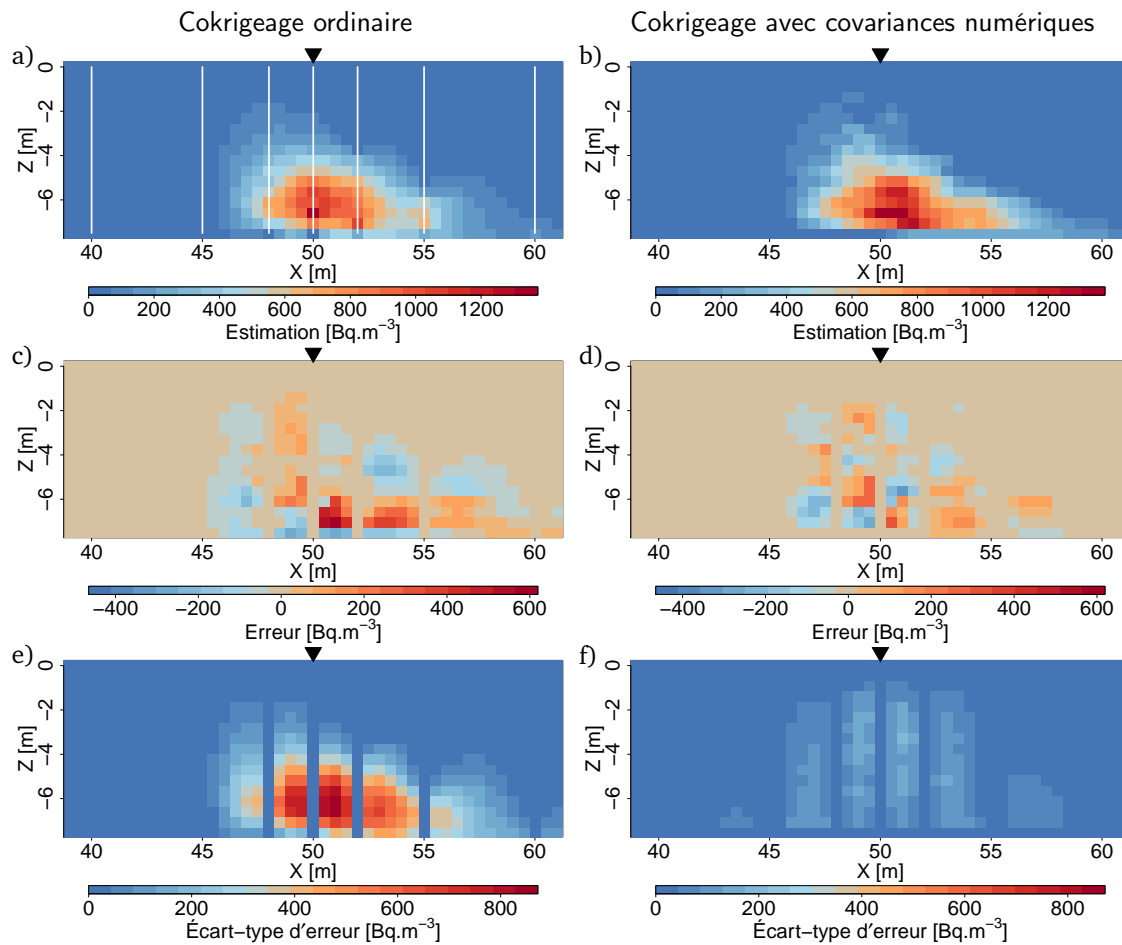


FIGURE 4.22 – Résultats d'estimation de Z_1 en utilisant Z_3 comme variable auxiliaire pour le scénario S1, rappelé sur la première carte. a) et b) Cartes d'estimation ; c) et d) cartes d'erreur d'estimation ; e) et f) cartes d'écart-type théorique d'erreur d'estimation.

TABEAU 4.9 – Erreurs moyennes obtenues par cokrigage ordinaire usuel (CoKO) et par cokrigage avec covariances numériques (CoKCN), pour l'estimation de Z_1 en utilisant d'une part Z_3 et d'autre part Z_5 comme variables auxiliaires et pour le scénario S1. Les points tels que $Z_1^{\text{ref}} < 1 \text{ Bq.m}^{-3}$ et $Z_1^* < 1 \text{ Bq.m}^{-3}$ (pour toutes les méthodes) ont été ôtés du calcul (soit 585 points conservés sur 765).

	CoKO avec Z_3	CoKCN avec Z_3	CoKO avec Z_5	CoKCN avec Z_5
EAM [Bq.m^{-3}]	40	24	43	26
REQM [Bq.m^{-3}]	81	55	93	59
ERM [-]	-6.8	-0.9	-5.2	-0.9

Cokrigage de Z_3

Les valeurs de Z_3 estimées par cokrigage avec covariances numériques en utilisant Z_5 comme variable auxiliaire sont légèrement plus élevées que celles estimées par cokrigage ordinaire (Figure 4.23). Les écart-types d'erreur d'estimation sont nettement plus élevés pour le cokrigage ordinaire que pour le cokrigage avec covariances numériques. Les cartes obtenues en utilisant Z_1 comme variable auxiliaire sont semblables à celles de la Figure 4.23. Les erreurs moyennes (Tableau 4.10) dépendent peu de la variable

auxiliaire utilisée : qu'il s'agisse de Z_1 ou Z_5 , les erreurs sont plus faibles pour le krigage avec covariances numériques.

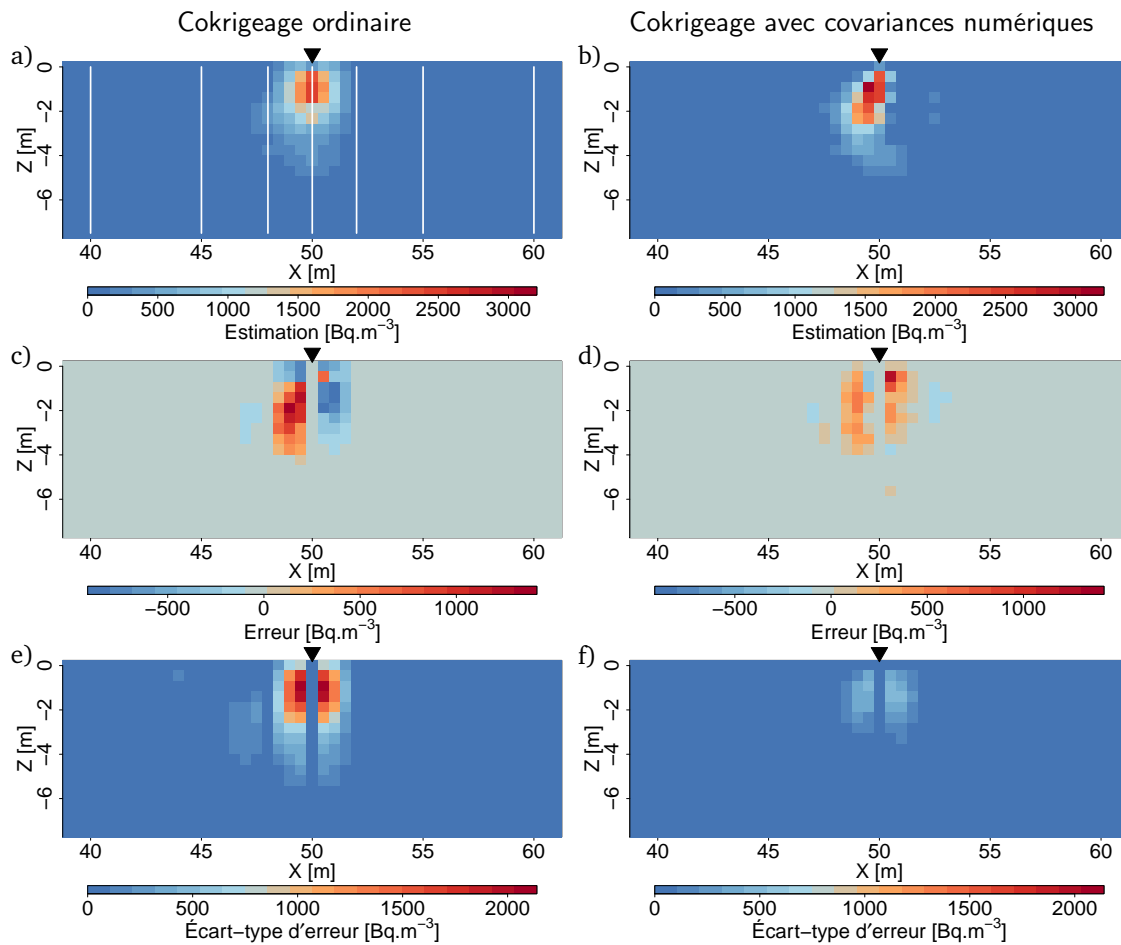


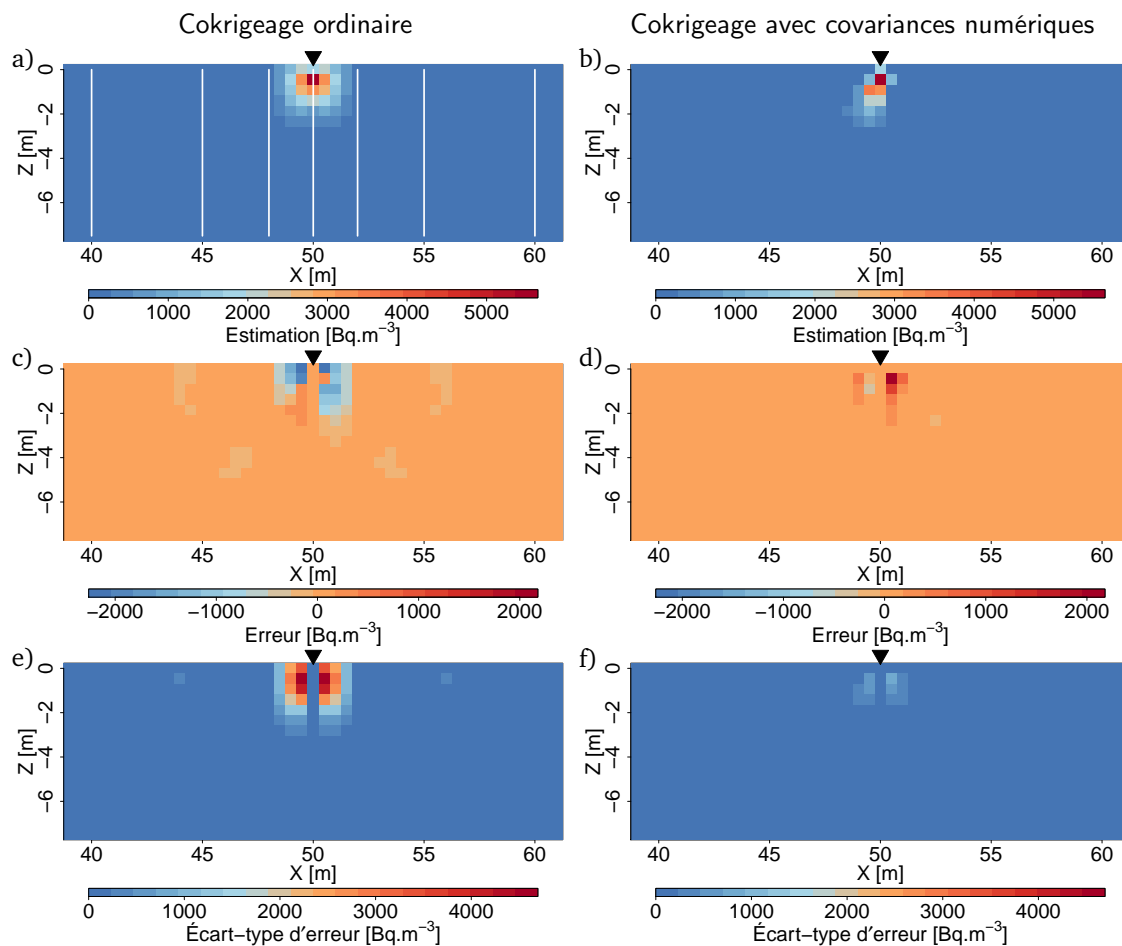
FIGURE 4.23 – Résultats d'estimation de Z_3 en utilisant Z_5 comme variable auxiliaire.

TABEAU 4.10 – Erreurs moyennes obtenues pour l'estimation de Z_3 en utilisant d'une part Z_1 et d'autre part Z_5 comme variables auxiliaires. Les points tels que $Z_3^{\text{ref}} < 1 \text{ Bq.m}^{-3}$ et $Z_3^* < 1 \text{ Bq.m}^{-3}$ ont été ôtés du calcul (soit 531 points conservés).

	CoKO avec Z_1	CoKCN avec Z_1	CoKO avec Z_5	CoKCN avec Z_5
EAM [Bq.m^{-3}]	59	26	62	24
REQM [Bq.m^{-3}]	193	112	195	99
ERM [-]	-5.6	-0.2	-6.9	-0.5

Cokrigage de Z_5

Le panache de référence simulé avec $R=5$ est très concentré et peu étendu. Par conséquent, les erreurs d'estimation sont très localisées et assez fortes en valeur absolue (Figure 4.24). Les écart-types d'erreur de krigage sont plus faibles pour le cokrigage avec covariances numériques. Les cartes obtenues en utilisant Z_1 comme variable auxiliaire sont proches de celles de la Figure 4.24. Quelle que soit la variable auxiliaire utilisée, les erreurs moyennes (Tableau 4.11) sont inférieures pour le cokrigage avec covariances numériques. Contrairement à ce qui est attendu, ces erreurs ne sont pas plus faibles lorsque la variable auxiliaire est Z_3 plutôt que Z_1 . Mais les erreurs moyennes sont assez peu représentatives ici, puisque le panache est concentré sur une dizaine de mailles.

FIGURE 4.24 – Résultats d'estimation de Z_5 en utilisant Z_3 comme variable auxiliaire.TABLEAU 4.11 – Erreurs moyennes obtenues pour l'estimation de Z_5 en utilisant d'une part Z_1 et d'autre part Z_3 comme variables auxiliaires. Les points tels que $Z_5^{\text{ref}} < 1 \text{ Bq.m}^{-3}$ et $Z_5^* < 1 \text{ Bq.m}^{-3}$ ont été ôtés du calcul (soit 437 points conservés).

	CoKO avec Z_1	CoKCN avec Z_1	CoKO avec Z_3	CoKCN avec Z_3
EAM [Bq.m^{-3}]	71	19	64	20
REQM [Bq.m^{-3}]	284	116	261	134
ERM [-]	-18.3	-0.1	-14.9	-0.1

4.5.4 Discussion

Cet exemple souligne l'intérêt de coupler la géostatistique et les simulations d'écoulement et de transport lorsque la contamination est liée à plusieurs radionucléides. En effet, le cokrigage avec covariances numériques permet d'améliorer les estimations. Cette méthode pourrait par exemple être utilisée dans le cas où un radionucléide très mobile est déjà sorti du domaine de modélisation, alors qu'un radionucléide peu mobile est encore très concentré autour de la source.

Cependant, si les radionucléides en présence ont des comportements physico-chimiques trop différents les uns des autres, le cokrigage avec covariances numériques est plus complexe à mettre en œuvre : les simulations d'écoulement et de transport doivent être capables de reproduire la migration des deux radionucléides, même si ceux-ci ne sont pas modélisés de la même manière.

4.6 Synthèse du chapitre

Un exemple d'application des méthodes combinant la géostatistique et les simulations d'écoulement et de transport a été détaillé dans ce chapitre. Les méthodes sont mises en œuvre sur un cas synthétique à deux dimensions, constitué par la migration d'un panache de tritium dans la zone non saturée d'une formation sableuse.

Ce premier exemple permet de mettre en évidence plusieurs points. D'abord, l'apport d'information de nature physique *via* les simulations d'écoulement et de transport améliore les estimations (krigeage avec variogrammes numériques ou krigeage avec dérive externe), tant sur des critères quantitatifs (erreurs moyennes, courbes de sélectivité, *etc.*) que qualitatifs (allure des panaches). L'intérêt principal du krigeage avec variogrammes numériques est de reproduire convenablement la non-stationnarité du phénomène étudié, puisque les variogrammes numériques dépendent des deux points d'appuis et ne sont pas moyennés par classes de distances. De plus, une méthode de quantification des incertitudes a été proposée. Elle n'est pas basée sur l'écart-type théorique d'erreur de krigeage, ni sur la réalisations de simulations conditionnelles, mais sur un cokrigeage d'indicatrice de dépassement de seuil.

Il est intéressant de souligner que le krigeage avec variogrammes numériques et le krigeage avec dérive externe sont d'autant plus performants que peu de données sont disponibles. En effet, au premier abord, l'une des limites évidentes de la méthode est le nombre de simulations nécessaires, et par conséquent les coûts de modélisation puis de calcul associés. Cependant, ces coûts sont à comparer aux coûts des prélèvements et analyses dans ce type de situation, ainsi qu'aux coûts des remédiations et surveillances de sites. Par exemple, une à deux semaines de calcul sont nécessaires pour passer de l'estimation par krigeage ordinaire à l'estimation par krigeage avec variogrammes numériques, mais cela permet de diviser l'erreur moyenne par trois lorsque quatre sondages sont disponibles (scénario S2), sans avoir à effectuer de forages ou d'analyses supplémentaires.

Un critère objectif de choix entre le krigeage avec dérive externe et le krigeage avec variogrammes numériques est donné par le nombre de simulations réalisables. Si le temps pouvant être consacré à la simulation numérique ne permet pas de réaliser plus d'une centaine à quelques centaines de simulations, le krigeage avec dérive externe est plus robuste.

Le krigeage avec covariances numériques a également été mis en œuvre sur un cas multivariable. La méthode se montre aussi intéressante dans ce cas, même si cela requiert la réalisation d'un nombre plus important de simulations, et éventuellement une modélisation un peu plus complexe.

Par ailleurs, dans ce chapitre, les estimations ont été réalisées à une date fixée, sans prendre en compte l'évolution temporelle de la contamination, et sans chercher à prévoir le développement des panaches. Ce point est abordé au Chapitre 5, qui traite de l'estimation spatio-temporelle sur ce même cas synthétique.

Enfin, dans cet exemple, seuls les champs de paramètres de Mualem-van Genuchten sont supposés inconnus, et par conséquent le modèle d'écoulement et de transport est déjà très contraint. Il est légitime de questionner la performance des méthodes lorsque le modèle est moins contraint. Ce point sera abordé dans la Partie III, avec des applications à des jeux de données réels.

Chapitre 5

Estimations spatio-temporelles et prévisions

Dans ce chapitre, le cas synthétique est utilisé pour réaliser des estimations spatio-temporelles et des prévisions de la migration de panaches de tritium. Les méthodes spatio-temporelles ont été présentées aux Chapitres 1 (Section 1.2) et 2 (Section 2.2.4). Le jeu de simulations d'écoulement et de transport est ici désigné par le terme ensemble de particules, pour reprendre la terminologie de l'assimilation de données.

5.1 Position du problème et ensemble de particules

5.1.1 Position du problème

On suppose que des données d'activité en tritium sont disponibles en plusieurs sondages et à plusieurs dates et que des simulations d'écoulement et de transport peuvent être générées aux mêmes dates, ainsi qu'à d'autres dates antérieures ou postérieures. En pratique, sur le cas synthétique traité, les panaches simulés sont disponibles tous les deux mois pendant six ans après le déversement de tritium.

Plusieurs situations sont envisagées :

1. aucune donnée n'est disponible à la date à laquelle l'estimation est recherchée, mais des données sont disponibles à des dates antérieures. Il s'agit alors de faire une prévision du développement du panache, et le krigeage spatio-temporel est comparé au filtre particulaire et au filtre de Kalman d'ensemble (Section 5.3).
2. des données d'activité sont disponibles à la date à laquelle l'estimation est recherchée ainsi qu'à des dates antérieures. Dans ce cas, le krigeage spatio-temporel est comparé au krigeage spatial utilisant uniquement les données à la date d'estimation (Section 5.4).
3. la localisation de la source de contamination n'est pas connue précisément. Dans ce cas, une estimation à des dates antérieures aux premières observations disponibles est recherchée par krigeage spatio-temporel, dans le but de localiser plus précisément la source de contamination (Section 5.5).

5.1.2 Référence et scénarios d'échantillonnage dans l'espace

La référence utilisée est la même que celle du Chapitre 3 (Figure 5.1). Une référence supplémentaire est également utilisée (la référence stationnaire construite dans l'Annexe D) ; les résultats associés sont donnés en annexe (Annexe F).

Les méthodes de filtres sont coûteuses en calculs, notamment car il est nécessaire d'interrompre les simulations à chaque date où des observations doivent être assimilées, pour corriger les particules. Afin de limiter les temps de calcul, seuls les scénarios d'échantillonnage S1 (sept sondages traversant l'intégralité de la zone non saturée) et S2 (quatre sondages traversant l'intégralité de la zone non saturée) sont maintenus.

5.1.3 Ensemble de particules

Comme au Chapitre 3, les particules diffèrent par les champs de paramètres de Mualem-van Genuchten utilisés en entrée du code MELODIE. La Figure 5.2 présente les trajectoires des centres de gravité de quelques particules et de la référence. Cette figure souligne à nouveau la variabilité (en terme de forme et de longueur des trajectoires) existante au sein de l'ensemble de particules, même si la seule incertitude de modélisation est liée à la variabilité spatiale des paramètres de Mualem-van Genuchten.

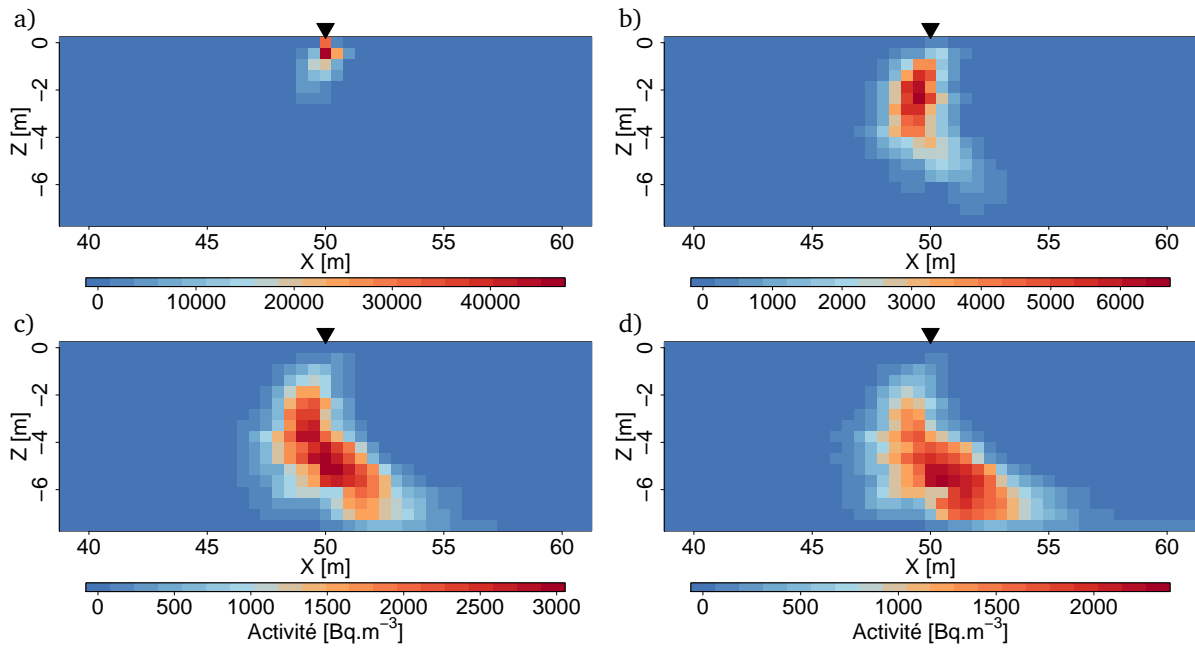


FIGURE 5.1 – Référence à plusieurs dates. a) Un an, b) deux ans, c) trois ans et d) quatre ans après le déversement de tritium.

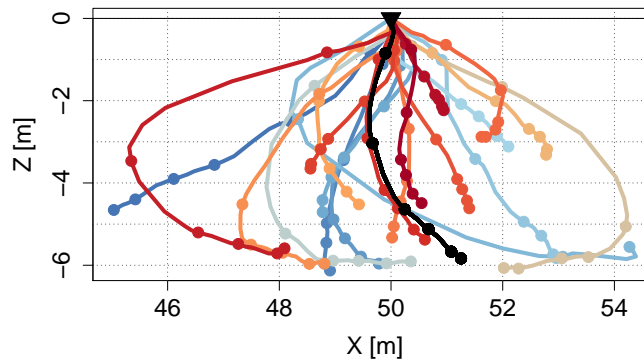


FIGURE 5.2 – Trajectoires des centres de gravité de vingt particules (en couleur) et de la référence (en noir). Chaque point indique la position des centres de gravité un an, deux ans, trois ans, quatre ans, cinq ans et six ans après le déversement de tritium.

5.2 Mise en œuvre des méthodes spatio-temporelles

Trois méthodes ont été décrites au Chapitre 1. Elles sont détaillées ici pour l'application au cas synthétique étudié.

5.2.1 Scénario d'échantillonnage temporel pour les méthodes de prévision

Pour mettre en œuvre les méthodes de prévision, un scénario d'échantillonnage temporel est construit (Figure 5.3). Les observations d'activité sont disponibles trois ans (date notée t_1) et trois ans et deux mois (date notée t_2) après le déversement de tritium. Des prévisions sont recherchées à deux dates :

1. deux mois après les dernières observations (date notée t_3 , avec $t_3 = t_2 + 2$ mois) ;
2. un an après (date notée t_4 , avec $t_4 = t_3 + 1$ an).

Les conditions aux limites (charges et gradients hydrauliques, flux d'eau entrant dans le système par la surface) sont connues jusqu'à t_1 . Quelques tests ont montré que de faibles modifications du gradient hydraulique ou des charges imposées avaient peu d'influence sur la migration du panache dans la zone non saturée. Ainsi, seul le flux correspondant à l'entrée d'eau par le haut du système (excédant hydrique) est

considéré comme inconnu après t_1 . Une chronique de ce flux est disponible pour une station située près de Poitiers pendant 55 ans à un pas journalier. Afin de construire des scénarios de conditions aux limites différents, des morceaux de chronique d'excédant hydrique sont tirés à des dates aléatoires (mais correspondant à la même période de l'année à partir de t_1).

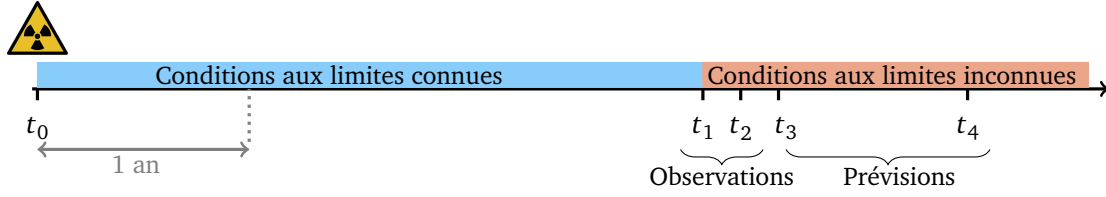


FIGURE 5.3 – Schéma du scénario d'échantillonnage temporel.

5.2.2 Filtre particulaire

L'intérêt des méthodes particulières réside dans la possibilité d'assimiler des données au fur et à mesure de leur obtention. Ainsi, pour réaliser les prévisions, on considère dans un premier temps que l'on a uniquement les données à t_1 , puis que l'on assimile les données à t_2 , de la façon suivante :

1. à partir de l'ensemble $X_{t_1, m=1 \dots M}^{(m)}$ (ensemble de particules utilisé au Chapitre 4) et des données à t_1 , des poids $\omega_{t_1}^{(m)}$ sont calculés (Complément 5.1).
2. l'ensemble $X_{t_1}^{(m)}$ est ré-échantillonné suivant les poids $\omega_{t_1}^{(m)}$ pour obtenir 100 particules équiprobables formant l'ensemble $\tilde{X}_{t_1, i=1 \dots 100}^{(i)}$. Au sein de cet ensemble, certaines particules peuvent être identiques.
3. l'ensemble $\tilde{X}_{t_1}^{(i)}$ est propagé en utilisant le code MELODIE, avec des conditions aux limites variables. On obtient l'ensemble $X_{t_2}^{(i)}$, ainsi que les ensembles $X_{t_3}^{(i)}$ et $X_{t_4}^{(i)}$.
4. avec les données à t_2 , des poids $\omega_{t_2}^{(i)}$ sont calculés. L'ensemble $X_{t_2}^{(i)}$ est ré-échantillonné pour obtenir l'ensemble $\tilde{X}_{t_2}^{(i)}$, équiprobable.
5. ce nouvel ensemble est propagé avec le code MELODIE, pour obtenir les ensembles à t_3 et t_4 .

En principe, les prévisions à t_3 et t_4 issues de l'étape (5) sont de meilleure qualité que celles issues de l'étape (3), puisqu'une étape supplémentaire d'assimilation de données a été réalisée.

5.2.3 Filtre de Kalman d'ensemble

La démarche pour réaliser des prévisions par filtre de Kalman d'ensemble est la suivante :

1. un ensemble initial $X_{t_1}^{(i)}$ composé de 100 particules (tirées parmi les 500 particules présentant les poids $\omega_{t_1}^{(i)}$ les plus élevés) est conditionné par krigeage simple aux données à t_1 pour obtenir un ensemble $\tilde{X}_{t_1}^{(i)}$ (en considérant la même variance d'erreur de mesure que celle utilisée pour calculer les poids $\omega_{t_1}^{(i)}$).
2. cet ensemble est propagé avec MELODIE, pour obtenir l'ensemble $X_{t_2}^{(i)}$ et les ensembles $X_{t_3}^{(i)}$ et $X_{t_4}^{(i)}$.
3. avec les données à t_2 , l'ensemble $X_{t_2}^{(i)}$ est corrigé pour obtenir l'ensemble $\tilde{X}_{t_2}^{(i)}$.
4. cet ensemble est ensuite propagé pour obtenir les ensembles à t_3 et t_4 .

En principe, les prévisions issues de l'étape (4) sont de meilleures qualité que celles issues de l'étape (2), car des informations supplémentaires sont prises en compte entre les deux étapes.

Pour rappel, la principale différence entre le filtre particulaire et le filtre de Kalman d'ensemble réside dans l'étape de correction : dans le filtre particulaire, les particules formant l'ensemble corrigé sont sélectionnées, et non modifiées. Au contraire, dans le filtre de Kalman d'ensemble, on travaille toujours avec les mêmes particules, mais elles sont conditionnées par krigeage à chaque étape.

Calcul des poids et ré-échantillonnage

Usuellement, les poids associés aux particules dans les méthodes de filtres particulières sont calculés par vraisemblance (Équation (1.30)). Cependant, dans le cas d'une contamination radiologique, les valeurs d'activité mises en jeu diffèrent de plusieurs ordres de grandeur, ce qui conduit à des poids ainsi calculés très mal répartis et à une dégénérescence du système dès la première itération. De plus, lorsque l'on calcule des poids de cette manière, on privilégie très fortement les particules proches des observations en terme de valeurs d'activité (aux erreurs de mesure près). Pourtant, toute particule se trouvant dans la zone proche des observations non nulles est intéressante, quelles que soient les valeurs d'activité. Ainsi, la distance entre observations est remplacée par une distance entre indicatrices de présence ou d'absence de contamination :

$$\omega_t^{(m)} = A \exp \left[-\frac{1}{2s} \|\mathbb{1}_{y_t > y_0} - \mathbb{1}_{\mathbf{B}_t p_t^{(m)} > y_0}\|^2 \right], \quad \text{avec } y_0 > 0 \quad (5.1)$$

où A est un coefficient permettant de s'assurer que la somme des poids vaut 1, s est la variance d'erreur de mesure (choisie égale à $100 \text{ Bq}^2 \cdot \text{m}^{-6}$ ici), $p_t^{(m)}$ est la particule m à la date t , y_t sont les observations à la date t , \mathbf{B}_t est l'opérateur permettant de passer de la particule aux observations sur cette particule et y_0 est un seuil déterminant la présence ou l'absence de contamination.

Le ré-échantillonnage consiste à obtenir N particules équiprobables à partir d'un ensemble initial de M particules (avec $M \geq N$) : soit N_m le nombre de fois où la particule m est tirée, avec $\sum_{m=1}^M N_m = N$. Les N_m sont tirés suivant une loi multinomiale de paramètre N et de probabilités $\{\omega^{(m)}, m = 1 \dots M\}$. On obtient ainsi N nouvelles particules, chacune de poids $1/N$.

5.2.4 Krigeage spatio-temporel avec covariances numériques

Les covariances numériques sont préférées aux variogrammes numériques, car elles sont ici plus stables numériquement et conduisent à des erreurs de prévision plus faibles. Elles sont calculées à partir de 500 particules (dont celles utilisées pour les filtres). L'excédant hydrique est supposé inconnu à partir de t_1 , comme pour les méthodes de filtre. La même variance d'erreur de mesure est prise en compte. Les conditions d'universalité considérées ici sont les plus simples (Équation (1.24)), *i.e.* elles ne changent pas selon la date. Ce point est discuté dans la Section 5.4.

Les résultats de prévision obtenus avec des covariances spatio-temporelles numériques ne sont pas comparés aux résultats obtenus avec des modèles de covariances spatio-temporelles usuels. En effet, le phénomène étudié ne remplit pas les conditions pour utiliser des modèles séparables ou symétriques, et il faudrait travailler avec des modèles de covariances paramétriques plus complexes.

5.3 Prévisions

Les trois méthodes (filtre particulière, filtre de Kalman d'ensemble et krigeage spatio-temporel) sont mises en œuvre pour réaliser des prévisions aux dates t_3 et t_4 , pour les scénarios d'échantillonnage S1 et S2. Les ensembles de particules utilisés pour S1 et S2 ne sont pas les mêmes, puisque les données assimilées sont différentes. Seuls les résultats détaillés pour le scénario S2 sont donnés ici. Les résultats pour le scénario S1 et pour la référence supplémentaire sont présentés dans l'Annexe F.

Pour les méthodes de filtres, les résultats sont comparés avant et après l'étape d'assimilation des données à t_2 , et les moyennes empiriques des ensembles sont utilisées. Ce dernier point est discuté dans la Section 5.3.3.

5.3.1 Indicateurs de performance

Les trois méthodes sont comparées qualitativement par les cartes de prévision et d'erreur de prévision. Plusieurs indicateurs quantitatifs sont aussi calculés :

- l'erreur absolue moyenne (EAM) sur le domaine ;
- la racine de l'erreur quadratique moyenne (REQM) sur le domaine ;
- le rapport entre la somme de l'activité sur la prévision et la somme de l'activité sur la référence (S), ce qui permet de quantifier la sous-estimation ($S < 1$) ou la surestimation ($S > 1$) globale de la prévision.

5.3.2 Résultats

Filtre particulaire

Les cartes de prévision et d'erreur de prévision sont présentées sur la Figure 5.4, après l'étape d'assimilation des données à t_2 . Les contours du panache sont lissés. Les erreurs moyennes ont tendance à diminuer au cours du temps (Tableaux 5.1a et 5.1d), ce qui est dû à l'atténuation des valeurs d'activité, et non au fait que les prévisions à t_4 sont de meilleure qualité qu'à t_3 . L'activité totale est légèrement surestimée ($S > 1$).

Filtre de Kalman d'ensemble

Les cartes de prévision et d'erreur de prévision sont présentées sur la Figure 5.5, après l'étape d'assimilation des données à t_2 . Le panache obtenu à t_4 est plus lissé qu'à t_3 . Les erreurs moyennes diminuent au cours du temps (Tableaux 5.1b et 5.1e). L'activité totale est surestimée ($S > 1$).

Krigeage spatio-temporel

Les cartes de prévision et d'erreur de prévision sont présentées sur la Figure 5.6. La localisation globale du panache est bien prédite, même si le contour du panache est lissé. Les erreurs moyennes augmentent de t_3 à t_4 (Tableaux 5.1c et 5.1f). L'activité globale est légèrement sous-estimée ($S < 1$) et cette sous-estimation est plus marquée à t_4 qu'à t_3 .

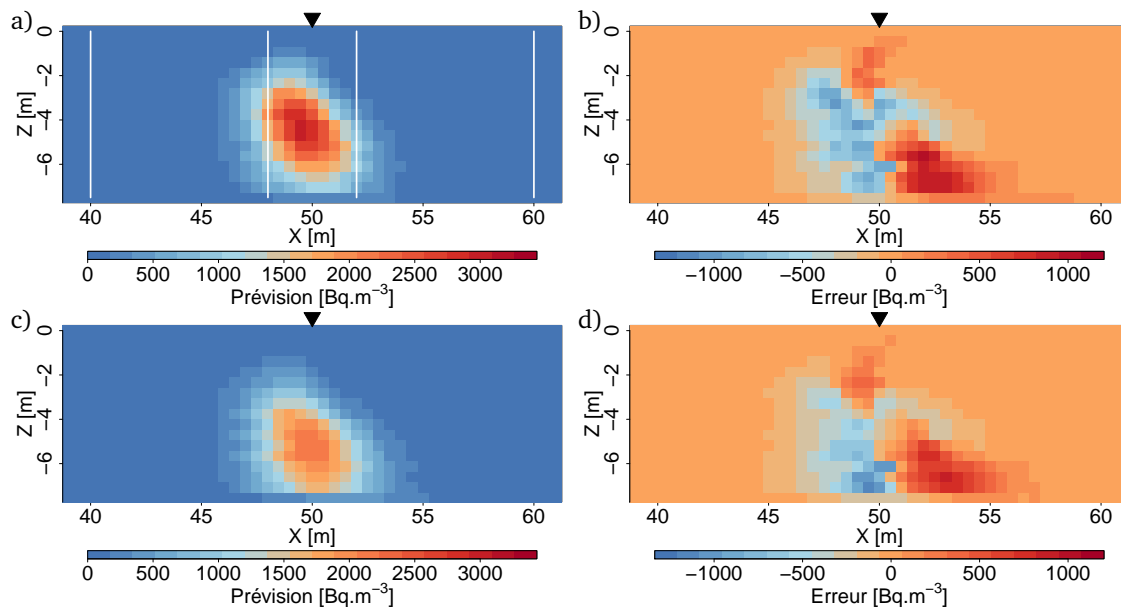


FIGURE 5.4 – Prévisions et erreurs de prévision par filtre particulaire pour le scénario S2 et pour les dates t_3 (a et b) et t_4 (c et d).

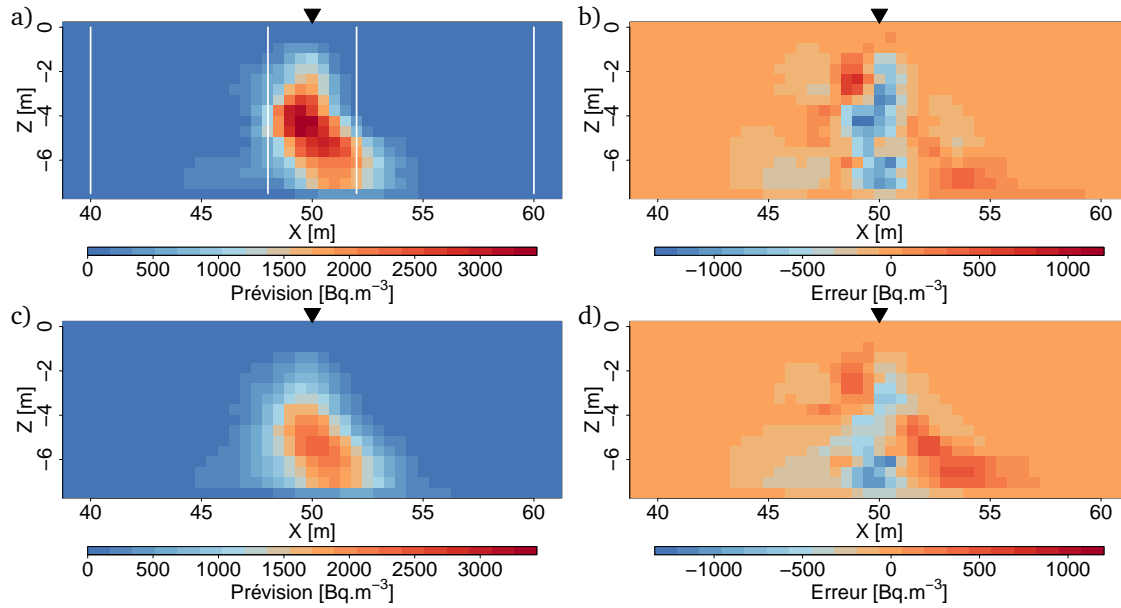


FIGURE 5.5 – Prévisions et erreurs de prévision par filtre de Kalman d'ensemble pour les dates t_3 (a et b) et t_4 (c et d).

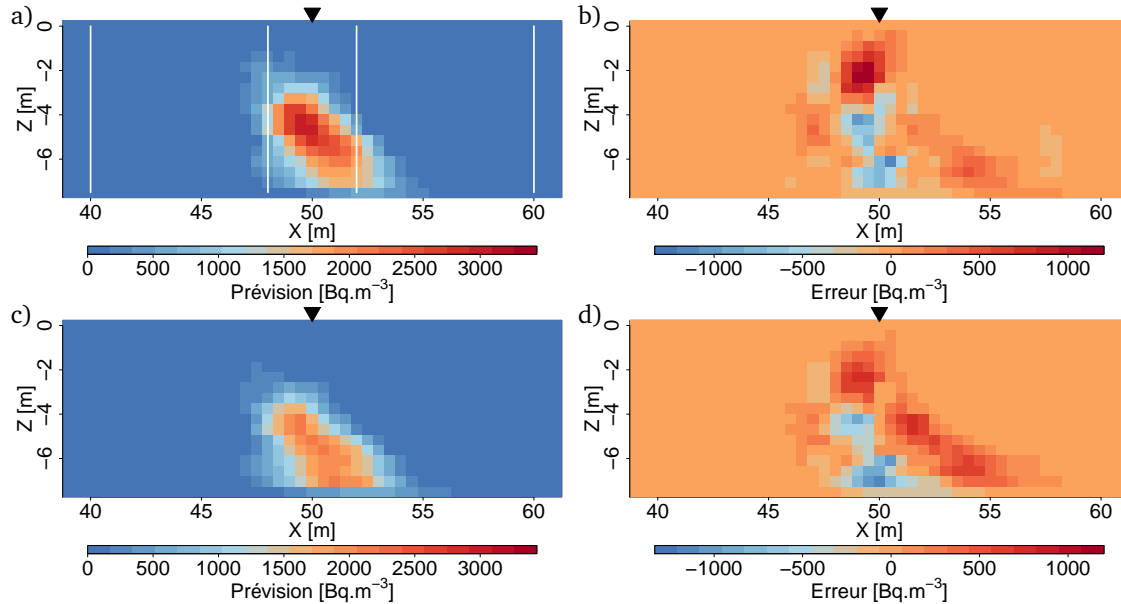


FIGURE 5.6 – Prévisions et erreurs de prévision par krigeage spatio-temporel pour les dates t_3 (a et b) et t_4 (c et d).

5.3.3 Comparaison des méthodes et discussion

Comparaison des méthodes

Les erreurs obtenues avec les trois méthodes sont comparées dans le Tableau 5.1. Les erreurs moyennes sont légèrement plus faibles pour le krigeage spatio-temporel que pour le filtre de Kalman d'ensemble, et supérieures pour le filtre particulière. Lorsque le nombre de données augmente (scénario S1, Annexe F), les écarts entre les méthodes sont moins marqués.

Les panaches obtenus par krigeage et par filtre de Kalman d'ensemble sont moins lissés que ceux obtenus par filtre particulière. En revanche, les prévisions obtenues par krigeage sous-estiment la contamination totale ($S < 1$), ce qui n'est pas le cas pour celles obtenues par filtres. Ainsi, même si les erreurs de prévision sont en moyenne inférieures pour le krigeage, les filtres permettent de mieux reproduire le niveau global d'activité, ce qui est intéressant en terme de risque. Pour le scénario S1 (Annexe F), le krigeage surestime

TABLEAU 5.1 – Erreurs obtenues aux dates t_3 et t_4 par filtre particulaire (FP) et par filtre de Kalman d'ensemble (EnKF) avant et après l'assimilation des données à t_2 , ainsi que par krigeage.

(a) FP à t_3			(b) EnKF à t_3		(c) Krigeage à t_3
	Avant	Après	Avant	Après	
EAM [Bq.m ⁻³]	131	121	90	90	69
REQM [Bq.m ⁻³]	257	250	199	205	181
S [-]	1.15	1.19	1.27	1.25	0.97

(d) FP à t_4			(e) EnKF à t_4		(f) Krigeage à t_4
	Avant	Après	Avant	Après	
EAM [Bq.m ⁻³]	113	102	84	81	78
REQM [Bq.m ⁻³]	226	214	170	171	189
S [-]	1.12	1.17	1.24	1.19	0.91

très légèrement la contamination globale, comme les autres méthodes.

Pour la référence supplémentaire (Annexe F), le filtre de Kalman d'ensemble conduit à des erreurs plus faibles que les autres méthodes, mais à une sous-estimation plus importante de la contamination globale dans le cas du scénario S1.

En résumé, toutes références et tous scénarios considérés, le krigeage spatio-temporel et le filtre de Kalman d'ensemble sont plus performants que le filtre particulaire.

Intérêt de l'assimilation de données

Pour les filtres, les indicateurs quantitatifs ont été calculés avant et après l'étape d'assimilation des données à t_2 (Tableau 5.1) : comme attendu, l'étape d'assimilation de donnée réduit les erreurs moyennes, puisque le modèle est ajusté au regard de nouvelles observations. Ainsi, bien qu'assez coûteuse en temps de calcul, l'assimilation de données est utile pour améliorer les prévisions. Cette étape semble d'autant plus intéressante que le nombre de données est faible : pour le scénario S1 (Annexe F), les erreurs ne diminuent pas autant après l'assimilation des données à t_2 .

Prévision par une autre grandeur que la moyenne empirique

Précédemment, les prévisions sont données par la moyenne empirique des ensembles de particules. D'autres grandeurs peuvent être utilisées, comme la médiane, ou encore le panache central de l'ensemble (Chapitre 4, Section 4.2.1). Pour le filtre particulaire, les performances des prévisions par la médiane à la place de la moyenne dépendent de l'indicateur d'erreur considéré. En revanche, la prévision par la médiane améliore les résultats pour le filtre de Kalman d'ensemble. La prévision par le panache central de l'ensemble obtenu par filtre particulaire dégrade les prévisions, mais celui de l'ensemble obtenu par filtre de Kalman d'ensemble améliore les prévisions.

5.4 Krigeage spatial ou krigeage spatio-temporel ?

Le krigeage spatio-temporel permet également de faire des estimations en utilisant des observations à plusieurs dates, alors que le krigeage spatial ne prend en compte que les observations à la date de l'estimation. Ainsi, il est intéressant de comparer les estimations obtenues par krigeage spatial avec variogrammes numériques (utilisation des observations à une date) et par krigeage spatio-temporel avec covariances numériques (utilisation d'une série temporelle d'observations).

5.4.1 Scénario

L'objectif est d'estimer l'activité en tritium à la date notée précédemment t_4 , soit 4 ans et 4 mois après le déversement de tritium. La localisation de la source, la date du déversement et les conditions aux limites

sont supposées parfaitement connues et les mêmes particules (celles du Chapitre 4) sont utilisées pour toutes les estimations. Plusieurs cas de figure sont comparés, en terme de disponibilité des données, et pour les scénarios d'échantillonnage S1 et S2 :

1. les données d'activité ne sont disponibles qu'à la date t_4 (estimation par krigeage avec variogrammes numériques) ;
2. les données d'activité sont disponibles tous les deux mois depuis un an et demi avant t_4 (Figure 5.7).

Dans ce cas, les estimations sont réalisées par krigeage spatio-temporel avec covariances numériques. Par ailleurs, diverses variantes des conditions de non biais sont mises en œuvre pour l'estimation par krigeage spatio-temporel. Ces variantes ont été données dans les Équations (1.24), (1.25) et (1.26). Pour rappel, elles reviennent à considérer des moyennes différentes, ou non, selon les dates. Elles sont notées par la suite C1, C2 et C3.

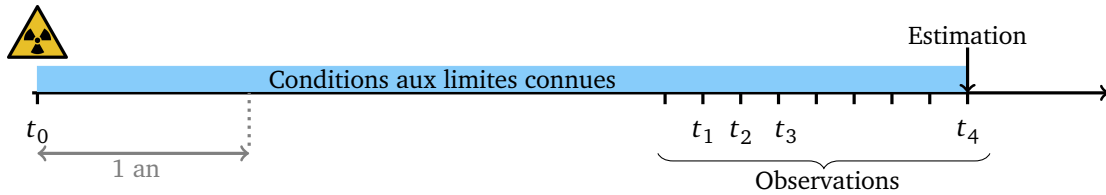


FIGURE 5.7 – Schéma du scénario d'échantillonnage temporel.

5.4.2 Résultats

Les erreurs absolues moyennes sont plus faibles pour le krigeage avec variogrammes numériques (spatial seulement) que pour les estimations spatio-temporelles (Figure 5.8), sauf lorsque l'on ne considère que quelques dates d'observation supplémentaires (quatre pour le scénario S1, une pour le scénario S2). Ainsi, avoir des observations à des dates antérieures ne diminue pas (ou très peu) les erreurs d'estimation, et leur prise en compte nécessite la mise en place d'une méthode d'estimation plus complexe, notamment en terme de construction du système de krigeage et de temps de calcul.

Par ailleurs, les résultats obtenus avec les différentes variantes de conditions de non biais sont presque identiques (Figure 5.8).

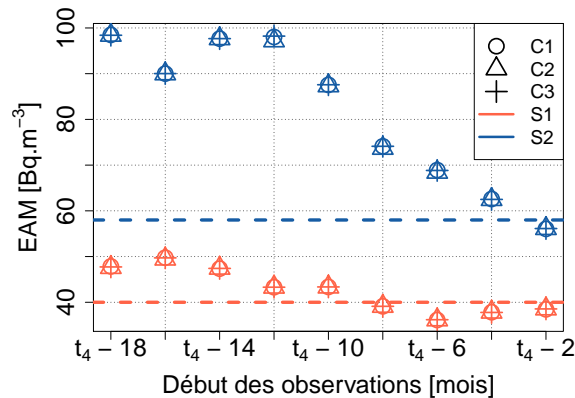


FIGURE 5.8 – Erreur absolue moyenne en fonction de la date de début des observations (entre un an et demi avant t_4 et deux mois avant t_4), pour les scénarios d'échantillonnage S1 et S2. Les lignes en pointillés donnent les erreurs absolues moyennes obtenues par krigeage spatial. Les symboles représentent les résultats obtenus avec diverses conditions de non biais (C1, C2 ou C3).

5.5 Recherche de la localisation de la source

Le krigeage spatio-temporel peut être utilisé pour déterminer la localisation de la source de contamination lorsque celle-ci est mal connue. Les covariances spatio-temporelles numériques sont particulièrement intéressantes, puisqu'elles ne sont pas symétriques par rapport au temps, et permettent donc de reproduire des phénomènes non réversibles. Cependant, pour mettre en œuvre cette méthode, il aurait fallu générer plusieurs centaines de nouvelles simulations, ce qui a été exclu au vu du temps de calcul que cela demande (ce temps de calcul a été utilisé pour les applications sur cas réels de la Partie III). Ainsi, seuls quelques résultats préliminaires sont présentés ici.

5.5.1 Scénario et jeu de simulations

La localisation de la source de contamination n'est pas connue précisément, contrairement aux scénarios précédents. On suppose qu'elle est située quelque part sur un segment de 10 m de large, entre $x=47$ m et $x=57$ m (ce segment est volontairement non symétrique autour de la localisation correcte de la source, en $x=50$ m). Les mesures d'activité en tritium sont disponibles quatre ans après le déversement. L'activité en tritium est alors estimée à des dates plus proches du début de la contamination, dans le but de localiser plus précisément la source de contamination (Figure 5.9).

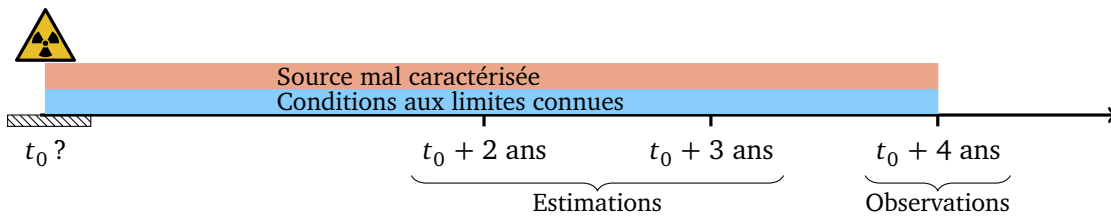


FIGURE 5.9 – Schéma du scénario de recherche de la source.

Afin de tester l'approche sans générer de nouvelles simulations, la position du terme source est rendue inconnue par une translation horizontale d'une distance aléatoire (entre 3 m à gauche et 7 m à droite) des panaches déjà simulés. À noter qu'il s'agit d'une approche approximative, puisque les simulations de panaches ont été obtenues à partir de champs de paramètres de Mualem-van Genuchten conditionnés par des sondages à des distances fixes de la source. En translatant les panaches simulés, les sondages utilisés pour ce conditionnement sont également décalés, et l'on ne peut plus considérer que les champs de paramètres de Mualem-van Genuchten sont conditionnés.

De même, pour introduire une incertitude sur la date du déversement, les panaches simulés sont décalés d'une ou deux dates de sortie, antérieures ou postérieures (ce qui correspond à une incertitude de quatre mois). Ensuite, ce jeu de simulations modifiées est utilisé pour calculer les covariances numériques spatio-temporelles utilisées pour le krigeage.

5.5.2 Résultats

Les résultats sont donnés pour le scénario d'échantillonnage spatial S1 (sept sondages), mais les conclusions restent valables pour le scénario S2 (quatre sondages). L'estimation obtenue trois ans après le déversement (Figure 5.10b), *i.e.* un an avant les observations, a une forme proche de celle de la référence (Figure 5.10a), bien que les valeurs d'activités soient sous-estimées et les erreurs assez fortes (Figure 5.10c). La localisation du panache est également satisfaisante. La carte d'écart-type théorique de krigeage (Figure 5.10d) souligne l'étalement des localisations des sources de tritium sur un segment au sein du jeu de simulations.

Lorsque l'on se rapproche de la date du déversement (Figure 5.11), les erreurs augmentent très fortement, mais la localisation et la forme du panache restent cohérentes. Les fortes erreurs sont en partie liées à la concentration des valeurs d'activité autour de la source.

Ainsi, si l'objectif est de localiser la source de contamination, la méthode donne des résultats corrects. Cependant, les données utilisées pour obtenir les résultats de la Figure 5.10 présentent des fortes valeurs

aux sondages situés en $x=48$ m, $x=50$ m et $x=52$ m, avec un échantillonnage symétrique à droite et à gauche de la vraie source. Ainsi, la qualité des résultats obtenus pourrait autant être liée à cet échantillonnage qu'à la méthode employée : il semble difficilement concevable qu'avec des données fortes entre $x=48$ m et $x=52$ m, les estimations indiquent une source localisée hors de cet intervalle.

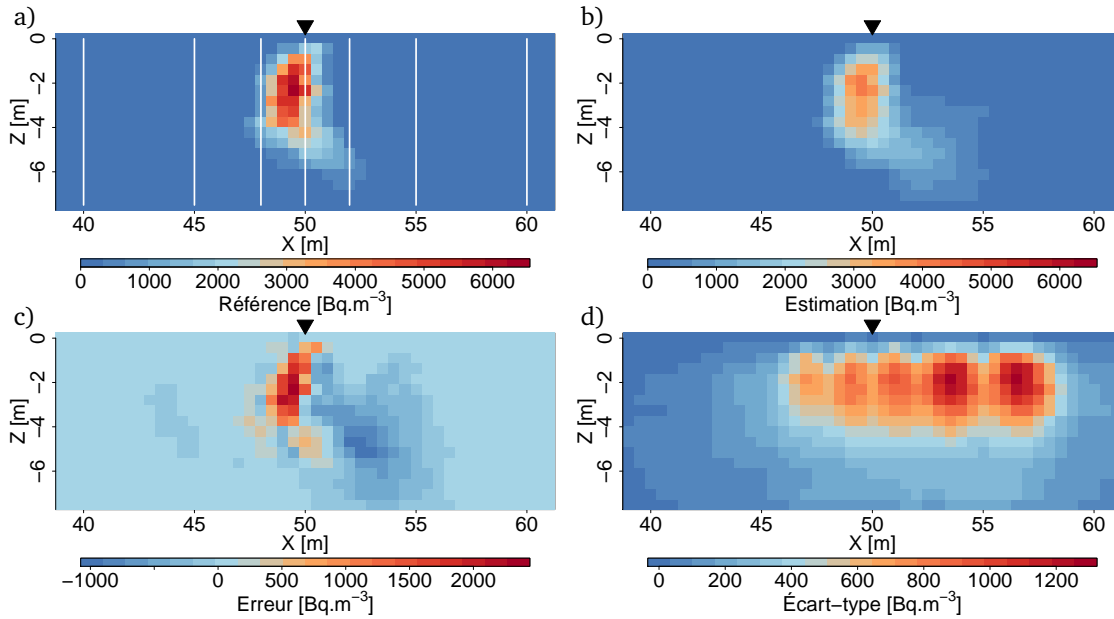


FIGURE 5.10 – Résultats d'estimation trois ans après le déversement. a) Référence et position des sondages ; b) estimation ; c) erreur d'estimation ; d) écart-type théorique. L'échelle de couleur est la même pour les cartes a) et b). La position des sondages est rappelée sur la carte a), mais ces données ne sont pas utilisées pour l'estimation : on utilise les données à cette position, mais quatre ans après le déversement.

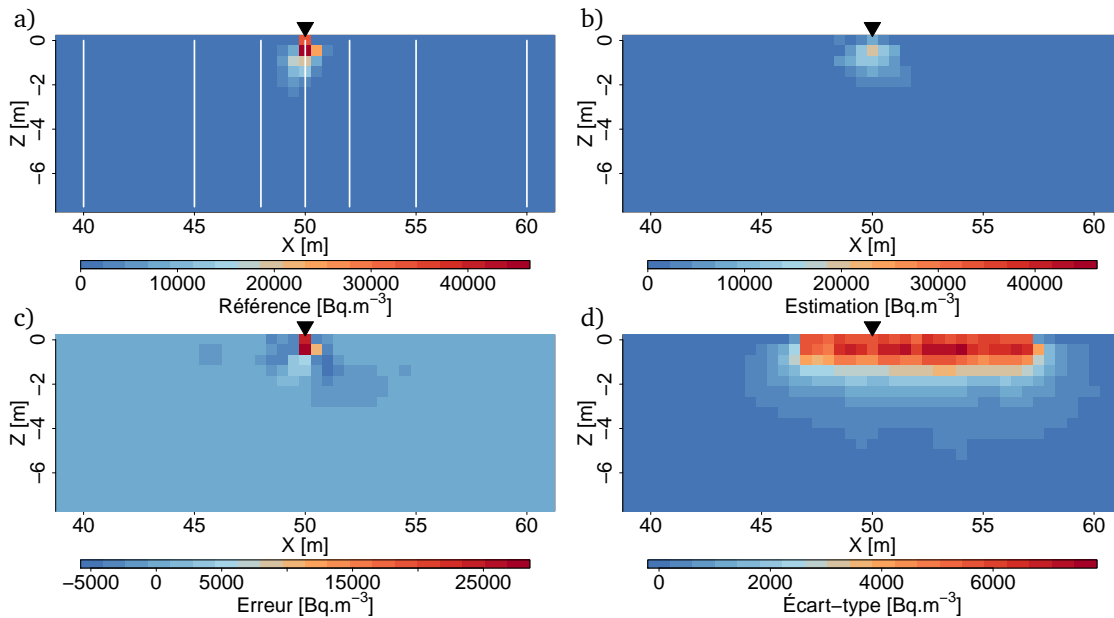


FIGURE 5.11 – Résultats d'estimation deux ans après le déversement. a) Référence et positions des sondages ; b) estimation ; c) erreur d'estimation ; d) écart-type théorique.

Pour discuter ce dernier point et tester la robustesse de la méthode, les estimations ont été effectuées sur une autre référence (une simulation du jeu). Cette référence est également issue d'une source de tritium située en $x=50$ m, mais les données présentent des valeurs fortes autour de $x=45$ m (Figure 5.12a). Une estimation par krigeage est réalisée trois ans après le déversement, avec les mêmes simulations modifiées que précédemment. À nouveau, les erreurs sont très élevées (Figure 5.12c), mais la localisation globale de

l'estimation (Figure 5.12b) est cohérente : le panache estimé se rapproche de la localisation de la source.

Finalement, le krigeage spatio-temporel permet de localiser approximativement la source, ou au moins de restreindre la zone où celle-ci peut se trouver. Cependant, dans la zone non saturée, les panaches ne migrent que de quelques mètres, même pour les plus rapides. Ainsi, dans la plupart des cas, les données sont proches de l'aplomb de la source. Cette approche aurait probablement plus d'intérêt pour des panaches ayant migré beaucoup plus, en considérant par exemple la zone saturée.

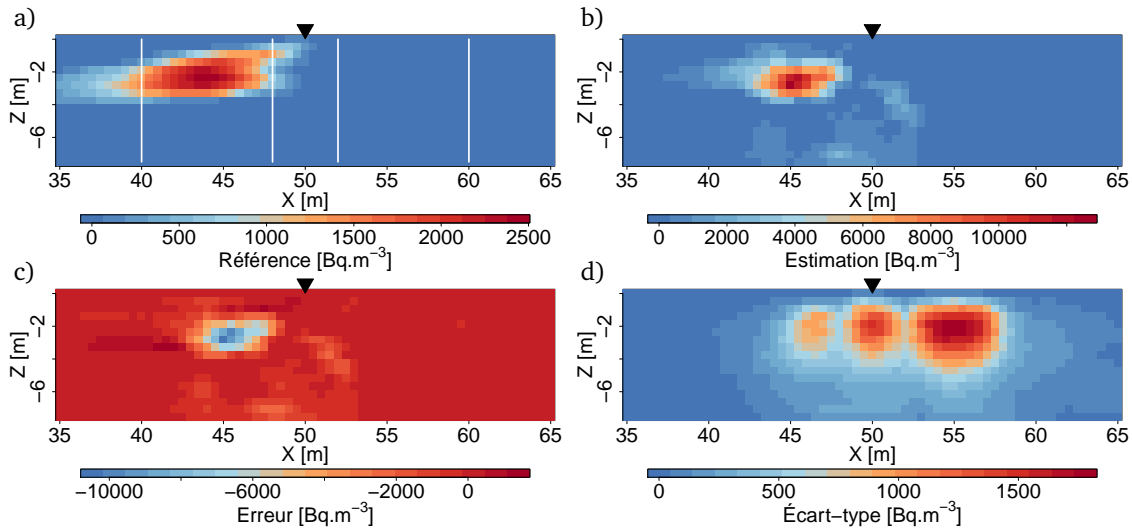


FIGURE 5.12 – Résultats d'estimation trois ans après le déversement pour une autre référence. a) Référence et positions des sondages ; b) estimation ; c) erreur d'estimation ; d) écart-type théorique.

5.6 Synthèse du chapitre

Dans ce chapitre, les méthodes combinant les simulations d'écoulement et de transport et la géostatistique, notamment le krigeage avec covariances numériques, ont été étendues à un contexte spatio-temporel. Les performances du krigeage spatio-temporel avec covariances numériques sont comparables aux performances de méthodes classiques d'assimilation de données, notamment le filtre de Kalman d'ensemble. Ces méthodes permettent de prédire convenablement la localisation et l'intensité d'un panache de tritium, à plus ou moins long terme (deux mois ou un an ici).

La mise en place de ces méthodes est plus complexe que les méthodes exclusivement spatiales. En effet, la taille des systèmes de krigeage augmente très rapidement lorsque l'on ajoute des dates d'observations. Par ailleurs, la mise en place des méthodes de filtres requiert la possibilité d'interrompre les simulations, de corriger les sorties de ces simulations (étapes de correction des filtres), puis de reprendre les simulations à partir de ces particules modifiées.

Enfin, les résultats préliminaires de la Section 5.5 montrent que le krigeage spatio-temporel avec covariances numériques permet de restreindre la zone où peut se trouver la source d'une contamination, lorsqu'elle n'est pas connue précisément. Cette méthode aurait pu être comparée aux méthodes de lissage particulière, mais cela aurait également nécessité la génération de nouvelles simulations.

Troisième partie

Applications à deux jeux de données réels acquis autour de la centrale de Tchernobyl

Présentation des sites

Les données exploitées aux Chapitres 6 et 7 ont été acquises au voisinage de la centrale de Tchernobyl. Le jeu de données du Chapitre 6 est issu d'une expérience de traçage au chlore 36 (^{36}Cl) ayant pour but l'étude du transport de radionucléides dans la zone non saturée d'un sol sableux typique de l'environnement de la centrale. Les données du Chapitre 7 sont des mesures d'activité en strontium 90 (^{90}Sr) dans une nappe, à proximité d'une zone d'enfouissement de déchets radioactifs. Cette présentation succincte reprend les principaux éléments de contexte communs aux deux jeux de données.

L'accident de la centrale nucléaire de Tchernobyl

L'accident de la centrale nucléaire de Tchernobyl a eu lieu le 26 avril 1986 en Ukraine, à une centaine de kilomètres de Kiev. L'explosion accidentelle du réacteur 4 (Figure 1), lors d'un essai technique, a entraîné l'émission brutale de produits radioactifs dans l'atmosphère (environ 12 milliards de Becquerels, [Site de l'IRSN (d)]). La zone d'exclusion, la plus contaminée par les dépôts de particules de combustible, couvre un rayon de 30 km autour de la centrale [Site de l'IRSN (e)]. Aujourd'hui, la reconquête de cette zone est évoquée, mais cela passe par une évaluation précise du risque d'exposition à des matières radioactives. Cette évaluation concerne principalement des sites de stockage, car environ 90% des déchets de décontamination ont été stockés au sein de cette zone [Site de l'IRSN (f)]. Ces déchets, représentant plus d'un million de mètres cubes, sont entreposés dans des casemates en béton ou dans des tranchées.



FIGURE 1 – Image satellite de la zone entourant le réacteur 4 (image Google Earth - 2020).

Le stockage immédiat des déchets de faible à moyenne activité

Les déchets de faible à moyenne activité (entre 1 et 100 kBq.g⁻¹), constitués principalement de débris végétaux et de matériaux de construction, sont stockés dans des tranchées [[Site de l'IRSN \(e\)](#)]. Environ 800 tranchées ont été creusées entre 1986 et 1987 dans une formation sableuse [[IRSN, 2016](#)]. Elles mesurent quelques mètres de large, quelques centaines de mètres de long, sont profondes de 2 à 4 m et se situent dans un rayon de 10 km autour de la centrale. Ces tranchées ne sont pas équipées de structures étanches et ne constituent donc pas une barrière efficace pour empêcher la migration des contaminants, surtout que la base des tranchées est régulièrement inondée par la remontée du niveau de la nappe [[Kashparov et al., 2012](#)]. Par ailleurs, peu de description des tranchées (localisation précise, dimensions précises, inventaire, etc.) ont été fournies au moment de leur creusement [[Kashparov et al., 2004](#)], ce qui rend d'autant plus difficile la quantification des risques environnementaux associés à ces ouvrages.

L'expérience de traçage

Une expérience de traçage au ³⁶Cl a été réalisée en octobre 2000 au bord de la rivière Prypiat (Figure 1). Le site est une zone sableuse n'ayant pas été remaniée suite à l'accident de 1986. L'objectif de cette expérience est d'obtenir la distribution d'un traceur dans le sol et d'améliorer les connaissances sur les écoulements et le transport à l'échelle de l'expérience. Les données de cette expérience sont exploitées au Chapitre 6.

Le projet Epic et la tranchée T22

Pour évaluer l'impact des tranchées sur l'environnement, l'IRSN, en coopération avec deux instituts ukrainiens, a équipé et entretient depuis 1999 un site expérimental autour de l'une de ces tranchées, la T22 [[Site de l'IRSN \(g\)](#)]. Il s'agit du projet Epic (Experimental Platform in Chernobyl, Figure 1), qui vise à étudier les mécanismes de transport des radionucléides dans les sols et les nappes, et à développer des modèles prédictifs de la migration des radionucléides. Le site est ainsi équipé de nombreux capteurs, dont un réseau d'une centaine de piézomètres permettant de suivre en continu depuis 2000 l'évolution du niveau de la nappe et de mesurer ponctuellement les niveaux de radioactivité [[Kashparov et al., 2012](#)]. Les activités en ⁹⁰Sr mesurées par ces piézomètres sont exploitées au Chapitre 7.

Le site de Prypiat Zaton

Le site de Prypiat Zaton (Figure 1) est un analogue géologique du site Epic. Des analyses granulométriques ont été effectuées sur quelques centaines d'échantillons de sable de cette zone. Ces données sont utilisées pour déterminer certains des paramètres des modèles d'écoulement et de transport présentés aux Chapitres 6 et 7.

Chapitre 6

Application à une expérience de traçage

Ce chapitre présente l'application des méthodes de combinaison des simulations numériques et de la géostatistique à un jeu de données d'activité en ^{36}Cl . Ce jeu de données, dense dans l'espace et pour lequel les conditions du traçage sont décrites assez précisément, permet de construire un modèle d'écoulement et de transport bien contraint puis d'appliquer et de valider les méthodes sur un cas tridimensionnel. Enfin, cette application est un cas représentatif d'une étude de sites et sols pollués, lorsque la source de contamination est bien identifiée.

6.1 Présentation de l'expérience et des données

La description de l'expérience et toutes les données sont issues de travaux réalisés par des équipes de l'IRSN et l'UIAR (Ukrainian Institute of Agricultural Radiology) [Kashparov *et al.*, 2006].

6.1.1 Expérience de traçage

L'expérience de traçage a été effectuée sur une zone de quelques mètres carrés, sur un sol plat et sans végétation. Des observations réalisées dans des fosses montrent que le sol est organisé en couches horizontales, avec des alternances de niveaux sableux et de niveaux limoneux. Le traceur utilisé est le ^{36}Cl , considéré comme un traceur parfait n'interagissant pas avec le milieu.

Le dispositif consiste en deux carrés de 3 m et 0,7 m de côté (Figure 6.1a). Le petit carré, dans lequel est injecté le ^{36}Cl , est placé au centre du grand. Une quantité d'eau correspondant à une hauteur d'eau de 3 cm est injectée dans les deux carrés, en trois périodes de dix minutes (soit 1 cm par période), pour garantir les mêmes conditions d'humidification dans le carré où le ^{36}Cl est injecté et dans son environnement immédiat. En revanche, la quantité précise de ^{36}Cl injectée est inconnue.

Vingt-deux heures après l'injection, une fosse est creusée sur un côté du petit carré, légèrement à l'extérieur de celui-ci (Figure 6.1b, plan I). Soixante-dix-sept prélèvements sont effectués sur la section verticale dégagée, suivant une grille pseudo-régulière (mailles de 15 cm ou 20 cm). Au total, sept plans verticaux sont ainsi dégagés (Figure 6.1b) et un volume d'un mètre cube est échantillonné (539 échantillons).

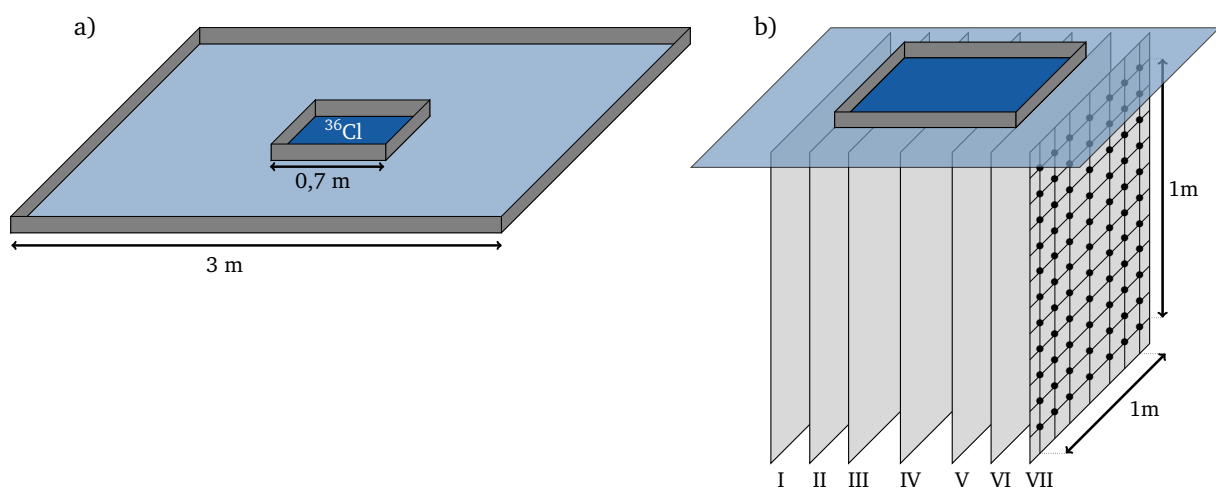


FIGURE 6.1 – a) Schéma du dispositif. b) Plan d'échantillonnage. Pour des raisons de lisibilité, la grille détaillée n'est représentée que sur le plan VII.

6.1.2 Données issues de l'expérience de traçage

Plusieurs types de données sont issues de l'expérience de traçage, dont des mesures d'activité et de surface spécifique.

Activités en ^{36}Cl

Les mesures d'activité montrent que le ^{36}Cl est peu présent en bordure du domaine de l'expérience (plans I et VII) et qu'il est irrégulièrement réparti sur les plans II à VI, sur lesquels il atteint une profondeur de 40 à 60 cm (Figure 6.2a). Sur les 539 points de prélèvement, 197 présentent une activité non nulle, allant jusqu'à quelques dizaines de milliers de Becquerels par kilogramme de sable (Figure 6.2b). La somme de ces données est égale à 2,1 MBq, et cette valeur est utilisée comme estimateur de l'activité en ^{36}Cl ayant été injectée au début de l'expérience. Cet estimateur constitue une limite basse, puisqu'une partie de ce qui a été injecté a pu sortir de la zone échantillonnée.

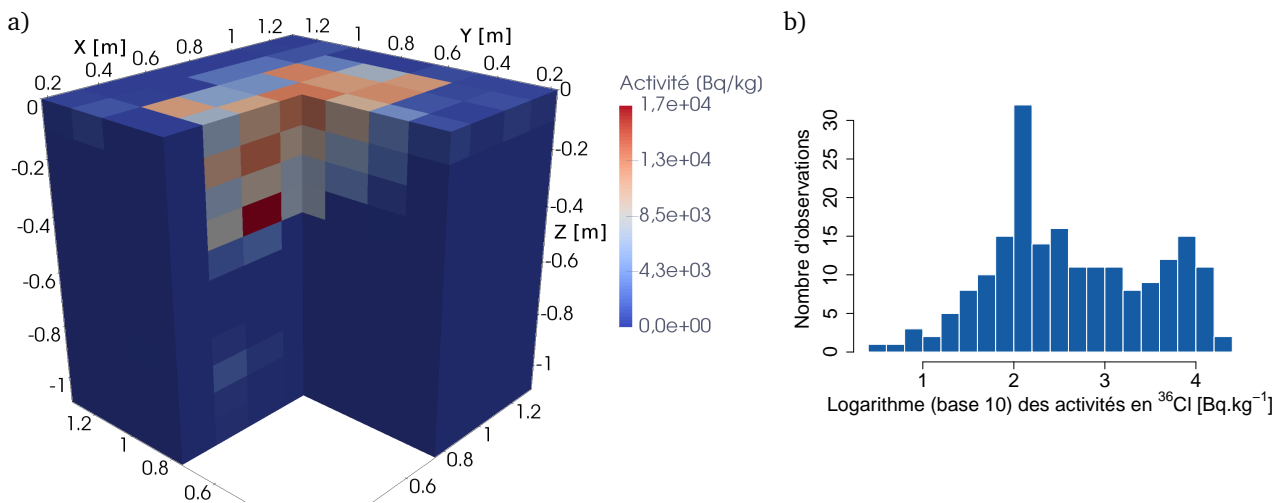


FIGURE 6.2 – a) Représentation partielle des données en trois dimensions. b) Histogramme du logarithme des données non nulles.

Surfaces spécifiques et paramètres du modèle de Mualem-van Genuchten

La surface spécifique (surface totale des particules formant l'échantillon divisé par sa masse) a été mesurée en 193 points répartis sur les plans I, II et IV. Ces données sont utilisées pour estimer les paramètres du modèle de Mualem-van Genuchten dans la zone de traçage (Complément 6.1).

Passage des surfaces spécifiques aux paramètres du modèle de Mualem-van Genuchten

Les données granulométriques de Prypiat Zaton (Annexe G) servent d'intermédiaire pour passer des surfaces spécifiques aux paramètres de Mualem-van Genuchten. Les données granulométriques (proportions de particules dont le rayon appartient à un intervalle donné) permettent de calculer la surface spécifique par la relation suivante (sous l'hypothèse de particules sphériques) :

$$S_s = \frac{3}{\rho \sum_{\text{classes } i} p_i r_i} \quad (6.1)$$

où S_s [L 2 .M $^{-1}$] est la surface spécifique de l'échantillon, ρ [M.L $^{-3}$] sa masse volumique et $\sum_{\text{classes } i} p_i r_i$ [L] la moyenne des rayons des particules de l'échantillon, pondérés par la proportion des particules de chaque classe de rayons. Par ailleurs, le passage des données granulométriques

aux paramètres de Mualem-van Genuchten s'effectue à l'aide de la fonction de pédotransfert Rosetta3. Ainsi, pour chaque échantillon de Prypiat Zaton, les surfaces spécifiques et les paramètres de Mualem-van Genuchten correspondants sont calculés. Ensuite, pour chaque échantillon du site de traçage, on recherche l'échantillon de Prypiat Zaton dont la surface spécifique est la plus proche. Ce plus proche voisin fournit une estimation des paramètres de Mualem-van Genuchten pour les échantillons du site de traçage.

Cette approche peut paraître simpliste, puisqu'elle ne prend notamment pas en compte l'agencement spatial des mesures de surfaces spécifiques. Cependant, elle permet de conserver une cohérence dans les structures spatiales exhibées d'une part par les mesures de surface spécifique, et d'autre part par les paramètres de Mualem-van Genuchten ainsi estimés (Annexe H). Par ailleurs, la texture du sol à Prypiat Zaton est proche de celle de la zone de traçage ; la méthode utilisée permet en pratique de sélectionner les données de Prypiat Zaton qui correspondent le mieux aux textures du site de traçage.

Les moyennes et variances des paramètres de Mualem-van Genuchten sont données dans le Tableau 6.1. Au regard des résultats de l'analyse de sensibilité du Chapitre 3 (Section 3.2.3), le paramètre α est celui dont la variabilité influence le plus la migration d'un panache dans la zone non saturée : des variances de l'ordre de 0,1 pour $\log(K_s)$ et de 0,0015 pour $\log(n)$ sont respectivement comparables à des variances de 0,02 et 0,06 pour $\log(\alpha)$. Les modèles de variogrammes ajustés (Annexe H) sont des combinaisons de variogrammes exponentiels, sphériques et d'effets de pépite, avec des portées de quelques dizaines de centimètres.

TABLEAU 6.1 – Valeurs moyennes et variances logarithmiques des paramètres de Mualem-van Genuchten.

	K_s [m.s ⁻¹]	α [m ⁻¹]	n [-]
Moyenne	$6,3 \cdot 10^{-6}$	0,65	1,63
Variance logarithmique	0,10	0,13	0,0015

6.2 Simulations d'écoulement et de transport avec MELODIE

Un modèle tridimensionnel d'écoulement et de transport est construit, entièrement en zone non saturée, pour reproduire l'expérience de traçage ; 150 simulations sont générées à partir de ce modèle.

6.2.1 Domaine de modélisation et paramétrisation

Le domaine de modélisation est un pavé de 2 m de hauteur (discrétisation 0,0625 m) et de base 1,5 m par 1,5 m (discrétisation 0,1 m par 0,1 m). Un flux d'eau correspondant aux conditions expérimentales est imposé sur la surface supérieure du pavé, avec un flux de ³⁶Cl de 33 Bq.s⁻¹ pendant trois périodes de dix minutes, ce qui correspond à une activité totale injectée de 2,1 MBq. Les paramètres de transport utilisés sont donnés dans le Tableau 6.2 et les paramètres de Mualem-van Genuchten dans le Tableau 6.3.

Dans cette modélisation, la principale incertitude provient des paramètres de Mualem-van Genuchten, et de leur variabilité. Ainsi, les paramètres K_s et α sont représentés par des champs aléatoires lognormaux, corrélés (coefficient de corrélation 0,9) et de variogrammes exponentiels avec des portées variables (Tableau 6.4). Les portées identifiées sur les données granulométriques (Annexe H) sont d'environ 40 cm dans la direction horizontale et 30 cm dans la direction verticale, avec des variations de l'ordre de 10 cm selon le pas utilisé pour calculer le variogramme expérimental (l'effet de pépite n'est pas pris en compte dans cette modélisation). Deux séries ont été générées avec les valeurs moyennes et variances de K_s et α identifiées (séries 1 et 2). Une dernière série a été générée avec une valeur moyenne de K_s plus élevée (série 3), pour avoir un front d'infiltration un peu plus profond et se rapprocher des observations.

Par ailleurs, le paramètre n est constant dans les simulations. En effet, la prise en compte de la variabilité spatiale de n entraîne une augmentation des durées de simulations, ainsi que des problèmes de

convergence. De plus, il a été montré au Chapitre 3 que la prise en compte de la variabilité spatiale des trois paramètres K_s , α et n avait peu d'influence sur les panaches simulés par rapport à la prise en compte de la variabilité d'un seul paramètre. Ainsi, garder le paramètre n à une valeur constante n'entraîne pas de diminution de la variabilité au sein des panaches de ^{36}Cl simulés.

TABLEAU 6.2 – Paramètres de transport : ω est la porosité totale, D_x et D_z sont les composantes horizontales et verticales du tenseur de dispersivité, d est le coefficient de diffusion moléculaire, λ est la constante de désintégration et R est le coefficient de retard.

ω [-]	D_x [m]	D_z [m]	d [m ² .s ⁻¹]	λ [s ⁻¹]	R [-]
0,38	5.10^{-2}	5.10^{-3}	$1,2.10^{-10}$	$7,3.10^{-14}$	1

TABLEAU 6.3 – Valeurs des paramètres de Mualem-van Genuchten.

θ_r [m ³ .m ⁻³]	θ_s [m ³ .m ⁻³]	K_s [m.s ⁻¹]	α [m ⁻¹]	n [-]
0,039	0,38	voir Tableau 6.4	voir Tableau 6.4	1,6

TABLEAU 6.4 – Paramètres des champs aléatoires lognormaux, pour K_s et α (μ et σ sont les moyennes et variances des champs de paramètres et a_x et a_z sont les portées horizontales et verticales).

Série	Effectif	$\mu(K_s)$ [m.s ⁻¹]	$\sigma^2(\log K_s)$	$\mu(\alpha)$ [m ⁻¹]	$\sigma^2(\log \alpha)$	a_x [m]	a_z [m]
1	50 simulations	6.10^{-6}	0,1	0,6	0,1	0,5	0,5
2	50 simulations	6.10^{-6}	0,1	0,6	0,1	0,5	0,2
3	50 simulations	6.10^{-5}	0,1	0,6	0,1	0,4	0,3

6.2.2 Résultats de simulations

Deux exemples de panaches simulés sont donnés sur la Figure 6.3. Ils ont une forme et une taille proches de celles du panache observé, mais sont plus homogènes que celui-ci (Figure 6.2a).

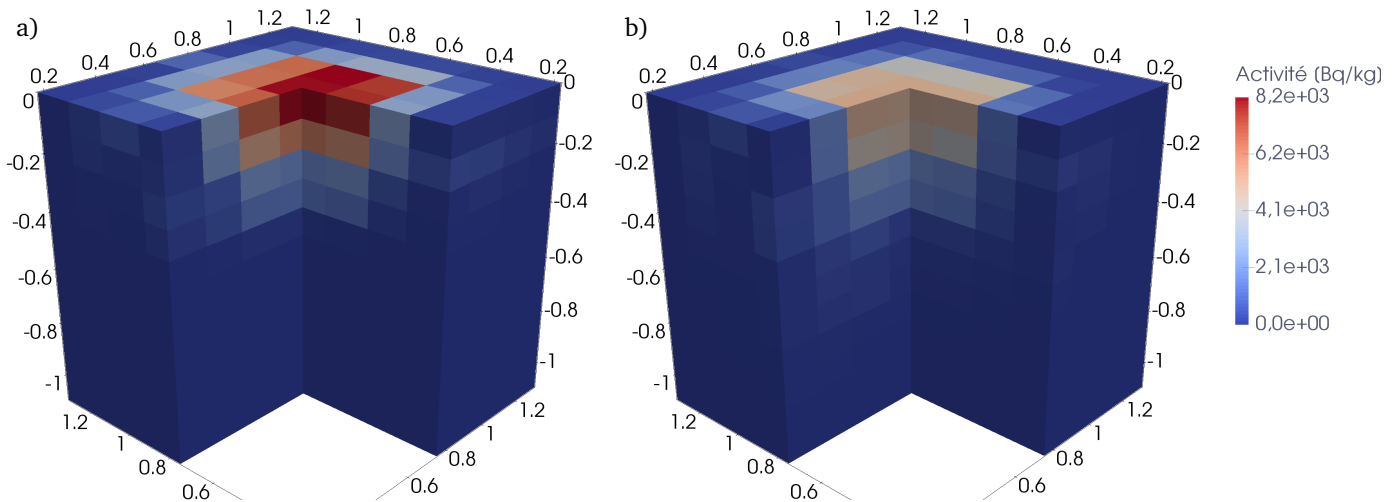


FIGURE 6.3 – Exemples de résultats de simulations. a) Panache de volume minimal ; b) panache de volume maximal.

Puisque les nœuds du maillage ne correspondent pas exactement aux localisations des observations, la comparaison entre observations et simulations est effectuée sur le nœud le plus proche de chaque observation. Les simulations reproduisent correctement les niveaux d'activité observés, malgré une légère sous-estimation (Figure 6.4). Par ailleurs, les panaches simulés sont en moyenne plus étendus que le panache observé : le volume moyen des panaches simulés est de 280 mailles (dans la zone délimitée par les plans I à VII), contre 192 pour le panache observé.

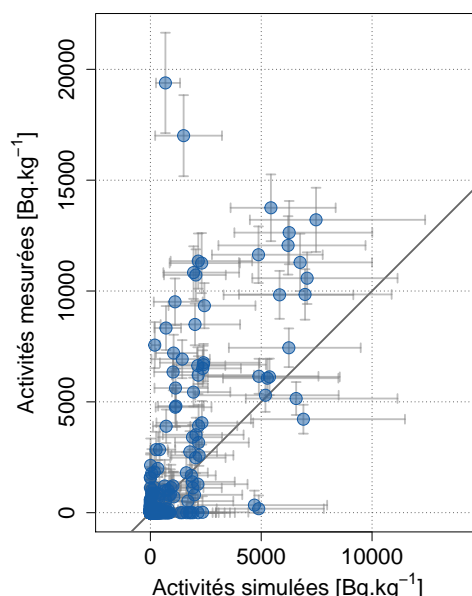


FIGURE 6.4 – Comparaison des activités observées et simulées. Les points représentent la valeur observée en fonction de la moyenne des simulations. Les barres sur les simulations donnent les valeurs minimales et maximales simulées. Les barres d'erreur sur les observations donnent l'incertitude de mesure.

Les panaches simulés avec les paramètres des séries 1 et 2 sont peu différents les uns des autres : comme au Chapitre 3, la variation des portées des champs de paramètres de Mualem-van Genuchten n'a pas d'influence significative sur les panaches simulés. Les différences sont plus notables entre les panaches des séries 1 et 2 d'une part et ceux de la série 3 d'autre part (augmentation de la valeur moyenne de K_s) : les panaches de la série 3 sont en moyenne 1,3 fois plus étendus que ceux des séries 1 et 2. Enfin, les simulations pourraient être améliorées en introduisant par exemple une hétérogénéité verticale plus marquée entre les niveaux sableux et limoneux.

6.3 Estimations de l'activité en ^{36}Cl

Puisque le bloc contenant les mesures d'activité est échantillonné de manière exhaustive, on se propose de réaliser des estimations avec différents scénarios d'échantillonnage, *i.e.* en supprimant certaines observations et en cherchant à les estimer. Trois méthodes sont comparées : le krigeage ordinaire avec un modèle de variogramme ajusté sur les données (KO), le krigeage avec dérive externe en utilisant la moyenne des simulations comme fonction de dérive (KDE) et le krigeage avec variogrammes numériques (KVN). Les estimations sont réalisées en voisinage unique. Pour le krigeage ordinaire usuel, les variogrammes ajustés pour chaque scénario sont linéaires. Les estimations sont réalisées sans tenir compte des erreurs de mesure.

6.3.1 Scénarios d'échantillonnage

Quatre scénarios d'échantillonnage sont construits, en faisant varier les nombres de sondages et de niveaux échantillonnés par sondage (Tableau 6.5). Lorsque trois niveaux sont échantillonnés, ils sont situés à 10, 50 et 90 cm de profondeur. Lorsque cinq niveaux sont échantillonnés, ils sont situés à 10, 30, 50, 70 et 90 cm de profondeur. Les positions des sondages sont données sur la Figure 6.5.

TABLEAU 6.5 – Tableau récapitulatif des scénarios d'échantillonnage.

Scénario	Nombre de sondages	Nombre de niveaux par sondage	Nombre d'observations
S5N3	5	3	15
S5N5	5	5	25
S9N3	9	3	27
S9N5	9	5	45

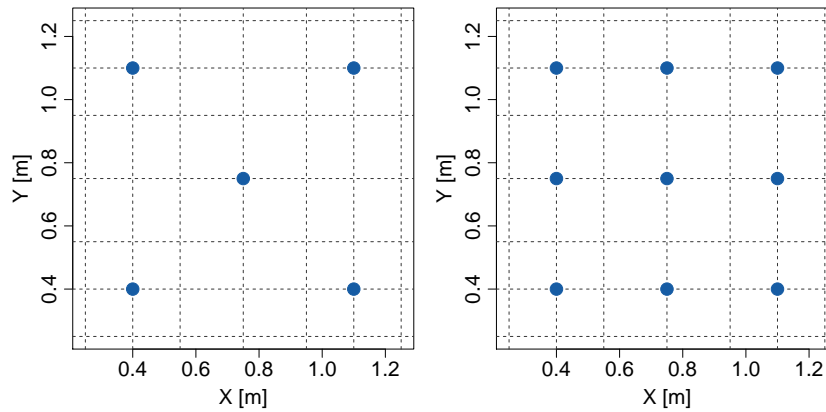


FIGURE 6.5 – Localisation des cinq et neuf sondages (vue de la surface supérieur du bloc). La grille tracée donne le plan de prélèvement à profondeur fixée (49 prélèvements).

6.3.2 Indicateurs de la qualité des estimations

Les représentations tridimensionnelles des estimations sont tracées, pour évaluer qualitativement les méthodes. Plusieurs indicateurs quantitatifs sont calculés sur l'ensemble du domaine :

- les erreurs sont quantifiées par l'erreur absolue moyenne ;
- les nuages de corrélation entre les valeurs d'activité de la référence et les activités estimées sont tracés. Des espérances conditionnelles empiriques sont également calculées, afin de quantifier le biais conditionnel de l'estimateur (voir Chapitre 4, Section 4.3.1) ;
- la proportion de données bien estimées est calculée. Cette proportion est définie comme le nombre de points du domaine tels que $Z_{\inf}(x) \leq Z^*(x) \leq Z_{\sup}(x)$, où $Z^*(x)$ est l'activité estimée en x , et $Z_{\inf}(x)$ (respectivement $Z_{\sup}(x)$) est l'activité mesurée en x , à laquelle l'incertitude de mesure est enlevée (respectivement ajoutée), divisé par le nombre de points du domaine (539). Les incertitudes sur les estimations ne sont pas prises en compte.

6.3.3 Résultats

Les activités estimées sont plus faibles pour le krigeage ordinaire que pour les deux autres méthodes (Figure 6.6). Les trois méthodes conduisent à des panaches estimés plus homogènes et plus symétriques que le panache réel (Figure 6.2a). Cette tendance est moins marquée pour le krigeage avec variogrammes numériques, pour les scénarios avec neuf sondages (Figure 6.6i et l). Pour le krigeage ordinaire et le krigeage avec dérive externe, les estimations dépendent peu du nombre de niveaux observés par sondage (trois ou cinq).

Quel que soit le scénario d'échantillonnage, le krigeage ordinaire donne les erreurs moyennes les plus élevées (Tableau 6.6). Le krigeage avec dérive externe donne les erreurs les plus faibles. Il est intéressant de remarquer que les erreurs absolues moyennes augmentent avec le nombre d'observations. Cela est dû aux scénarios d'échantillonnage particuliers considérés ici, et il ne s'agit pas d'une tendance générale : quand le nombre de sondages augmente (avec des localisations aléatoires), les erreurs moyennes diminuent (voir Section 6.3.4, Figure 6.8).

Par ailleurs, l'erreur moyenne sur les simulations (moyenne de l'écart entre activités mesurées et activités simulées sur le domaine) est de l'ordre de celle obtenue par krigeage avec dérive externe.

TABLEAU 6.6 – Erreurs absolues moyennes [Bq.kg^{-1}] pour les trois méthodes d'estimation et les simulations.

	S5N3	S5N5	S9N3	S9N5
KO	710	776	732	810
KDE	581	621	637	710
KVN	684	692	666	770
Simulations (moyenne)	630			

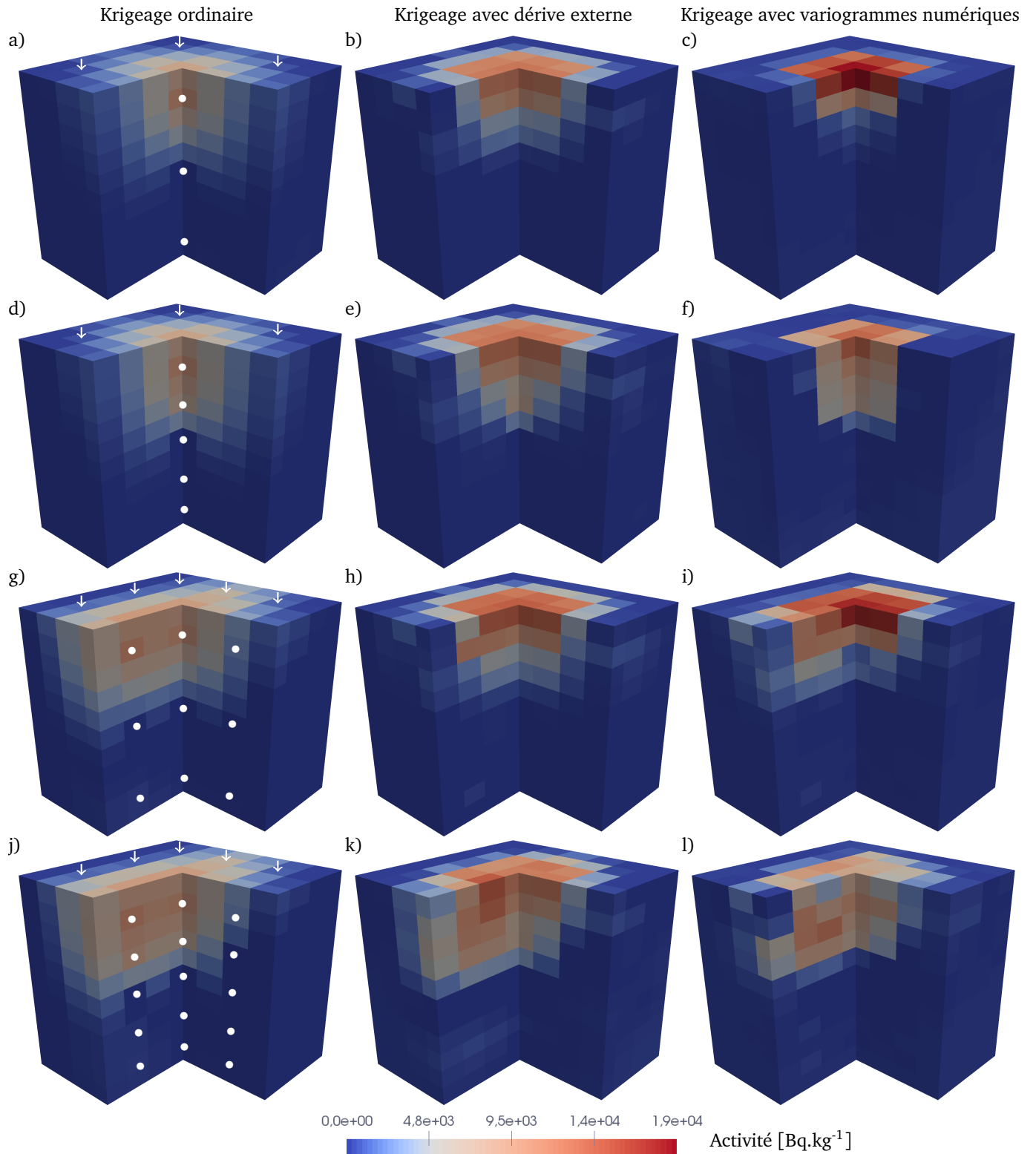


FIGURE 6.6 – Représentation tridimensionnelle des activités estimées, pour les trois méthodes et les quatre scénarios. De a) à c) scénario S5N3 ; de d) à f) scénario S5N5 ; de g) à i) scénario S9N3 ; de j) à l) scénario S9N5. Les flèches blanches rappellent les positions des sondages et les points indiquent les observations. Les échelles des axes X, Y et Z sont les mêmes que sur la Figure 6.2a.

Les nuages de corrélation entre les valeurs observées et estimées montrent qu'avec cinq sondages, le krigeage ordinaire a tendance à sous-estimer les valeurs d'activité (Figure 6.7a et b), ce qui est moins le cas avec neuf sondages (Figure 6.7c et d). Au contraire, le krigeage avec dérive externe et le krigeage avec variogrammes numériques ont tendance à surestimer les activités.

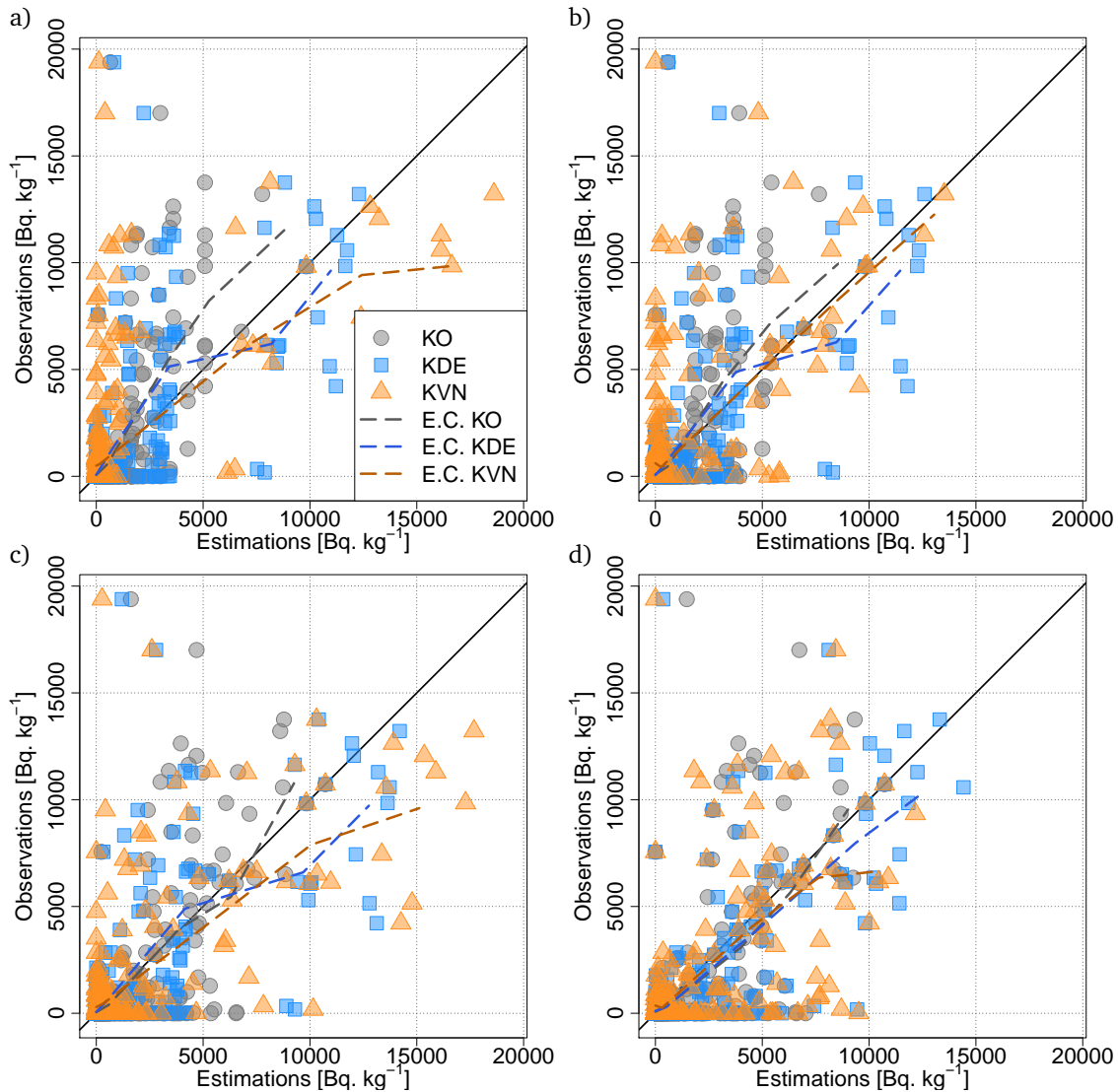


FIGURE 6.7 – Nuages de corrélation entre activités observées et estimées, avec les espérances conditionnelles empiriques (E.C.). a) Scénario S5N3 ; b) scénario S5N5 ; c) scénario S9N3 ; d) scénario S9N5.

Enfin, les proportions de données bien estimées sont plus élevées pour le krigeage avec variogrammes numériques que pour les deux autres méthodes d'estimation, quel que soit le scénario d'échantillonnage (Tableau 6.7). Ces proportions sont plus faibles pour le krigeage ordinaire. La proportion de points bien simulés est de l'ordre de celles obtenues par krigeage avec dérive externe.

TABLEAU 6.7 – Proportions [%] de données bien estimées pour les trois méthodes d'estimation et les simulations (moyenne sur les simulations des proportions de points bien simulés).

	S5N3	S5N5	S9N3	S9N5
KO	75	69	75	73
KDE	83	80	81	76
KVN	86	83	82	81
Simulations (moyenne)	80			

6.3.4 Influence du nombre de sondages

Les résultats ont été détaillés de manière précise pour quatre scénarios d'échantillonnage. Cependant, il est intéressant d'évaluer les performances des méthodes pour d'autres scénarios. Ainsi, les estimations ont été réalisées avec différents nombres de sondages N , placés aléatoirement dans le bloc. On suppose que cinq échantillons sont disponibles par sondage, toujours aux mêmes profondeurs (10, 30, 50, 70 et 90 cm). À nombre de sondages fixé (entre 5 et 15), 500 combinaisons de N sondages parmi 49 sont tirées aléatoirement. La qualité des estimations est quantifiée par l'erreur absolue moyenne (Figure 6.8). Quelle que soit la méthode d'estimation, les erreurs moyennes ainsi que les écarts entre les premiers et derniers déciles diminuent quand le nombre de sondages augmente. Cette baisse est cependant plus marquée pour le krigeage ordinaire que pour les deux autres méthodes. Le krigeage avec dérive externe donne toujours les erreurs moyennes et les écarts entre déciles les plus faibles. Par ailleurs, les écarts entre les méthodes diminuent lorsque le nombre de sondages augmente : l'apport des simulations est d'autant plus intéressant que le nombre de sondages observés est faible.

Ces erreurs absolues moyennes sont comparées à celles obtenues en ne considérant que les simulations : en moyenne, les erreurs sur les simulations sont plus faibles que celles obtenues par estimation, quel que soit le nombre de sondages observés (Figure 6.8).

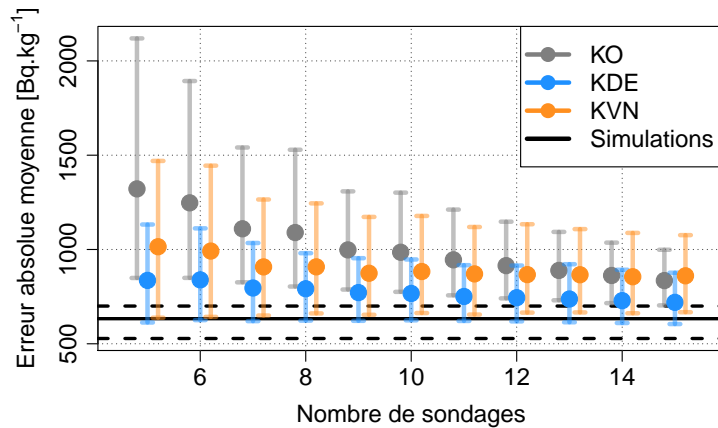


FIGURE 6.8 – Erreurs absolues moyennes en fonction du nombre de sondages. Les points donnent la moyenne des erreurs sur les 500 tirages, et les barres d'erreur le premier et le dernier décile. La ligne noire en trait plein donne l'erreur moyenne sur les simulations seules, les lignes en pointillés donnent les erreurs minimales et maximales sur les simulations.

6.3.5 Influence d'une valeur forte

La valeur d'activité la plus élevée (près de 20000 Bq.kg⁻¹) est observée sur la surface du bloc, assez excentrée. Elle n'a pas été échantillonnée dans les scénarios S5N3, S5N5, S9N3 et S9N5, et fait partie des activités estimées. Cependant, aucune des méthodes d'estimation ne la reproduit convenablement (triplet de points dans le coin supérieur gauche des nuages de corrélation de la Figure 6.7). Il semble cohérent que les méthodes d'estimation utilisant les simulations ne la reproduisent pas mieux que le krigeage ordinaire, puisque les simulations ne mettent pas en évidence de fortes valeurs d'activité sur la surface du bloc. Au vu des connaissances sur le dispositif expérimental et la réalisation du traçage, il semble difficile de générer des simulations reproduisant cette valeur forte.

6.4 Synthèse du chapitre

Le jeu de données d'activité en ^{36}Cl utilisé dans ce chapitre est intéressant par sa densité d'échantillonnage. En effet, celui-ci permet de valider l'application des méthodes combinant la géostatistique et les simulations d'écoulement et de transport sur un cas tridimensionnel en zone non saturée, sur un jeu de données réel, et non sur des données synthétiques comme au Chapitre 4. Un modèle d'écoulement et de transport est construit et reproduit convenablement les niveaux d'activité. Cent-cinquante simulations sont générées à partir de ce modèle, avec des champs de paramètres de Mualem-van Genuchten variables. Ces simulations sont utilisées pour estimer des valeurs d'activité en ^{36}Cl .

Les estimations réalisées par krigeage avec dérive externe et krigeage avec variogrammes numériques sont comparées à celles réalisées par krigeage ordinaire avec un modèle de variogramme usuel. Sur les quatre scénarios d'échantillonnage dont les résultats sont détaillés à la Section 6.3.3, le krigeage avec dérive externe conduit aux erreurs d'estimation les plus faibles. Selon le scénario d'échantillonnage, les erreurs absolues moyennes sont diminuées de 12% à 19% pour le krigeage avec dérive externe, et de 4% à 10% pour le krigeage avec variogrammes numériques par rapport au krigeage ordinaire. Cette tendance est confirmée pour d'autres scénarios d'échantillonnage : lorsque cinq sondages sont placés aléatoirement dans le bloc, les erreurs diminuent de 37% et de 23% en moyenne pour le krigeage avec dérive externe et le krigeage avec variogrammes numériques, par rapport au krigeage ordinaire. Lorsque dix sondages sont observés, ces diminutions sont respectivement de 22% et 10%. Ainsi, les méthodes de combinaison se montrent particulièrement intéressantes lorsque peu de données sont disponibles, ce qui avait déjà été mis en évidence sur le cas synthétique du Chapitre 4. Lorsque l'on considère la proportion de données bien estimées, le krigeage avec variogrammes numériques apparaît plus performant que le krigeage avec dérive externe.

Enfin, les erreurs moyennes obtenues sur les simulations seules sont plus faibles que celles obtenues par estimation. Ainsi, si le seul indicateur considéré est l'erreur moyenne, la combinaison de la géostatistique et des simulations semble surtout améliorer les résultats par rapport à la géostatistique seule (krigeage ordinaire avec variogramme usuel), mais pas par rapport aux simulations seules. Les proportions de points bien estimés nuancent ce résultat : dans ce cas, la combinaison de la géostatistique et des simulations permet d'avoir des résultats plus fiables que les simulations seules d'une part, et que la géostatistique seule d'autre part.

Il serait possible d'améliorer les modélisations d'écoulement et de transport, en simulant par exemple un milieu plus hétérogène : en effet, le passage par les surfaces spécifiques, ainsi que la conversion des textures en paramètres hydrodynamiques par la fonction de pédotransfert Rosetta3 tendent à lisser la variabilité des paramètres. Les variances des champs simulés pourraient être augmentées. Deux milieux aux propriétés différentes pourraient également être envisagés (milieu sableux et milieu plus limoneux). Cela permettrait peut être d'obtenir des panaches simulés moins lisses et plus hétérogènes, encore plus proches du panache observé que les simulations actuelles.

Chapitre 7

Application à un site d'enfouissement de déchets radioactifs : la tranchée T22

Ce chapitre présente l'application des méthodes de combinaison des simulations numériques et de la géostatistique sur un jeu de données d'activité en ^{90}Sr (Complément 7.1) mesurées dans la nappe à l'aval de la tranchée T22. Contrairement au Chapitre 6, le contexte (source de contamination, niveau piézométrique, interactions entre le radionucléide et le milieu, *etc.*) est ici moins bien connu, ce qui conduit à la construction d'un modèle d'écoulement et de transport moins contraint.

Les estimations d'activité dans la nappe sont effectuées dans deux cadres différents. Dans un premier temps, on se place dans un cadre spatial, et l'on quantifie la contamination dans la nappe à une date fixée. Cela permet d'une part d'évaluer les performances des méthodes d'estimation lorsque les simulations ne reproduisent pas aussi bien les données que dans les cas précédents (Chapitres 4 et 6) et d'autre part de mettre en place et de valider une méthodologie de quantification de la contamination en zone saturée. Dans un second temps, on se place dans un cadre temporel, et l'on estime l'activité en ^{90}Sr en différents points de la nappe sur une période d'une dizaine d'années pour évaluer les flux de ^{90}Sr ayant traversé différentes sections de la nappe. Cela permet d'estimer la diminution du stock de ^{90}Sr durant cette période.

Complément 7.1

Le strontium 90 (^{90}Sr)

Le ^{90}Sr se désintègre par rayonnement β en yttrium 90 avec une période radioactive de 28,8 ans. Son activité est de 5,1 TBq par gramme de ^{90}Sr pur. Son origine est anthropique, principalement due à des essais nucléaires atmosphériques ainsi qu'à des rejets accidentels, comme ceux de Tchernobyl. Sa nocivité vient de son rayonnement β d'énergie importante et de sa capacité à remplacer son analogue chimique, le calcium, dans les organismes.

7.1 Présentation du site et des données

7.1.1 Présentation générale du site

La tranchée T22 se situe au centre du Site Epic (Figure 7.1a), à quelques kilomètres à l'ouest du réacteur 4. Un réseau de piézomètres est implanté autour de la T22 (Figure 7.1b).

L'aquifère étudié dans ce travail est constitué de deux formations :

- des sables éoliens sur 4 à 5 m, relativement homogènes et constitués de grains fins à moyens [Dewière *et al.*, 2004 ; Bugai *et al.*, 2012].
- des dépôts alluviaux sur 20 à 30 m, constitués de sables plus hétérogènes, présentant une fraction argileuse pouvant atteindre 8%, et organisés en un ensemble de niveaux horizontaux de texture et d'épaisseur variables [Bugai *et al.*, 2002 ; Dewière *et al.*, 2004].

Cet aquifère contient une nappe libre dont le niveau varie entre 0 et 4 m de profondeur [Nguyen, 2017]. L'écoulement régional est dirigé vers la rivière Prypiat. Il est orienté nord/nord-est, avec un gradient hydraulique moyen de 0,001 à 0,0015 m.m^{-1} .

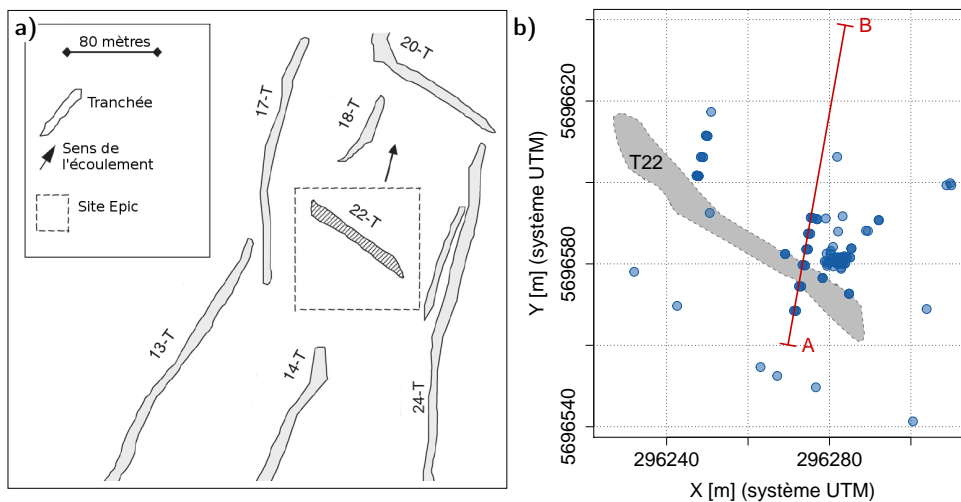


FIGURE 7.1 – a) Schéma de l'environnement du site Epic (modifié d'après [Kashparov et al. \[2012\]](#)). b) Carte de l'implantation des piézomètres. Les contours de la T22 sont approximatifs.

7.1.2 Données d'activité dans la nappe

La migration du ^{90}Sr depuis la tranchée T22 a été mise en évidence grâce au réseau de piézomètres (Figure 7.1b). Dans la suite de ce travail, on s'intéresse uniquement à la section AB, qui est parallèle à la direction principale de l'écoulement. À partir de ce stade, toutes les coordonnées le long de cette section sont données dans un repère local, le point A étant localisé à l'abscisse 0, et le point B à l'abscisse 80 m.

Le long de la section AB, six groupes de piézomètres (implantés en flûte de pan) sont distingués, permettant chacun des prélèvements à quatre profondeurs. Par commodité, ils sont regroupés et notés P1-P2-P3-P4-P5-P6, d'amont en aval (Figure 7.2). Des activités en ^{90}Sr dans la nappe sont donc disponibles en 24 points de prélèvement. Ces points ont été renommés $Pi-j$, où i est le numéro du piézomètre et j le niveau de prélèvement (niveau 1 le plus proche de la surface, niveau 4 le plus profond).

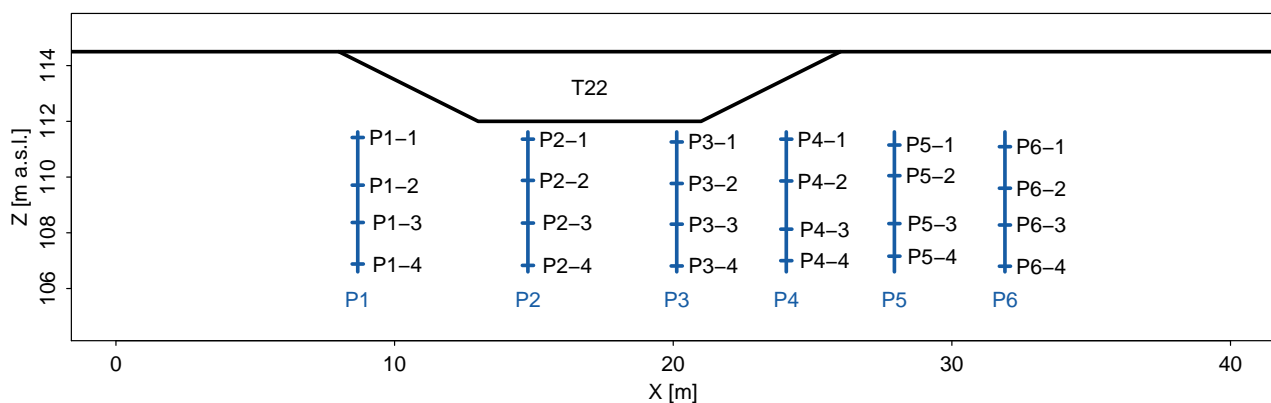


FIGURE 7.2 – Piézomètres le long de la section AB. Les coordonnées verticales sont en mètres au-dessus du niveau de la mer [m a.s.l.].

Pour chaque point de prélèvement, entre 12 et 25 mesures sont disponibles entre 2000 et 2015, correspondant à une activité volumique moyenne sur la hauteur crépinée (40 cm). Elles ne sont pas nécessairement réalisées de manière synchrone pour tous les piézomètres : les données des 24 points ne sont disponibles simultanément qu'à cinq dates (Figure 7.3).

Ces données mettent en évidence une variabilité spatiale des activités en ^{90}Sr . Les valeurs d'activité les plus fortes sont localisées sur le niveau 1 des piézomètres P3 et P4 et sur le niveau 2 des piézomètres P4, P5 et P6, avec des activités de plusieurs centaines à quelques milliers de Becquerels par litre d'eau (Figures 7.3 et 7.4). Ainsi, le panache de ^{90}Sr semble migrer légèrement en profondeur de l'amont vers l'aval, ce qui est

cohérent avec la composante verticale du gradient hydraulique mise en évidence localement [Bugai *et al.*, 2012]. Enfin, les activités observées sont assez nettement inférieures au piézomètre P6 par rapport aux piézomètres P4 et P5, ce qui semble indiquer une diminution de la quantité de ^{90}Sr traversant ce piézomètre, qui peut être liée à une rétention du ^{90}Sr entre l'amont et l'aval, ou éventuellement une modification locale du gradient hydraulique.

Ces données mettent également en évidence une variabilité temporelle de l'activité en ^{90}Sr marquée à chaque point de prélèvement (Figure 7.4), bien que les niveaux moyens varient fortement d'un point de prélèvement à l'autre. Cependant, au vu de la fréquence d'échantillonnage (une à deux mesures par an), il n'est pas garanti que la variabilité temporelle observée soit représentative de la variabilité réelle.

Il est important de garder à l'esprit qu'il existe de nombreuses incertitudes sur ces données. Elles sont issues d'une base de données consolidée, mais elles ont été acquises par des instituts différents, avec des méthodes différentes, et l'on dispose d'assez peu d'informations à ce sujet. Les incertitudes de mesure associées à ces données ne sont par exemple pas quantifiées de manière systématique : lorsqu'elles sont précisées, elles valent 5% à 10% de l'activité mesurée. Enfin, la position des piézomètres et les limites inférieures et supérieures des crépines sont connues de manière fiable et assez précise, mais la position relative des piézomètres par rapport à la tranchée est plus incertaine.

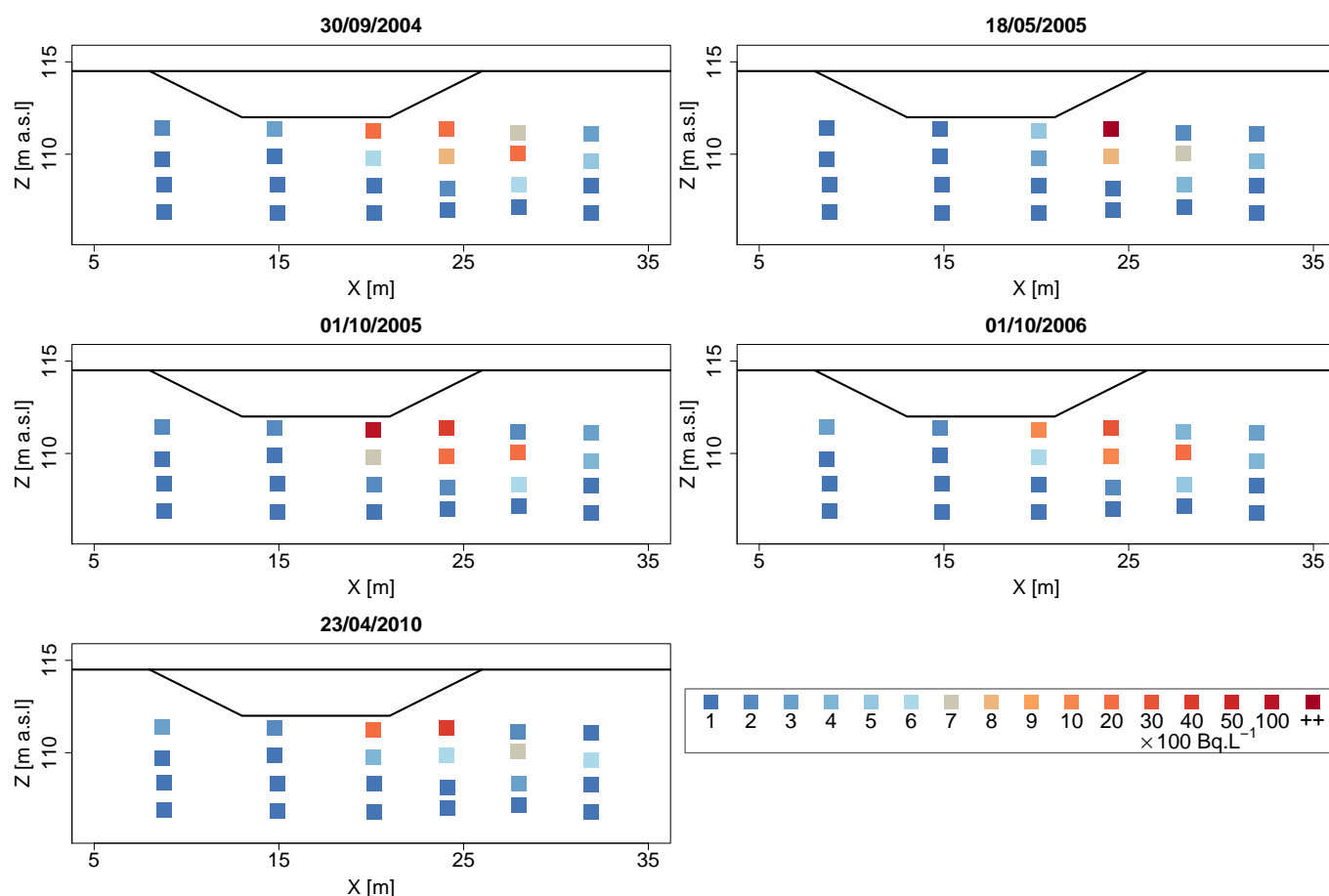


FIGURE 7.3 – Représentation des données aux cinq dates pour lesquelles des mesures d'activité sont disponibles aux 24 points de prélèvement. L'échelle de couleurs est identique pour toutes les dates.

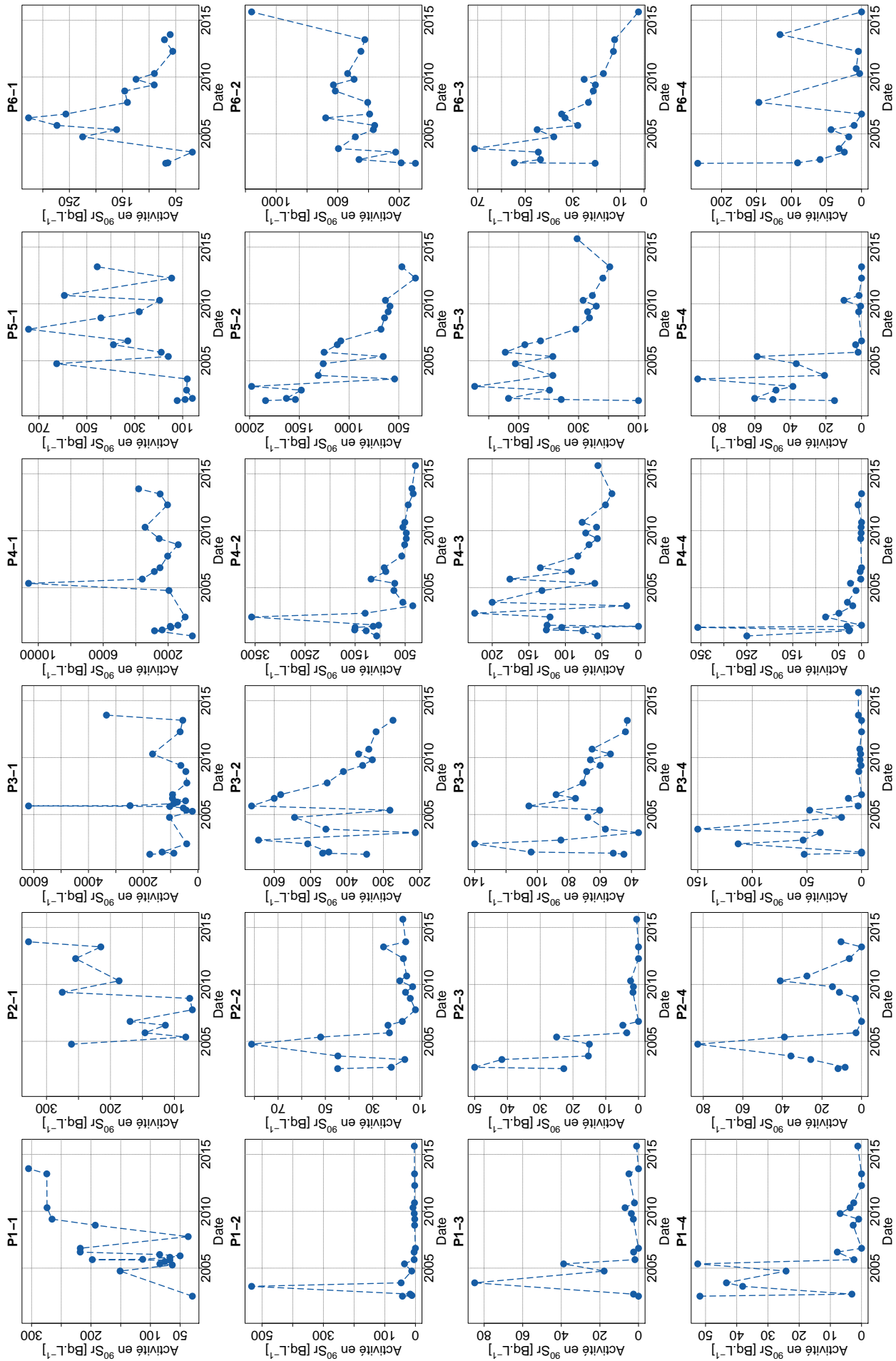


FIGURE 7.4 – Représentation des données aux 24 points de prélèvement. Les points sont reliés pour faciliter la lecture mais cela ne donne pas d'informations sur l'évolution de l'activité entre deux mesures.

7.2 Simulations d'écoulement et de transport avec MELODIE

Un modèle d'écoulement et de transport est construit pour simuler les activités en ^{90}Sr dans la nappe, le long de la section AB. Cette section étant parallèle à la direction principale de l'écoulement, le modèle construit est bidimensionnel. Ce choix permet de réduire nettement la durée des simulations.

7.2.1 Domaine de modélisation et conceptualisation

Cette section présente les concepts utilisés pour réaliser la modélisation. La plupart des données ont été acquises sur le site Epic. Elles ne sont pas présentées exhaustivement ici (voir par exemple [Kashparov et al. \[2012\]](#) ; [Bugai et al. \[2012\]](#) ; [Nguyen \[2017\]](#) ; [Kashparov et al. \[2019\]](#) pour plus de détails).

Domaine de modélisation

Le domaine de modélisation est une section de 80 m de longueur et de 10 m de profondeur, discrétisée en mailles triangulaires de 0,5 m (base) par 0,3 m (hauteur). Les conditions aux limites sont décrites sur le schéma de la Figure 7.5 : des charges hydrauliques (variables dans le temps) sont imposées aux limites amont et aval du domaine, un flux de percolation variable est imposé sur la partie supérieure du domaine, et la limite inférieure du domaine est imperméable, avec une conductivité hydraulique équivalente imposée sur la dernière rangée de mailles (voir le paragraphe "Propriétés hydrauliques").

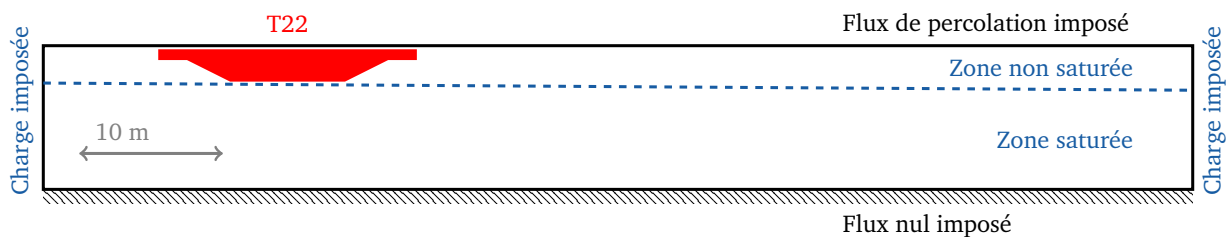


FIGURE 7.5 – Schéma du domaine de modélisation et des conditions aux limites.

Niveau piézométrique

Les charges hydrauliques imposées sont données par l'évolution du niveau piézométrique, connu depuis 1999 grâce à l'instrumentation du site Epic. Un modèle de l'évolution de ce niveau avant 1999 a été construit à l'IRSN. Il est basé sur le modèle GARDÉNIA, développé au BRGM [[Thiéry, 2014](#)], et qui simule de manière simplifiée les principaux mécanismes du cycle de l'eau dans un bassin versant.

Les observations et la modélisation (Figure 7.6) montrent que le niveau piézométrique varie de manière saisonnière sur une amplitude de 2 m au plus. Il est parfois au-dessus de la base de la T22, notamment à partir de 1999. Ainsi, la tranchée est régulièrement inondée, ce qui pourrait favoriser le transport de contaminants de la tranchée vers la nappe.

Stock de ^{90}Sr à l'origine de la contamination

Plusieurs campagnes de mesures ont été réalisées pour caractériser la contamination dans la tranchée [[Bugai et al., 2002](#) ; [Kashparov et al., 2012](#) ; [Nguyen, 2017](#)] : des mesures d'activité en ^{137}Cs (césium 137) et quelques mesures d'activité en ^{154}Eu (europium 154) et en ^{90}Sr ont été effectuées. Les mesures d'activité en ^{137}Cs sont utilisées pour estimer la distribution de l'activité dans la tranchée en 2000 [[Nguyen, 2017](#)] et des relations empiriques (Complément 7.2) permettent de convertir le stock de ^{137}Cs en 2000 en stock initial de ^{90}Sr (au moment du creusement de la tranchée). Des simulations conditionnelles du stock de ^{137}Cs en 2000 ont été réalisées lors de la thèse de L. Nguyen [[Nguyen, 2017](#)]. L'une de ces simulations est ainsi convertie en stock initial de ^{90}Sr et utilisée comme source de contamination dans le modèle d'écoulement et de transport (Figure 7.7).

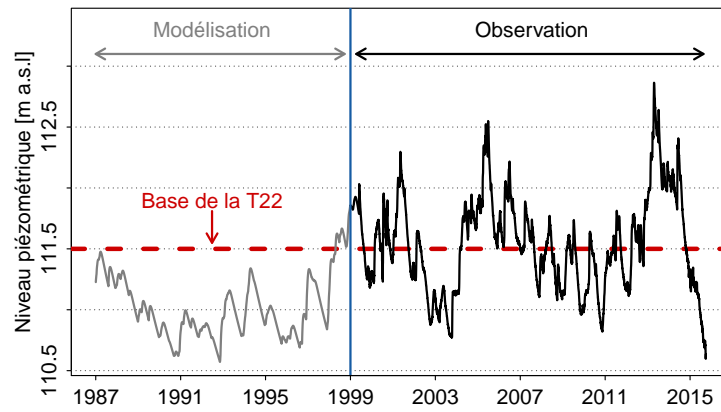
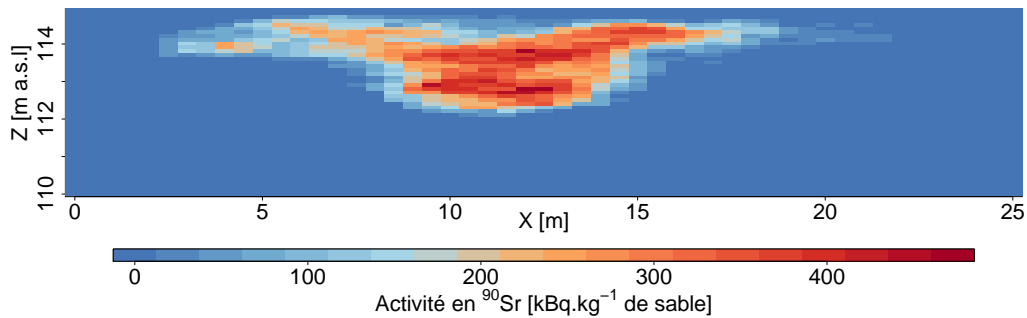


FIGURE 7.6 – Chronique piézométrique modélisée (avant 1999) et observée (après 1999).


 FIGURE 7.7 – Exemple de stock initial de ^{90}Sr le long du profil AB, obtenu par conversion d'une simulation de ^{137}Cs [Nguyen, 2017]. Les coordonnées longitudinales sont dans un repère local.

En réalité, seule une fraction du stock de ^{90}Sr est disponible à la migration dans l'environnement. Un modèle conceptuel de l'évolution de la fraction mobile a été établi (Complément 7.3) ; il décrit la disponibilité du stock au cours du temps dans le modèle d'écoulement et de transport.

Enfin, la migration du ^{90}Sr et son interaction avec le milieu sont modélisées par une approche de type K_d (Complément 7.4). Les valeurs du coefficient K_d estimées pour les sables éoliens sont de l'ordre de 0,5 à 5 mL.g^{-1} , et un ordre de grandeur au-dessus pour les sables alluviaux [Nguyen, 2017 ; Bugai *et al.*, 2020].

Conversion du stock de ^{137}Cs en 2000 en stock de ^{90}Sr en 1986

Il n'y a pas de corrélation directe entre les activités en ^{137}Cs et en ^{90}Sr en 2000, car le ^{90}Sr est nettement plus mobile que le ^{137}Cs . Cependant, il existe une corrélation entre les activités en ^{154}Eu et en ^{137}Cs en 2000, puisque les deux radionucléides sont peu mobiles [Bugai *et al.*, 2002] :

$$[^{154}\text{Eu}_{(2000)}] = \frac{1}{145} [^{137}\text{Cs}_{(2000)}] \quad (7.1)$$

Il existe également une corrélation entre les activités en ^{90}Sr et en ^{154}Eu dans les particules de combustible du réacteur accidenté en 1986 [Kuriny *et al.*, 1993] :

$$[^{90}\text{Sr}_{(1986)}] = \frac{100}{3} [^{154}\text{Eu}_{(1986)}] \quad (7.2)$$

Les Équations (7.1) et (7.2), associées à la loi de décroissance de ^{154}Eu , permettent de convertir le stock de ^{137}Cs en 2000 en stock initial de ^{90}Sr .

Modèle de libération du ^{90}Sr

La tranchée T22 contient des déchets divers dont la principale source de radioactivité est constituée de particules de combustible nucléaire. De manière conceptuelle, trois types de particules de combustible sont distingués [Kashparov *et al.*, 2012, 2019] :

1. des particules relativement instables, dans une matrice UO_{2+x} ;
2. des particules relativement stables, dans une matrice UO_2 ;
3. des particules très stables, dans une matrice ZrU_yO_x .

La dissolution de ces particules est modélisée par une cinétique d'ordre 1. La fraction du stock qui passe sous forme mobilisable, pour chaque type de particule de combustible i , est donnée par :

$$\Delta\text{PC}_i(t) = E_i (1 - e^{-(k_i - \lambda)t}), \quad i \in \{1, 2, 3\} \quad (7.3)$$

où E_i [-] est la fraction initiale des particules de type i (Tableau 7.1), k_i [T^{-1}] est la constante de dissolution des particules de type i (Tableau 7.1), λ [T^{-1}] est la constante de décroissance radioactive du ^{90}Sr et t [T] est le temps. L'évolution des fractions de ^{90}Sr mobilisable et de chaque type de particules est donnée sur la Figure 7.8.

TABLEAU 7.1 – Fractions initiales et constantes de dissolution de chaque type de particule de combustible [Kashparov *et al.*, 2012].

	Type 1	Type 2	Type 3
Fraction initiale [%]	21 ± 15	57 ± 15	19 ± 10
Constante de dissolution [an^{-1}]	$0,28 \pm 0,05$	$0,018 \pm 0,001$	≈ 0

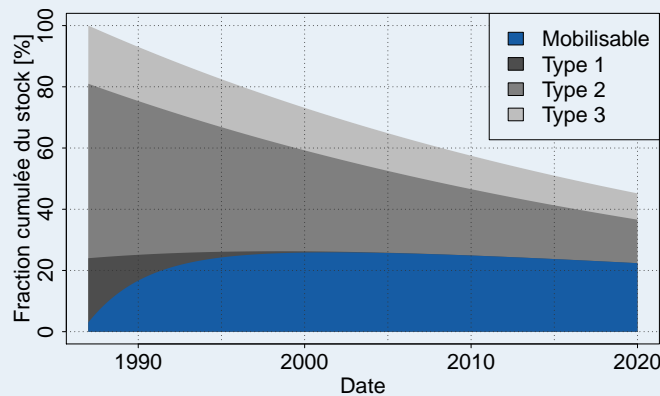


FIGURE 7.8 – Évolution de la fraction cumulée du stock initial répartie entre le ^{90}Sr mobilisable et les particules de combustible des trois types.

L'approche K_d

L'approche K_d est acceptable ici car la sorption du ^{90}Sr est réversible (si les conditions hydrogéo-chimiques sont relativement stables) [Nguyen, 2017]. Le coefficient K_d [$\text{L}^3 \cdot \text{M}^{-1}$] est un paramètre empirique, qui donne la distribution d'un élément entre la fraction contenue dans la phase solide (fraction sorbée) et la fraction contenue dans la phase liquide (fraction en solution). Dans les codes d'écoulement et de transport comme MELODIE, on utilise un coefficient de retard noté R [-] :

$$R = 1 + \frac{1 - \omega}{\omega} \rho_s K_d \quad (7.4)$$

où ω [-] est la porosité du milieu et ρ_s [$\text{M} \cdot \text{L}^{-3}$] la densité de la phase solide. On retrouve que pour un traceur parfait comme le tritium, la fraction sorbée est nulle, et le coefficient R vaut alors 1.

Propriétés hydrauliques

Dans les sables éoliens, la conductivité hydraulique est isotrope et de l'ordre de 4.10^{-5} à $6.10^{-5} \text{ m.s}^{-1}$ [Bugai *et al.*, 2002 ; Dewièrre *et al.*, 2004]. Dans les sables alluviaux, elle est anisotrope : de l'ordre de $1.10^{-5} \text{ m.s}^{-1}$ dans la direction horizontale et $4.10^{-7} \text{ m.s}^{-1}$ dans la direction verticale [Bugai *et al.*, 2002].

Pour caractériser plus précisément la conductivité hydraulique à saturation, des données granulométriques de Prypiat Zaton et du site Epic (Annexe G) sont exploitées. Ces données sont converties en paramètres de Mualem-van Genuchten par la fonction de pédotransfert Rosetta3, et les valeurs de paramètres obtenues sont comparées à celles de la littérature, ainsi qu'à une interprétation des données granulométriques réalisée précédemment par l'IRSN (Tableau 7.2). Les valeurs moyennes sont cohérentes avec celles de la littérature, mais diffèrent assez nettement d'un site à l'autre et d'une méthode à l'autre. L'évolution des valeurs de K_s avec la profondeur met en évidence l'augmentation de la variabilité de la conductivité hydraulique avec la profondeur (Figure 7.9a) ; le modèle de variogramme ajusté est un modèle sphérique avec une portée de d'une dizaine de mètres (Figure 7.9b).

TABLEAU 7.2 – Comparaison des valeurs moyennes de K_s (notées μ) et des variances du logarithme de K_s (notées σ^2), pour l'interprétation par l'IRSN et la conversion par Rosetta3.

	Sables éoliens		Sables alluviaux	
	$\mu(K_s) [\text{m.s}^{-1}]$	$\sigma^2(\log K_s)$	$\mu(K_s) [\text{m.s}^{-1}]$	$\sigma^2(\log K_s)$
Site Epic (IRSN)	$5,2.10^{-5}$	0,02	$1,0.10^{-5}$	0,2
Site Epic (Rosetta3)	$9,8.10^{-5}$	0,007	$1,6.10^{-5}$	0,3
Prypiat Zaton (IRSN)	$6,1.10^{-5}$	0,03	$2,1.10^{-6}$	0,1
Prypiat Zaton (Rosetta3)	$1,2.10^{-4}$	0,04	$4,9.10^{-6}$	0,2

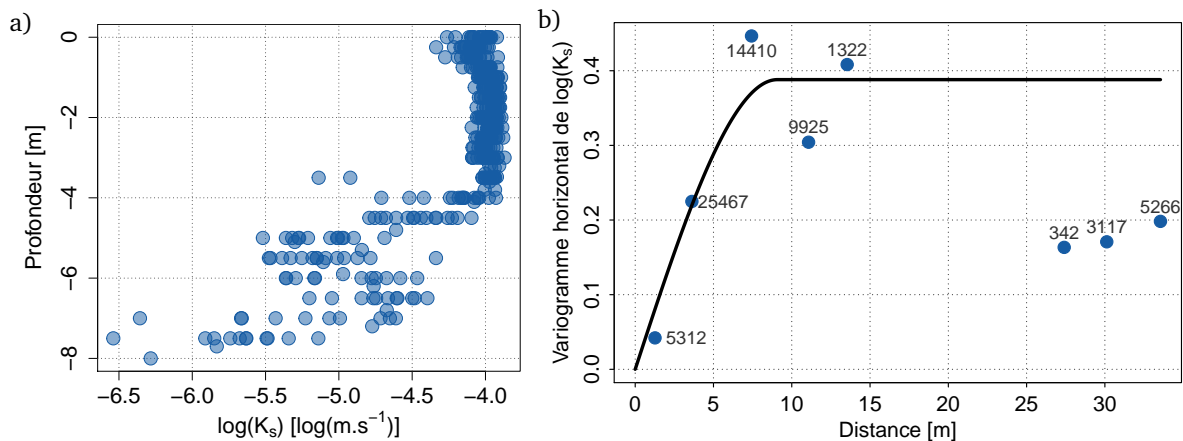


FIGURE 7.9 – a) Profil du logarithme de la conductivité hydraulique à saturation, pour le site Epic et avec des données obtenues par Rosetta3, sans distinguer les formations éoliennes et alluviales. b) Variogramme de ces valeurs dans la direction horizontale (le nombre de paires de points est indiqué).

Dans le modèle d'écoulement et de transport, la conductivité hydraulique à saturation est modélisée par un champ aléatoire de moyenne et de variance non stationnaires (Figure 7.10 et Tableau 7.4). Un champ aléatoire gaussien stationnaire Y_s est d'abord généré. Il est ensuite modifié pour obtenir un champ non stationnaire dans la direction verticale : $Y_{Ns}(z) = A(z) + B(z)Y_s(z)$, où les fonctions A et B sont des polynômes permettant de reproduire les valeurs de conductivité souhaitées. Le champ Y_{Ns} est le logarithme de K_s .

La rangée inférieure des mailles du champ de conductivité hydraulique est modifiée pour prendre en compte la vingtaine de mètres de sables alluviaux se situant sous le domaine de modélisation : une conductivité hydraulique équivalente de $3.10^{-3} \text{ m.s}^{-1}$ y est imposée.

Les autres paramètres du modèle de Mualem-van Genuchten (α , n , θ_r et θ_s) sont également obtenus par conversion des données de texture (Tableau 7.3). Ces valeurs sont comparées à celles mesurées dans la zone non saturée du Site Epic par l'IRSN [Ardois *et al.*, 2007].

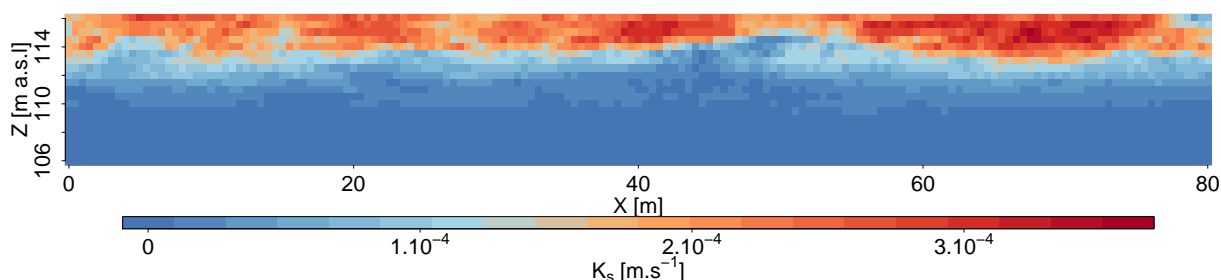


FIGURE 7.10 – Exemple de champ de conductivité hydraulique à saturation.

TABLEAU 7.3 – Paramètres de Mualem-van Genuchten moyens pour les sables éoliens.

Méthode et site	α [m ⁻¹]	n [-]	θ_r [m ³ .m ⁻³]	θ_s [m ³ .m ⁻³]
Ardois <i>et al.</i> [2007] (Site Epic)	4,3	2,4	0,04	0,29
Rosetta3 (Site Epic)	3,8	3,8	0,05	0,33
Rosetta3 (Prypiat Zaton)	3,7	3,9	0,05	0,34

Synthèse

Un modèle d'écoulement et de transport à deux dimensions est construit en reprenant les éléments présentés précédemment. Les paramètres utilisés pour construire ce modèle sont résumés dans le Tableau 7.4.

TABLEAU 7.4 – Récapitulatif des paramètres du modèle d'écoulement et de transport. Les paramètres variables d'une simulation à l'autre sont en bleu (voir Section 7.2.2).

	Paramètres		Valeurs
Écoulement	Paramètres du modèle de Mualem-van Genuchten	α	3,7 m ⁻¹
		n	3,8
		θ_r	0,04
		θ_s	0,4
	K_s (champ aléatoire)	$\mu(K_s)$ (haut)	1.10 ⁻⁴ m.s ⁻¹
		$\mu(K_s)$ (bas)	2.10 ⁻⁶ m.s ⁻¹
		$\sigma^2(\log K_s)$ (haut)	0,02
		$\sigma^2(\log K_s)$ (bas)	0,5
		Portée horizontale	10 m
		Portée verticale	3 m
	Gradient hydraulique		0,015 m.m ⁻¹
	Charges hydrauliques		exemple Figure 7.6
Transport	Coefficient de diffusivité		1,2.10 ⁻¹⁰ m ² .s ⁻¹
	Dispersivité	Horizontale	0,05 m
		Verticale	0,005 m
	Porosité		0,4
	Porosité efficace	Zone non saturée	0,4
		Zone saturée	0,04
	Coefficient de retard	Tranchée	12,2
		Aquifère	6 ou 12
Terme source	Stock de ⁹⁰ Sr initial		exemple Figure 7.7
	Libération du stock		exemple Figure 7.8
	Constante de décroissance radioactive		7,63.10 ⁻¹⁰ s ⁻¹

7.2.2 Principales incertitudes de modélisation

Plusieurs incertitudes de modélisation sont prises en compte, et certains paramètres de modélisation varient d'une simulation à l'autre :

- les champs de conductivité hydraulique à saturation sont différents pour chaque simulation d'écoulement et de transport. Ils correspondent à des réalisations différentes de la même fonction aléatoire.
- le modèle de variation du niveau piézométrique avant 1999 est tiré parmi 100 réalisations équiprobables, et la chronique complète (modélisation et observations) peut être remontée ou abaissée de la hauteur d'une maille. L'incertitude sur la position relative du fond de la tranchée par rapport au niveau de la nappe est ainsi prise en compte.
- le coefficient de retard dans les sables est soit égal à 6, soit à 12.
- la même simulation de stock initial de ^{90}Sr est utilisée comme terme source, mais la largeur de la tranchée est égale à 8 m, 13 m ou 18 m (troncature de la simulation de la Figure 7.7). Cela prend en compte l'incertitude sur le stock initial et sur la géométrie de la tranchée.
- les fractions initiales de chaque type de particules de combustible sont tirées parmi cinq possibilités pour le triplet de fractions initiales $\{E_1; E_2; E_3\}$, correspondant aux valeurs données dans le Tableau 7.1, ainsi qu'aux incertitudes maximales et intermédiaires sur E_1 . Cela conduit à différentes fractions de stock initial ayant pu être libéré en 2020 : 34%, 39%, 43%, 48% ou 52%.

Les variations du coefficient de retard (R) et de la largeur de la tranchée (L) sont les incertitudes ayant le plus d'influence sur les résultats de simulation d'écoulement et de transport. Six groupes de simulations sont ainsi distingués (au sein desquels les champs de conductivité hydraulique à saturation, les chroniques piézométriques et les fractions initiales de particules de combustible varient) :

Groupe 1 $R=6$ et $L=18$ m ;

Groupe 2 $R=12$ et $L=18$ m ;

Groupe 3 $R=6$ et $L=13$ m ;

Groupe 4 $R=12$ et $L=13$ m ;

Groupe 5 $R=6$ et $L=8$ m ;

Groupe 6 $R=12$ et $L=8$ m.

7.2.3 Résultats de simulations

Six cents simulations sont générées à partir du modèle d'écoulement et de transport. Chaque simulation est effectuée en quelques heures. Un exemple de panache de ^{90}Sr simulé est donné sur la Figure 7.11.

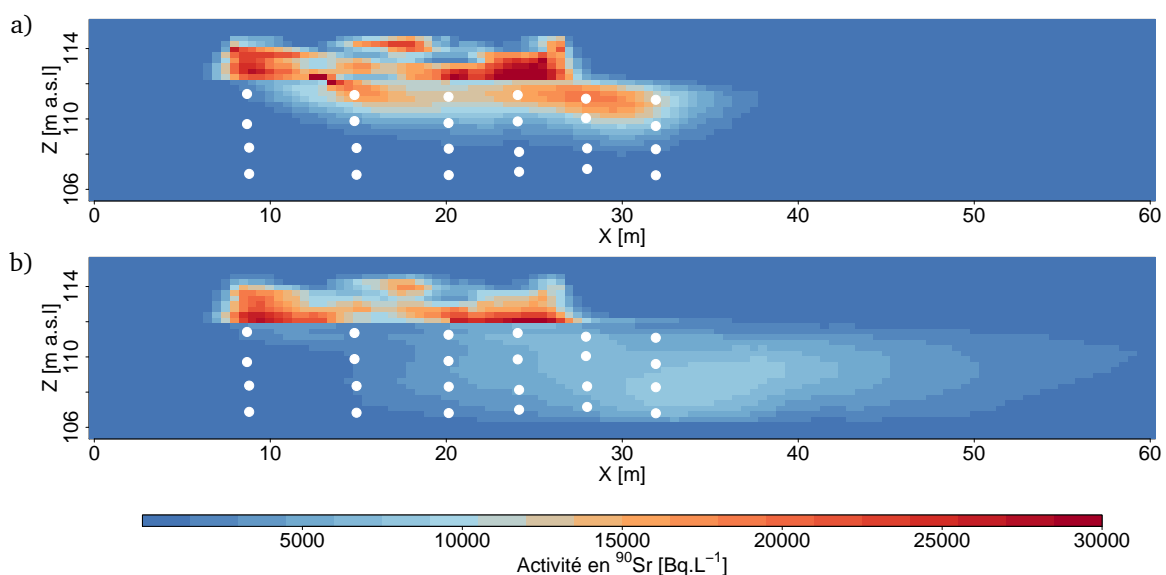


FIGURE 7.11 – Exemple d'une simulation de la migration d'un panache de ^{90}Sr , issue du groupe 1. a) En janvier 2000 ; b) en janvier 2010. Les points blancs marquent la position des 24 points de prélèvement.

Les simulations présentent des valeurs d'activité plus élevées que les observations (Figure 7.12). Les différents groupes donnent lieu à des comportements médians variés : les simulations réalisées avec le coefficient de retard le plus élevé (groupes 2, 4 et 6) sont en général plus proches des observations (en terme de valeur d'activité), mais présentent une dynamique trop lente, ce qui est par exemple visible aux niveaux 3 et 4 des piézomètres P4-P5-P6. La dynamique des simulations est cohérente avec les observations, lorsque le plus faible coefficient de retard est utilisé (groupes 1, 3 et 5). Les différences dues à la variation de la largeur de la tranchée sont surtout visibles au piézomètre P1.

7.3 Combinaison du krigeage et de simulations d'écoulement et de transport : cadre spatial

Dans un premier temps, les estimations sont réalisées dans un cadre spatial, pour quantifier la contamination dans la nappe à un instant donné. Les méthodes mises en œuvre peuvent être utilisées dans le cas de la remédiation du site par exemple. L'activité en ^{90}Sr dans la nappe est estimée aux cinq dates auxquelles les activités sont connues aux 24 points de prélèvement (Figure 7.3). Trois méthodes sont comparées : le krigeage ordinaire (KO) avec un modèle de variogramme ajusté sur les données, le krigeage avec dérive externe (KDE) (moyenne du jeu de simulations numériques en dérive) et le krigeage avec variogrammes numériques (KVN). Les estimations sont réalisées en voisinage unique. Pour le krigeage ordinaire et le krigeage avec dérive externe, les modèles de variogramme ajustés sont linéaires, avec une anisotropie géométrique entre la direction horizontale et la direction verticale.

7.3.1 Critères de comparaison des méthodes

Des critères qualitatifs et quantitatifs permettent de comparer les méthodes. Les résultats détaillés sont présentés ici pour la première date d'intérêt (30/09/2004). Pour les autres dates, les résultats sont donnés dans l'Annexe I.

Critères qualitatifs

Les critères qualitatifs évaluent la capacité des méthodes d'estimation à reproduire la cohérence des phénomènes physiques connus (sens d'écoulement de la nappe, localisation de la tranchée, etc.). Les deux critères suivants sont examinés :

- les cartes d'estimation et d'écart-type théorique d'erreur d'estimation ;
- les profils d'activité par piézomètre (*i.e.* la répartition de l'activité sur un segment vertical), sans utiliser les données du piézomètre où l'activité est estimée.

Critères quantitatifs

Contrairement aux Chapitres 4, 5 et 6, la valeur réelle de l'activité n'est connue qu'aux points de prélèvement et il faut procéder par validation croisée pour évaluer les performances des méthodes d'estimation. Pour cela, les données sont retirées par piézomètre et non par point de prélèvement (retrait des quatre niveaux d'un piézomètre à la fois). Les critères suivants sont calculés :

- le classement des méthodes par rapport aux erreurs absolues obtenues en enlevant successivement chaque piézomètre. Pour chaque point de prélèvement et chaque date, les méthodes sont classées de l'erreur absolue la plus faible (rang 1) à l'erreur absolue la plus élevée (rang 4). Ce classement est ensuite moyenné par date ;
- les erreurs absolues moyennes en enlevant un ou plusieurs piézomètres à la fois. Ces erreurs sont moyennées sur le nombre de combinaisons possibles de piézomètres observés, correspondant au coefficient binomial $\binom{6}{k}$, où k est le nombre de piézomètres observés.

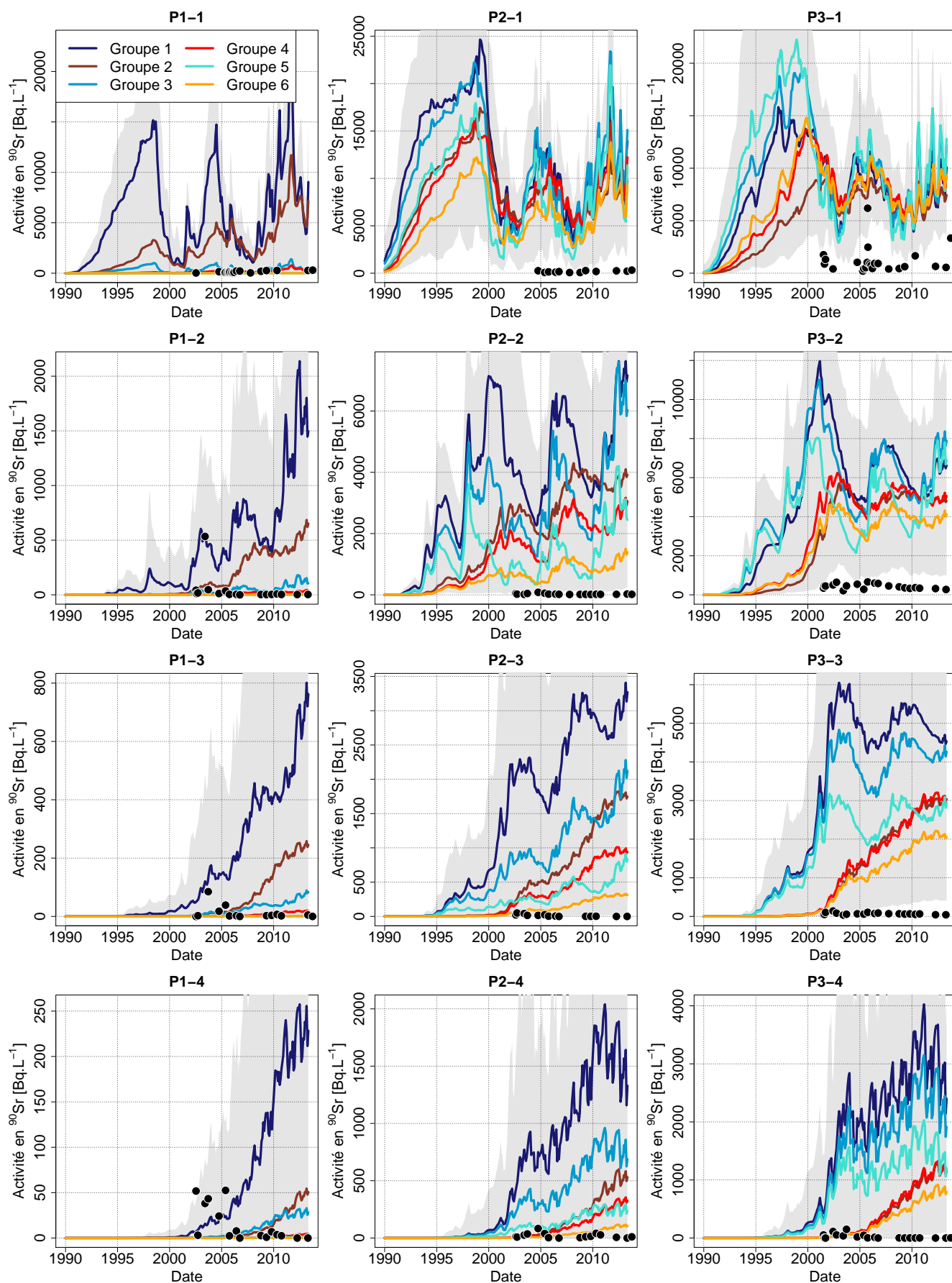


FIGURE 7.12 – Comparaison des observations et des simulations. La zone grisée correspond à l'enveloppe des simulations. Les groupes sont représentés par leur chronique médiane. Les observations sont en noir.

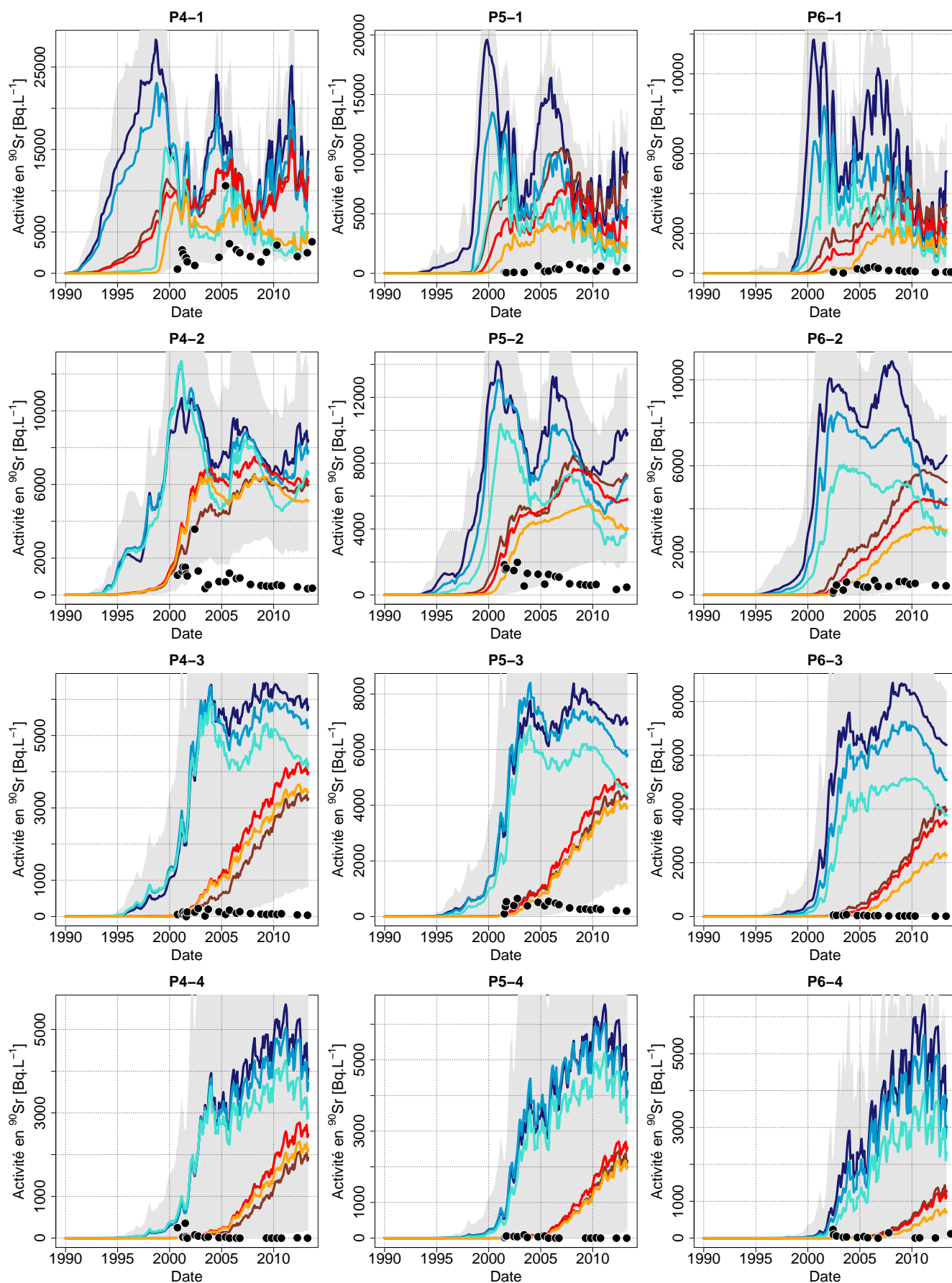


FIGURE 7.12 – Suite.

7.3.2 Résultats

Les formes des panaches sont assez semblables quelle que soit la méthode d'estimation et cohérentes avec le sens d'écoulement de la nappe (Figure 7.13). Les activités estimées sont plus élevées pour le krigeage avec variogrammes numériques que pour les deux autres méthodes, surtout près de la tranchée. Les écart-types théoriques sont proches pour le krigeage ordinaire et le krigeage avec dérive externe, mais différent pour le krigeage avec variogrammes numériques : ils sont nettement plus élevés à proximité de la tranchée pour le krigeage avec variogrammes numériques, mais ne présentent pas d'effets de bord sur les limites latérales et inférieures du domaine. Pour le krigeage ordinaire et le krigeage avec dérive externe, les écart-types théoriques de krigeage pourraient être corrigés par un effet proportionnel (comme au Chapitre 4).

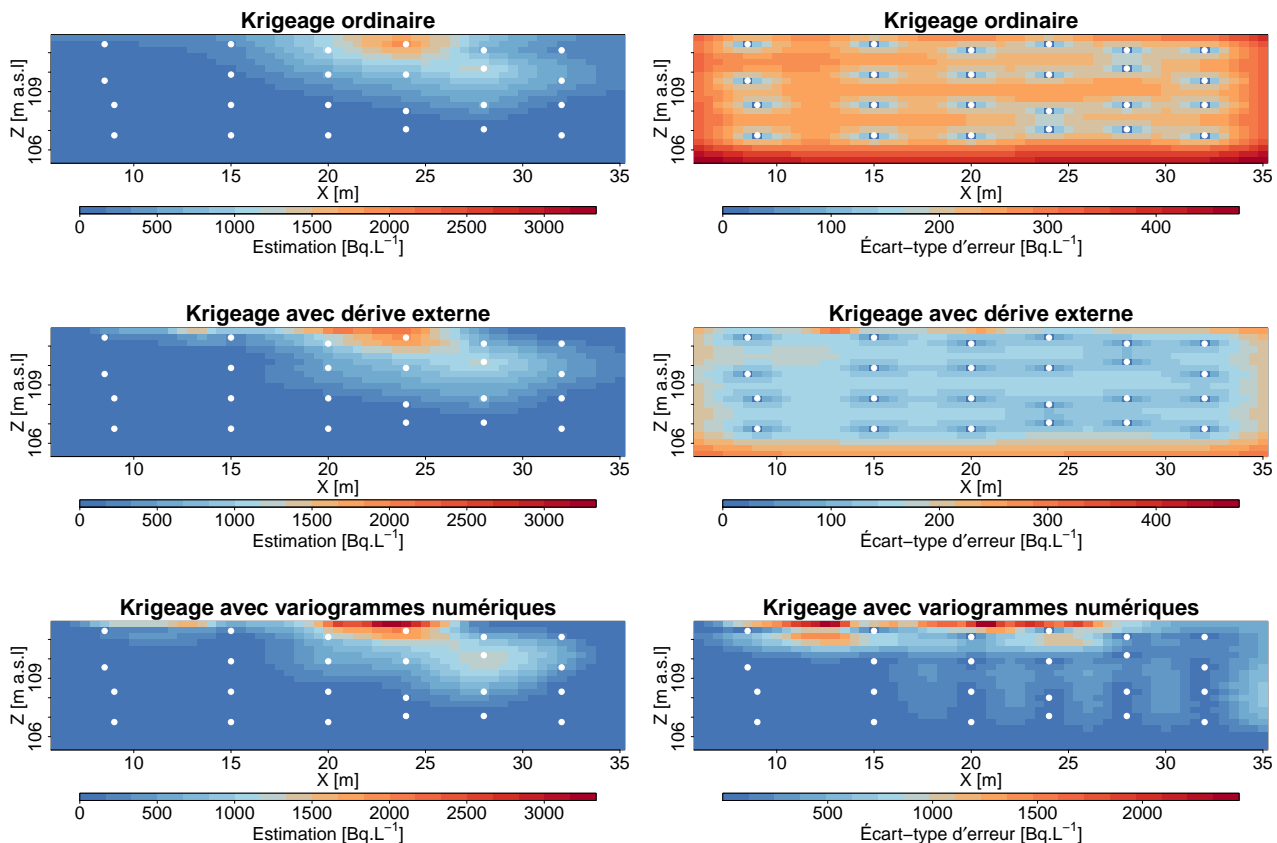


FIGURE 7.13 – Cartes d'estimation (colonne de gauche) et cartes d'écart-type théorique d'erreur d'estimation (colonne de droite) pour les trois méthodes, au 30/09/2004. Les échelles de couleurs sont identiques pour les cartes d'estimations mais différent pour les cartes d'écart-type. Les points blancs rappellent la localisation des observations.

Les profils estimés montrent que les performances des méthodes varient d'un piézomètre à l'autre (Figure 7.14). Par exemple, le krigeage ordinaire donne des résultats satisfaisants aux piézomètres P1 et P2. En revanche, aux piézomètres P5 et P6, les profils estimés par krigeage avec variogrammes numériques sont en meilleure adéquation avec les observations.

Le classement par rang du Tableau 7.5 ne permet pas de discriminer facilement les méthodes : sur les rangs globaux, le krigeage avec dérive externe apparaît légèrement meilleur, mais toutes les méthodes sont assez proches. Cependant, si l'on considère chaque date indépendamment, les écarts entre méthodes sont plus importants. Par exemple, pour le 18/05/2005, le krigeage avec variogrammes numériques est assez nettement plus performant que les deux autres méthodes. Le 01/10/2006 et le 23/04/2010, le krigeage avec dérive externe est plus performant que les autres méthodes.

Le dernier critère calculé est l'erreur absolue moyenne en fonction du nombre de piézomètres disponibles. Comme attendu, ces erreurs absolues moyennes diminuent lorsque le nombre de piézomètres observés augmente (Figure 7.15). Lorsque seuls deux ou trois piézomètres sont observés, les écarts entre

le krigeage ordinaire et les méthodes combinant géostatistique et simulations numériques sont plus importants que lorsque quatre ou cinq piézomètres sont observés. Cela souligne l'intérêt des méthodes proposées lorsque peu de données sont disponibles. Cette information est synthétisée pour toutes les dates d'intérêt dans le Tableau 7.6. Ce dernier critère permet de mieux discriminer les méthodes. En effet, quel que soit le nombre de piézomètres observés, le krigeage ordinaire est la méthode la plus mal classée.

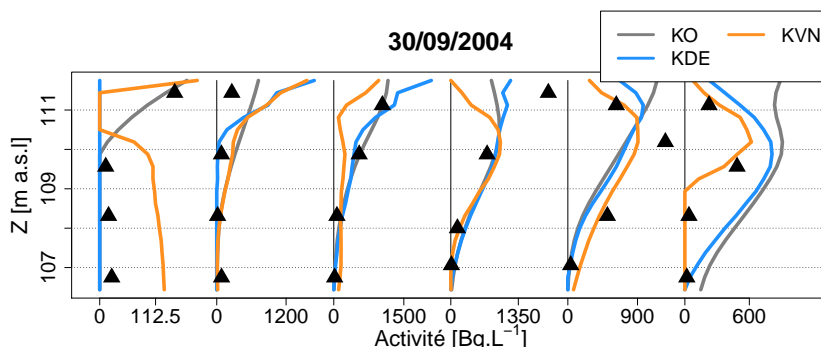


FIGURE 7.14 – Profils estimés le 30/09/2004. Les six piézomètres sont représentés d'amont en aval (de gauche à droite). Les observations sont représentées par les triangles noirs.

TABLEAU 7.5 – Rang moyen de chaque méthode à chaque date.

	30/09/2004	18/05/2005	01/10/2005	01/10/2006	23/04/2010	Rang global
KO	2	2	2.1	1.8	1.9	2
KDE	1.8	2	1.8	1.6	1.6	1.8
KVN	2	1.6	1.7	2.2	2.2	1.9

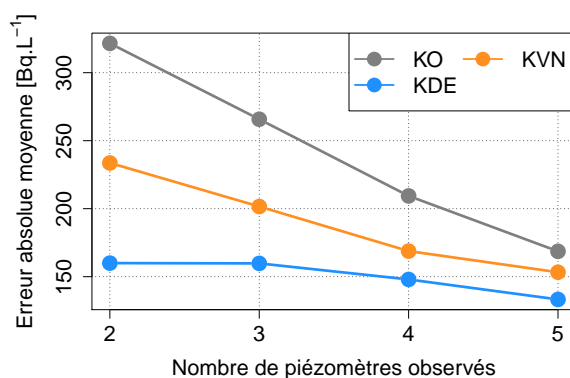


FIGURE 7.15 – Erreurs absolues moyennes en fonction du nombre de piézomètres observés (erreurs moyennées sur le nombre de scénarios pour lesquels deux, trois, quatre ou cinq piézomètres sont observés) le 30/09/2004.

TABLEAU 7.6 – Rang moyen de chaque méthode en fonction du nombre de piézomètres observés, toutes dates confondues.

	2 piézo.	3 piézo.	4 piézo.	5 piézo.	Rang global
KO	3	3	3	2.4	2.8
KDE	1.4	1.8	1.6	1.6	1.6
KVN	1.6	1.2	1.4	2	1.6

7.3.3 Discussion

Apport des simulations

Bien que les simulations d'écoulement et de transport surestiment fortement les observations, elles reproduisent convenablement la structure spatiale du panache de ^{90}Sr . Ainsi, leur prise en compte améliore les estimations. Par ailleurs, comme pour le cas synthétique, la combinaison des simulations d'écoulement et de transport et de la géostatistique est d'autant plus intéressante que le nombre de piézomètres observés est faible. Cela justifie le coût des simulations. Cependant, il n'est pas évident de conclure quant à la robustesse des méthodes : en effet, leur classement change d'une date à l'autre. D'autre part, la distinction des simulations en six groupes (Section 7.2.2) n'a pas été exploitée pour obtenir ces résultats : l'utilisation d'un seul des six groupes ne permet pas d'améliorer les estimations, même s'il s'agit du groupe de simulations les plus proches des observations. Cela s'explique par deux raisons : en n'utilisant qu'un seul groupe, on réduit nettement le nombre de simulations disponibles. De plus, la variabilité entre simulations diminue et devient trop faible pour fournir de bons résultats.

Par ailleurs, beaucoup de données n'ont pas été exploitées : seules les dates auxquelles tous les piézomètres ont été prélevés sont ici considérées. Plus de résultats auraient pu être obtenus si les piézomètres avaient tous été prélevés aux mêmes dates, et non à quelques mois d'intervalle comme c'est souvent le cas. Pour exploiter toutes les données, on se placera dans un cadre temporel (Section 7.4).

Enfin, cet exemple montre que les méthodes mises en place en zone non saturée (cas synthétique et expérience de traçage) sont également intéressantes en zone saturée. Si des données récentes étaient disponibles, une estimation de l'activité en ^{90}Sr dans la nappe pourrait être fournie, afin d'aider aux réflexions sur des stratégies de remédiation ou de surveillance du site.

Comparaison du krigeage avec dérive externe et du krigeage avec variogrammes numériques

Les résultats du Tableau 7.6 ne permettent pas de conclure sur la meilleure méthode à employer entre krigeage avec dérive externe et krigeage avec variogrammes numériques. Pour les comparer plus précisément, deux autres indicateurs sont étudiés :

- le nombre de simulations nécessaires à l'obtention d'un résultat robuste.
- la qualité de l'estimation sur une simulation tirée dans le jeu de simulations et randomisée. Cela revient à se placer dans le cas où tous les processus physico-chimiques sont correctement modélisés et où le modèle d'écoulement et de transport est bien calibré, à quelques incertitudes près.

Pour étudier l'influence du nombre de simulations, les estimations ont été réalisées en tirant aléatoirement N simulations dans le jeu (entre 20 et 300). Pour chaque valeur de N , 50 tirages sont réalisés. Les résultats sont présentés pour le 30/09/2004 et le 01/10/2005 sur la Figure 7.16. Les erreurs obtenues par krigeage avec dérive externe sont presque indépendantes du nombre de simulations, alors que celles obtenues par krigeage avec variogrammes numériques diminuent nettement lorsque le nombre de simulations augmente (surtout pour moins d'une centaine de simulations). La variabilité des erreurs obtenues par krigeage avec variogrammes numériques est également plus importante lorsque le nombre de simulations est faible. Ainsi, lorsque peu de simulations sont disponibles (moins de 100) l'estimation par krigeage avec dérive externe est plus fiable que celle par krigeage avec variogrammes numériques, même si l'erreur moyenne par krigeage avec variogrammes numériques peut être plus faible ; cela confirme les résultats obtenus sur le cas synthétique (Chapitre 4). À très faible nombre de simulations (20 à 40), le krigeage ordinaire donne de meilleures estimations que le krigeage avec variogrammes numériques.

Pour le second indicateur, 60 simulations de référence sont tirées aléatoirement dans le jeu de simulations. Elles sont supposées connues aux 24 points de prélèvement et aux cinq dates d'intérêt. Chaque référence est estimée par krigeage avec dérive externe et krigeage avec variogrammes numériques en utilisant 100 autres simulations. Les erreurs absolues moyennes sont obtenues par validation croisée, en enlevant un piézomètre à la fois. En moyenne, les erreurs absolues sont plus faibles pour le krigeage avec variogrammes numériques que pour le krigeage avec dérive externe (Figure 7.17). L'écart interquartile est également plus faible pour le krigeage avec variogrammes numériques. Cette expérience montre que lorsque le modèle d'écoulement et de transport reproduit la réalité de manière fiable, le krigeage avec variogrammes numériques est préférable au krigeage avec dérive externe, même avec peu de simulations.

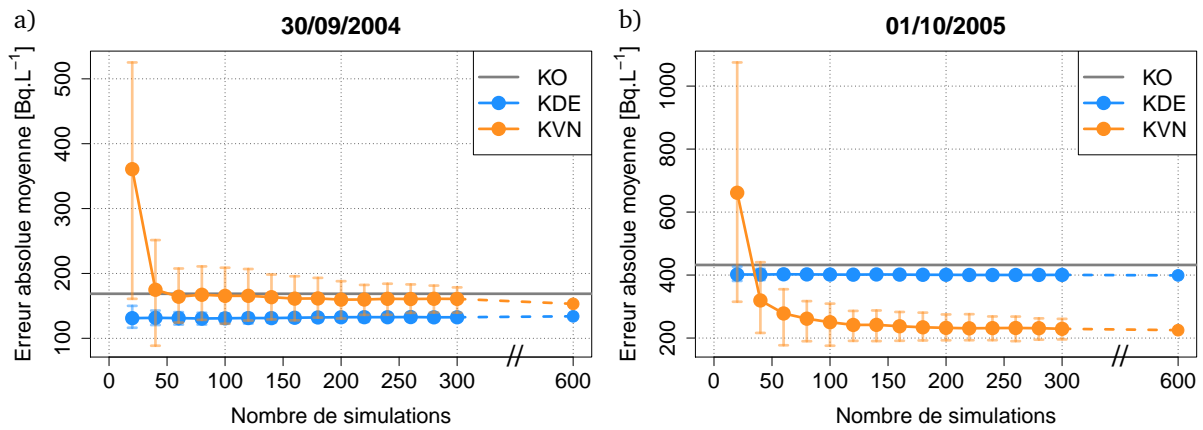


FIGURE 7.16 – Erreur absolue moyenne en fonction du nombre de simulations utilisées. a) Le 30/09/2004 ; b) le 01/10/2005. Les points représentent la moyenne des 50 tirages à nombre de simulations fixé, et les barres d'erreurs donnent les valeurs du premier et du neuvième décile. L'erreur obtenue par krigeage ordinaire est rappelée.

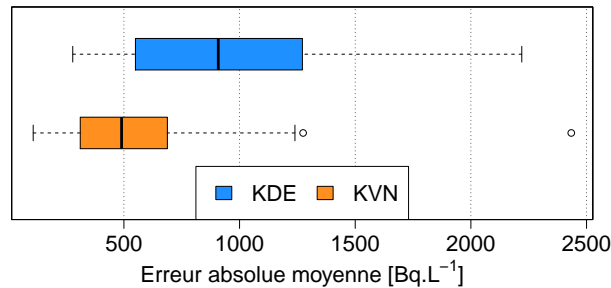


FIGURE 7.17 – Répartition des erreurs absolues moyennes sur 60 simulations tirées aléatoirement et estimées par krigeage avec dérive externe et krigeage avec variogrammes numériques, pour le 30/09/2004. À noter que ces erreurs sont nettement supérieures à celles obtenues sur le jeu de données réelles ; cela est dû aux fortes valeurs présentées par les simulations d'écoulement et de transport par rapport aux observations.

Lorsque le modèle à base physique ne reproduit pas parfaitement la réalité, ce qui est le cas ici, il pourrait être intéressant de mettre en œuvre la méthode de *physics-informed cokriging* proposée par [Yang et al. \[2019\]](#). Cette méthode permet d'améliorer le krigeage avec covariances numériques, en prenant en compte à la fois les covariances numériques obtenues à partir de simulations et la covariance des observations. Elle est basée sur des approches multi-fidélité classiques [[Kennedy et O'Hagan, 2000](#)] (introduction d'une corrélation entre des sorties de codes haute-fidélité et basse-fidélité). L'information haute-fidélité est alors constituée par les observations et l'information basse-fidélité par les simulations d'écoulement et de transport. Les tests effectués montrent que cette méthode est prometteuse sur ce jeu de données, puisque l'on a une bonne corrélation entre les observations et les simulations.

Variantes du krigeage avec variogrammes numériques

Dans le Chapitre 4, trois variantes du krigeage avec variogrammes numériques ont été mises en place. Elles sont également testées ici :

1. le krigeage avec covariances numériques donne des résultats légèrement moins bons que le krigeage avec variogrammes numériques ;
2. la modification des conditions de non-biais pour prendre en compte la moyenne des simulations n'améliore pas de manière évidente les estimations ;
3. la régularisation des variogrammes numériques (par moyenne sur les points voisins) ne permet pas d'améliorer de manière sensible ni systématique les estimations. En revanche, la randomisation de l'origine de la grille permet d'améliorer sensiblement les estimations, pour toutes les dates sauf le

01/10/2005 (Figure 7.18). Cependant, cette approche ne permet pas de diminuer la variabilité de l'erreur à faible nombre de simulations, ce qui en limite l'intérêt.

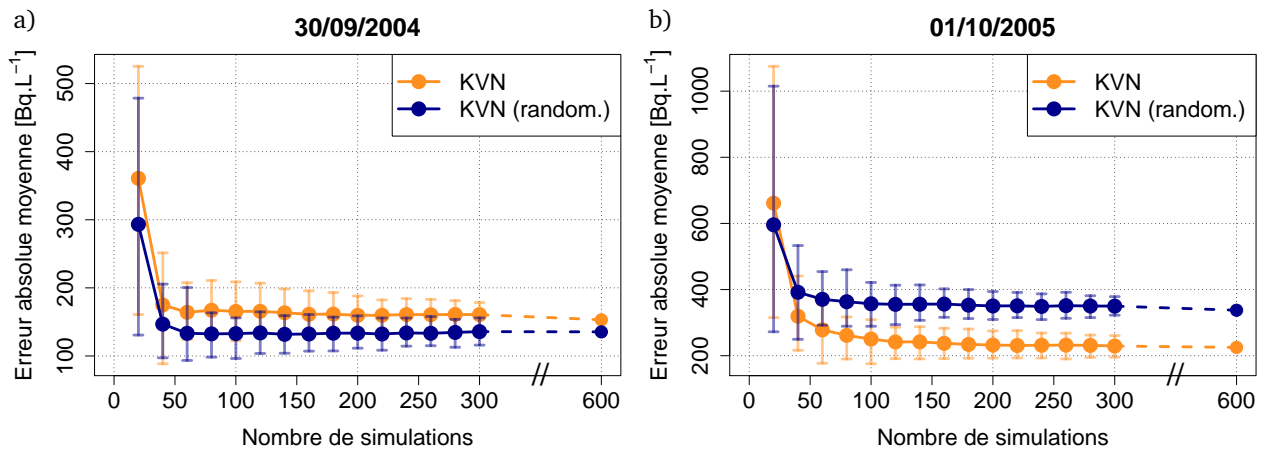


FIGURE 7.18 – Erreur absolue moyenne en fonction du nombre de simulations utilisées, avec ou sans randomisation de l'origine de la grille des simulations. a) Le 30/09/2004 ; b) le 01/10/2005. Les points représentent la moyenne des cinquante tirages à nombre de simulation fixé, et les barres d'erreurs donnent les valeurs du premier et du neuvième décile.

7.4 Combinaison du krigeage et de simulations d'écoulement et de transport : cadre temporel

Les estimations précédentes permettent de tracer des cartes de l'activité en ^{90}Sr dans la nappe à des dates données. Cependant, leur intérêt opérationnel est limité, car trop peu de données sont disponibles pour établir des cartes actuelles, voire prévisionnelles. Dans cette section, les méthodes d'estimation sont mises en place dans un cadre temporel (reconstruction des chroniques d'activité en une localisation précise). Ces estimations sont ensuite utilisées pour évaluer le flux d'activité en ^{90}Sr passé par quelques sections de la nappe.

7.4.1 Recherche d'une modélisation adaptée

Position du problème

On se place dans un cadre exclusivement temporel, c'est-à-dire que les chroniques d'activité à chaque point de prélèvement sont estimées indépendamment les unes des autres. Dans ce cadre, pour chaque point de prélèvement x , les variogrammes numériques entre deux dates t et t' sont calculés comme suit :

$$\gamma_x^{\text{num}}(t, t') = \frac{1}{P} \sum_{p=1}^P \frac{1}{2} [Z_p(x, t) - Z_p(x, t')]^2 \quad (7.5)$$

Contrairement au cas spatial, l'utilisation de ces variogrammes dans un système de krigeage ordinaire ne conduit pas à des estimations satisfaisantes. En effet, les systèmes sont mal conditionnés, ce qui conduit à des valeurs d'activité estimées parfois extrêmement élevées ou à des fluctuations très rapides. Cela semble incohérent au vu de la dynamique du système hydrogéologique étudié.

Krigeage avec variogrammes numériques modifiés

Puisque les matrices de variogrammes numériques (ou de covariances) manquent de stabilité, des méthodes d'estimation robuste de matrices aléatoires, comme la méthode de Ledoit-Wolf [Ledoit et Wolf,

2004] (Complément 7.5), ont été adaptées au cas étudié ici. Le prior choisi est un modèle de variogramme linéaire, proportionnel à la moyenne des simulations :

$$\gamma_x^p(t, t') = K \cdot \overline{Z(x, t)} \cdot \overline{Z(x, t')} |t - t'| \quad (7.6)$$

où le coefficient K est ajusté sur le variogramme expérimental des données. Le modèle de variogramme numérique modifié est le suivant :

$$\gamma_x^{LW}(t, t') = (1 - \alpha) \gamma_x^{num}(t, t') + \alpha \gamma_x^p(t, t') \quad (7.7)$$

où le coefficient α est ajusté par validation croisée.

Estimation robuste de Ledoit-Wolf (*shrinkage*)

Même si la matrice de covariance empirique $\hat{\Sigma}$ est un bon estimateur de la covariance réelle, son inversion est parfois imprécise, voire impossible, pour des raisons numériques. Une méthode de transformation de la matrice est proposée par Ledoit et Wolf [2004]. Elle consiste à réduire le rapport entre la plus grande et la plus petite valeur propre de la matrice de covariance :

$$\Sigma_{LW} = (1 - \alpha) \hat{\Sigma} + \alpha \frac{\text{Tr}(\hat{\Sigma})}{n} I_n \quad \text{avec} \quad 0 \leq \alpha \leq 1 \quad (7.8)$$

où α est le paramètre de *shrinkage* à ajuster (pour conserver un bon équilibre entre le biais introduit et la réduction de variance) et n est le nombre de colonnes de la matrice $\hat{\Sigma}$. Cette méthode correspond à une moyenne pondérée entre la matrice de covariance empirique et un prior (matrice diagonale). Cependant, l'application directe de cette méthode à la matrice des covariances numériques obtenue à partir des simulations MELODIE revient à effectuer une estimation avec un effet de pépité presque pur. Un autre prior a donc été recherché (Équation (7.6)).

Application sur une référence connue

Puisque les données de validation sont peu nombreuses, cette méthode est d'abord validée sur une référence connue. Cette référence est une simulation MELODIE tirée dans le jeu et modifiée par ajout d'une perturbation gaussienne structurée (covariance gaussienne) dans le temps, de longueur de corrélation d'un an (Figure 7.19). Cette modification permet d'ajouter une variabilité qui n'est pas simulée par le modèle d'écoulement et de transport.

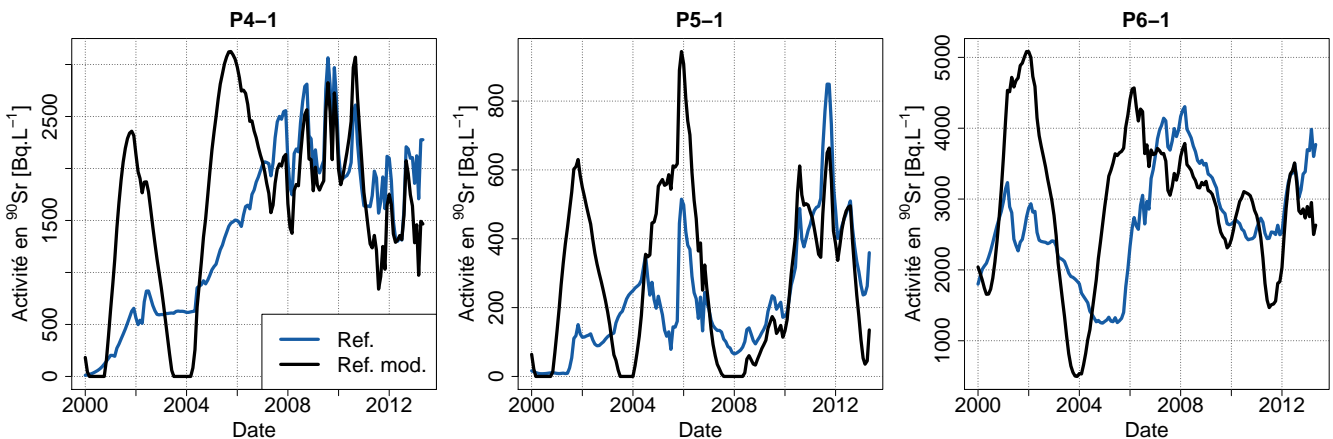


FIGURE 7.19 – Exemple d'une simulation (Ref.) modifiée (Ref. mod.) aux points P4-1, P5-1, P6-1.

Cinq références sont ainsi générées et les chroniques sont estimées aux 24 points d'intérêt par trois méthodes : le krigeage avec variogrammes numériques (KVN), le krigeage avec variogrammes numériques

modifiés (KVN*) et le krigeage avec dérive externe (KDE), avec la moyenne des simulations non modifiées en dérive. Le krigeage ordinaire est exclu car il conduit à des chroniques beaucoup trop lissées.

Les résultats pour une référence et six points particuliers (niveaux 1 et 2 de P4, P5 et P6, qui présentent les plus fortes valeurs d'activité sur les données) sont présentés sur la Figure 7.20. Les fluctuations rapides des chroniques estimées par krigeage avec variogrammes numériques sont par exemple visibles aux points P4-1 et P5-1. Les erreurs absolues moyennes obtenues en considérant les 24 points, pour chacune des cinq références, sont données dans le Tableau 7.7. Quelle que soit la référence, les erreurs les plus faibles sont obtenues par krigeage avec dérive externe. La modification des variogrammes numériques permet d'améliorer les résultats.

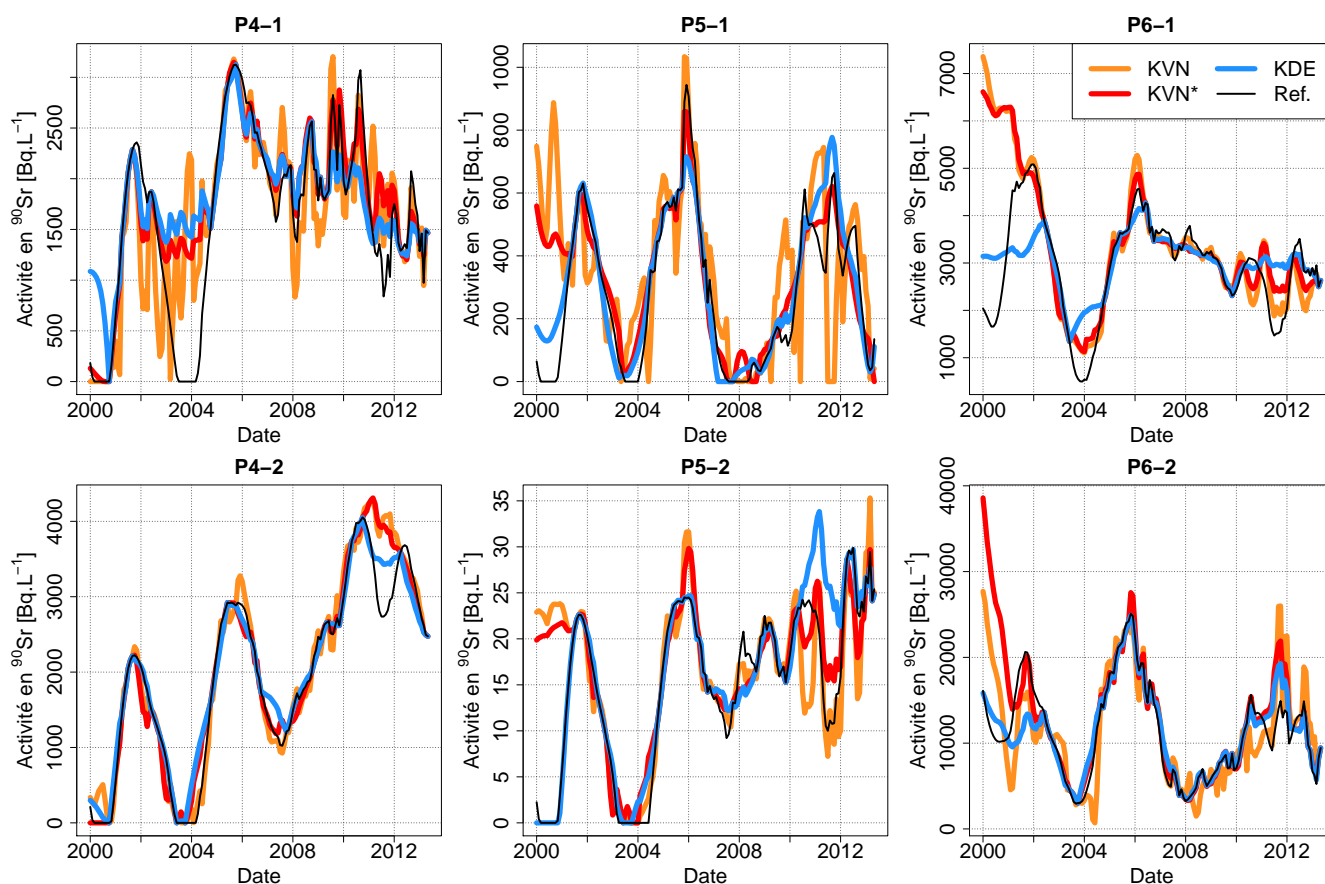


FIGURE 7.20 – Résultats d'estimation sur une simulation modifiée pour six points particuliers.

TABEAU 7.7 – Erreurs absolues moyennes [Bq.L^{-1}] obtenues pour les trois méthodes sur cinq références.

	KVN	KVN*	KDE
Test 1	1260	699	661
Test 2	1195	863	523
Test 3	812	650	543
Test 4	2325	1441	1263
Test 5	810	621	579

Il est à noter que la modification des conditions de non-biais pour prendre en compte la non-stationnarité de la moyenne ne modifie pas les résultats d'estimation. Il en est de même pour la régularisation des variogrammes numériques par moyenne sur les plus proches voisins et pour la sélection de certaines simulations (en ne prenant par exemple que celles avec le coefficient de retard le plus faible).

Au vu des résultats obtenus sur les références modifiées, seuls le krigeage avec dérive externe et le krigeage avec variogrammes numériques modifiés sont mis en œuvre sur le jeu de données réelles.

7.4.2 Estimation des chroniques aux 24 points de prélèvement

Les chroniques sont estimées par les deux méthodes aux 24 points de prélèvement entre 2000 et 2013 (Figure 7.21) avec un pas de temps mensuel, correspondant au pas de sortie des simulations MELODIE. Sur certains points, comme par exemple P2-4, P4-1 ou P6-1, les deux méthodes donnent des résultats assez proches. Pour d'autres points, comme par exemple P2-1, P3-1 ou P5-3, le krigeage avec dérive externe lisse très fortement la chronique, ce qui paraît peu réaliste. Par ailleurs, les écart-types théoriques d'erreur de krigeage sont en général plus faibles pour le krigeage avec variogrammes numériques modifiés que pour le krigeage avec dérive externe.

Pour comparer les performances des méthodes, les erreurs absolues moyennes sont calculées par validation croisée, en retirant un tiers des observations à la fois. Ces erreurs sont ensuite moyennées sur les quatre niveaux d'un piézomètre (Tableau 7.8). Quel que soit le piézomètre, les erreurs moyennes sont plus faibles pour le krigeage avec dérive externe que pour le krigeage avec variogrammes numériques modifiés. Cependant, les chroniques obtenues par krigeage avec dérive externe semblent parfois moins réalistes que par krigeage avec variogrammes numériques modifiés, c'est pourquoi les deux résultats sont conservés dans la suite. La robustesse des méthodes est complexe à évaluer, étant donné le faible nombre de données disponibles à chaque point de prélèvement.

TABLEAU 7.8 – Erreurs absolues moyennes [Bq.L⁻¹] par piézomètre obtenues par validation croisée.

	P1	P2	P3	P4	P5	P6	Moyenne
KDE	39	27	258	533	138	81	179
KVN*	41	39	411	741	173	102	251

7.4.3 Estimation des flux d'activité par piézomètre

Les résultats précédents sont utilisés pour estimer la quantité de ⁹⁰Sr ayant traversé chaque piézomètre au cours du temps. Pour cela, les estimations obtenues entre les quatre niveaux de prélèvement d'un même piézomètre sont interpolées linéairement, pour obtenir une estimation de l'activité en ⁹⁰Sr le long d'un profil. À partir de l'estimation de l'activité en ⁹⁰Sr sur chaque maille d'un piézomètre, le flux à la date t passant par la section du piézomètre P_j est calculé comme suit :

$$F_{P_j}(t) = \sum_{i=1}^{n_{P_j}(t)} Z_i^*(t) K_{s_i} \frac{dh}{dx} dz \quad (7.9)$$

où $n_{P_j}(t)$ [-] est le nombre de mailles le long du piézomètre P_j qui sont sous le niveau piézométrique à la date t , $Z_i^*(t)$ [T⁻¹.L⁻³] est l'activité volumique estimée au point i à la date t , K_{s_i} [L.T⁻¹] est la conductivité hydraulique à saturation moyenne à la profondeur du point i , $\frac{dh}{dx}$ [L.L⁻¹] est le gradient hydraulique et dz [L] est la hauteur d'une maille (Tableau 7.9).

TABLEAU 7.9 – Paramètres utilisés pour calculer le flux passant par chaque piézomètre.

$n_{P_j}(t)$ [-]	K_{s_i} [m.s ⁻¹]	$\frac{dh}{dx}$ [m.m ⁻¹]	dz [m]
entre 20 et 25	entre 1.10 ⁻⁴ et 1.10 ⁻⁶	0,0015	0,3125

Les flux cumulés ainsi calculés sont donnés pour les trois piézomètres les plus à l'aval (P4-P5-P6) sur la Figure 7.22. Ils diffèrent peu selon la méthode d'estimation des chroniques. Le flux le plus important est obtenu au piézomètre P4, et l'on observe que seule une fraction de l'activité passée par le piézomètre P4 se retrouve en P5 et P6 (flux divisé par deux de P4 à P5 puis de P5 à P6). Cela souligne qu'une partie de l'activité est retenue et stockée par le milieu entre les piézomètres. Cette rétention n'est pas prise en compte dans le modèle d'écoulement et de transport utilisé, ce qui explique en partie que les simulations surestiment nettement les activités mesurées. Par ailleurs, la simulation est réalisée sur un domaine à deux dimensions borné, et les éventuelles pertes d'activité sous le domaine de modélisation ou perpendiculairement à la section (contournement de P5 et P6 par exemple) ne sont pas prises en compte.

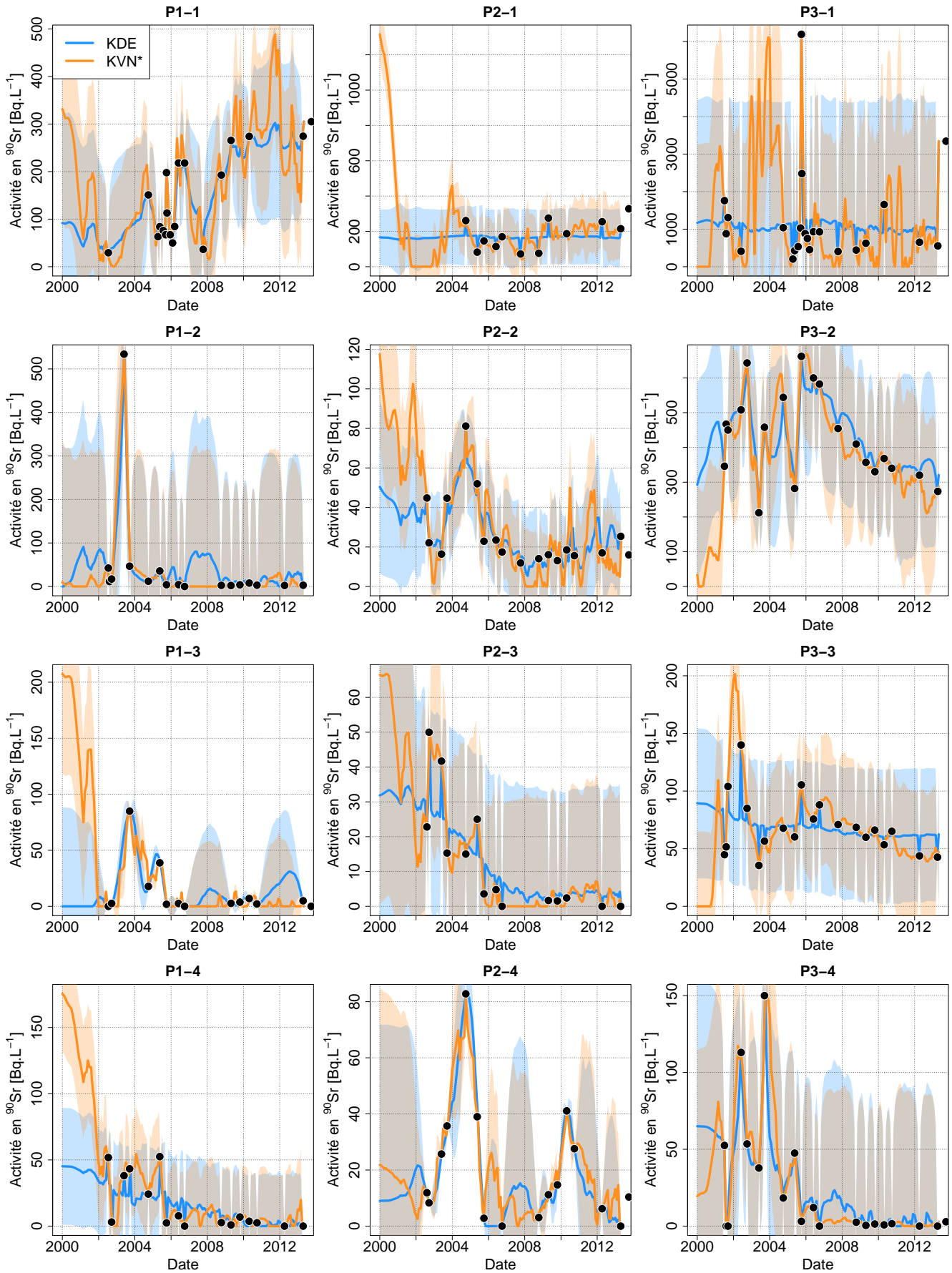


FIGURE 7.21 – Estimation de l'activité en ^{90}Sr dans la nappe aux 24 points de prélèvement. Les zones colorées indiquent les intervalles $[Z^*(t) - 2\sigma_K(t); Z^*(t) + 2\sigma_K(t)]$.

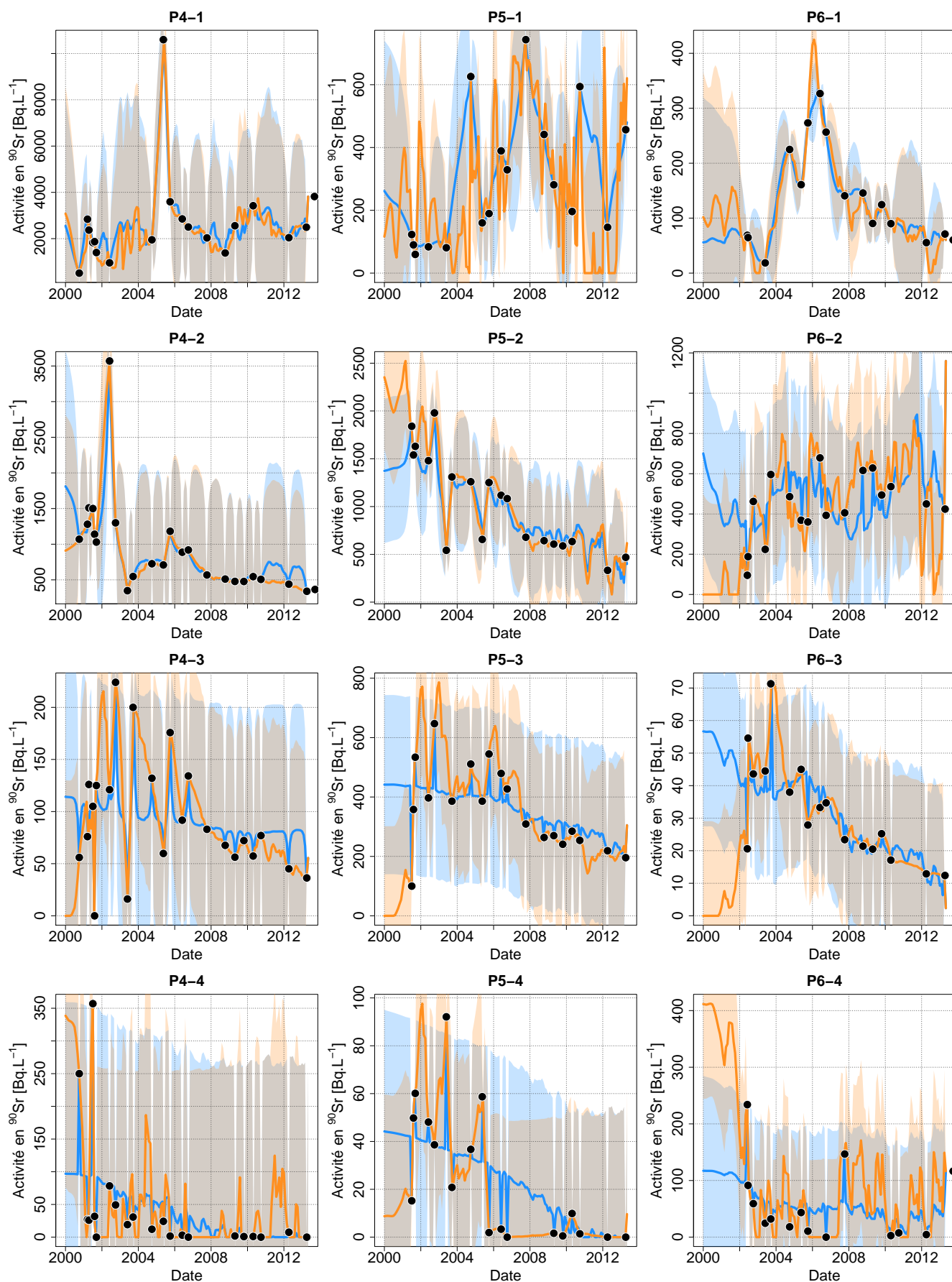


FIGURE 7.21 – Suite.

Le flux cumulé ayant traversé la section du piézomètre P4 entre 2000 et 2013 représente environ 1% du stock total initial de ^{90}Sr et 2% du stock mobilisable. Cet ordre de grandeur est légèrement inférieur à celui de la diminution du stock calculée sur la période 1987-2000 par Kashparov *et al.* [2012], de $7 \pm 5\%$.

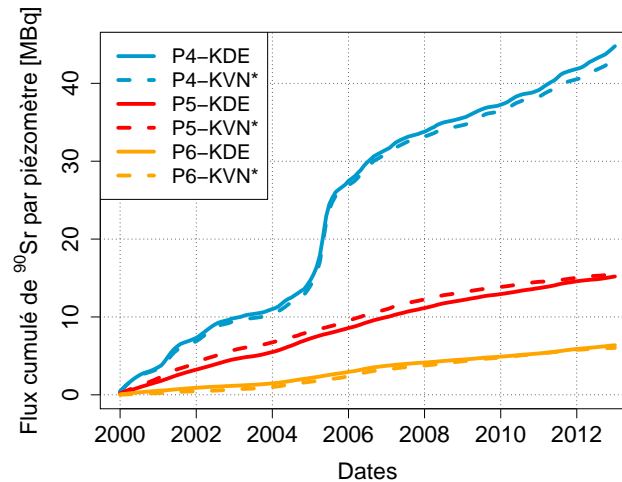


FIGURE 7.22 – Flux cumulés entre 2000 et 2013 sur la hauteur des piézomètres P4, P5 et P6.

7.4.4 Discussion

Pour évaluer les flux, les estimations de chroniques d'activités obtenues en 24 points, indépendamment les unes des autres, ont été combinées par interpolation linéaire. D'autres méthodes auraient pu être envisagées.

Krigeage spatio-temporel

Bien que performant sur le cas synthétique du Chapitre 5, le krigeage spatio-temporel avec variogrammes numériques ne donne pas de résultats satisfaisants sur ce jeu de données. Les cartes d'activité obtenues ne présentent pas une structure cohérente, physiquement réaliste, et les systèmes à résoudre sont numériquement instables. De plus, les différences d'échantillonnage dans le temps et l'espace, ainsi que le grand nombre de données non synchrones, rendent la modélisation complexe. Les méthodes spatio-temporelles ont donc été écartées. Cependant, le krigeage spatio-temporel avec variogrammes numériques a quand même été mis en œuvre sur une référence issue du jeu de simulations et donne de bons résultats. Ainsi, si le modèle d'écoulement et de transport reproduit de manière très satisfaisante la réalité, le krigeage spatio-temporel peut être une méthode performante.

Cela souligne l'intérêt de pouvoir acquérir des données plus régulièrement, dans l'espace et surtout dans le temps, pour construire des modèles plus complexes, mais également valider les modèles actuels.

Variables utiles

Plutôt que d'estimer les chroniques aux quatre niveaux de chaque piézomètre, puis de les combiner pour calculer les flux, le krigeage pourrait être réalisé directement sur les flux. En effet, les simulations d'écoulement et de transport permettent de calculer les flux à n'importe quelle section, et par conséquent des variogrammes numériques ou des fonctions de dérive. Cependant, le problème des données non synchrones se pose à nouveau : les mesures d'activité en ^{90}Sr sont rarement disponibles à la même date pour les quatre niveaux d'un piézomètre.

7.5 Synthèse du chapitre

Les méthodes combinant la géostatistique et les simulations d'écoulement et de transport ont été mises en œuvre sur un jeu de données réel, constitué par des mesures d'activité en ^{90}Sr dans la nappe sous-jacente à une tranchée d'enfouissement de déchets radioactifs. Un modèle d'écoulement et de transport a été construit à partir de données bibliographiques et d'études réalisées précédemment à l'IRSN. Ce modèle simule la migration de ^{90}Sr à partir de la tranchée, entre 1987 et 2013, à un pas de temps mensuel. Il reproduit convenablement la dynamique d'évolution de l'activité en ^{90}Sr mais surestime nettement les activités mesurées. Six cents simulations ont été générées à partir de ce modèle, en prenant en compte une partie des incertitudes de modélisation (stock initial, coefficient de retard, *etc.*).

Même si les simulations surestiment les activités mesurées, elles reproduisent convenablement la structure spatiale du panache de ^{90}Sr , et par conséquent leur utilisation améliore les estimations par rapport au krigeage ordinaire classique, d'autant plus que le nombre de piézomètres disponibles est faible. Ainsi, l'intérêt des méthodes combinant la géostatistique et les simulations d'écoulement et de transport est mis en évidence sur un cas réel pour lequel la modélisation ne reproduit pas aussi bien la réalité que dans le cas du traçage (Chapitre 6) et dans les cas synthétiques de la Partie II. Par ailleurs, l'estimation est ici réalisée dans la zone saturée.

Dans un cadre temporel, le krigeage avec variogrammes numériques n'est pas aussi performant que dans un cadre spatial. Cependant, une modification de la méthode donne des estimations satisfaisantes, utilisées pour estimer la quantité de ^{90}Sr ayant traversé un piézomètre au cours du temps. Les ordres de grandeurs obtenus sont cohérents avec la littérature.

Finalement, même si la modélisation à base physique ne semble pas suffisamment maîtrisée sur cet exemple, elle permet d'obtenir des résultats intéressants. Avec les informations disponibles sur le site et le code utilisé, l'amélioration de ce modèle apparaît complexe. Par exemple, de nombreux phénomènes pouvant influencer la migration du ^{90}Sr ne sont pas pris en compte (prélèvement racinaire, périodes d'enneigement, inondations, *etc.*), mais peu de données sont disponibles sur ces phénomènes. Par ailleurs, les résultats d'estimation de flux suggèrent la rétention d'une partie du ^{90}Sr entre les piézomètres P4 et P5. Pour modéliser correctement cette rétention, le coefficient de retard pourrait être augmenté. Cependant, une augmentation trop importante du coefficient de retard conduit à un lissage très fort du signal simulé et donc à une perte de la variabilité des valeurs d'activité en un point donné. Les simulations ne reproduiraient alors plus la variabilité temporelle observée au sein des données. Ainsi, le modèle, par sa construction, ne peut pas simuler correctement les niveaux d'activité et les fluctuations temporelles d'activité. Pour cela, il faudrait par exemple travailler avec un coefficient de retard variable dans l'espace.

De plus, une augmentation de la densité de mesure d'activité en ^{90}Sr , dans l'espace et dans le temps, permettrait de mieux valider les modélisations actuelles, voire d'obtenir des modélisations plus fiables.

Conclusion

Rappel des objectifs et de la démarche

L'objectif de ce travail est de proposer des méthodes permettant d'améliorer la caractérisation de la contamination dans les sols et les nappes autour d'installations industrielles, en combinant l'estimation géostatistique et les simulations d'écoulement et de transport. Le bénéfice attendu de cette combinaison est d'améliorer le réalisme et la précision des estimations géostatistiques, par la prise en compte d'informations de nature physico-chimique *via* des simulations d'écoulement et de transport.

Deux méthodes sont mises en œuvre, sur un cas synthétique puis sur deux cas réels. La première méthode est un krigeage avec dérive externe, où la dérive est donnée par la moyenne d'un jeu de simulations d'écoulement et de transport. La seconde méthode est un krigeage ordinaire, avec des variogrammes numériques calculés à partir d'un jeu de simulations.

Un cas synthétique simulant la migration d'un contaminant dans la zone non saturée d'un aquifère sableux est construit. Des estimations d'activité en tritium sont réalisées, dans des cadres spatiaux (estimation de l'activité à une date fixée) et spatio-temporels (prévision de la migration d'un panache).

Ensuite, les méthodes sont appliquées à deux jeux de données réels, acquis au voisinage de la centrale de Tchernobyl. Le premier jeu de données est constitué par une expérience de traçage au ^{36}Cl . Le second jeu de données est un ensemble de valeurs d'activité en ^{90}Sr , mesurées dans la nappe à l'aval d'une tranchée d'enfouissement de déchets radioactifs. Les activités en ^{90}Sr sont estimées dans un cadre spatial, pour quantifier la contamination dans la nappe, et dans un cadre temporel, pour estimer la quantité de ^{90}Sr ayant migré de la tranchée vers la nappe au cours du temps.

Principaux résultats et discussion

Influence de la variabilité spatiale des paramètres hydrodynamiques de la zone non saturée

L'étude de sensibilité du Chapitre 3 montre qu'à moyennes fixées, la modification des variances (dans des plages réalistes) des champs de paramètres hydrodynamiques (modèle de Mualem-van Genuchten) influence significativement le transport de contaminants en zone non saturée [Pannecoucke *et al.*, 2019]. Ainsi, la seule incertitude sur la variabilité des champs de paramètres de Mualem-van Genuchten (tous autres phénomènes parfaitement modélisés) conduit à une caractérisation imprécise de la contamination. Par exemple, les surfaces moyennes des panaches augmentent de 50% entre un champ de faible variance et un champ de forte variance. Cela souligne l'importance de caractériser la variabilité à petite échelle des paramètres de Mualem-van Genuchten (mise en œuvre de méthodes de mesure *in situ* [Léger *et al.*, 2014 ; Houzé *et al.*, 2019]), puis de prendre en compte cette variabilité dans les modèles d'écoulement et de transport, ce qui n'est pas toujours fait en pratique [Bugai *et al.*, 2012 ; Testoni *et al.*, 2017 ; Kwicklis *et al.*, 2019].

De plus, cette analyse oriente la construction d'un cas synthétique pour lequel la délimitation de la contamination est imprécise : même avec les paramètres de Mualem-van Genuchten comme seules inconnues, les panaches simulés présentent des formes, tailles et localisations significativement différentes, ce qui permet d'évaluer l'intérêt des méthodes de combinaison proposées.

Évaluation des méthodes sur le cas synthétique

Des activités en tritium sont estimées pour comparer les méthodes de combinaison au krigeage ordinaire usuel. Les estimations sont plus précises et plus fiables avec les méthodes de combinaison qu'avec la géostatistique seule : les erreurs d'estimation sont plus faibles, la distribution des valeurs d'activité est mieux reproduite et les panaches sont visuellement plus réalistes [Pannecoucke *et al.*, 2020a]. Cette amélioration est d'autant plus marquée que le nombre d'observations disponibles est faible.

Plus précisément, le krigeage avec variogrammes numériques est plus fiable que le krigeage avec dérive externe lorsque les 1700 simulations d'écoulement et de transport sont utilisées. En revanche, lorsqu'une centaine de simulations sont utilisées, le krigeage avec dérive externe est plus performant. Par ailleurs, le conditionnement par estimation géostatistique (krigeage avec dérive externe ou krigeage avec variogrammes numériques) réduit les erreurs par rapport aux simulations numériques seules.

Trois méthodes de quantification des incertitudes sur les surfaces contaminées et leur activité sont comparées : des méthodes usuelles basées sur les écart-types théoriques d'erreur de krigeage et sur le tracé de cartes de probabilités de dépassement de seuil à partir de simulations géostatistiques [Demougeot-Renard et de Fouquet, 2004 ; de Fouquet *et al.*, 2011 ; Boudreault *et al.*, 2016], et une méthode proposée dans ce travail, consistant à estimer par cokrigeage avec covariances numériques une indicatrice de dépassement de seuil de contamination. Les trois méthodes donnent des résultats équivalents, mais la quantification des incertitudes reste sujette à de nombreux choix arbitraires.

Cela montre par ailleurs que les covariances numériques peuvent être utilisées pour générer des réalisations de fonctions aléatoires, bien que certaines hypothèses restent discutables. Des covariances croisées numériques entre indicatrices de dépassement de seuil et activités en tritium peuvent également être calculées, ce qui est intéressant en présence de non-stationnarités complexes à modéliser.

Ensuite, le krigeage avec covariances numériques est étendu à un cas multivariable, pour estimer les activités de plusieurs radionucléides avec des coefficients de retard différents. À nouveau, l'utilisation de simulations d'écoulement et de transport pour modéliser les structures spatiales non stationnaires, simples et croisées, améliore les estimations.

Enfin, le krigeage avec covariances numériques spatio-temporelles est comparé à deux méthodes de filtre (filtre particulaire et filtre de Kalman d'ensemble), pour prévoir la migration du panache de tritium. Les performances du krigeage avec covariances numériques et du filtre de Kalman d'ensemble sont proches, ce qui est attendu au vu des similarités entre les deux approches : les covariances spatiales utilisées dans l'étape de correction du filtre de Kalman d'ensemble et les covariances spatio-temporelles utilisées dans le krigeage sont calculées à partir du même ensemble. Des tests préliminaires soulignent par ailleurs que le krigeage avec covariances numériques spatio-temporelles peut être utilisé pour rechercher une source de contamination.

Ainsi, ce cas synthétique démontre qu'il est intéressant de combiner la géostatistique et les simulations d'écoulement et de transport pour caractériser une contamination en zone non saturée. Les méthodes mises en œuvre dans la littérature en zone saturée (krigeage avec dérive externe [Rivest *et al.*, 2008], covariances empiriques [Roth *et al.*, 1998 ; Schwede et Cirpka, 2010 ; Yang *et al.*, 2018]) sont ici implémentées en zone non saturée. Le krigeage avec dérive externe et le krigeage avec variogrammes numériques sont comparés précisément sur un exemple d'estimation d'activités en tritium. Le krigeage avec variogrammes ou covariances numériques est étendu à de nombreuses configurations : cas multivariable (plusieurs radionucléides et cokrigeage indicatrice/activité), cas spatio-temporel pour faire des prévisions. De manière générale, l'utilisation de covariances numériques facilite la modélisation des structures spatiales entre une fonction aléatoire et ses transformées, ou dans le cas multivarié. De plus, de nombreuses variantes des méthodes sont étudiées sur ce cas synthétique (influence des conditions de non biais, lissage des variogrammes et covariances numériques, modification des fonctions de dérive). Ce cas synthétique permet enfin d'étudier l'influence du nombre de données et du nombre de simulations numériques sur la qualité des estimations.

Application aux données de Tchernobyl

Ensuite, se pose la question de la transposition des méthodes à un cas réel, pour lequel la modélisation d'écoulement et de transport est moins bien maîtrisée (autres sources d'incertitudes que les seuls champs de paramètres hydrodynamiques).

Le premier cas réel est issu d'une expérience, ce qui permet d'avoir de nombreuses informations sur les conditions hydrogéologiques avant et pendant l'expérience ainsi que de nombreuses mesures d'activité. Un modèle d'écoulement et de transport est construit, à trois dimensions et en zone non saturée, et 150 simulations sont générées ; elles reproduisent convenablement les activités observées. Les résultats d'estimation montrent que le krigeage avec dérive externe et le krigeage avec variogrammes numériques sont plus précis que le krigeage ordinaire ou que les simulations seules. Le krigeage avec dérive externe conduit aux erreurs moyennes les plus faibles, mais le krigeage avec variogrammes numériques conduit à une proportion de points bien estimés plus importante.

Ainsi, ce premier jeu de données permet de valider les méthodes de combinaison de la géostatistique et des simulations numériques sur un cas réel tridimensionnel, en zone non saturée, pour lequel le modèle d'écoulement et de transport est relativement bien contraint et reproduit bien les observations.

Le second cas réel diffère plus du cas synthétique que le premier : il s'agit de mesures d'activité effectuées principalement en zone saturée et dans un environnement plus complexe. En effet, les processus qui gouvernent la migration du ^{90}Sr sont moins bien caractérisés, tout comme la source de contamination. Un modèle d'écoulement et de transport à deux dimensions, reproduisant les principaux phénomènes connus et décrits dans la littérature, est construit et 600 simulations sont générées. Ces simulations surestiment nettement les valeurs d'activité mesurées à l'aval de la tranchée T22. Cependant, elles reproduisent convenablement la variabilité, spatiale et temporelle, des activités mesurées. Ainsi, leur intégration dans une modélisation géostatistique permet d'améliorer les estimations d'activité en ^{90}Sr , à la fois dans un cadre spatial (quantification de la contamination dans la nappe) et dans un cadre temporel (estimation de la quantité de radionucléides ayant traversé une section de la nappe sur une période).

Dans un cadre spatial, les performances du krigeage avec dérive externe et du krigeage avec variogrammes numériques dépendent de la date à laquelle l'activité dans la nappe est estimée. Cependant, les estimations réalisées par krigeage avec dérive externe sont plus robustes : elles dépendent peu du nombre de simulations. Ainsi, même si le krigeage avec variogrammes numériques conduit parfois à des erreurs plus faibles, le krigeage avec dérive externe apparaît ici plus fiable. En revanche, si l'on suppose que le modèle d'écoulement et de transport reproduit presque parfaitement la réalité (estimation d'une référence tirée parmi les 600), le krigeage avec variogrammes numériques est plus précis que le krigeage avec dérive externe.

Synthèse et recommandations

Les résultats obtenus sur le cas synthétique et les cas réels montrent l'intérêt de la combinaison de l'estimation géostatistique et des simulations d'écoulement et de transport, pour caractériser des contaminations dans diverses situations : zone saturée et zone non saturée ; deux dimensions et trois dimensions ; radionucléides non réactifs (tritium et ^{36}Cl) et radionucléide réactif (^{90}Sr) ; cadre spatial, temporel ou spatio-temporel ; cas multivariable.

Le cas synthétique a permis d'implémenter les méthodes, d'en démontrer la faisabilité et l'intérêt, de les comparer et d'étudier de nombreuses variantes, tandis que les deux jeux de données réels ont permis d'éprouver les méthodes sur des problèmes concrets, où la modélisation d'écoulement et de transport est moins maîtrisée.

Ce travail conduit également à la construction de trois modèles d'écoulement et de transport, qui sont assez génériques et pourront être utilisés dans des travaux futurs : le cas synthétique à deux dimensions avec une zone non saturée et une zone saturée, un modèle à trois dimensions en zone non saturée pour l'expérience de traçage, et un modèle à deux dimensions de la tranchée T22.

Les performances des méthodes de combinaison dépendent du nombre de simulations d'écoulement et de transport disponibles et de leur réalisme. Pour le cas synthétique, 1700 simulations reproduisant tous les phénomènes régissant la migration du tritium sont disponibles : le krigeage avec variogrammes numériques est alors plus performant ; avec quelques dizaines de simulations seulement, le krigeage avec dérive externe est meilleur. Pour l'expérience de traçage, 150 simulations reproduisant convenablement les phénomènes régissant le transport du ^{36}Cl sont générées. Le krigeage avec dérive externe donne alors des erreurs d'estimation légèrement plus faibles que le krigeage avec variogrammes numériques. Enfin,

dans le cas de la tranchée T22, 600 simulations sont réalisées, mais elles ne reproduisent qu'une partie des phénomènes régissant la migration du ^{90}Sr et les incertitudes de modélisation sont nombreuses. Dans ce cas, si une centaine de simulations sont disponibles, le krigeage avec dérive externe est plus performant ; si plus d'une centaine de simulations sont disponibles, les performances des méthodes dépendent des dates auxquelles les activités sont estimées.

En pratique, pour réaliser une estimation par krigeage avec dérive externe, quelques dizaines de simulations suffisent. Pour le krigeage avec variogrammes numériques, une centaine de simulations est un minimum.

L'une des difficultés dans la mise en œuvre opérationnelle des méthodes proposées est de trouver un équilibre entre le temps consacré aux simulations d'écoulement et de transport (conceptualisation et construction du modèle, étude exploratoire de l'espace des paramètres, temps de calcul pour obtenir suffisamment de réalisations) et leur apport dans l'estimation. L'exemple de la tranchée T22 est représentatif de cette difficulté. En effet, beaucoup de temps a été passé à la construction du modèle et à sa calibration pour tenter de se rapprocher des observations, mais les simulations réalisées surestiment nettement les observations. Le modèle aurait alors pu être complexifié et calibré à nouveau pour tenter d'améliorer les simulations. Pourtant, ces simulations, même imparfaites, permettent d'améliorer les estimations et ont donc un apport réel.

Il faut également garder à l'esprit qu'ici, les simulations donnent des tendances sur la structure spatiale ou spatio-temporelle des phénomènes et sont une source d'information dans un modèle probabiliste. Elles ne sont pas réalisées dans l'objectif de calibrer un modèle, même si elles sont d'autant plus intéressantes qu'elles fournissent des éléments de réflexion sur les mécanismes régissant les écoulements et le transport. Leur utilisation comme source d'information sur les tendances justifie la construction de modèles assez conceptuels (approche K_d par exemple). Le conditionnement géostatistique compense certaines lacunes de modélisation, ce qui est visible sur le cas de la tranchée T22.

Enfin, pour que les méthodes de combinaison aient un apport conséquent, il faut être dans un cas où il n'est pas possible de construire un modèle d'écoulement et de transport permettant de fournir des éléments de décision sûrs, et où le nombre de données disponibles est trop réduit par rapport à la complexité du phénomène pour que la géostatistique usuelle soit précise. Cependant, il est nécessaire d'avoir suffisamment d'informations sur le site, ou de pouvoir en collecter relativement simplement, pour ne pas passer trop de temps à construire le modèle conceptuel.

Perspectives

À l'issue de ce travail, plusieurs perspectives sont envisagées.

Amélioration des modélisations sur les cas réels

De nouvelles simulations de l'expérience de traçage au ^{36}Cl pourraient être effectuées avec un milieu plus hétérogène, en augmentant la variance des champs ou en simulant deux milieux différents (un milieu sableux et un milieu plus limoneux). La quantité de ^{36}Cl injectée pourrait également être randomisée. Il serait intéressant de voir si une complexification de la modélisation améliore les estimations par rapport à ce qui a été fait dans cette thèse.

Dans le cas de la tranchée T22, les simulations surestiment nettement les observations, mais permettent tout de même d'améliorer les résultats d'estimation. Le modèle d'écoulement et de transport pourrait être complexifié, dans l'objectif de mieux comprendre et caractériser les phénomènes régissant le transport du ^{90}Sr de la tranchée vers la nappe, puis la migration dans la nappe. Ainsi, la modélisation de la rétention du ^{90}Sr pourrait être modifiée, avec par exemple un champ de coefficient de retard variable. De plus, il serait intéressant de comparer les résultats obtenus dans ce travail sur la section AB à ceux obtenus au niveau d'autres piézomètres (une autre section parallèle à AB par exemple). Par ailleurs, mettre en place des capteurs permettant d'obtenir des mesures de l'activité en ^{90}Sr plus denses dans le temps et/ou l'espace permettrait de mieux calibrer les modèles et valider les estimations.

Caractérisation de la source de contamination

Les résultats préliminaires du Chapitre 5 montrent que le krigeage avec covariances numériques peut être utilisé pour rechercher une source de contamination. Il s'agit d'une perspective intéressante d'un point de vue opérationnel (présence d'une contamination à proximité d'un site industriel étendu, dont l'origine précise est inconnue). Pour cela, un cas synthétique pourrait être construit en zone saturée (pour avoir de plus grandes distances entre la source et la zone de forte activité), en prenant par exemple en compte une incertitude sur la localisation et l'intensité de la source. Ce type de cas d'étude permettrait également de travailler sur des distances plus grandes que ce qui a été étudié dans cette thèse (jusqu'à quelques centaines de mètres par exemple).

Contaminations liées à plusieurs éléments

Le cas de contaminations liées à plusieurs radionucléides aux comportements plus différents que ce qui a été traité au Chapitre 4 pourrait être approfondi. En effet, il est assez courant de détecter un radionucléide mobile, mais de ne pas détecter les radionucléides moins mobiles, encore situés près de la source. De manière qualitative, les vitesses relatives de migration des différents radionucléides sont connues, mais les niveaux de contamination de chaque radionucléide pourraient être mieux quantifiés grâce aux méthodes proposées.

Optimisation d'échantillonnage

Les méthodes d'estimation proposées permettent d'adapter l'échantillonnage [Bechler *et al.*, 2013 ; Romary *et al.*, 2014]. Par exemple, les écart-types théoriques de krigeage pourraient permettre de déterminer les localisations optimales pour effectuer des nouvelles mesures dans le but de réduire les incertitudes sur une contamination détectée. Un autre exemple consisterait à utiliser les estimations spatio-temporelles pour choisir où implanter les piézomètres pour la surveillance d'un site, connaissant l'évolution probable de la contamination.

Généralisation à d'autres contextes

Des milieux géologiques différents pourraient être étudiés, avec par exemple d'autres types d'hétérogénéités (formations géologiques contrastées, milieu fracturé, *etc.*). En effet, le milieu géologique simulé dans le cas synthétique et les cas réels étudiés dans cette thèse est relativement homogène à l'échelle macroscopique ; les hétérogénéités modélisées dans ce travail sont des hétérogénéités intra-faciès.

Il serait également intéressant de mettre en place les méthodes sur un cas présentant des contrastes d'occupation des sols importants, ce qui est le cas de la plupart des sites industriels (présence de structures imperméables en surface). En effet, la présence de bâti empêche l'échantillonnage mais modifie également les écoulements : quelques simulations préliminaires réalisées au cours de la thèse montrent que la migration des panaches dans la zone non saturée est influencée par la présence d'une zone imperméable à proximité de la source.

Finalement, les méthodes sont ici implémentées pour estimer des contaminations radiologiques dans les sols ou les nappes, mais elles peuvent être étendues à d'autres contextes de pollution (contaminations chimiques, pollution de l'air) ou d'autres domaines des géosciences (températures des océans, phénomènes météorologiques, modélisation de gisements ou de réservoirs). Par exemple, ces méthodes pourraient permettre de travailler sur d'autres géométries, comme des graphes (étude de la qualité de l'eau le long d'un réseau hydrographique [Lefebvre Polus, 2010], étude de la qualité de l'air le long d'un réseau routier [Beauchamp *et al.*, 2018]), ou toute autre géométrie sur laquelle des simulations à base physique peuvent être réalisées. Elles pourraient également permettre de modéliser des systèmes géologiques complexes, comme des systèmes chenalisés méandriformes (modélisation à base physique avec le logiciel FLUMY [MINES ParisTech/ARMINES, 2019]) ou de fournir des aides à la décision concernant l'exploitation et l'impact environnemental de gisements d'uranium exploités par lixiviation *in situ* (modélisation du transport réactif avec le logiciel HYTEC [van der Lee *et al.*, 2003 ; Lagneau *et al.*, 2019]).

ANNEXES

Annexe A

Vignette RGeostats

Cette annexe présente une vignette RGeostats détaillant l'utilisation de la fonction `my_cov_func` pour réaliser un krigeage avec covariances numériques. Elle a été réalisée avec la version 11.2.13 du package RGeostats.

Kriging with an external covariance function

1 Introduction

This vignette illustrates the use of an external covariance function with **RGeostats** in a kriging procedure. For some geostatistical methods, such as kriging with numerical covariances or variograms (see for example Pannecoucke et al., 2020. Combining geostatistics and simulations of flow and transport to characterize contamination within the unsaturated zone. *Science of the Total Environment*), the use of an external covariance function is needed.

In those applications, the classical covariance model is replaced by a covariance matrix externally built by the user.

```
library(RGeostats)
constant.define('asp',1)
```

2 Construction of a synthetic test case

For this vignette, a synthetic case is built based on simulations. The reference case consists in a gaussian stationary simulation, to which a drift is added.

2.1 Grid

The grid is a 100×100 square. A database called `dbRef` is created to store the reference.

```
# definition of the grid
x0 = 0 ; dx = 1 ; nx = 100

# database containing the grid
dbRef = db.create(nx=rep(nx+1,2),x0=rep(x0,2),dx=rep(dx,2))
dbRef = db.locate(dbRef,2:3,'x')
```

2.2 Reference

A reference is built based on a gaussian simulation with a stationary model of covariance.

```
# construction of a stationary covariance model for the reference
C = 30 ; a = 3
model.ref = model.create(vartype='Gaussian',range=C,sill=a,ndim=2)

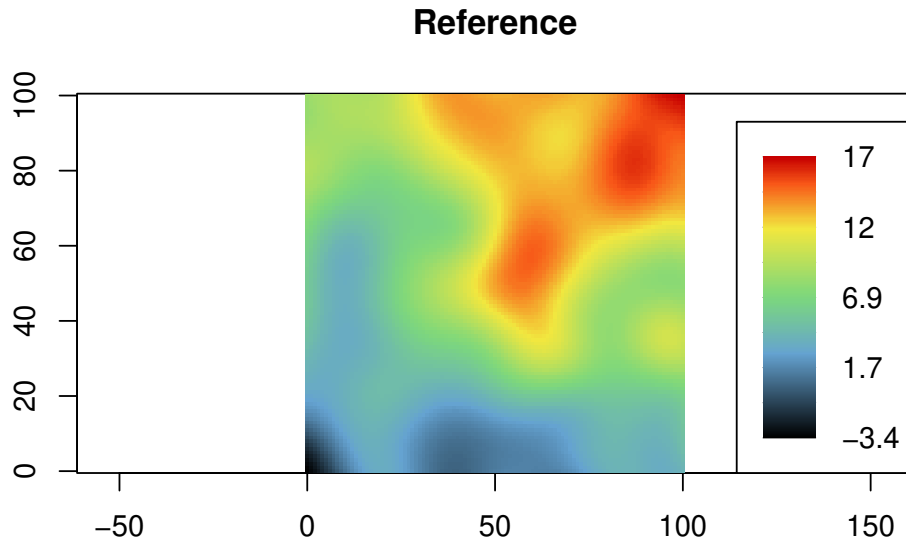
# construction of a unique neighborhood
unique.neigh = neigh.create(type=0)

# simulation of the reference (turning bands simulation)
dbRef = simtub(,dbout=dbRef,model=model.ref,neigh=unique.neigh,nbtuba=800,seed=2020)
dbRef = db.rename(dbRef,'Simu.V1.S1','Ref')
```

The drift consists in a linear combination of the x and y coordinates.

```
xx = seq(x0,nx,by=dx)
drift = as.vector(outer(.05*xx,.1*xx,'+'))
```

```
# addition of the drift to the reference and plot
dbRef[, 'Ref'] = dbRef[, 'Ref'] + drift
plot(dbRef, title='Reference', pos.legend=1)
```



2.3 Observations

The reference is sampled to implement the estimation method: 50 random observations are assumed to be known, while the rest of the reference is supposed to be unknown. The observation are stored in a database called `dbObs`.

```
# observations
nObs = 50 # number of observations
set.seed(12345)
indObs = sort(sample(seq(dbRef$nech), nObs, replace=F)) # sampling of the reference
dbObs = db.reduce(db=dbRef, ranks=indObs)
```

3 Kriging with an external covariance function

3.1 Simulations

The first step to perform kriging with an external covariance function is to generate a set of simulations. In this case, 200 simulations are generated. The stationary model is supposed to be known, but the drift coefficients are assumed to be unknown. The set of simulations is stored in the matrix called `SIMU`.

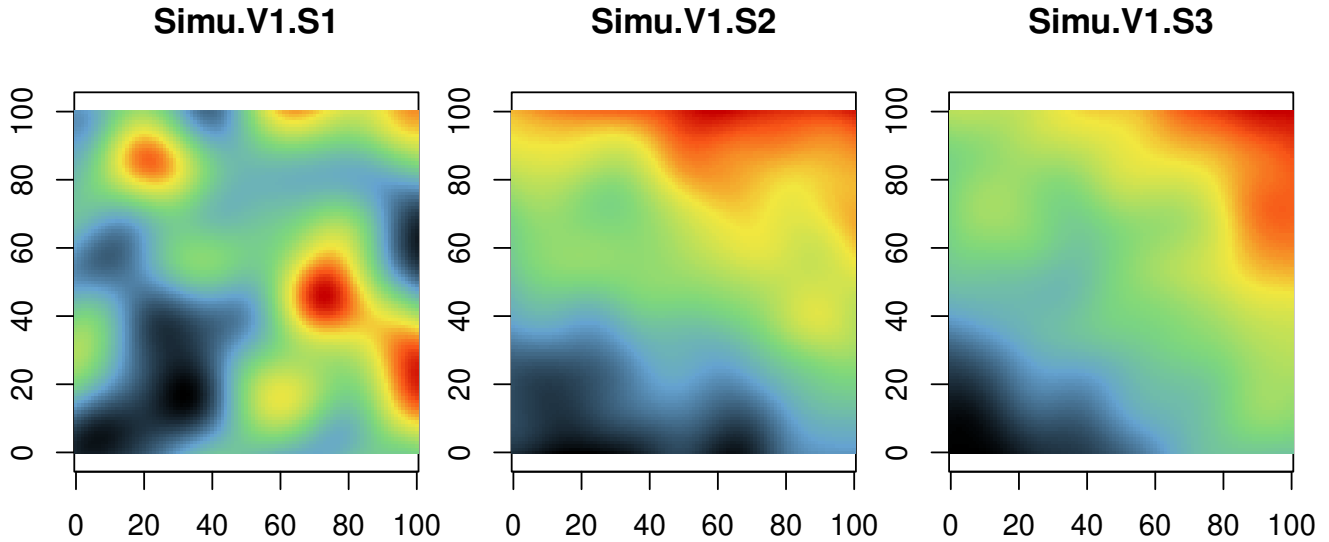
```
# generation of 200 stationary realizations on the grid using model.ref
nSimu = 200
dbSimu = simtub(,dbout=dbRef,model=model.ref,neigh=unique.neigh,nbsimu=nSimu,nbtuba=800)

driftRandom = function(){
  # the drift coefficients are drawn from a uniform distribution between 0 and 0.25
  a = runif(1,0,.25)
  b = runif(1,0,.25)
  return(as.vector(outer(a*xx,b*xx,'+')))
}

# generation of 200 drifts
DRIFT = NULL
for (i in seq(nSimu)){
  DRIFT = cbind(DRIFT,driftRandom())
}
```

```
# addition of the drifts to the stationary simulations
dbSimu[, 'Simu.V1.S*'] = dbSimu[, 'Simu.V1.S*'] + DRIFT
SIMU = as.matrix(dbSimu[, 'Simu.V1.S*'])
```

```
# example of some simulations
plot(dbSimu, name='*.S1')
plot(dbSimu, name='*.S2')
plot(dbSimu, name='*.S3')
```



3.2 Covariance matrices

Two covariance matrices and one vector are computed to perform kriging with an external covariance function:

- `cxx`: the observation-observation matrix, which must be positive-definite (*i.e* a variogram matrix is not allowed),
- `cx0`: the observation-target matrix,
- `c00`: the $C(0)$ vector needed to compute kriging error variance.

In this example, the covariance matrices are computed based on the set of simulations. The simulations are assumed to be stored in a matrix in which each column is a simulation.

```
numericalCov = function(pairOfPoints, simulationTable){
  # This function returns the covariance value between the points stored in pairOfPoints,
  # based on the simulations stored in simulationTable.
  # The indexes in pairOfPoints refer to the rows of simulationTable.
  index1 = pairOfPoints[1]
  index2 = pairOfPoints[2]
  mean1 = mean(simulationTable[index1,])
  mean2 = mean(simulationTable[index2,])
  return(mean((simulationTable[index1,]-mean1)*(simulationTable[index2,]-mean2)))
}

# pairs of observation points, to compute cxx
pairObs = expand.grid(indObs, indObs)
cxx = matrix(apply(pairObs, 1, FUN=numericalCov, simulationTable=SIMU), nrow=nObs)
# pairs of observations and targets, to compute cx0
pairTarget = expand.grid(indObs, seq(dbRef$nech))
cx0 = matrix(apply(pairTarget, 1, FUN=numericalCov, simulationTable=SIMU), nrow=nObs)
# pairs of targets to compute c00
pairs = cbind(seq(dbRef$nech), seq(dbRef$nech))
c00 = apply(pairs, 1, FUN=numericalCov, simulationTable=SIMU)
```

3.3 Kriging

In order to use `cxx` and `cx0` instead of a classical model, we must define the external covariance function `my_cov_func`. This function is called internally during the kriging procedure (here via the `def_cov` function). Its prototype contains the following arguments:

- `dist`: the distance between two points (not used),
- `db1`: the type of db (1 if `dbin` or 2 if `dbout`) of the first point,
- `iech1`: the index of the first point in `db1`,
- `db2`: the type of db of the second point,
- `iech2`: the index of the second point in `db2`,
- `incr`: not used,
- `x1`: the coordinate vector of the first point (not used),
- `x2`: the coordinate vector of the second point (not used).

Finally the function `my_cov_func` needs to access three variables: `cxx`, `cx0`, `c00`.

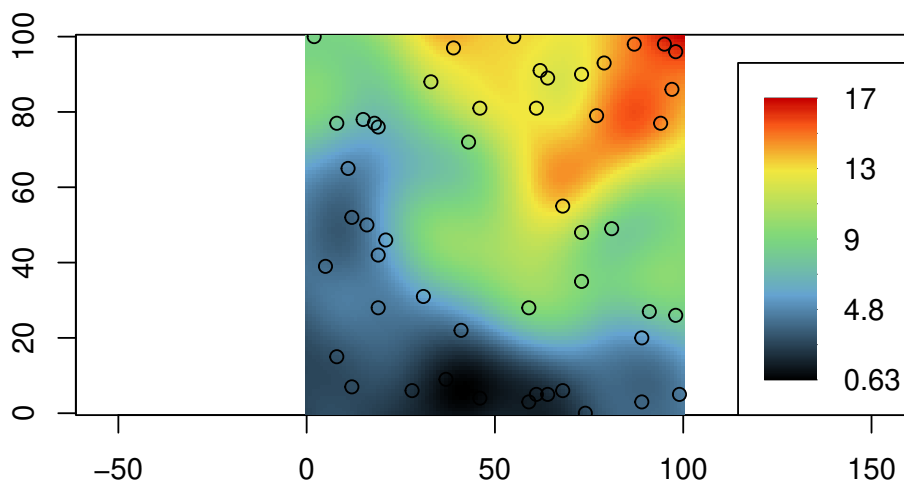
```
def_cov <- function(cxx,cx0,c00) {
  function(dist,db1=NA,iech1=NA,db2=NA,iech2=NA,incr=NA,x1=NA,x2=NA,...){
    result = 0
    if (db1 == 1 && db2 == 1)
      result = cxx[iech1,iech2]
    else if (db1 == 1 && db2 == 2)
      result = cx0[iech1,iech2]
    else if (db1 == 2 && db2 == 1)
      result = cx0[iech2,iech1]
    else if (db1 == 2 && db2 == 2)
      result = c00[iech1]
    else
      cat('db1=',db1,' db2=',db2,'\n')
    result
  }
}
```

```
my_cov_func <- def_cov(cxx=cxx,cx0=cx0,c00=c00)
```

Finally, we perform ordinary kriging on a unique neighborhood.

```
dbRef = kriging(dbin=dbObs,dbout=dbRef,model=my_cov_func,neigh=unique.neigh,uc='1',radix='KNC')
plot(dbRef,name='KNC.*.estim',pos.legend=1,title='Kriging estimate (KNC)')
points(dbObs[,2],dbObs[,3]) # add the observation locations
```

Kriging estimate (KNC)

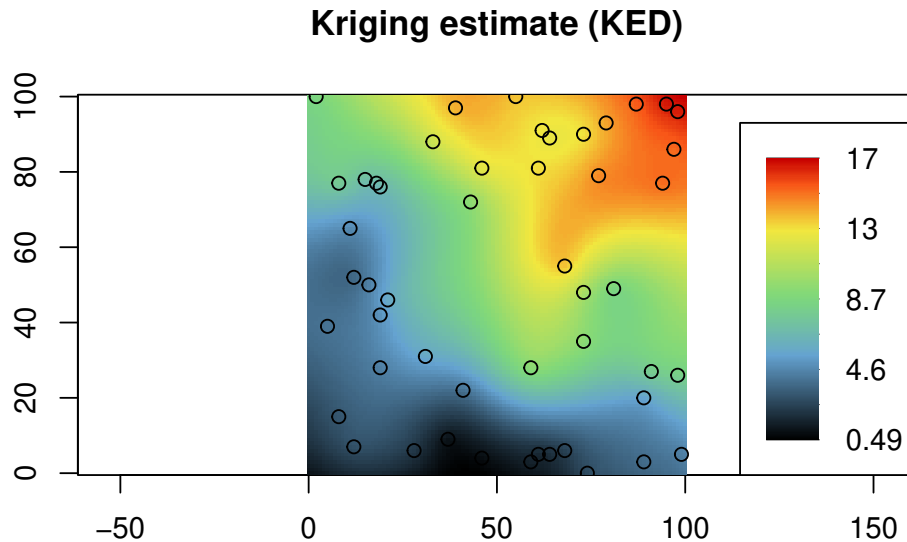


```
# computation of the error
errorKNC = dbRef[, 'Ref'] - dbRef[, 'KNC.*.estim']
dbRef = db.add(dbRef, errorKNC, loctype='z')
```

4 Comparison with kriging with an external drift

The results obtained previously are compared with the results obtained with kriging with an external drift, with the drift function used to build the reference.

```
# databases preparation
dbRef = db.add(dbRef, drift, loctype='f')
dbObs = db.add(dbObs, drift[indObs], loctype='f')
# computation of experimental variogram and fitting of a model
varioED = vario.calc(dbObs, uc=c('1', 'f1'))
modelED = model.auto(varioED, title='Variogram model', draw=F)
# kriging with an external drift
dbRef = kriging(dbin=dbObs, dbout=dbRef, model=modelED, neigh=unique.neigh, radix='KED', uc=c('1', 'f1'))
plot(dbRef, name='KED.*.estim', pos.legend=1, title='Kriging estimate (KED)')
points(dbObs[, 2], dbObs[, 3]) # add the observation locations
```



```
# computation of the error
errorKED = dbRef[, 'Ref'] - dbRef[, 'KED.*.estim']
dbRef = db.add(dbRef, errorKED, loctype='z')
```

The mean absolute errors and root mean squared errors are finally computed for the two methods. Kriging with numerical covariance leads to smaller errors than kriging with an external drift.

```
MAE_KNC = mean(abs(errorKNC))
MAE_KED = mean(abs(errorKED))
print(paste('MAE for KNC:', round(MAE_KNC, 2)))
print(paste('MAE for KED:', round(MAE_KED, 2)))

RMSE_KNC = sqrt(mean(errorKNC^2))
RMSE_KED = sqrt(mean(errorKED^2))
print(paste('RMSE for KNC:', round(RMSE_KNC, 2)))
print(paste('RMSE for KED:', round(RMSE_KED, 2)))
```

```
## [1] "MAE for KNC: 0.48"
## [1] "MAE for KED: 0.61"
## [1] "RMSE for KNC: 0.79"
## [1] "RMSE for KED: 0.92"
```


Annexe B

Influence des portées des champs de paramètres de Mualem-van Genuchten

Cette annexe présente les résultats concernant l'influence des portées des champs de paramètres de Mualem-van Genuchten sur la migration de panaches de tritium (Figures B.1 et B.2 pour la portée horizontale et Figures B.3 et B.4 pour la portée verticale).

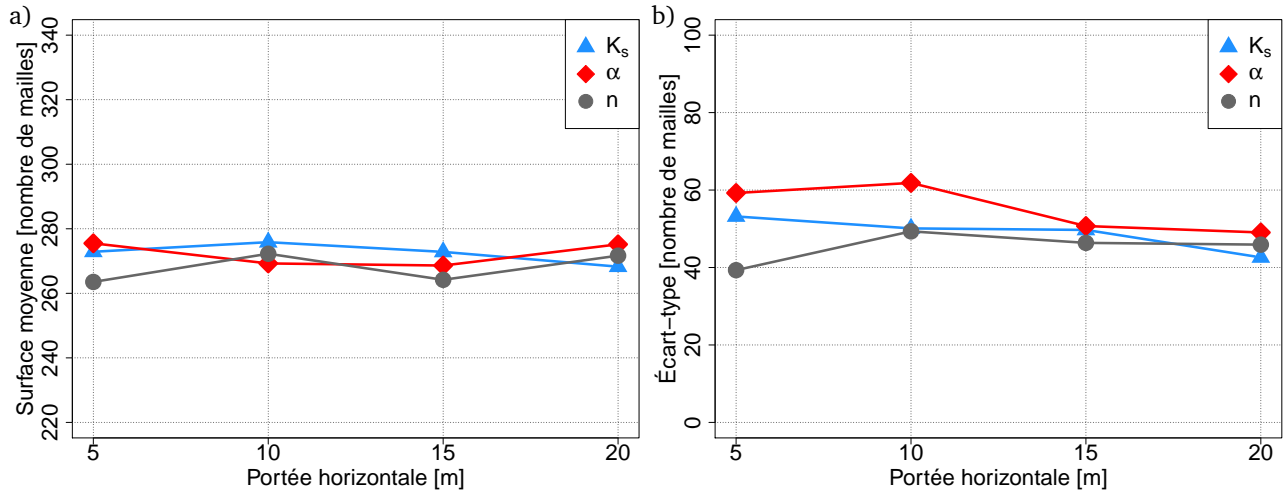


FIGURE B.1 – a) Moyenne et b) écart-type des surfaces des panaches par série de simulations en fonction de la portée horizontale des champs de paramètres. Les valeurs de variance sont fixées à $\sigma^2(\log K_s)=0,2$, $\sigma^2(\log \alpha)=0,04$ et $\sigma^2(\log n)=0,001$. La portée verticale vaut 3 m. L'échelle de l'axe des abscisses est la même que celle de la Figure 3.8.

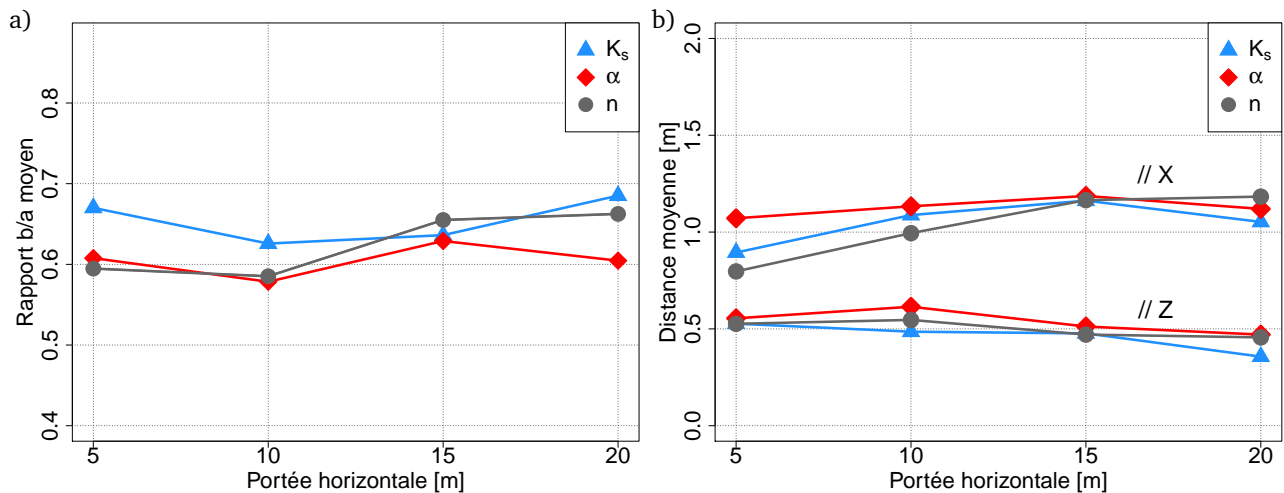


FIGURE B.2 – a) Rapport moyen entre les petits axes et les grands axes des ellipses équivalentes aux panaches et b) distance moyenne entre les centres de gravité des panaches de chaque série de simulations et le centre de gravité du panache simulé dans le milieu homogène en fonction de la portée horizontale des champs de paramètres. Les échelles des axes des abscisses sont les mêmes que celles des Figures 3.9 et 3.11.

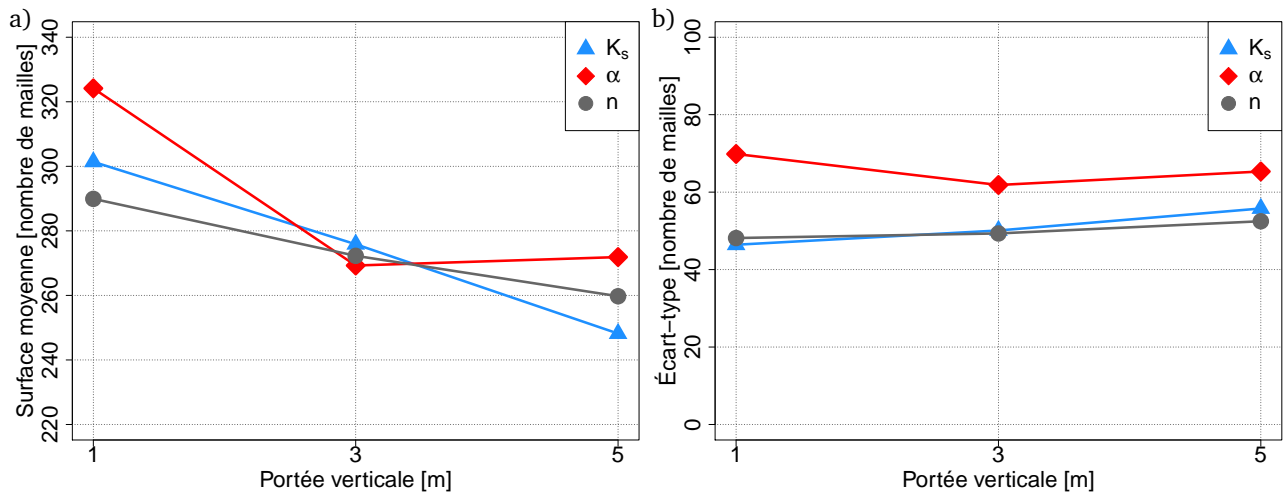


FIGURE B.3 – a) Moyenne et b) écart-type des surfaces des panaches par série de simulations en fonction de la portée verticale des champs de paramètres. La portée horizontale vaut 10 m.

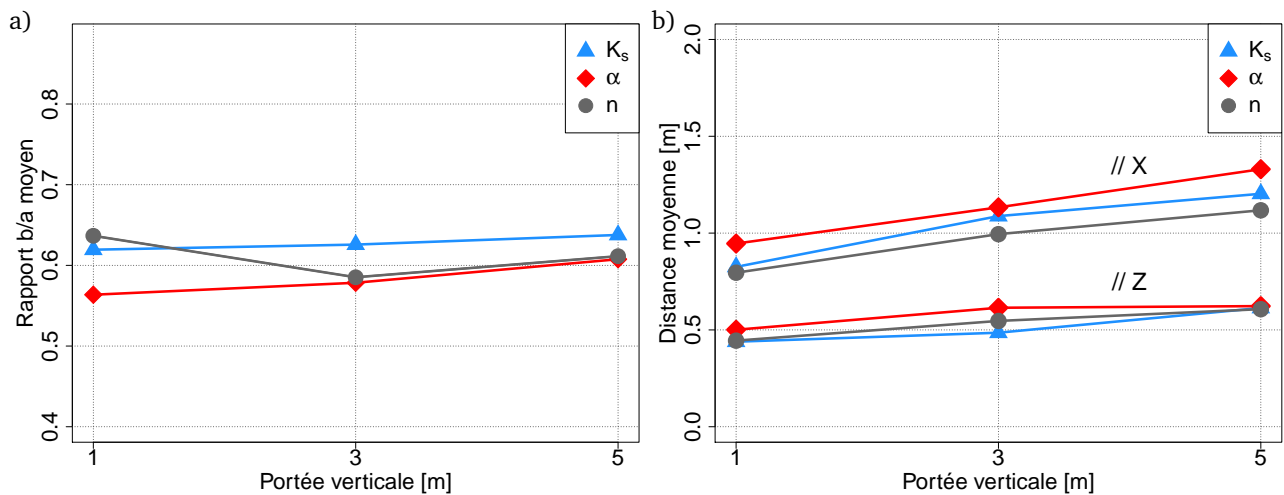


FIGURE B.4 – a) Rapport moyen entre les petits axes et les grands axes des ellipses équivalentes aux panaches et b) distance moyenne entre les centres de gravité des panaches de chaque série de simulations et le centre de gravité du panache simulé dans le milieu homogène en fonction de la portée horizontale des champs de paramètres.

Annexe C

Résultats complémentaires sur l'influence de la variabilité des champs de paramètres du modèle de Mualem-van Genuchten

Changement de milieu

Pour étudier l'effet de la texture moyenne, quelques séries de simulations ont été réalisées en changeant les propriétés moyennes du milieu. Au Chapitre 3, les propriétés du milieu homogène équivalent utilisées sont celles du sable limoneux (Tableau 3.1, ligne 2). Ce milieu moyen est ici remplacé par un milieu de texture plus fine (limon silto-argileux, Tableau 3.1, ligne 1).

Des séries de 100 simulations sont générées pour chaque paramètre de Mualem-van Genuchten avec différentes valeurs de variances et des portées fixées à 10 m dans la direction horizontale et 3 m dans la direction verticale. Le radionucléide utilisé est du tritium, et aucun autre paramètre des simulations n'est modifié par rapport au Chapitre 3. Les résultats obtenus dans les deux milieux sont assez proches (Figures C.1 et C.2), sauf pour le paramètre n : lorsque n est fortement variable, les panaches sont en moyenne moins étendus et moins allongés dans le limon silto-argileux que dans le sable limoneux.

Changement de radionucléide

Le tritium, considéré comme un traceur parfait n'interagissant pas avec le milieu, est remplacé dans quelques simulations par un radionucléide réactif qui interagit avec le milieu. Cette interaction est quantifiée par le coefficient de retard R , qui vaut ici 3. La constante de désintégration utilisée est celle du ^{90}Sr ($\lambda=7,63 \cdot 10^{-10} \text{ s}^{-1}$) et le coefficient de retard est compatible avec les propriétés du ^{90}Sr .

Des séries de 100 simulations de la migration de ce radionucléide sont générées dans des milieux identiques à ceux du Chapitre 3. Comme attendu, les panaches sont moins étendus dans le cas du radionucléide réactif que dans le cas du tritium (Figure C.3). Ils sont également plus circulaires et moins distants les uns des autres (Figure C.4). Cependant, les mêmes tendances sont observées : quand les variances des champs de paramètres augmentent, les surfaces moyennes et les distances au centre du panache simulé dans le milieu homogène augmentent. Les panaches sont également plus allongés.

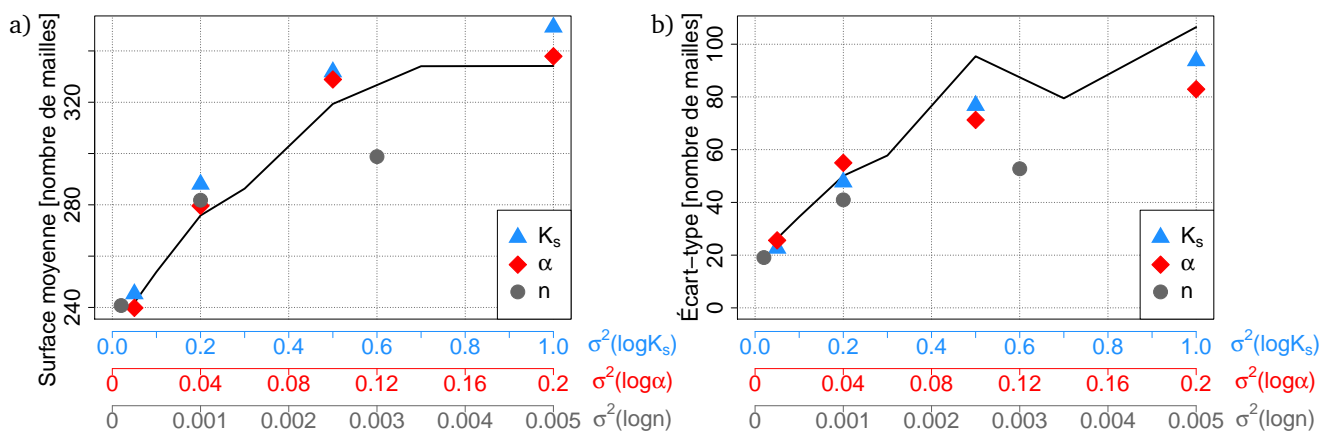


FIGURE C.1 – a) Moyenne et b) écart-type des surfaces des panaches par série de simulations en fonction de la variance des champs de paramètres. Les figurés correspondent aux simulations réalisées dans le limon silto-argileux. La ligne noire correspond aux résultats obtenus pour K_s dans le sable limoneux.

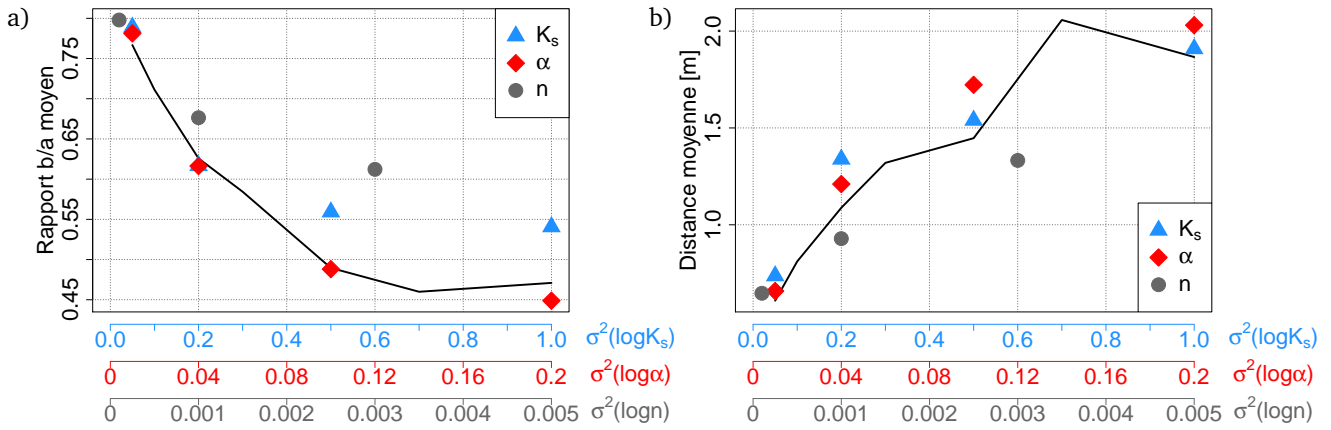


FIGURE C.2 – a) Rapport moyen entre les petits axes et les grands axes des ellipses équivalentes aux panaches et b) distance moyenne dans la direction horizontale entre les centres de gravité des panaches de chaque série de simulations et le centre de gravité du panache simulé dans le milieu homogène en fonction de la variance des champs de paramètres.

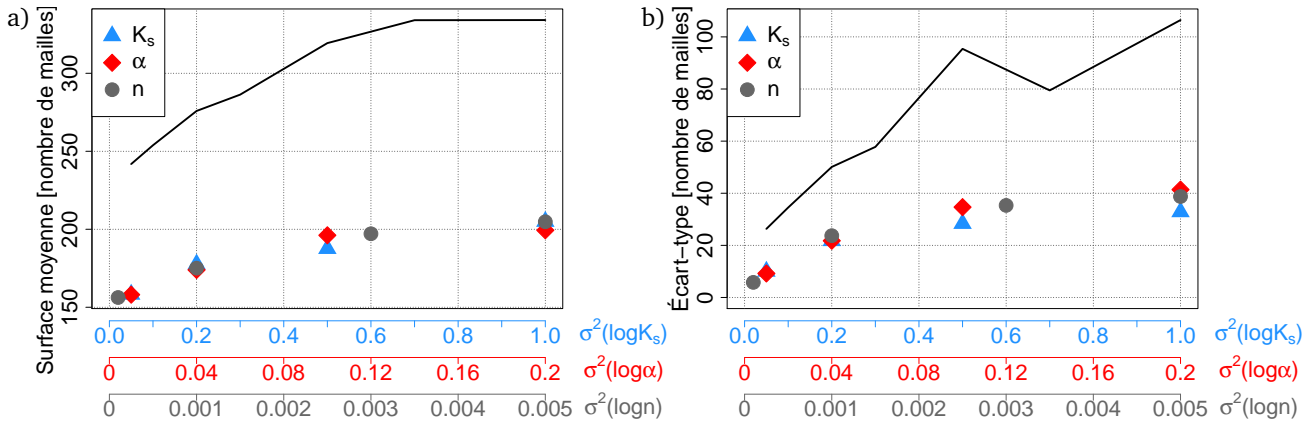


FIGURE C.3 – a) Moyenne et b) écart-type des surfaces des panaches par série de simulations en fonction de la variance des champs de paramètres. Les figurés correspondent aux simulations réalisées avec le radionucléide réactif. La ligne noire correspond aux résultats obtenus pour K_s avec du tritium.

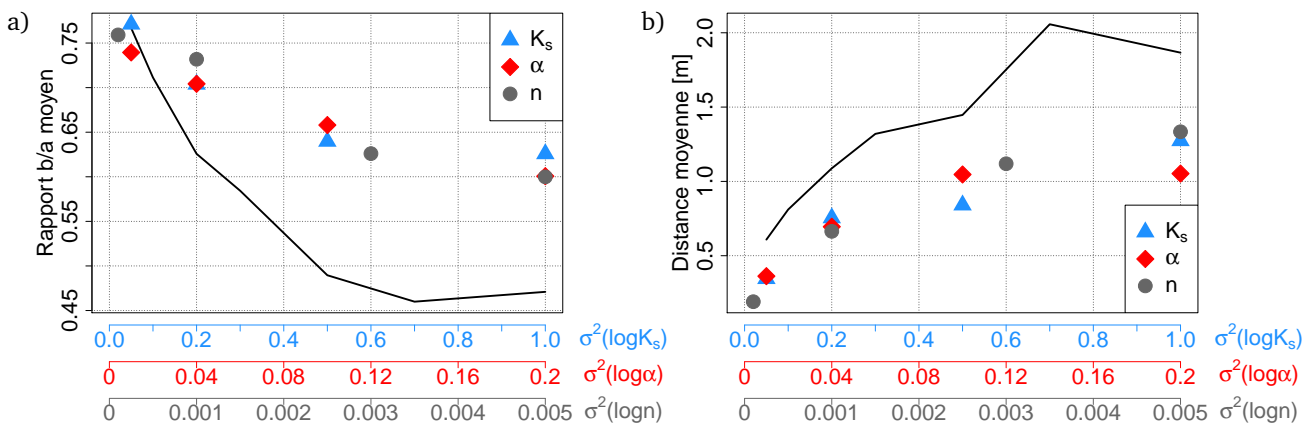


FIGURE C.4 – a) Rapport moyen entre les petits axes et les grands axes des ellipses équivalentes aux panaches et b) distance moyenne dans la direction horizontale entre les centres de gravité des panaches de chaque série de simulations et le centre de gravité du panache simulé dans le milieu homogène en fonction de la variance des champs de paramètres.

Annexe D

Construction de références supplémentaires

Cette annexe présente la construction de deux autres références, en complément de la référence construite au Chapitre 3 (Section 3.3).

Référence stationnaire additionnelle

Une autre réalisation d'un triplet de champs de proportions de sable, de limon et d'argile est générée comme décrit au Chapitre 3 (Section 3.3), avec les mêmes paramètres de variogramme et de distribution. Ces champs sont identiques aux champs de référence aux huit sondages servant à échantillonner la texture. Cela permet d'utiliser le même jeu de simulations pour mettre en œuvre les méthodes d'estimation sur cette référence additionnelle et sur la référence du Chapitre 3. Les champs de texture sont convertis en paramètres de Mualem-van Genuchten par Rosetta3. Le panache qui en résulte est donné sur la Figure D.1.

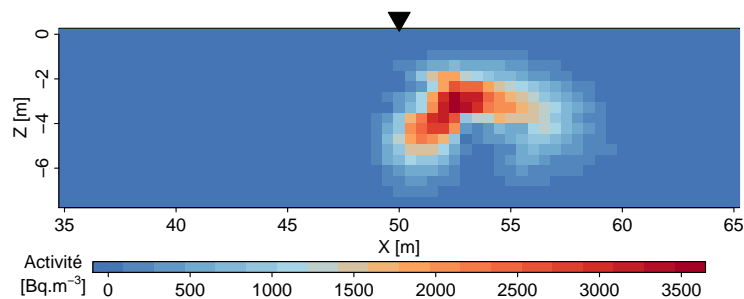


FIGURE D.1 – Panache de tritium simulé avec MELODIE, cinq ans après le rejet (zoom dans la zone non saturée).

Référence non stationnaire

La référence dite non stationnaire est générée à partir de champs de texture non stationnaires, mais identiques aux champs stationnaires aux huit sondages servant à échantillonner la texture. Ces champs sont obtenus en ajoutant un champ non stationnaire (Figure D.2) aux champs de texture initialement stationnaires, puis en appliquant une anamorphose qui permet de conserver la bonne distribution des textures. Les champs de proportions de sable, de limon et d'argile ainsi obtenus sont donnés sur la Figure D.3. La transformation non stationnaire conduit à une proportion de sable plus forte vers la gauche et vers le bas du domaine, et inversement pour les proportions de limon et d'argile. Ces champs sont convertis en paramètres de Mualem-van Genuchten par Rosetta3. Enfin, en appliquant les mêmes conditions initiales et conditions aux limites que précédemment, on obtient avec le code MELODIE le panache de tritium de la Figure D.4.

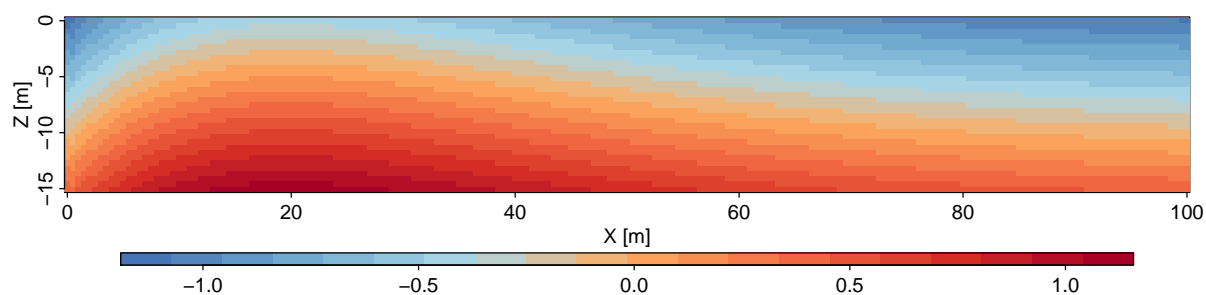


FIGURE D.2 – Champ ajouté au champ stationnaire pour le rendre non stationnaire dans l'espace.

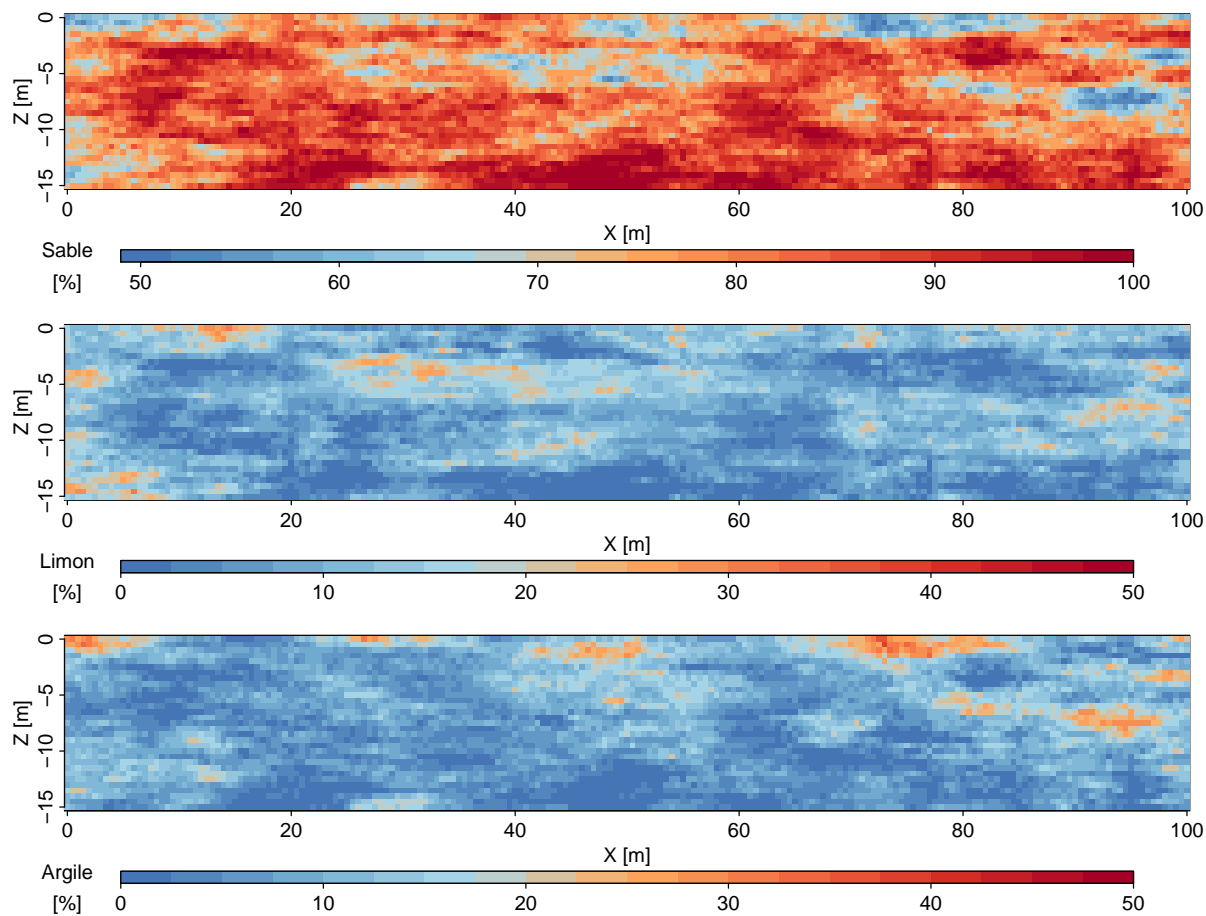


FIGURE D.3 – Champs de proportions de sable, de limon et d'argile.

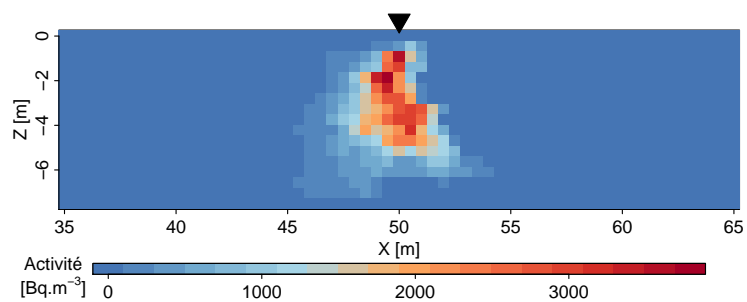


FIGURE D.4 – Panache de tritium simulé avec MELODIE, cinq ans après le rejet (zoom dans la zone non saturée).

Annexe E

Résultats sur les références supplémentaires

Cette annexe présente les résultats d'estimations sur les références supplémentaires construites à l'Annexe D. Le même jeu de simulations est utilisé quelle que soit la référence.

Référence stationnaire additionnelle

La référence stationnaire additionnelle (Figure D.1) présente une forme concave, qui n'est reproduite que par krigeage avec variogrammes numériques pour le scénario S1 (Figure E.1). Pour les autres méthodes et le scénario S2, celle-ci n'est pas reproduite. Quel que soit le scénario d'échantillonnage, les écart-types d'erreur sont plus faibles pour le krigeage avec variogrammes numériques que pour les autres méthodes (Figures E.1 et E.3). Les erreurs moyennes ne sont pas toujours ordonnées de la même manière : pour le scénario S1, le krigeage avec variogrammes numériques donne des erreurs absolues et quadratiques plus faibles, suivi par le krigeage ordinaire, alors que le krigeage avec dérive externe conduit à une erreur relative plus faible (Tableau E.1). Pour le scénario S2, le krigeage avec variogrammes numériques donne les erreurs moyennes les plus faibles, suivi du krigeage ordinaire (sauf pour l'erreur relative) (Tableau E.2).

Les nuages de corrélation entre les valeurs d'activité de la référence et les valeurs estimées montrent que pour le scénario S2, les activités estimées sont assez éloignées de la référence, et qu'il existe un biais conditionnel (Figures E.2a et E.4a). Les courbes de sélectivité montrent que le krigeage avec variogrammes numériques reproduit mieux la distribution d'activité (Figures E.2b et E.4b). Les proportions de faux-négatifs sont proches pour toutes les méthodes, voire légèrement plus faibles pour le krigeage avec variogrammes numériques (Figures E.2c et E.4c). Les proportions de faux-positifs sont plus faibles pour le krigeage avec variogrammes numériques que pour les deux autres méthodes dès que le seuil de contamination dépasse 20 Bq.m^{-3} (Figures E.2d et E.4d).

Référence non stationnaire additionnelle

Cette référence est donnée sur la Figure D.4. Pour le scénario S1, les cartes d'estimation obtenues par les trois méthodes sont semblables (Figure E.5). Elles le sont également pour le scénario S2, mais elles présentent toutes des valeurs d'activité très faibles (Figure E.7). Pour le scénario S1, les écart-types d'erreur sont plus faibles pour le krigeage avec variogrammes numériques que pour les autres méthodes. Le krigeage avec variogrammes numériques conduit par ailleurs aux erreurs moyennes les plus faibles (Tableau E.3), alors que pour le scénario S2, le krigeage avec dérive externe conduit aux erreurs les plus faibles (Tableau E.4).

Les nuages de corrélation entre les valeurs d'activité de la référence et les valeurs d'activité estimées soulignent la sous-estimation des activités quelle que soit la méthode pour le scénario S2 (Figure E.8). Les courbes de sélectivité montrent que le krigeage avec variogrammes numériques reproduit mieux la distribution d'activité pour le scénario S1, et que le krigeage avec dérive externe la reproduit mieux pour le scénario S2 (Figures E.6b et E.8b). Les proportions de faux-négatifs sont proches pour les trois méthodes (Figures E.6c et E.8c). Les proportions de faux-positifs sont plus faibles pour le krigeage avec variogrammes numériques (Figures E.6d et E.8d).

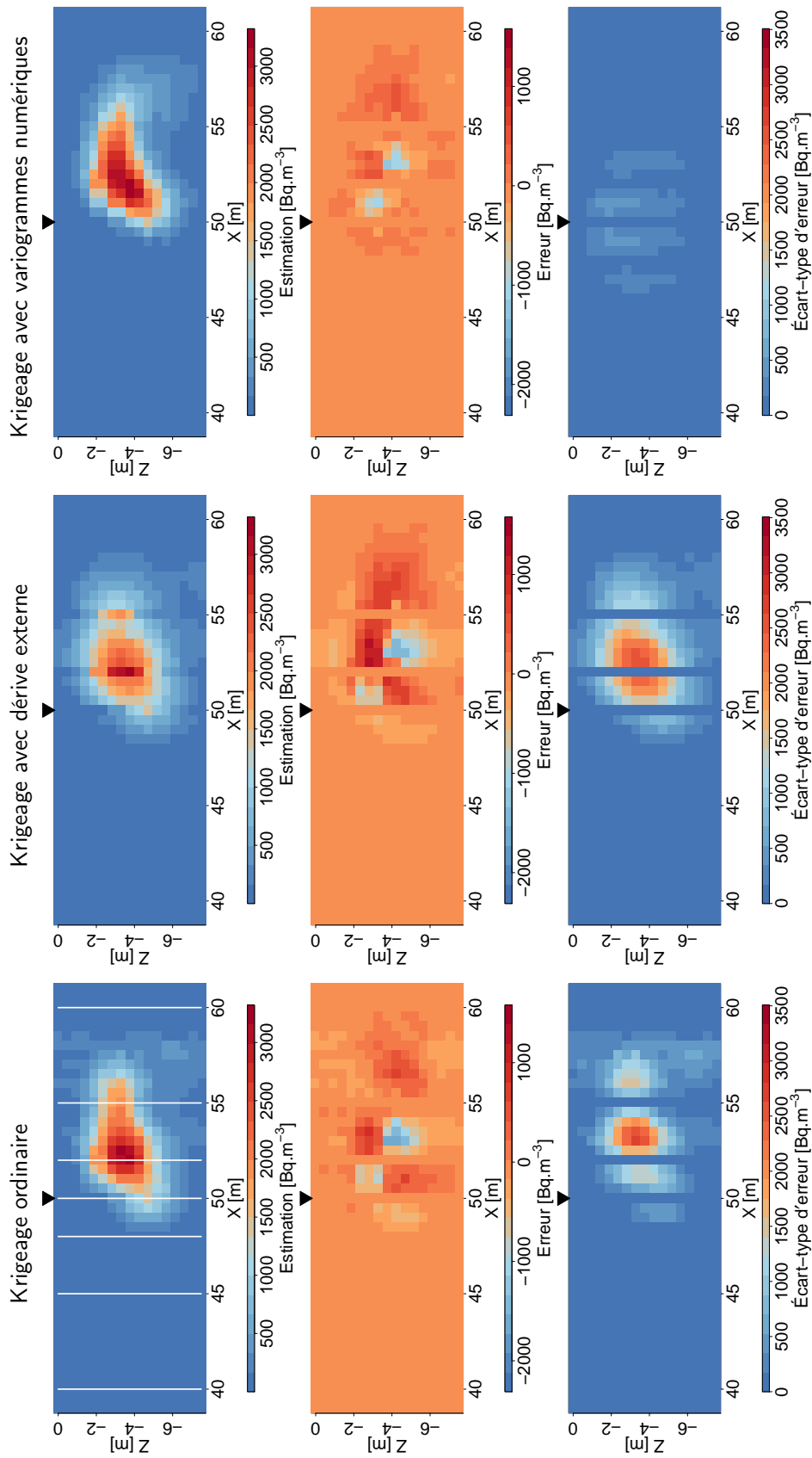


FIGURE E.1 – Cartes d'estimation, d'erreur d'estimation et d'écart-type d'erreur d'estimation obtenues avec les trois méthodes pour le scénario S1 (rappelé sur la première carte). La même échelle de couleur est utilisée sur chaque ligne. Les erreurs négatives correspondent à des sous-estimations et les erreurs positives à des surestimations.

TABLEAU E.1 – Erreurs moyennes obtenues avec les trois méthodes pour le scénario S1.

	KO	KDE	KVN
EAM [Bq.m^{-3}]	107	130	59
REQM [Bq.m^{-3}]	228	271	147
ERM [-]	-3.8	-1.4	-2.6

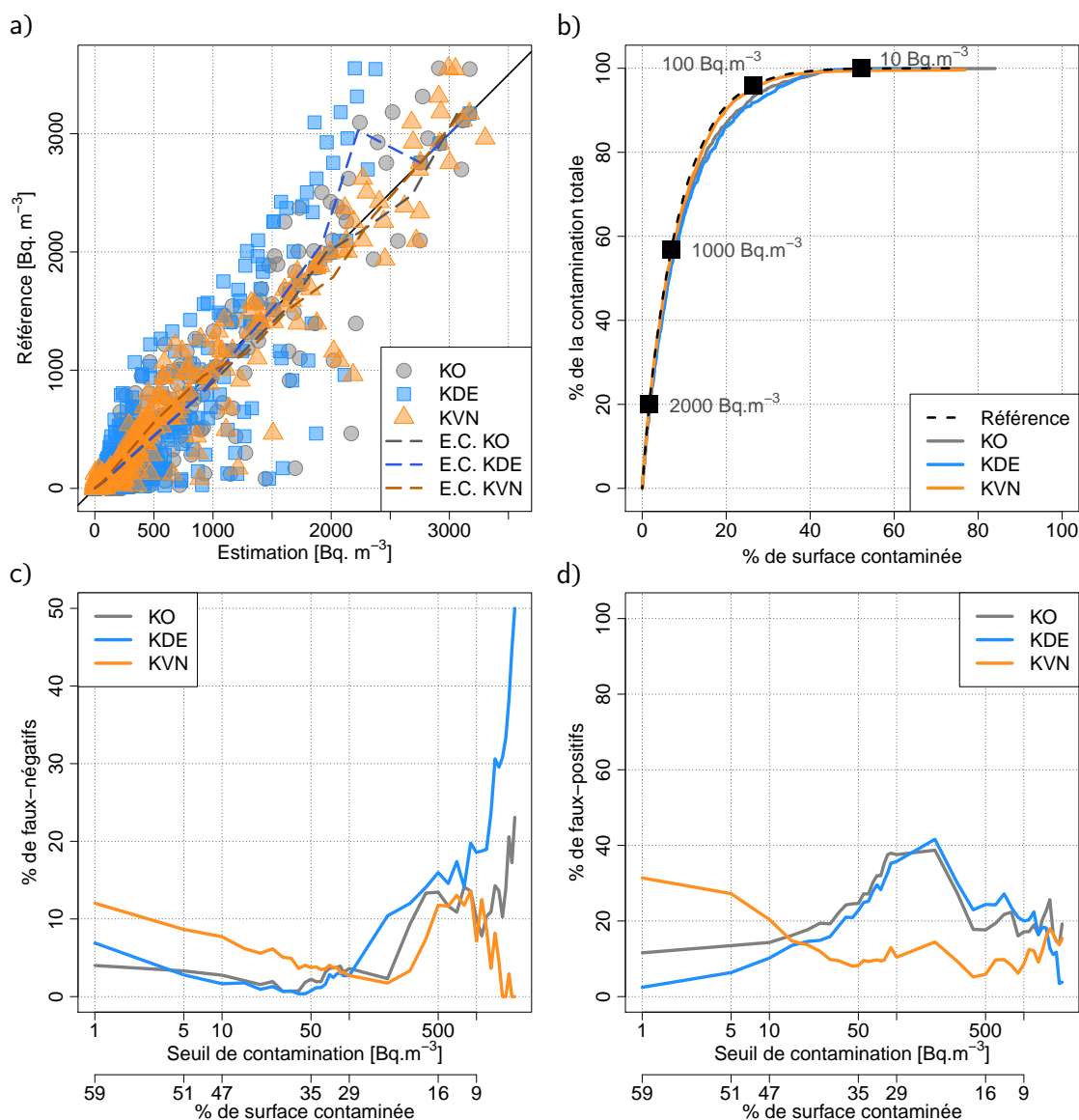


FIGURE E.2 – Indicateurs de performance pour le scénario S1 : a) comparaison entre les valeurs d'activité de la référence et les activités estimées ; b) courbes de sélectivité ; c) proportion de faux-négatifs par rapport à la surface réellement contaminée ; d) proportion de faux-positifs par rapport à la surface réellement contaminée. Sur les graphiques c) et d), les axes des abscisses secondaires donnent la surface contaminée (en % du domaine de modélisation) de la référence en fonction du seuil de contamination. Les échelles des axes des ordonnées sont par ailleurs choisies pour être identiques sur les Figures E.2 et E.4.

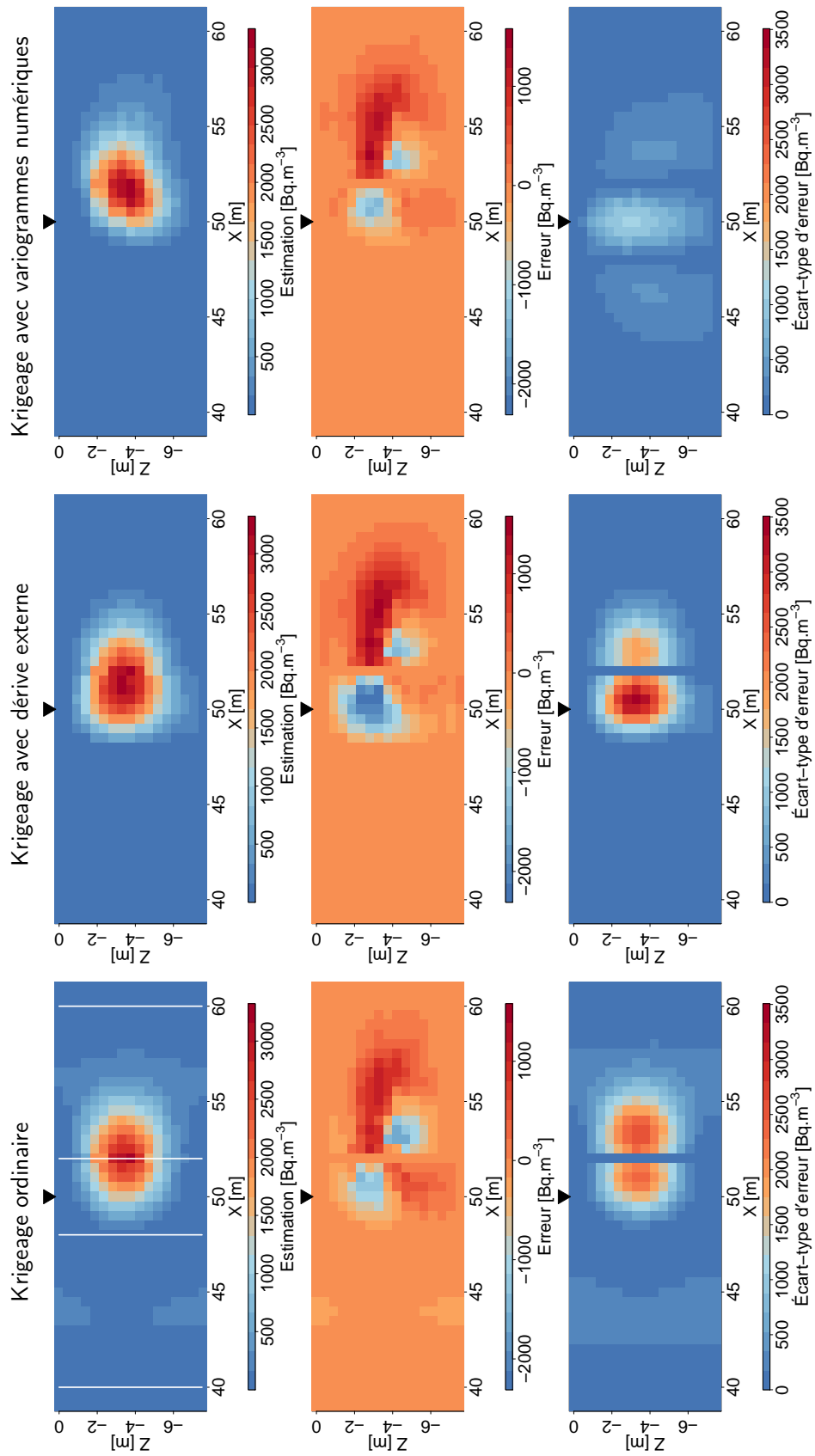


FIGURE E.3 – Cartes d'estimation, d'erreur d'estimation et d'écart-type d'erreur pour le scénario S2. Les échelles de couleur sont les mêmes que sur la Figure E.1.

TABLEAU E.2 – Erreurs moyennes pour le scénario S2.

	KO	KDE	KVN
EAM [Bq.m^{-3}]	192	208	128
REQM [Bq.m^{-3}]	337	461	304
ERM [-]	-37.6	-3.7	-3.4

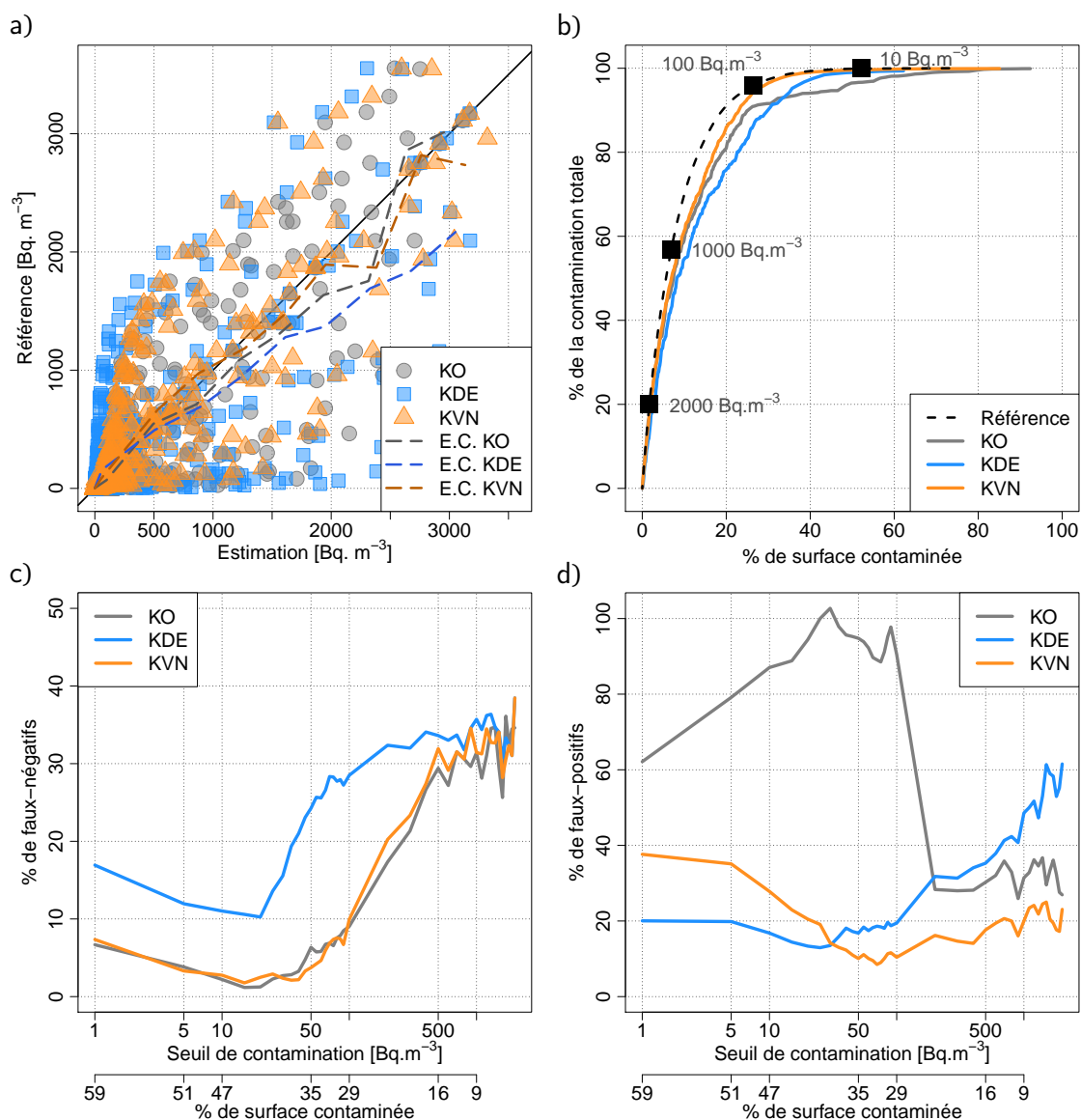


FIGURE E.4 – Indicateurs de performance pour le scénario S2 : a) comparaison entre les valeurs d'activité de la référence et les activités estimées ; b) courbes de sélectivité ; c) proportion de faux-négatifs par rapport à la surface réellement contaminée ; d) proportion de faux-positifs par rapport à la surface réellement contaminée.

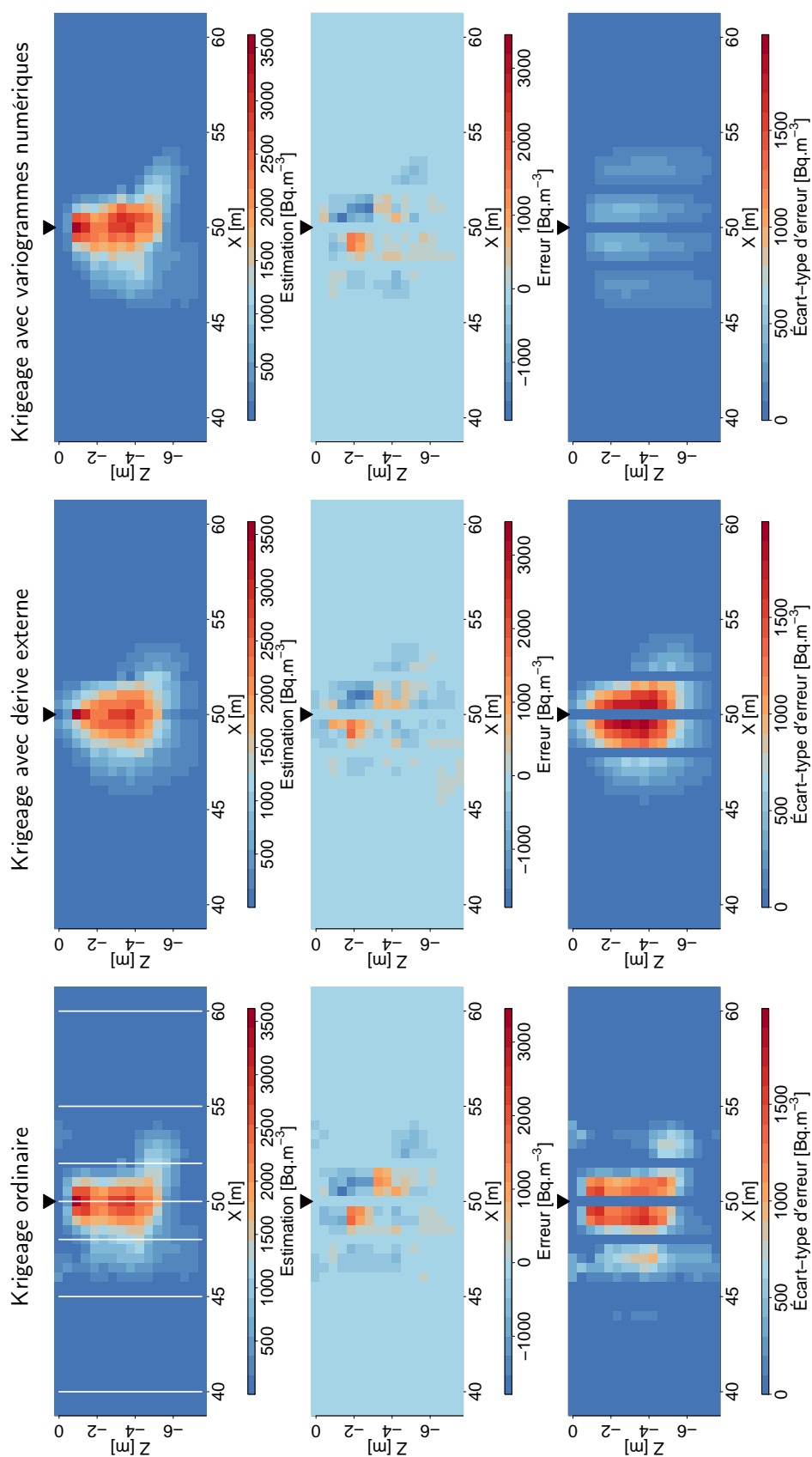


FIGURE E.5 – Cartes d'estimation, d'erreur d'estimation et d'écart-type d'erreur d'estimation pour le scénario S1.

TABLEAU E.3 – Erreurs moyennes obtenues pour le scénario S1.

	KO	KDE	KVN
EAM [Bq.m^{-3}]	100	81	67
REQM [Bq.m^{-3}]	239	214	206
ERM [-]	-15.6	-8.6	-3.4

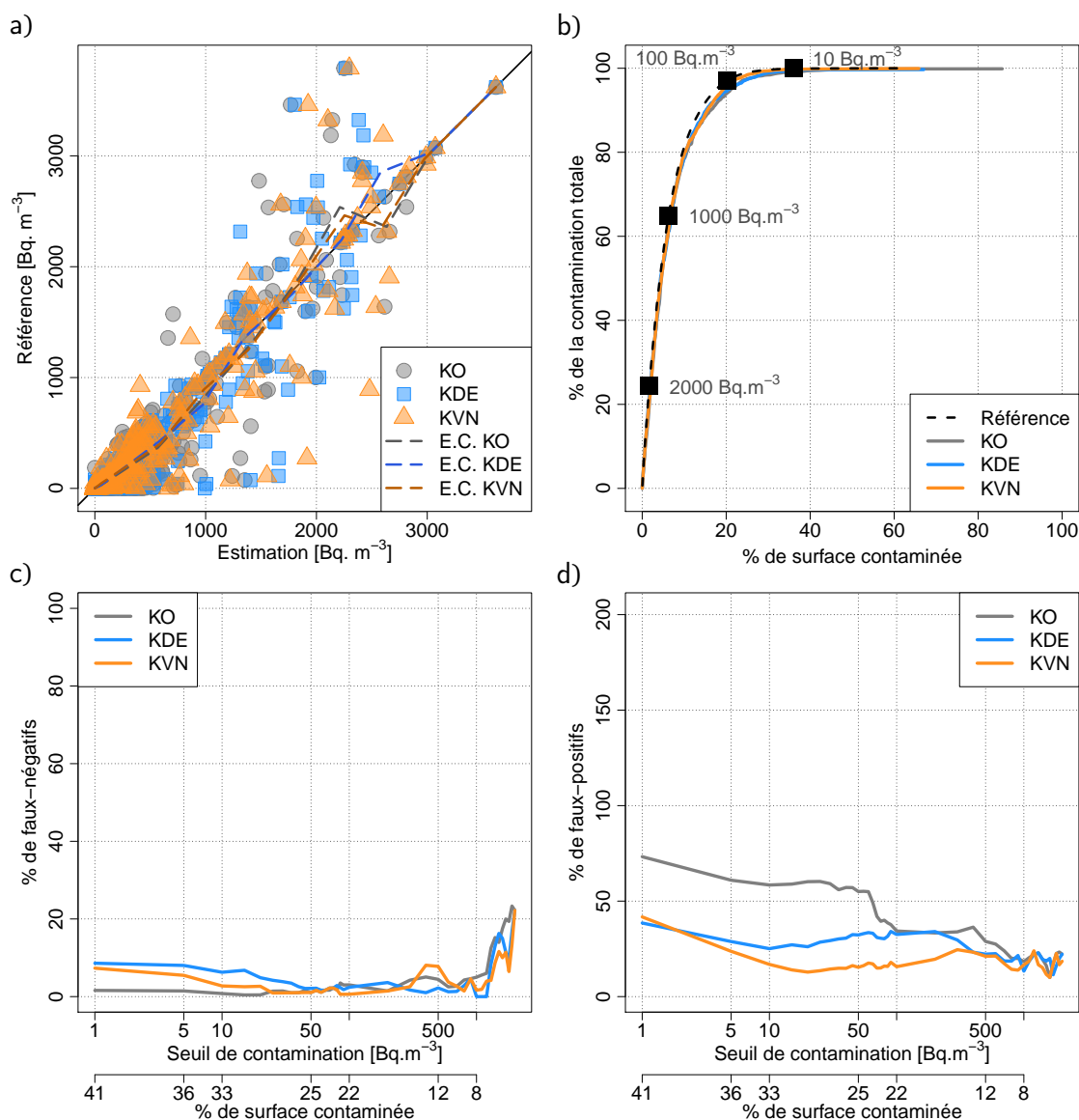


FIGURE E.6 – Indicateurs de performance pour le scénario S1 : a) comparaison entre les valeurs d'activité de la référence et les activités estimées ; b) courbes de sélectivité ; c) proportion de faux-négatifs par rapport à la surface réellement contaminée ; d) proportion de faux-positifs par rapport à la surface réellement contaminée.

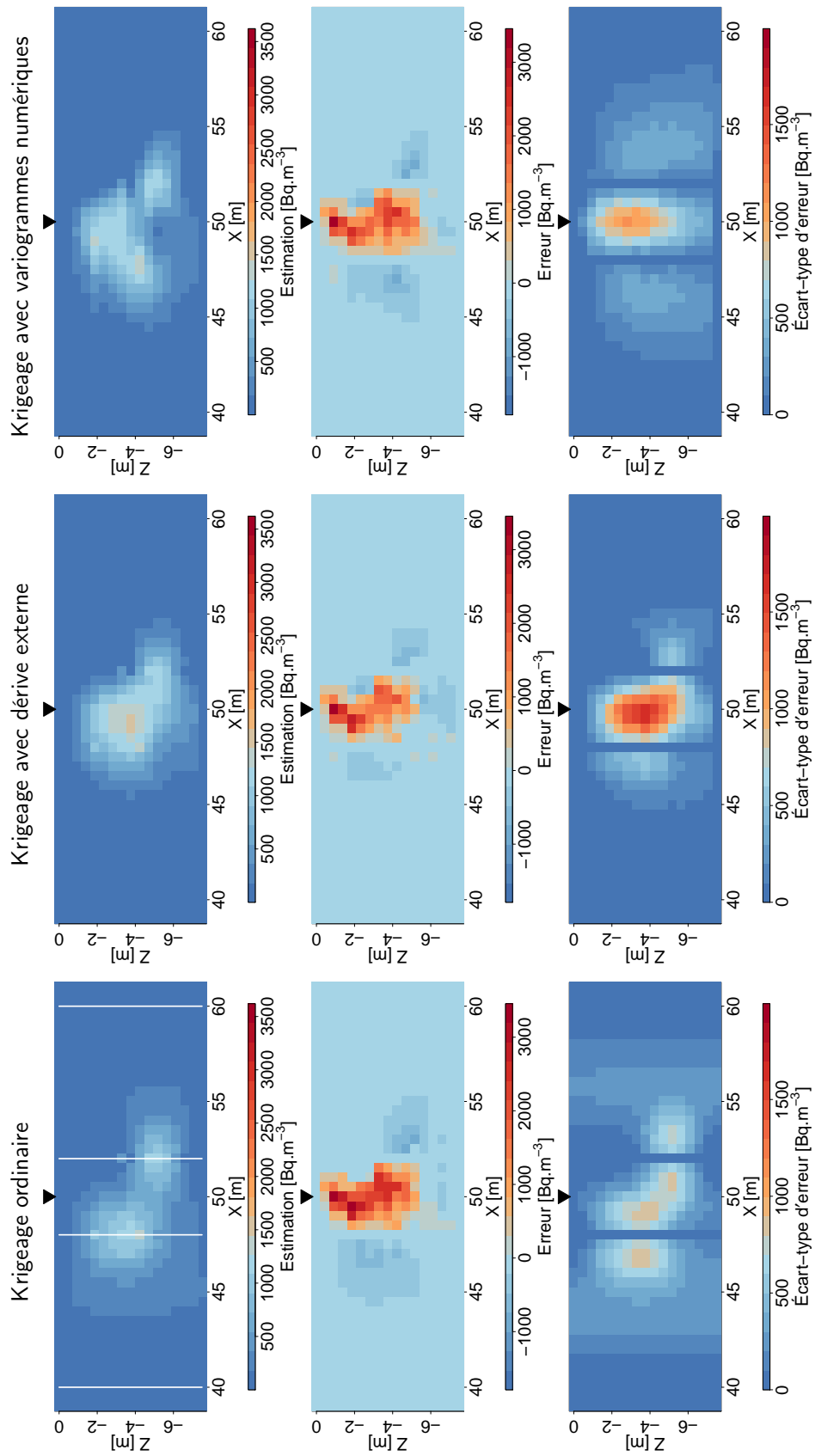


FIGURE E.7 – Cartes d'estimation, d'erreur d'estimation et d'écart-type d'erreur pour le scénario S2. Les échelles de couleur sont les mêmes que sur la Figure E.5.

TABLEAU E.4 – Erreurs moyennes obtenues pour le scénario S2.

	KO	KDE	KVN
EAM [Bq.m ⁻³]	243	138	159
REQM [Bq.m ⁻³]	512	380	425
ERM [-]	-66.9	-15.6	-9.2

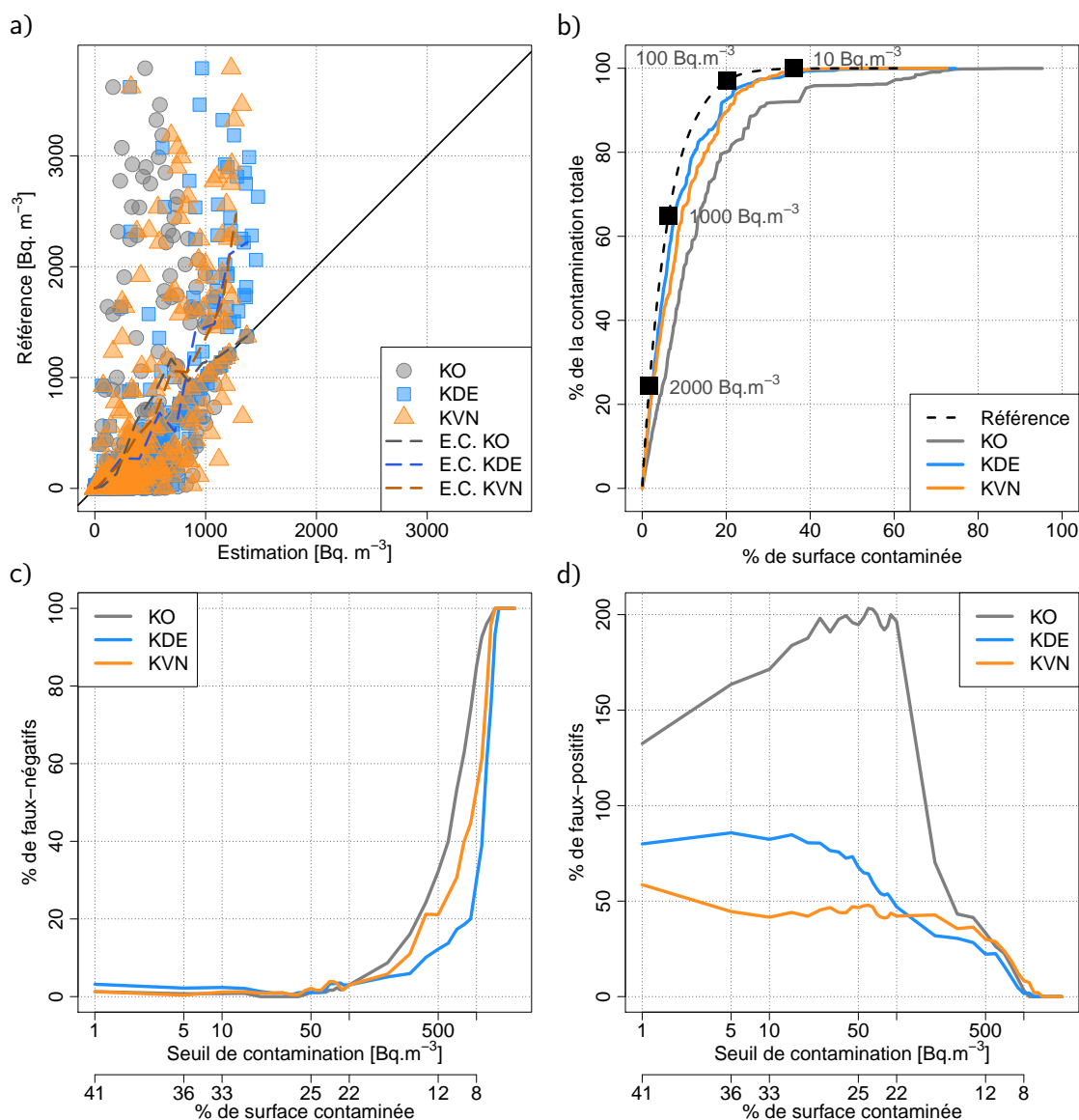


FIGURE E.8 – Indicateurs de performance pour le scénario S2 : a) comparaison entre les valeurs d'activité de la référence et les activités estimées ; b) courbes de sélectivité ; c) proportion de faux-négatifs par rapport à la surface réellement contaminée ; d) proportion de faux-positifs par rapport à la surface réellement contaminée.

Influence de la méthode de construction des champs de paramètres

Comme pour la référence principale, les estimations par krigeage avec variogrammes numériques ont été réalisées en utilisant uniquement les simulations obtenues à partir de l'approche 1, puis uniquement celles obtenues à partir de l'approche 2.

Les erreurs moyennes obtenues pour la référence stationnaire additionnelle sont données dans le Tableau E.5. Les résultats sont très différents selon le jeu de simulations utilisé : les erreurs sont beaucoup plus élevées en utilisant uniquement les simulations obtenues par l'approche 1. Pour la référence non stationnaire (Tableau E.6), les résultats sont assez proches quel que soit le jeu de simulations utilisé (erreurs très légèrement plus faibles avec l'approche 1 qu'avec l'approche 2).

TABLEAU E.5 – Erreurs moyennes obtenues par krigeage avec variogrammes numériques en utilisant uniquement les simulations générées par l'approche 1 (App. 1) puis uniquement celles générées par l'approche 2 (App. 2) pour la référence stationnaire additionnelle.

(a) Scénario S1			(b) Scénario S2	
	App. 1	App. 2	App. 1	App. 2
EAM [Bq.m ⁻³]	238	58	514	142
REQM [Bq.m ⁻³]	527	134	1259	328
ERM [-]	-4.5	-4.5	-6.4	-2.4

TABLEAU E.6 – Erreurs moyennes obtenues par krigeage avec variogrammes numériques en utilisant uniquement les simulations générées par l'approche 1 (App. 1) puis uniquement celles générées par l'approche 2 (App. 2) pour la référence non stationnaire.

(a) Scénario S1			(b) Scénario S2	
	App. 1	App. 2	App. 1	App. 2
EAM [Bq.m ⁻³]	70	76	152	171
REQM [Bq.m ⁻³]	211	217	416	429
ERM [-]	-2.4	-5.3	-4.6	-19.3

Annexe F

Résultats complémentaires de prévisions

Cette annexe présente des résultats complémentaires de prévisions par filtre particulaire, filtre de Kalman d'ensemble et krigeage spatio-temporel avec covariances numériques, pour le scénario S1 (référence principale) et pour les scénarios S1 et S2 (référence supplémentaire construite à partir de champs de paramètres de Mualem-van Genuchten stationnaires - Annexe D).

Référence principale - scénario S1

Comme pour le scénario S2, les panaches obtenus par krigeage sont moins lissés que ceux obtenus par filtres (Figure F1). À la date t_3 , les erreurs obtenues par filtre de Kalman d'ensemble et par krigeage sont très proches, et plus faibles que celles obtenues par filtre particulaire (Tableau F1). À la date t_4 , les erreurs moyennes sont proches pour les trois méthodes, sauf pour le critère S , qui souligne la nette surestimation de la contamination totale pour le filtre particulaire.

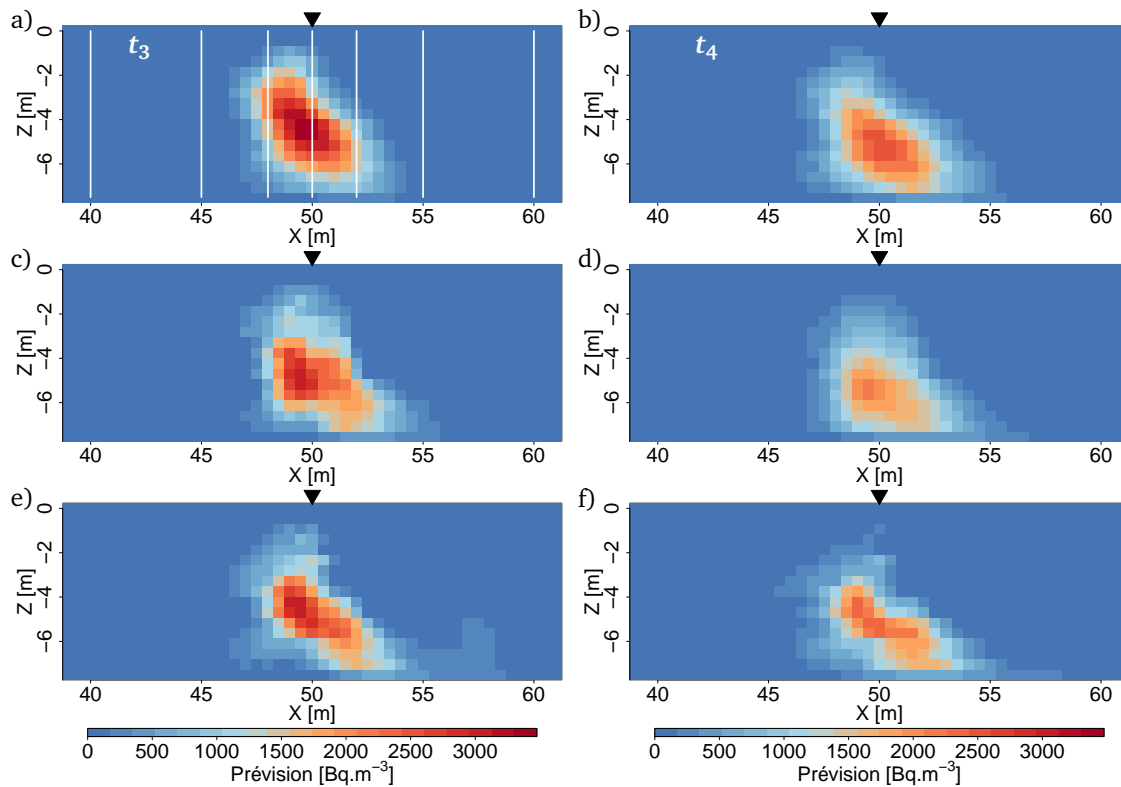


FIGURE F1 – Prévisions obtenues par filtre particulaire (a et b), filtre de Kalman d'ensemble (c et d) et krigeage spatio-temporel (e et f) pour le scénario S1 et pour les dates t_3 et t_4 .

TABLEAU F.1 – Erreurs moyennes obtenues par filtre particulaire (FP) avant et après l'assimilation des données à t_2 , par filtre de Kalman d'ensemble (EnKF) avant et après l'assimilation des données à t_2 et par krigeage pour le scénario S1 aux dates t_3 et t_4 .

(a) FP à t_3			(b) EnKF à t_3		(c) Krigeage à t_3
	Avant	Après	Avant	Après	
EAM [Bq.m^{-3}]	105	102	58	63	60
REQM [Bq.m^{-3}]	252	248	156	178	143
S [-]	1.23	1.24	1.09	1.1	1.05

(d) FP à t_4			(e) EnKF à t_4		(f) Krigeage à t_4
	Avant	Après	Avant	Après	
EAM [Bq.m^{-3}]	84	76	65	67	74
REQM [Bq.m^{-3}]	199	179	156	162	160
S [-]	1.23	1.25	1.06	1.05	1.03

Référence supplémentaire

Scénario S1

Le filtre particulaire ne reproduit pas la forme du panache, quelle que soit la date de prévision (Figure F.2). Le filtre de Kalman d'ensemble et le krigeage reproduisent bien cette forme, au moins pour la date t_3 . Les erreurs moyennes les plus faibles sont obtenues avec le filtre de Kalman d'ensemble et les erreurs les plus élevées avec le filtre particulaire (Tableau F.2). Cependant, le filtre de Kalman d'ensemble conduit systématiquement à une sous-estimation de la contamination ($S < 1$), ce qui n'est pas le cas pour les autres méthodes.

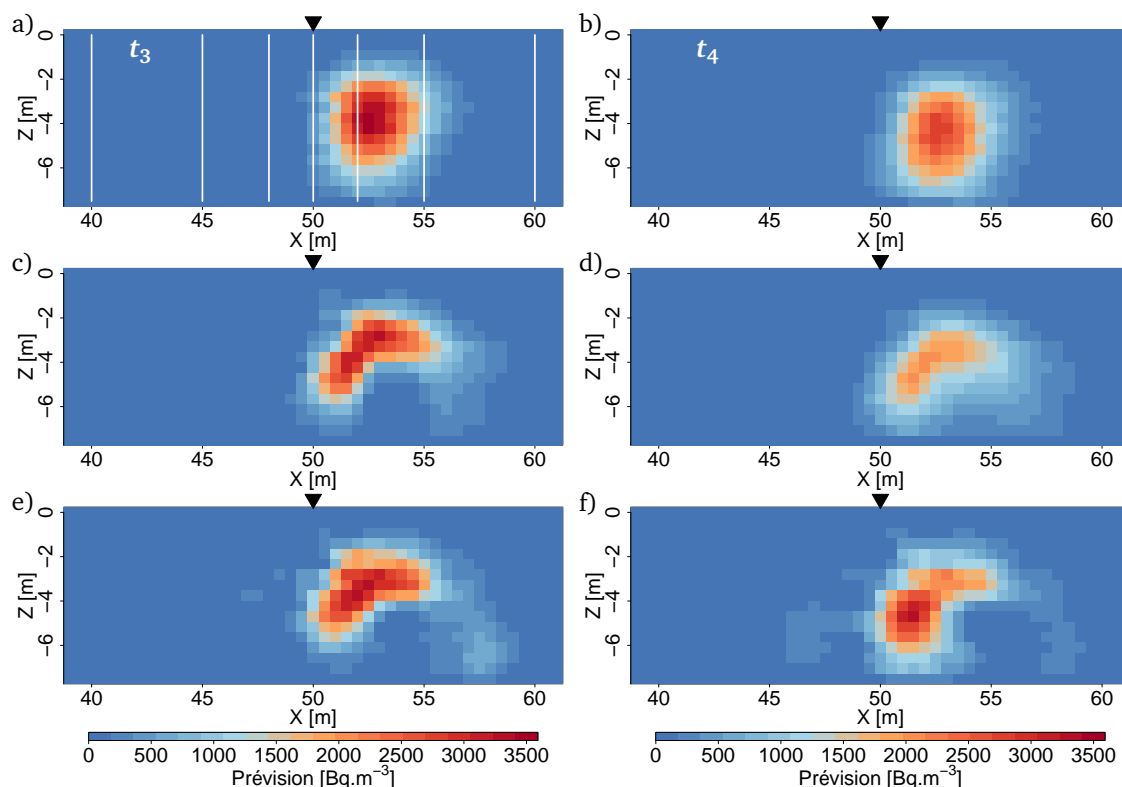


FIGURE F.2 – Prévisions obtenues par filtre particulaire (a et b), par filtre de Kalman d'ensemble (c et d) et par krigeage spatio-temporel (e et f) pour le scénario S1 et pour les dates t_3 et t_4 .

TABLEAU F.2 – Erreurs moyennes obtenues par filtre particulaire (FP) avant et après l'assimilation des données à t_2 , par filtre de Kalman d'ensemble (EnKF) avant et après l'assimilation des données à t_2 et par krigeage pour le scénario S1 et la référence supplémentaire aux dates t_3 et t_4 .

(a) FP à t_3			(b) EnKF à t_3			(c) Krigeage à t_3		
	Avant	Après	Avant	Après				
EAM [Bq.m^{-3}]	175	162	69	62			88	
REQM [Bq.m^{-3}]	419	406	152	137			184	
S [-]	1.06	1.08	0.94	0.96			1.05	

(d) FP à t_4			(e) EnKF à t_4			(f) Krigeage à t_4		
	Avant	Après	Avant	Après				
EAM [Bq.m^{-3}]	191	184	118	117			153	
REQM [Bq.m^{-3}]	436	429	252	248			296	
S [-]	1.01	1.04	0.91	0.93			1.09	

Scénario S2

Aucune des méthodes ne reproduit la forme du panache, quelle que soit la date de prévision, et les panaches apparaissent très lissés (Figure F.3). Comme pour le scénario S1, les erreurs moyennes les plus faibles sont obtenues avec le filtre de Kalman d'ensemble et les erreurs les plus élevées avec le filtre particulaire (Tableau F.3). Toutes les méthodes sous-estiment la contamination globale ($S < 1$), mais le filtre de Kalman d'ensemble la sous-estime un peu moins.

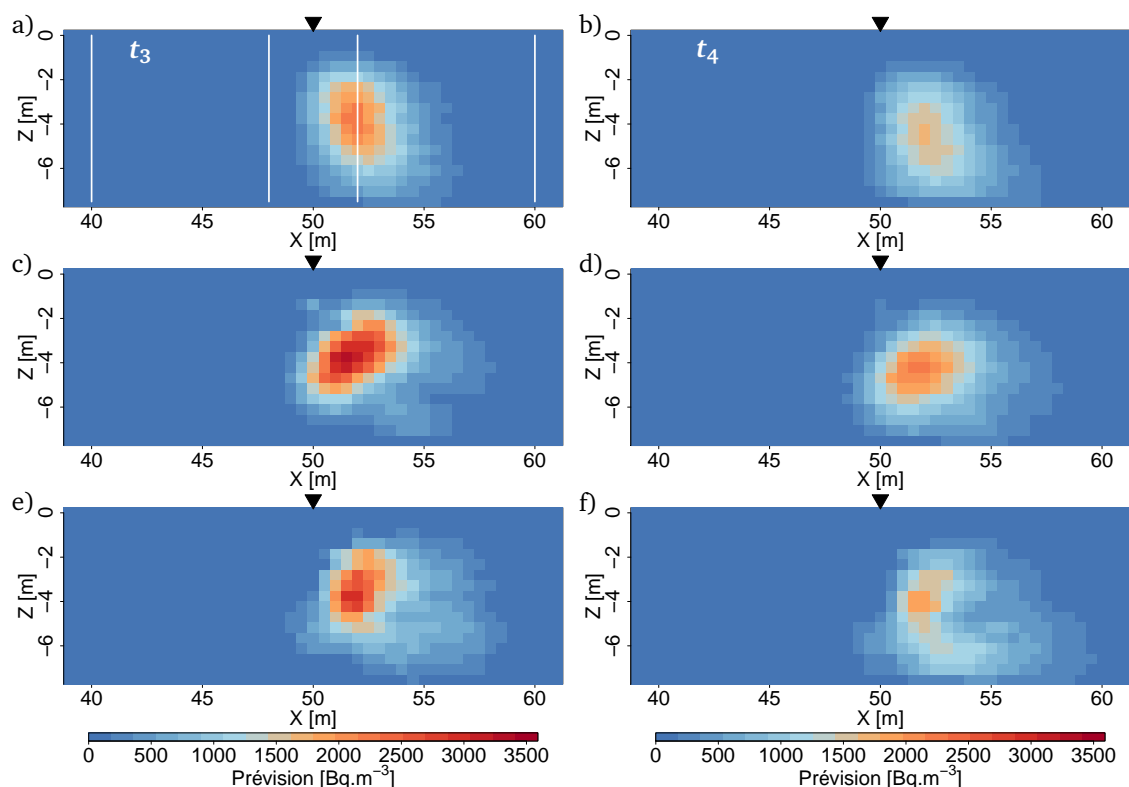


FIGURE F.3 – Prévisions obtenues par filtre particulaire (a et b), par filtre de Kalman d'ensemble (c et d) et par krigeage spatio-temporel (e et f) pour le scénario S2 et pour les dates t_3 et t_4 .

TABLEAU F.3 – Erreurs moyennes obtenues par filtre particulaire (FP) avant et après l'assimilation des données à t_2 , par filtre de Kalman d'ensemble (EnKF) avant et après l'assimilation des données à t_2 et par krigeage pour le scénario S2 et la référence supplémentaire aux dates t_3 et t_4 .

(a) FP à t_3			(b) EnKF à t_3		(c) Krigeage à t_3
	Avant	Après	Avant	Après	
EAM [Bq.m ⁻³]	198	186	127	123	125
REQM [Bq.m ⁻³]	429	403	285	272	286
S [-]	0.84	0.81	0.9	0.96	0.89
(d) FP à t_4			(e) EnKF à t_4		(f) Krigeage à t_4
	Avant	Après	Avant	Après	
EAM [Bq.m ⁻³]	197	187	144	149	168
REQM [Bq.m ⁻³]	423	401	312	321	361
S [-]	0.76	0.75	0.86	0.92	0.83

Annexe G

Analyses granulométriques à Prypiat Zaton et sur le site Epic

Cette annexe présente les schémas d'échantillonnage pour les données granulométriques à Prypiat Zaton et sur le site Epic.

Données de Prypiat Zaton

Les échantillons granulométriques de Prypiat Zaton ont été prélevés le long d'une coupe verticale d'environ 5 m de longueur et 2,5 m de profondeur (Figure G.1). Quelques sondages horizontaux ont également été effectués, sur une longueur de 2 m. Au total, 222 échantillons ont été prélevés, dont 97 dans les sables éoliens et 125 dans les sables alluviaux. Ce schéma d'échantillonnage permet de caractériser la variabilité de la conductivité hydraulique à l'échelle centimétrique voire pluri-centimétrique.

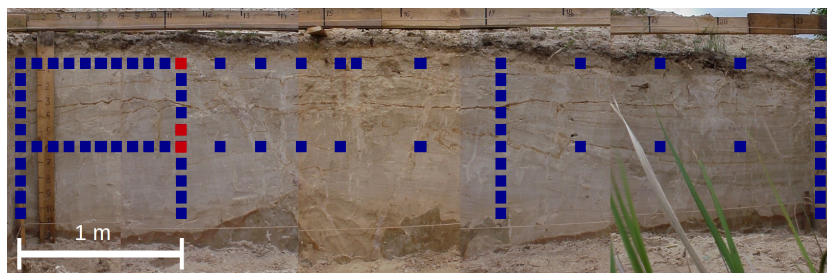


FIGURE G.1 – Partie supérieure de la coupe verticale dans laquelle ont été prélevés les échantillons. Chaque prélèvement est repéré par un carré bleu et les têtes des sondages horizontaux sont repérées par les carrés rouges.

Données du site Epic

Les données du site Epic sont des fractions granulométriques en une vingtaine de sondages de 6 à 8 m de profondeur, avec un prélèvement tous les 25 cm sur chaque sondage (Figure G.2). Il y a 368 échantillons en tout, dont 265 dans les sables éoliens et 103 dans les sables alluviaux. Ce schéma d'échantillonnage permet de quantifier la variabilité de la conductivité hydraulique à l'échelle métrique, voire pluri-métrique.

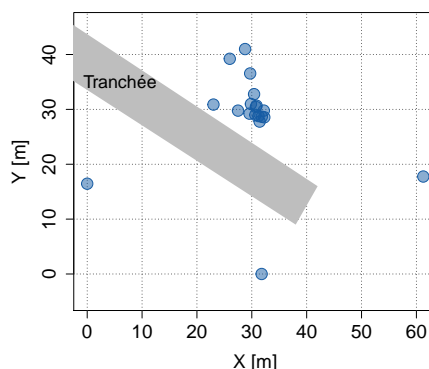


FIGURE G.2 – Localisation des sondages de prélèvement granulométrique sur le site Epic. La localisation et la forme de la tranchée sont approximatives. Les coordonnées sont projetées dans un repère local.

Annexe H

Analyse de la variabilité des paramètres de Mualem-van Genuchten (zone de traçage)

Cette annexe présente quelques analyses complémentaires menées sur les paramètres de Mualem-van Genuchten dans la zone du traçage au ^{36}Cl .

Variabilité des paramètres suivant la profondeur

La Figure H.1 présente l'évolution des paramètres K_s , α et n en fonction de la profondeur. Les valeurs des paramètres sont relativement homogènes d'un niveau à l'autre, sauf pour les niveaux situés à 0,8 et 0,9 m de profondeur, qui présentent des valeurs de paramètres plus élevées.

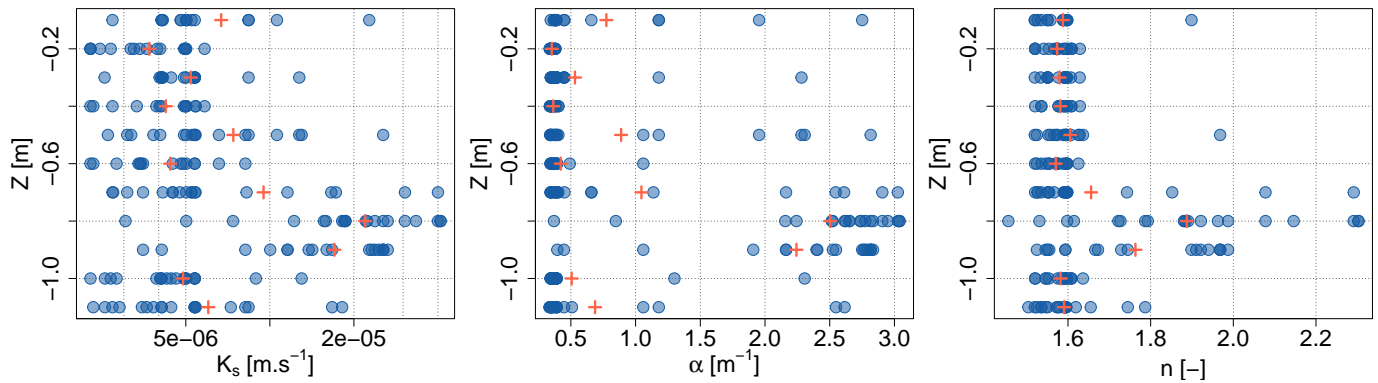


FIGURE H.1 – Paramètres K_s , α et n en fonction de la profondeur. Les croix indiquent les moyennes par classe de profondeur.

Ces variations sont cohérentes avec les variations des surfaces spécifiques mesurées dans la zone de traçage (Figure H.2).

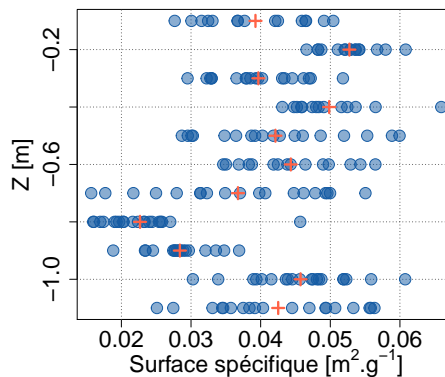


FIGURE H.2 – Évolution de la surface spécifique en fonction de la profondeur.

Variogrammes

Les variogrammes expérimentaux dans les directions horizontales et verticales, ainsi que les modèles ajustés sont présentés sur la Figure H.3, pour les paramètres K_s , α et n , ainsi que pour la surface spéci-

fique. Les variogrammes expérimentaux ont des allures similaires pour K_s , α et la surface spécifique. Les ajustements sont assez variables d'un paramètre à l'autre.

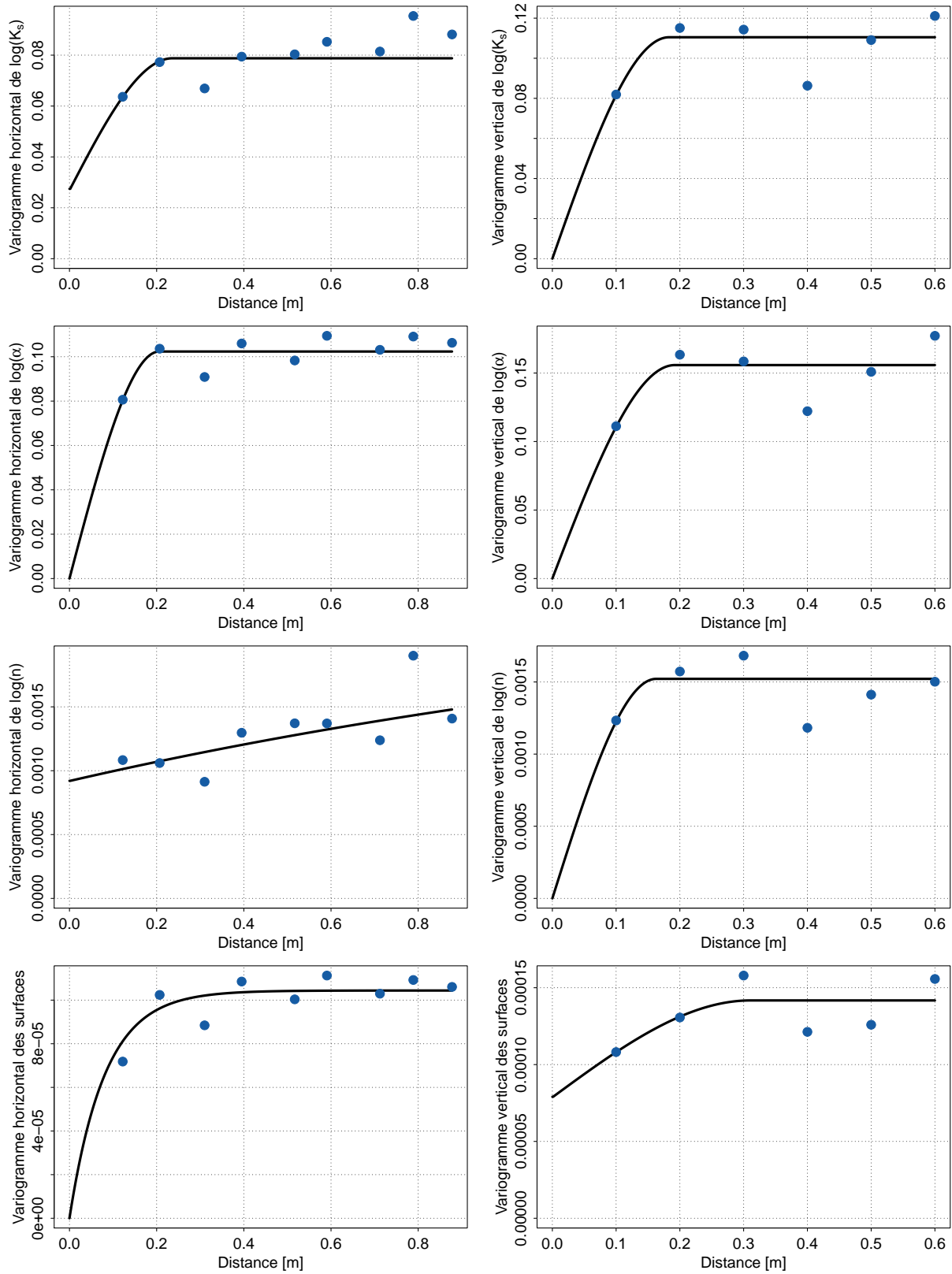


FIGURE H.3 – Variogrammes expérimentaux de $\log(K_s)$, $\log(\alpha)$, $\log(n)$ et de la surface spécifique dans les directions horizontales et verticales et exemples d'ajustement.

Annexe I

Résultats complémentaires sur les données de la T22

Cette annexe présente les résultats détaillés pour les dates non présentées au Chapitre 7.

Critères qualitatifs

Les cartes d'estimation et d'écart-type théorique d'estimation sont données sur les Figures I.1, I.2, I.3 et I.4. Pour les 18/05/2005, 01/10/2005 et 01/10/2006, la forme des panaches est très semblable quelle que soit la méthode ; les valeurs d'activité estimées sont cependant plus élevées pour le krigeage avec variogrammes numériques que pour les deux autres méthodes. Les formes des panaches diffèrent légèrement entre le krigeage ordinaire et le krigeage avec dérive externe d'une part et le krigeage avec variogrammes numériques d'autre part pour le 23/04/2010. Les valeurs d'écart-types théoriques sont comparables pour le krigeage ordinaire et le krigeage avec dérive externe, mais diffèrent nettement de celles obtenues par krigeage avec variogrammes numériques.

Les profils estimés sont donnés sur la Figure I.5. La capacité des méthodes d'estimation à reproduire ces profils dépend du piézomètre et de la date d'estimation. Le 18/05/2005, aucune méthode ne permet de reproduire la valeur forte observée au point P4-1. Le 1/10/2005, au piézomètre P4, les meilleures estimations sont obtenues par krigeage ordinaire et krigeage avec dérive externe, alors qu'au piézomètre P5, les meilleures estimations sont obtenues par krigeage avec variogrammes numériques.

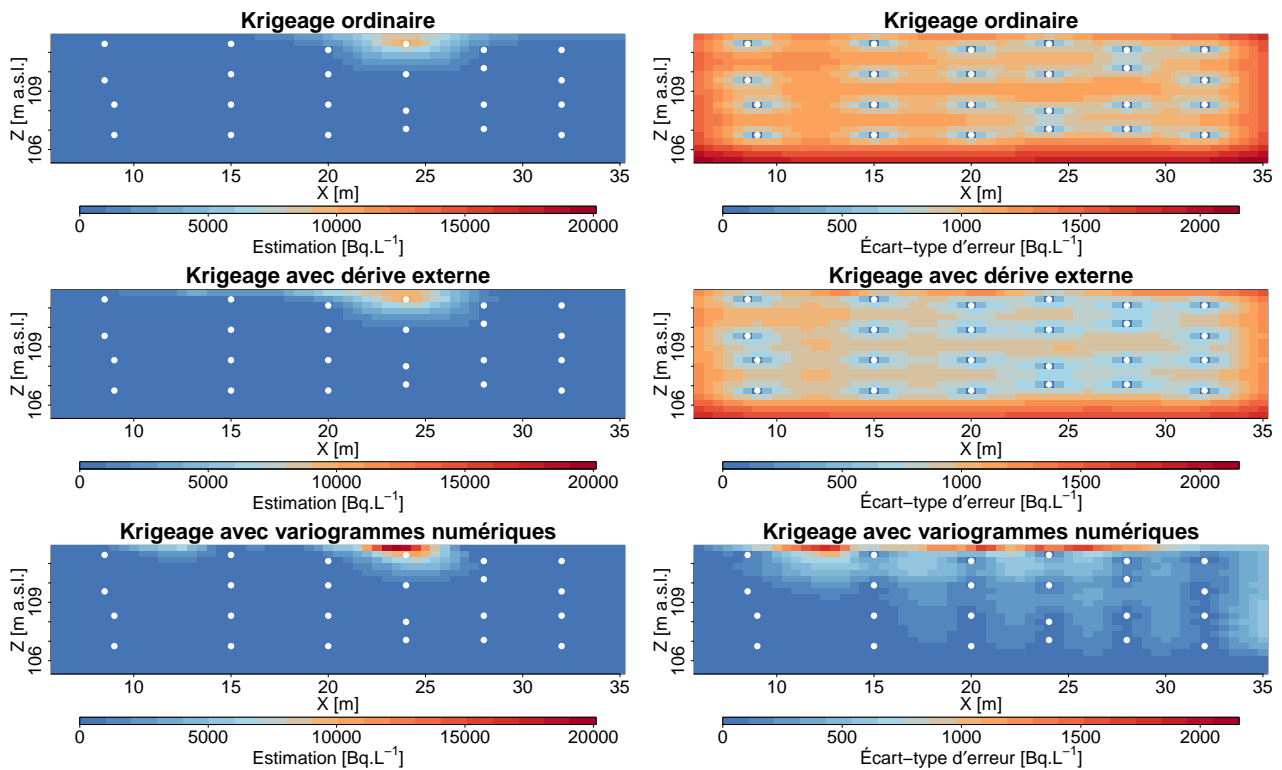


FIGURE I.1 – Cartes d'estimation et d'écart-type théorique d'erreur d'estimation au 18/05/2005.

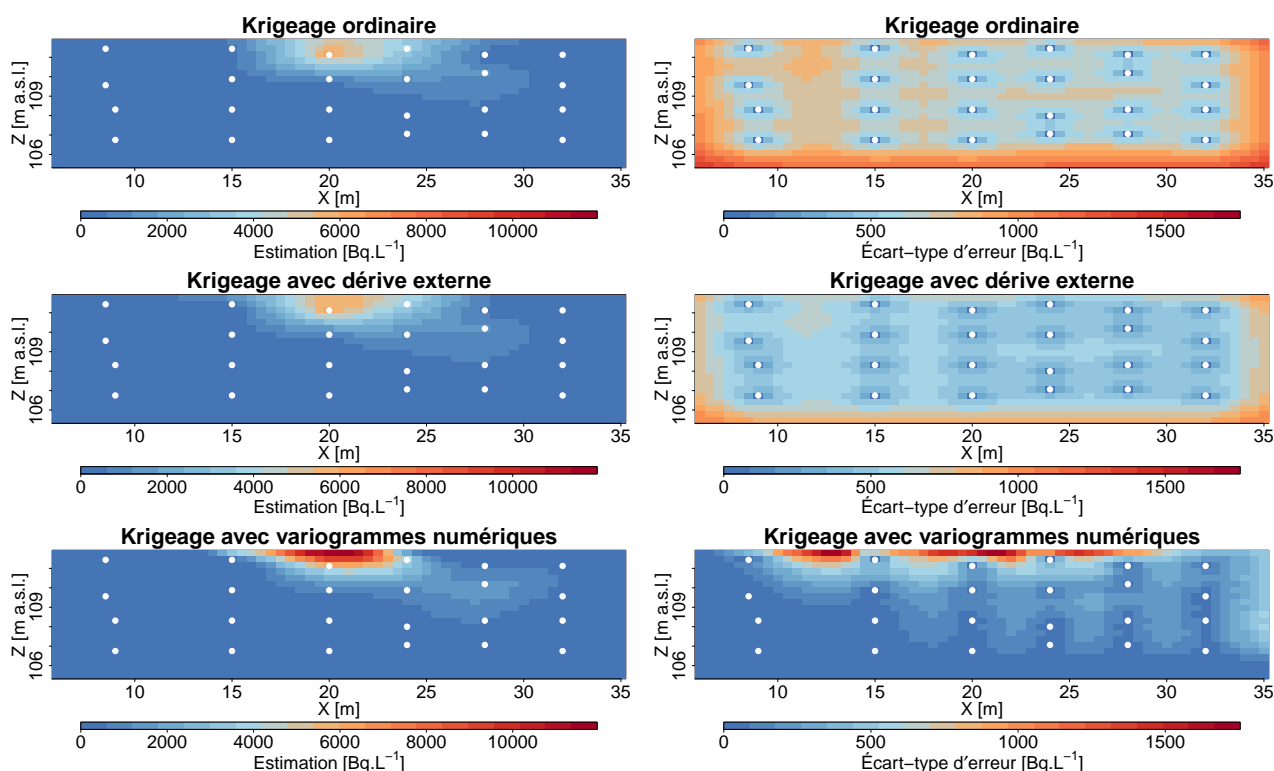


FIGURE I.2 – Cartes d'estimation et d'écart-type théorique d'erreur d'estimation au 01/10/2005.

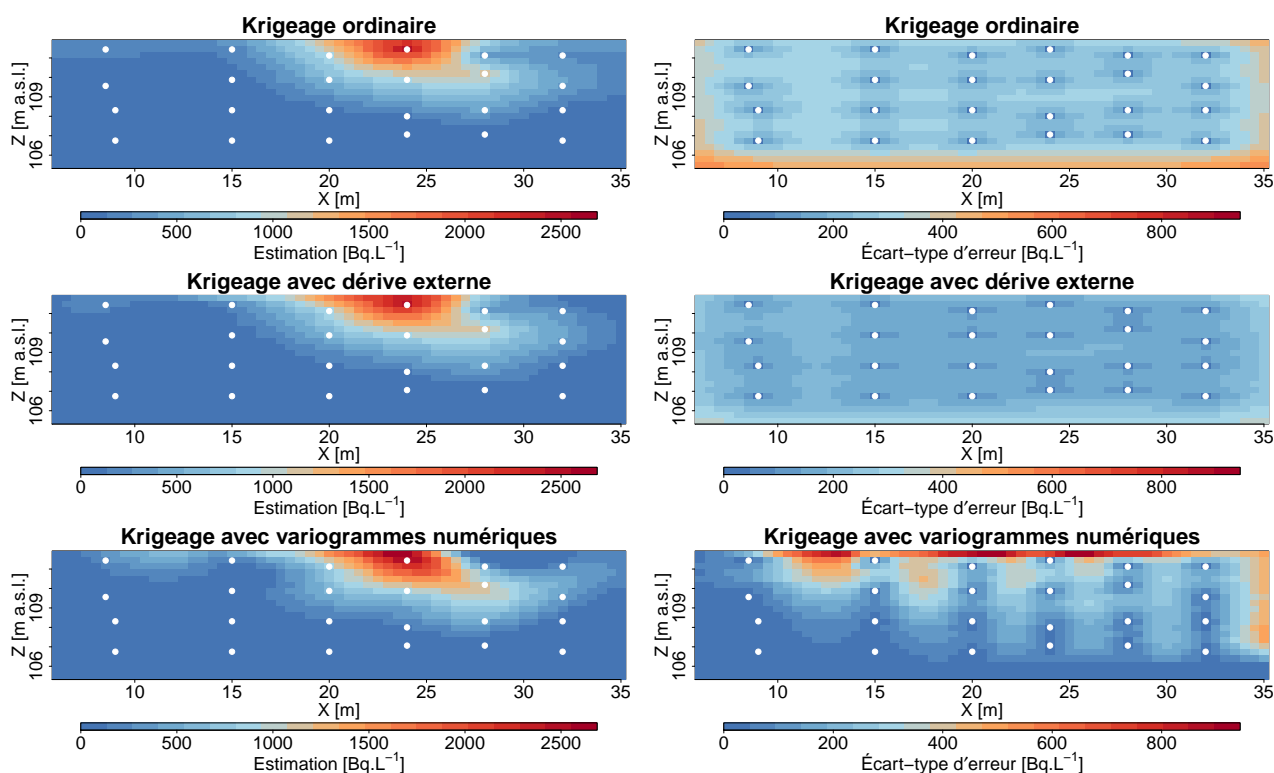


FIGURE I.3 – Cartes d'estimation et d'écart-type théorique d'erreur d'estimation au 01/10/2006.

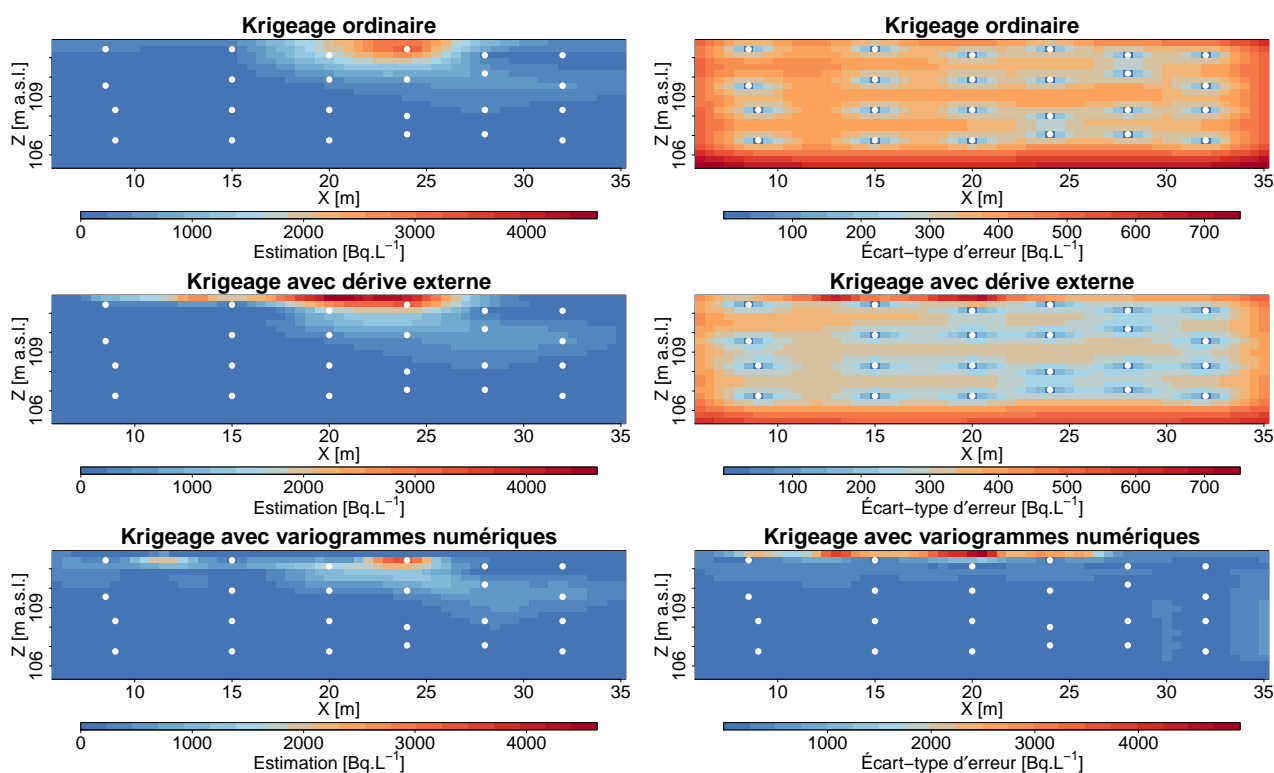


FIGURE I.4 – Cartes d'estimation et d'écart-type théorique d'erreur d'estimation au 23/04/2010.

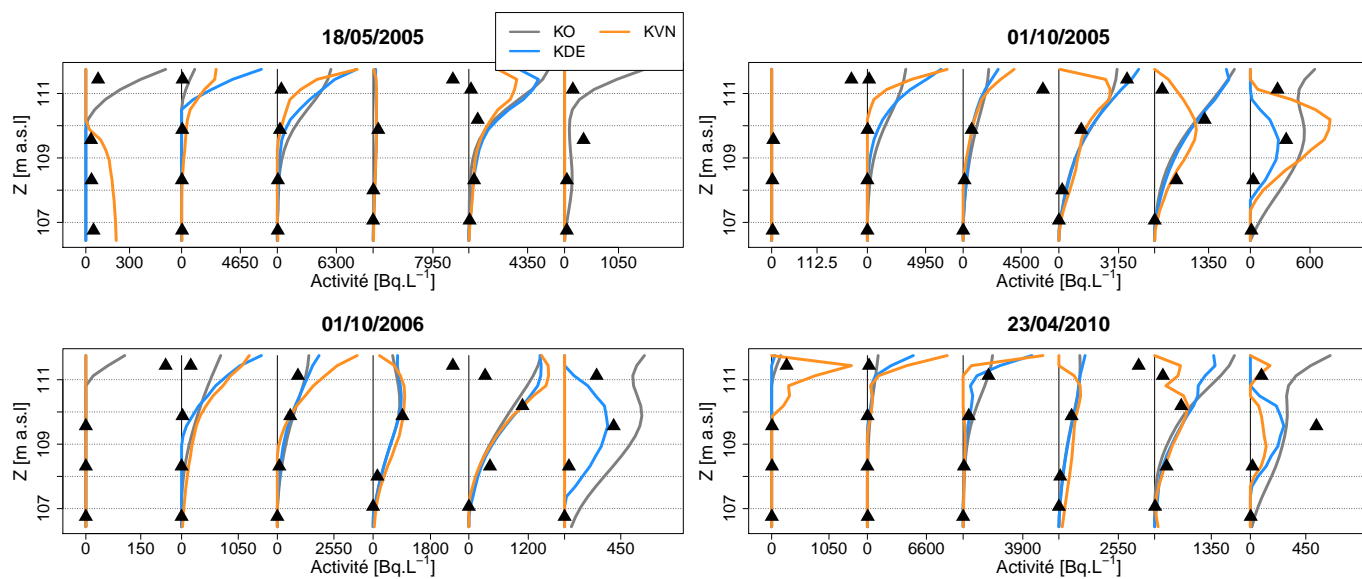


FIGURE I.5 – Profils estimés aux quatre dates.

Erreurs absolues moyennes en fonction du nombre de piézomètres observés

Quelle que soit la date considérée, les erreurs absolues moyennes les plus élevées sont obtenues par krigeage ordinaire, lorsque seuls deux, trois ou quatre piézomètres sont disponibles (Figure I.6). Lorsque cinq piézomètres sont disponibles, les performances des méthodes dépendent de la date considérée.

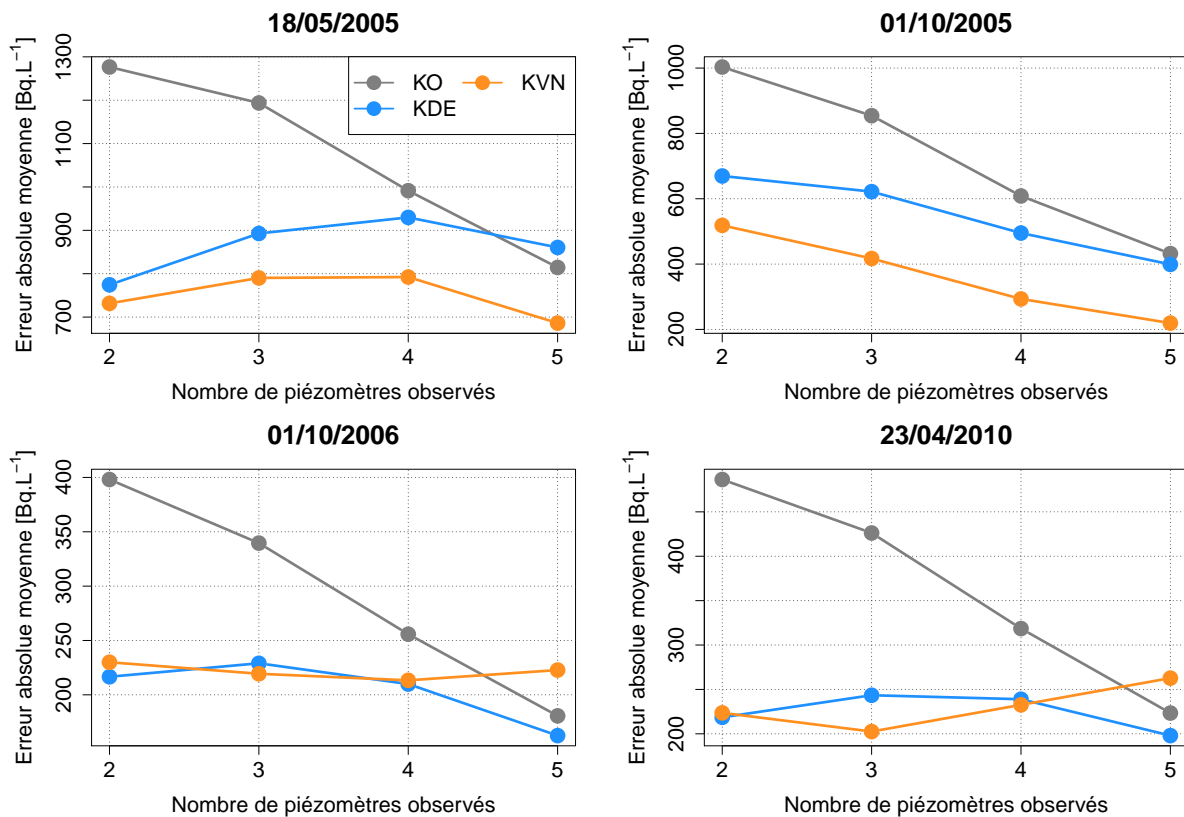


FIGURE I.6 – Erreurs absolues moyennes en fonction du nombre de piézomètres observés pour quatre dates.

Cas particulier du 18/05/2005

Le 18/05/2005, le point de prélèvement P4-1 présente une valeur d'activité en ⁹⁰Sr très supérieure à celles observées aux piézomètres voisins. Pour évaluer l'influence de cette forte valeur sur les estimations, celle-ci a été ôtée des données. Les profils ainsi estimés (Figure I.7) sont en meilleure adéquation avec les observations que sur la Figure I.5. Par ailleurs, comme attendu, les erreurs absolues moyennes sont plus faibles (Figure I.7), et le krigeage ordinaire devient aussi performant que les autres méthodes, quel que soit le nombre de piézomètres disponibles.

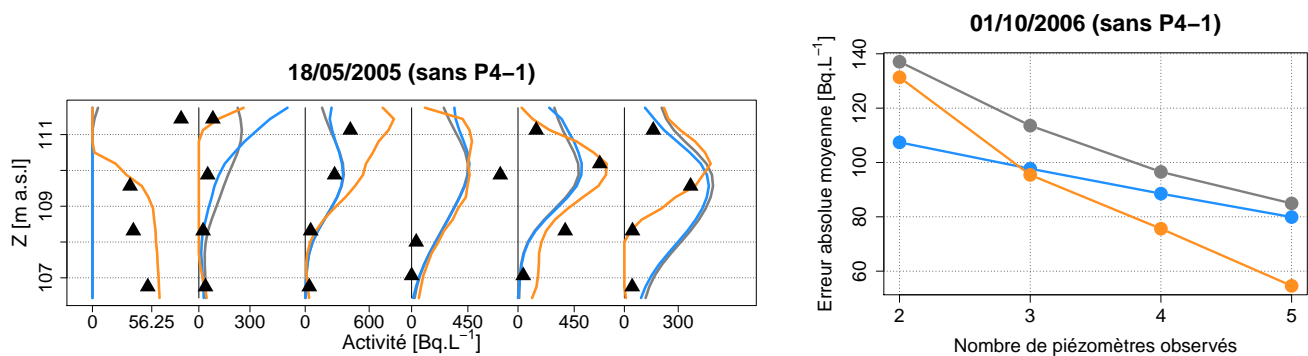


FIGURE I.7 – Profils estimés et erreur absolue moyenne en fonction du nombre de sondages le 18/05/2005 en enlevant la donnée du point de prélèvement P4-1.

Références

- J. Aitchison. The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(2): 139–177, 1982.
- A. Alcolea, J. Carrera, et A. Medina. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. *Advances in Water Resources*, 29: 1678–1689, 2006. doi: 10.1016/j.advwatres.2005.12.009.
- M. Alfaro. *Étude de la robustesse des simulations de fonctions aléatoires*. Thèse de Doctorat en Sciences et techniques minières - option Géostatistiques, École Nationale Supérieure des Mines de Paris, 1979. 165p.
- B. Amaziane, M. El Ossmani, et C. Serres. Numerical modeling of the flow and transport of radionuclides in heterogeneous porous media. *Computational Geosciences*, 12(4): 437–449, 2008. doi: 10.1007/s10596-008-9083-0.
- H. Amor, F. Benkhaldoun, M. Bourgeois, et M.-H. Le. Development of an adaptive mesh refinement strategy for the MELODIE software simulating flow and radionuclides transport in heterogeneous porous media. In *MAMERN VI–2015 Conference*, Pau, France, 2015.
- M. P. Anderson, W. W. Woessner, et R. J. Hunt. *Applied Groundwater Modeling*. Elsevier, 2015. ISBN 978-0-12-058103-0.
- W. M. Appels, A. M. Ireson, et S. L. Barbour. Impact of bimodal textural heterogeneity and connectivity on flow and transport through unsaturated mine waste rock. *Advances in Water Resources*, 112: 254–265, 2018. doi: 10.1016/j.advwatres.2017.12.008.
- C. Ardois, N. van Meir, et A. Martin-Garin. *Migration en milieu insaturé et dans les nappes superficielles - Rapport scientifique et technique IRSN*. 2007.
- M. J. Asher, B. F. W. Croke, A. J. Jakeman, et L. J. M. Peeters. A review of surrogate models and their application to groundwater modeling. *Water Resources Research*, 51(8): 5957–5973, 2015. doi: 10.1002/2015WR016967.
- M. T. Ayvaz. A hybrid simulation–optimization approach for solving the areal groundwater pollution source identification problems. *Journal of Hydrology*, 538: 161–176, 2016. doi: 10.1016/j.jhydrol.2016.04.008.
- N. H. Batjes, E. Ribeiro, A. van Oostrum, J. Leenaars, T. Hengl, et J. Mendes de Jesus. WoSIS: providing standardised soil profile data for the world. *Earth System Science Data*, 9(1): 1–14, 2017. doi: 10.5194/essd-9-1-2017.
- M. Beauchamp, L. Malherbe, C. de Fouquet, L. Létinois, et F. Tognet. A polynomial approximation of the traffic contributions for kriging-based interpolation of urban air quality model. *Environmental Modelling & Software*, 105: 132–152, 2018. doi: 10.1016/j.envsoft.2018.03.033.
- A. Bechler, T. Romary, N. Jeannée, et Y. Desnoyers. Geostatistical sampling optimization of contaminated facilities. *Stochastic Environmental Research and Risk Assessment*, 27(8): 1967–1974, 2013. doi: 10.1007/s00477-013-0731-0.
- J. P. Bennett, C. P. Haslauer, M. Ross, et O. A. Cirpka. An Open, Object-Based Framework for Generating Anisotropy in Sedimentary Subsurface Models. *Groundwater*, 57(3): 420–429, 2019. doi: 10.1111/gwat.12803.
- L. Bertino, G. Evensen, et H. Wackernagel. Combining geostatistics and Kalman filtering for data assimilation in an estuarine system. *Inverse Problems*, 18(1): 1–23, 2002. doi: 10.1088/0266-5611/18/1/301.

- K. Beven. Spatially Distributed Modeling: Conceptual Approach to Runoff Prediction. In D. Bowles et P. O'Connell, editors, *Recent Advances in the Modeling of Hydrologic Systems*, NATO ASI Series (Series C: Mathematical and Physical Sciences). Springer, Dordrecht, 1991. ISBN 978-94-010-5538-3.
- K. Beven. How to make advances in hydrological modelling. *Hydrology Research*, 50(6): 1481–1494, 2019. doi: 10.2166/nh.2019.134.
- J. Bevington, D. Piragnolo, P. Teatini, G. Vellidis, et F. Morari. On the spatial variability of soil hydraulic properties in a Holocene coastal farmland. *Geoderma*, 262: 294–305, 2016. doi: 10.1016/j.geoderma.2015.08.025.
- M. Bobbia, A. Cori, et C. de Fouquet. Représentativité spatiale d'une station de mesure de la pollution atmosphérique. *Pollution Atmosphérique*, 197: 63–75, 2008. doi: 10.4267/pollution-atmospherique.1265.
- P. Bogaert. Comparison of kriging techniques in a space-time context. *Mathematical Geology*, 28(1): 73–86, 1996. doi: 10.1007/BF02273524.
- F. E. Botros, T. Harter, Y. S. Onsoy, A. Tuli, et J. W. Hopmans. Spatial Variability of Hydraulic Properties and Sediment Characteristics in a Deep Alluvial Unsaturated Zone. *Vadose Zone Journal*, 8(2): 276–289, 2009. doi: 10.2136/vzj2008.0087.
- J.-P. Boudreault, J.-S. Dubé, et D. Marcotte. Quantification and minimization of uncertainty by geostatistical simulations during the characterization of contaminated sites: 3-D approach to a multi-element contamination. *Geoderma*, 264(Part A): 214–226, 2016. doi: 10.1016/j.geoderma.2015.10.019.
- F. Boulanger. *Modélisation et simulation de variables régionalisées par des fonctions aléatoires stables*. Thèse de Doctorat en Géostatistiques, École Nationale Supérieure des Mines de Paris, 1990. 380p.
- M. Bourotte, D. Allard, et E. Porcu. A flexible class of non-separable cross-covariance functions for multivariate space–time data. *Spatial Statistics*, 18(Part A): 125–146, 2016. doi: 10.1016/j.spasta.2016.02.004.
- D. Bugai, L. Dewière, V. Kashparov, et N. Ahamdach. Strontium-90 transport parameters from source term to aquifer in the Chernobyl Pilot Site. *Radioprotection*, 37(C1): 11–16, 2002. doi: 10.1051/radiopro/2002024.
- D. Bugai, A. Skalsky, S. Dzhepo, Y. Kubko, V. Kashparov, N. van Meir, D. Stammose, C. Simonucci, et A. Martin-Garin. Radionuclide migration at experimental polygon at Red Forest waste site in Chernobyl zone. Part 2: Hydrogeological characterization and groundwater transport modeling. *Applied Geochemistry*, 27(7): 1359–1374, 2012. doi: 10.1016/j.apgeochem.2011.09.028.
- D. Bugai, J. Smith, et M. A. Hoque. Solid-liquid distribution coefficients (Kd-s) of geological deposits at the Chernobyl Nuclear Power Plant site with respect to Sr, Cs and Pu radionuclides: A short review. *Chemosphere*, 242: 1–12, 2020. doi: 10.1016/j.chemosphere.2019.125175.
- F. Cadini, E. Tosoni, et E. Zio. Modeling the release and transport of ⁹⁰Sr radionuclides from a superficial nuclear storage facility. *Stochastic Environmental Research and Risk Assessment*, 30(2): 693–712, 2016. doi: 10.1007/s00477-015-1112-7.
- O. Cappe, S. J. Godsill, et E. Moulines. An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo. *Proceedings of the IEEE*, 95(5): 899–924, 2007. doi: 10.1109/JPROC.2007.893250.
- J. Carrera, A. Alcolea, A. Medina, J. Hidalgo, et L. J. Slooten. Inverse problem in hydrogeology. *Hydrogeology Journal*, 13(1): 206–222, 2005. doi: 10.1007/s10040-004-0404-7.
- R. Carrizo Vergara. *Développement de modèles géostatistiques à l'aide d'équations aux dérivées partielles stochastiques*. Thèse de Doctorat en Géostatistiques et Probabilités Appliquées, PSL Research University, 2018. 315p, english.

- I. Chen, Y. Chuo, et H. Ma. Uncertainty analysis of remediation cost and damaged land value for brownfield investment. *Chemosphere*, 220: 371–380, 2019. doi: 10.1016/j.chemosphere.2018.12.116.
- J.-P. Chilès. *Géostatistique des phénomènes non-stationnaires*. Thèse de doctorat, Université de Nancy-I, 1977. 146p.
- J.-P. Chilès et P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty, 2nd Edition*. Wiley series in probability and statistics. Wiley, 2012. ISBN 978-0-470-18315-1.
- G. Dagan. *Flow and transport in porous formation*. Springer, Berlin, Heidelberg, 1989. ISBN 978-3-540-51098-7.
- B. Datta, D. Chakrabarty, et A. Dhar. Identification of unknown groundwater pollution sources using classical optimization with linked simulation. *Journal of Hydro-environment Research*, 5(1): 25–36, 2011. doi: 10.1016/j.jher.2010.08.004.
- C. de Fouquet, Y. Benoît, C. Carpentier, et B. Fricaudet. Uncertainties on the extension of a polluted zone. In *Proceedings of the ASME 2011 14th International Conference on Environmental Remediation and Radioactive Waste Management*, Reims, France, 2011.
- G. de Marsily, F. Delay, J. Goncalves, P. Renard, V. Teles, et S. Violette. Dealing with spatial heterogeneity. *Hydrogeology Journal*, 13(1): 161–183, 2005. doi: 10.1007/s10040-004-0432-3.
- J.-P. Delhomme. Kriging in the hydrosiences. *Advances in Water Resources*, 1(5): 251–266, 1978. doi: 10.1016/0309-1708(78)90039-8.
- H. Demougeot-Renard et C. de Fouquet. Geostatistical Approach for Assessing Soil Volumes Requiring Remediation: Validation Using Lead-Polluted Soils underlying a Former Smelting Works. *Environmental Science & Technology*, 38(19): 5120–5126, 2004. doi: 10.1021/es0351084.
- A. J. Desbarats, C. E. Logan, M. J. Hinton, et D. R. Sharpe. On the kriging of water table elevations using collateral information from a digital elevation model. *Journal of Hydrology*, 255(1): 25–38, 2002. doi: 10.1016/S0022-1694(01)00504-2.
- L. Dewièrre, D. Bugai, C. Grenier, V. Kashparov, et N. Ahamdach. 90Sr migration to the geo-sphere from a waste burial in the Chernobyl exclusion zone. *Journal of Environmental Radioactivity*, 74(1): 139–150, 2004. doi: 10.1016/j.jenvrad.2004.01.019.
- J. Doherty. Ground Water Model Calibration Using Pilot Points and Regularization. *Groundwater*, 41(2): 170–177, 2003. doi: 10.1111/j.1745-6584.2003.tb02580.x.
- M. Donati et C. de Fouquet. How to describe organic contamination in soils: A model comparison for chlorinated solvent concentrations at industrial site scale. *Science of The Total Environment*, 633: 1480–1495, 2018. doi: 10.1016/j.scitotenv.2018.03.257.
- A. Doucet, N. de Freitas, et N. Gordon. An introduction to Sequential Monte Carlo Methods. In *Sequential Monte Carlo Methods in Practice*, Information Science and Statistics, X-Pferdchen. Springer-Verlag, New York, 2001. ISBN 978-1-4419-2887-0.
- G. Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer-Verlag, Berlin Heidelberg, 2006. ISBN 978-3-642-03710-8.
- A. Farchi et M. Bocquet. Review article: Comparison of local particle filters and new implementations. *Nonlinear Processes in Geophysics*, 25(4): 765–807, 2018. doi: 10.5194/npg-25-765-2018.
- E. Fetel et G. Caumon. Reservoir flow uncertainty assessment using response surface constrained by secondary information. *Journal of Petroleum Science and Engineering*, 60(3): 170–182, 2008. doi: 10.1016/j.petrol.2007.06.003.

- G. E. Fogg et Y. Zhang. Debates—Stochastic subsurface hydrology from theory to practice: A geologic perspective. *Water Resources Research*, 52(12): 9235–9245, 2016. doi: 10.1002/2016WR019699.
- F. Fouedjio. *Contributions à la modélisation et à l'inférence des fonctions aléatoires non-stationnaires de second ordre*. Thèse de Doctorat en Géostatistiques, École Nationale Supérieure des Mines de Paris, 2014. 190p.
- R. A. Freeze. A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media. *Water Resources Research*, 11(5): 725–741, 1975. doi: 10.1029/WR011i005p00725.
- T. Gneiting, M. Genton, et P. Guttorp. Geostatistical Space-Time Models, Stationarity, Separability and Full Symmetry. Technical Report 475, Department of Statistics, University of Washington, 2007.
- M. Herbst, B. Diekkrüger, et H. Vereecken. Geostatistical co-regionalization of soil hydraulic properties in a micro-scale catchment using terrain attributes. *Geoderma*, 132(1): 206–221, 2006. doi: 10.1016/j.geoderma.2005.05.008.
- A. Horta, P. Correia, L. M. Pinheiro, et A. Soares. Geostatistical Data Integration Model for Contamination Assessment. *Mathematical Geosciences*, 45(5): 575–590, 2013. doi: 10.1007/s11004-013-9466-1.
- C. Houzé, A. Saintenoy, L. Pannecoucke, M. Le Coz, C. Cazala, et C. de Fouquet. Evaluation of the spatial variations of the hydrodynamical parameters of soils with GPR. In *European Geosciences Union General Assembly, EGU*, Vienne, Austria, 2019.
- H. Hu, Z. Hu, K. Zhong, J. Xu, F. Zhang, Y. Zhao, et P. Wu. Satellite-based high-resolution mapping of ground-level PM_{2.5} concentrations over East China using a spatiotemporal regression kriging model. *Science of The Total Environment*, 672: 479–490, 2019. doi: 10.1016/j.scitotenv.2019.03.480.
- C. Huang, B. X. Hu, X. Li, et M. Ye. Using data assimilation method to calibrate a heterogeneous conductivity field and improve solute transport prediction with an unknown contamination source. *Stochastic Environmental Research and Risk Assessment*, 23(8): 1155–1167, 2008. doi: 10.1007/s00477-008-0289-4.
- H. Huang et Y. Sun. Visualization and assessment of spatio-temporal covariance properties. *Spatial Statistics*, 34: 1–18, 2017. doi: 10.1016/j.spasta.2017.11.004.
- S. Hörning, J. Sreekanth, et A. Bárdossy. Computational efficient inverse groundwater modeling using Random Mixing and Whittaker–Shannon interpolation. *Advances in Water Resources*, 123: 109–119, 2019. doi: 10.1016/j.advwatres.2018.11.012.
- INERIS, ADEME, et GeoSiPol. *Géostatistique appliquée aux sites et sols pollués - Manuel méthodologique et exemples d'applications*. 2005.
- IRSN. *MELODIE Modèle d'Evaluation à Long terme des Déchets Irradiants Enterrés - Notice théorique du code de calcul MELO version 5.0*. 2009.
- IRSN. *Comment gérer les déchets issus de la décontamination en cas de mise en œuvre d'un plan de reconquête de la zone d'exclusion ?* 2016.
- IRSN et ASN. *Gestion des sites potentiellement pollués par des substances radioactives*. 2011.
- K. Z. Jadoon, L. Weihermüller, B. Scharnagl, M. B. Kowalsky, M. Bechtold, S. S. Hubbard, H. Vereecken, et S. Lambot. Estimation of Soil Hydraulic Parameters in the Field by Integrated Hydrogeophysical Inversion of Time-Lapse Ground-Penetrating Radar Data. *Vadose Zone Journal*, 11(4): 701–706, 2012. doi: 10.2136/vzj2011.0177.
- R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME—Journal of Basic Engineering*, 82(Series D): 35–45, 1960. doi: 10.1115/1.3662552.
- V. Kashparov, N. Ahamdach, et L. Dewièrre. *Trench Source Term Characterization on the Chernobyl Pilot Site. Rapport IRSN*. 2004.

- V. Kashparov, V. Yoschenko, S. Levchuk, Y. Khomutinin, M. Lazarev, et N. van Meir. *Tracer experiments at the Beach site in 2000-2001. Technical report on the STCU project p-170*. 2006.
- V. Kashparov, V. Yoschenko, S. Levchuk, D. Bugai, N. van Meir, C. Simonucci, et A. Martin-Garin. Radionuclide migration in the experimental polygon of the Red Forest waste site in the Chernobyl zone – Part 1: Characterization of the waste trench, fuel particle transformation processes in soils, biogenic fluxes and effects on biota. *Applied Geochemistry*, 27(7): 1348–1358, 2012. doi: 10.1016/j.apgeochem.2011.11.004.
- V. Kashparov, B. Salbu, S. Levchuk, V. Protsak, I. Maloshtan, C. Simonucci, C. Courbet, H. L. Nguyen, N. Sanzharova, et V. Zabrotsky. Environmental behaviour of radioactive particles from Chernobyl. *Journal of Environmental Radioactivity*, 208-209: 1–12, 2019. doi: 10.1016/j.jenvrad.2019.106025.
- M. Katzfuss, J. R. Stroud, et C. K. Wikle. Understanding the Ensemble Kalman Filter. *The American Statistician*, 70(4): 350–357, 2016. doi: 10.1080/00031305.2016.1141709.
- M. C. Kennedy et A. O'Hagan. Predicting the Output from a Complex Computer Code When Fast Approximations Are Available. *Biometrika*, 87(1): 1–13, 2000. doi: 10.1093/biomet/87.1.1.
- J. P. C. Kleijnen. Kriging Metamodels and Their Designs. In *Design and Analysis of Simulation Experiments*, International Series in Operations Research & Management Science, pages 179–239. Springer International Publishing, Cham, 2015. doi: 10.1007/978-3-319-18087-8_5.
- C. Knudby et J. Carrera. On the relationship between indicators of geostatistical, flow and transport connectivity. *Advances in Water Resources*, 28(4): 405–421, 2005. doi: 10.1016/j.advwatres.2004.09.001.
- A. Kolovos, G. Christakos, D. T. Hristopulos, et M. L. Serre. Methods for generating non-separable spatio-temporal covariance models with potential environmental applications. *Advances in Water Resources*, 27(8): 815–830, 2004. doi: 10.1016/j.advwatres.2004.04.002.
- L. F. Konikow et J. W. Mercer. Groundwater flow and transport modeling. *Journal of Hydrology*, 100(1): 379–409, 1988. doi: 10.1016/0022-1694(88)90193-X.
- D. Krige. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52(6): 119–139, 1951.
- V. Kuriny, Y. Ivanov, V. Kashparov, N. Loshchilov, V. Protsak, E. Yudin, M. Zhurba, et A. Parshakov. Particle-associated Chernobyl fall-out in the local and intermediate zones. *Annals of Nuclear Energy*, 20(6): 415–420, 1993. doi: 10.1016/0306-4549(93)90067-Y.
- E. Kwicklis, Z. Lu, R. Middleton, T. Miller, S. Bourret, et K. Birdsell. Numerical Evaluation of Unsaturated-Zone Flow and Transport Pathways at Rainier Mesa, Nevada. *Vadose Zone Journal*, 18(1): 1–22, 2019. doi: 10.2136/vzj2019.01.0005.
- P. C. Kyriakidis et A. G. Journel. Geostatistical Space–Time Models: A Review. *Mathematical Geology*, 31(6): 651–684, 1999. doi: 10.1023/A:1007528426688.
- V. Lagneau, O. Regnault, et M. Descostes. Industrial Deployment of Reactive Transport Simulation: An Application to Uranium In situ Recovery. *Reviews in Mineralogy and Geochemistry*, 85(1): 499–528, 2019. doi: 10.2138/rmg.2019.85.16.
- C. Lantuéjoul. *Geostatistical Simulation - Models and Algorithms*. Springer, Berlin, Heidelberg, 2002. ISBN 978-3-642-07582-7.
- G. V. Last, V. J. Rohay, F. J. Schelling, A. L. Bunn, M. A. Delamare, R. L. Dirkes, R. D. Hildebr, J. G. Morse, B. A. Napier, R. G. Riley, L. Soler, et P. D. Thorne. A comprehensive and systematic approach to developing and documenting conceptual models of contaminant release and migration at the Hanford Site. *Stochastic Environmental Research and Risk Assessment*, 18(2): 109–116, 2004. doi: 10.1007/s00477-003-0144-6.

- M. Le Coz, G. Favreau, et S. D. Ousmane. Modeling increased groundwater recharge due to change from rainfed to irrigated cropping in a semiarid region. *Vadose Zone Journal*, 12(2): 1–12, 2013. doi: 10.2136/vzj2012.0148.
- L. Le Gratiet et J. Garnier. Recursive co-kriging model for Design of Computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, 4(5): 365–386, 2014. doi: 10.1615/Int.J.UncertaintyQuantification.2014006914.
- O. Ledoit et M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2): 365–411, 2004. doi: 10.1016/S0047-259X(03)00096-4.
- J. Lee et P. K. Kitanidis. Large-scale hydraulic tomography and joint inversion of head and tracer data using the Principal Component Geostatistical Approach (PCGA). *Water Resources Research*, 50(7): 5410–5427, 2014. doi: 10.1002/2014WR015483.
- E. Lefebvre Polus. *Apports croisés de la modélisation géostatistique et déterministe : exemple des nitrates et de l'oxygène dissous dans un réseau hydrographique*. Thèse de Doctorat en Géostatistiques, École Nationale Supérieure des Mines de Paris, 2010. 204p.
- L. Li, T. Romary, et J. Caers. Universal kriging with training images. *Spatial Statistics*, 14: 240–268, 2015. doi: 10.1016/j.spasta.2015.04.004.
- R. Li, L. Cui, Y. Meng, Y. Zhao, et H. Fu. Satellite-based prediction of daily SO₂ exposure across China using a high-quality random forest-spatiotemporal Kriging (RF-STK) model for health risk assessment. *Atmospheric Environment*, 208: 10–19, 2019. doi: 10.1016/j.atmosenv.2019.03.029.
- X. Li, L. Shi, Y. Zha, Y. Wang, et S. Hu. Data assimilation of soil water flow by considering multiple uncertainty sources and spatial-temporal features: a field-scale real case study. *Stochastic Environmental Research and Risk Assessment*, 32(9): 2477–2493, 2018. doi: 10.1007/s00477-018-1541-1.
- C. Liang, J. Chen, Y. Chien, et C. Chen. Spatial analysis of the risk to human health from exposure to arsenic contaminated groundwater: A kriging approach. *Science of The Total Environment*, 627: 1048–1057, 2018. doi: 10.1016/j.scitotenv.2018.01.294.
- M. Liang et D. Marcotte. A class of non-stationary covariance functions with compact support. *Stochastic Environmental Research and Risk Assessment*, 30(3): 973–987, 2016. doi: 10.1007/s00477-015-1100-y.
- F. Lindgren, H. Rue, et J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society*, 73(4): 423–498, 2011. doi: 10.1111/j.1467-9868.2011.00777.x.
- Z. Lu et D. Zhang. Stochastic analysis of transient flow in heterogeneous, variably saturated porous media: the van Genuchten-Mualem constitutive model. *Vadose Zone Journal*, 1: 137–149, 2002. doi: 10.2113/1.1.137.
- E. Léger, A. Saintenoy, et Y. Coquet. Hydrodynamic parameters of a sandy soil determined by ground-penetrating radar inside a single ring infiltrometer. *Water Resources Research*, 50(7): 5459–5474, 2014. doi: 10.1002/2013WR014226.
- R. Mackay et J. Morakinyo. A stochastic model of surface contaminant releases to support assessment of site contamination at a former industrial site. *Stochastic Environmental Research and Risk Assessment*, 20(3): 213–222, 2006. doi: 10.1007/s00477-005-0016-3.
- K. V. Mardia et R. J. Marshall. Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression. *Biometrika*, 71(1): 135–146, 1984. doi: 10.2307/2336405.
- G. Matheron. Principles of geostatistics. *Economic Geology*, 58(8): 1246–1266, 1963. doi: 10.2113/gsecongeo.58.8.1246.

- G. Matheron. La théorie des variables régionalisées, et ses applications. Cahiers du centre de Morphologie Mathématique de Fontainebleau, Fascicule 5, École Nationale Supérieure des Mines de Paris, 1971.
- G. Matheron. Estimer et choisir. Cahiers du centre de Morphologie Mathématique de Fontainebleau, Fascicule 7, École nationale supérieure des mines de Paris, 1978.
- M. I. McLean, L. Evers, A. W. Bowman, M. Bonte, et W. R. Jones. Statistical modelling of groundwater contamination monitoring data: A comparison of spatial and spatiotemporal methods. *Science of The Total Environment*, 652: 1339–1346, 2019. doi: 10.1016/j.scitotenv.2018.10.231.
- H. A. Michael, H. Li, A. Boucher, T. Sun, J. Caers, et S. M. Gorelick. Combining geologic-process models and geostatistics for conditional simulation of 3-D subsurface heterogeneity. *Water Resources Research*, 46(5): 1–20, 2010. doi: 10.1029/2009WR008414.
- MINES ParisTech/ARMINES. *FLUMY Project version 5.9. User’s guide*. 2019.
- MINES ParisTech/ARMINES. RGeostats: The Geostatistical R Package. Free download from: <http://rgeostats.free.fr/>, 2020.
- J. Morakinyo et R. Mackay. Geostatistical modelling of ground conditions to support the assessment of site contamination. *Stochastic Environmental Research and Risk Assessment*, 20(1-2): 106–118, 2006. doi: 10.1007/s00477-005-0015-4.
- Y. Mualem. A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resources Research*, 12(3): 513–522, 1976. doi: 10.1029/WR012i003p00513.
- C. Neukum et R. Azzam. Quantitative assessment of intrinsic groundwater vulnerability to contamination using numerical simulations. *Science of The Total Environment*, 408(2): 245–254, 2009. doi: 10.1016/j.scitotenv.2009.09.046.
- H. L. Nguyen. *Contribution à la quantification des incertitudes portées par la variabilité spatiale des déchets radioactifs enterrés à Tchernobyl*. Thèse de Doctorat, MINES ParisTech, PSL Research University, 2017. 148p.
- W. E. Nichols, G. V. Last, et C. T. Kincaid. Vadose zone modeling of dispersed waste sites in the framework of an integrated stochastic environmental transport and impacts assessment code for the Hanford Site. *Stochastic Environmental Research and Risk Assessment*, 19(1): 24–32, 2005. doi: 10.1007/s00477-004-0201-9.
- Nuclear Energy Agency. *Preparing for Decommissioning During Operation and After Final Shutdown*. Radioactive Waste Management. OECD, 2018. doi: 10.1787/9789264303126-en.
- OECD et NEA. *Nuclear Site Remediation and Restoration during Decommissioning of Nuclear Installations*. Radioactive Waste Management. OECD Publishing, 2014. doi: 10.1787/9789264222182-en.
- E. K. Paleologos, K. Papapetridis, et C. G. S. C. Kendall. Stochastic contaminant transport monitoring in heterogeneous sand and gravel aquifers of the United Arab Emirates. *Stochastic Environmental Research and Risk Assessment*, 29(5): 1427–1435, 2015. doi: 10.1007/s00477-014-0983-3.
- L. Pannecoucke, M. Le Coz, C. Houzé, A. Saintenoy, C. Cazala, et C. de Fouquet. Impact of spatial variability in hydraulic parameters on plume migration within unsaturated surficial formations. *Journal of Hydrology*, 574: 160–168, 2019. doi: 10.1016/j.jhydrol.2019.04.016.
- L. Pannecoucke, M. Le Coz, X. Freulon, et C. de Fouquet. Combining geostatistics and simulations of flow and transport to characterize contamination within the unsaturated zone. *Science of The Total Environment*, 699: 1–13, 2020a. doi: 10.1016/j.scitotenv.2019.134216.
- L. Pannecoucke, M. Le Coz, X. Freulon, et C. de Fouquet. Migration of Sr90 from a Chernobyl trench to groundwater: combining approaches to estimate a geochemical budget. 2020b. Article en préparation.

- N. G. Patil et S. K. Singh. Pedotransfer Functions for Estimating Soil Hydraulic Properties: A Review. *Pedosphere*, 26(4): 417–430, 2016. doi: 10.1016/S1002-0160(15)60054-6.
- G. S. H. Pau, Y. Zhang, et S. Finsterle. Reduced order models for many-query subsurface flow applications. *Computational Geosciences*, 17(4): 705–721, 2013. doi: 10.1007/s10596-013-9349-z.
- J. L. Pearce, S. L. Rathbun, M. Aguilar-Villalobos, et L. P. Naeher. Characterizing the spatiotemporal variability of PM_{2.5} in Cusco, Peru using kriging with external drift. *Atmospheric Environment*, 43(12): 2060–2069, 2009. doi: 10.1016/j.atmosenv.2008.10.060.
- M. Pereira. *Champs aléatoires généralisés définis sur des variétés riemanniennes: théorie et pratique*. Thèse de Doctorat en Géostatistiques et Probabilités Appliquées, PSL Research University, 2019. 269p, english.
- M. Raissi, P. Perdikaris, et G. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378: 686–707, 2019. doi: 10.1016/j.jcp.2018.10.045.
- B. S. RamaRao, A. M. LaVenue, G. de Marsily, et M. G. Marietta. Pilot Point Methodology for Automated Calibration of an Ensemble of conditionally Simulated Transmissivity Fields: 1. Theory and Computational Experiments. *Water Resources Research*, 31(3): 475–493, 1995. doi: 10.1029/94WR02258.
- P. M. Reed, T. R. Ellsworth, et B. S. Minsker. Spatial Interpolation Methods for Nonstationary Plume Data. *Groundwater*, 42(2): 190–202, 2004. doi: 10.1111/j.1745-6584.2004.tb02667.x.
- P. Renard et D. Allard. Connectivity metrics for subsurface flow and transport. *Advances in Water Resources*, 51: 168–196, 2013. doi: 10.1016/j.advwatres.2011.12.001.
- S. K. Reza, D. C. Nayak, T. Chattopadhyay, S. Mukhopadhyay, S. K. Singh, et R. Srinivasan. Spatial distribution of soil physical properties of alluvial soils: a geostatistical approach. *Archives of Agronomy and Soil Science*, 62(7): 972–981, 2015. doi: 10.1080/03650340.2015.1107678.
- M. Rivest et D. Marcotte. Kriging groundwater solute concentrations using flow coordinates and nonstationary covariance functions. *Journal of Hydrology*, 472-473(Supplement C): 238–253, 2012. doi: 10.1016/j.jhydrol.2012.09.027.
- M. Rivest, D. Marcotte, et P. Pasquier. Hydraulic head field estimation using kriging with an external drift: A way to consider conceptual model information. *Journal of Hydrology*, 361(3): 349–361, 2008. doi: 10.1016/j.jhydrol.2008.08.006.
- M. Rivest, D. Marcotte, et P. Pasquier. Sparse data integration for the interpolation of concentration measurements using kriging in natural coordinates. *Journal of Hydrology*, 416-417: 72–82, 2012. doi: 10.1016/j.jhydrol.2011.11.043.
- T. Romary, L. Malherbe, et C. de Fouquet. Optimal spatial design for air quality measurement surveys. *Environmetrics*, 25(1): 16–28, 2014. doi: 10.1002/env.2253.
- C. Roth. *Contribution de la géostatistique à la résolution du problème inverse en hydrogéologie*. Thèse de Doctorat en Géostatistiques, École Nationale Supérieure des Mines de Paris, 1995. 195p.
- C. Roth, J.-P. Chilès, et C. de Fouquet. Combining geostatistics and flow simulators to identify transmissivity. *Advances in Water Resources*, 21(7): 555–565, 1998. doi: 10.1016/S0309-1708(97)00019-5.
- D. Russo, I. Russo, et A. Laufer. On the spatial variability of parameters of the unsaturated hydraulic conductivity. *Water Resources Research*, 33(5): 947–956, 1997. doi: 10.1029/96WR03947.
- W. Rühaak, K. Bär, et I. Sass. Combining Numerical Modeling with Geostatistical Interpolation for an Improved Reservoir Exploration. *Energy Procedia*, 59: 315 – 322, 2014. doi: 10.1016/j.egypro.2014.10.383.

- N. Saby, D. Arrouays, L. Boulonne, C. Jolivet, et A. Pochot. Geostatistical assessment of Pb in soil around Paris, France. *Science of The Total Environment*, 367(1): 212–221, 2006. doi: 10.1016/j.scitotenv.2005.11.028.
- M. L. O. Salvaña et M. G. Genton. Nonstationary cross-covariance functions for multivariate spatio-temporal random fields. *Spatial Statistics*, 37: 1–22, 2020. doi: 10.1016/j.spasta.2020.100411.
- P. D. Sampson et P. Guttorp. Nonparametric Estimation of Nonstationary Spatial Covariance Structure. *Journal of the American Statistical Association*, 87(417): 108–119, 1992. doi: 10.2307/2290458.
- M. G. Schaap, F. J. Leij, et M. T. van Genuchten. ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology*, 251(3-4): 163–176, 2001. doi: 10.1016/S0022-1694(01)00466-8.
- M. G. Schaap, A. Nemes, et M. T. van Genuchten. Comparison of models for indirect estimation of water retention and available water in surface soils. *Vadose Zone Journal*, 3(4): 1455–1463, 2004. doi: 10.2113/3.4.1455.
- C. Scheidt et J. Caers. Representing Spatial Uncertainty Using Distances and Kernels. *Mathematical Geosciences*, 41(4): 397–419, 2009. doi: 10.1007/s11004-008-9186-0.
- R. L. Schwede et O. A. Cirpka. Use of steady-state concentration measurements in geostatistical inversion. *Advances in Water Resources*, 32(4): 607–619, 2009. doi: 10.1016/j.advwatres.2009.01.010.
- R. L. Schwede et O. A. Cirpka. Interpolation of Steady-State Concentration Data by Inverse Modeling. *Groundwater*, 48(4): 569–579, 2010. doi: 10.1111/j.1745-6584.2009.00668.x.
- S. Schädler, M. Morio, S. Bartke, R. Rohr-Zänker, et M. Finkel. Designing sustainable and economically attractive brownfield revitalization options using an integrated assessment model. *Journal of Environmental Management*, 92(3): 827–837, 2011. doi: 10.1016/j.jenvman.2010.10.026.
- S. Shlomi et A. M. Michalak. A geostatistical framework for incorporating transport information in estimating the distribution of a groundwater contaminant plume. *Water Resources Research*, 43: 1–12, 2007. doi: 10.1029/2006WR005121.
- Site de l'IRSIC. Wosis soil profile database. <https://www.isric.org/explore/wosis>. Consulté en février 2020.
- Site de l'IRSN (a). Les démantèlements d'installations nucléaires en France. https://www.irsn.fr/FR/connaissances/Installations_nucleaires/demantelement/demantelement-France-centrales-installations-nucleaires-EDF-recherche-militaire/Pages/0-demantelement-France-centrales-installations-nucleaires-sommaire.aspx. Consulté en janvier 2020.
- Site de l'IRSN (b). Les sites miniers d'uranium. <https://www.irsn.fr/FR/connaissances/Environnement/expertises-locales/sites-miniers-uranium/Pages/sommaire.aspx>. Consulté en avril 2020.
- Site de l'IRSN (c). Les sites et les sols pollués par des substances radioactives. <https://www.irsn.fr/FR/connaissances/Environnement/expertises-locales/sites-radiocontamines/Pages/sommaire.aspx>. Consulté en avril 2020.
- Site de l'IRSN (d). L'accident nucléaire de Tchernobyl et son déroulement. https://www.irsn.fr/FR/connaissances/Installations_nucleaires/Les-accidents-nucleaires/accident-tchernobyl-1986/2016-Tchernobyl-30ans-apres/Pages/1-Tchernobyl-2016-30ans-apres-accident-deroulement.aspx. Consulté en avril 2020.

- Site de l'IRSN (e). L'avenir de la zone d'exclusion de Tchernobyl et la contamination de l'environnement. https://www.irsnn.fr/FR/connaissances/Installations_nucleaires/Les-accidents-nucleaires/accident-tchernobyl-1986/2016-Tchernobyl-30ans-apres/Pages/4-Tchernobyl-2016-30ans-apres-accident-environnement-zone-exclusion.aspx. Consulté en avril 2020.
- Site de l'IRSN (f). Tchernobyl en 2016. https://www.irsnn.fr/FR/connaissances/Installations_nucleaires/Les-accidents-nucleaires/accident-tchernobyl-1986/2016-Tchernobyl-30ans-apres/Pages/0-Tchernobyl-2016-30ans-apres-Sommaire.aspx. Consulté en avril 2020.
- Site de l'IRSN (g). Le projet Epic : une plateforme expérimentale à Tchernobyl. <https://www.irsnn.fr/FR/Larecherche/outils-scientifiques/installations-moyens-experimentaux/Tchernobyl/Pages/Projet-EPIC-Une-plateforme-experimentale-a-Tchernobyl-1709.aspx>. Consulté en avril 2020.
- Site de l'USDA. Rosetta class average hydraulic parameters. <https://www.ars.usda.gov/pacific-west-area/riverside-ca/agricultural-water-efficiency-and-salinity-docs/model/rosetta-class-average-hydraulic-parameters/>. Consulté en février 2020.
- Site du Ministère de la Transition Écologique et Solidaire (a). Pollution des sols - BASOL. <https://basol.developpement-durable.gouv.fr>. Consulté en janvier 2020.
- Site du Ministère de la Transition Écologique et Solidaire (b). Sites et sols pollués. <https://www.ecologique-solidaire.gouv.fr/sites-et-sols-pollues>. Consulté en juin 2020.
- J. J. J. C. Snepvangers, G. B. M. Heuvelink, et J. A. Huisman. Soil water content interpolation using spatio-temporal kriging with external drift. *Geoderma*, 112(3): 253–271, 2003. doi: 10.1016/S0016-7061(02)00310-5.
- M. L. Stein. Space-Time Covariance Functions. *Journal of the American Statistical Association*, 100(469): 310–321, 2005. doi: 10.1198/016214504000000854.
- M. Stockmann, J. Schikora, D. A. Becker, J. Flügge, U. Noseck, et V. Brendler. Smart Kd-values, their uncertainties and sensitivities - Applying a new approach for realistic distribution coefficients in geochemical modeling of complex systems. *Chemosphere*, 187: 277–285, 2017. doi: 10.1016/j.chemosphere.2017.08.115.
- X. Tan, X. Wang, S. Khoshnevisan, X. Hou, et F. Zha. Seepage analysis of earth dams considering spatial variability of hydraulic parameters. *Engineering Geology*, 228: 260–269, 2017. doi: 10.1016/j.enggeo.2017.08.018.
- A. M. Tartakovsky, D. A. Barajas-Solano, et Q. He. Physics-Informed Machine Learning with Conditional Karhunen-Loève Expansions. *arXiv:1912.02248 [math.AP]*, pages 1–28, 2019.
- M. Taye, B. Simane, Y. G. Selssie, B. Zaitchik, et S. Setegn. Analysis of the Spatial Variability of Soil Texture in a Tropical Highland: The Case of the Jema Watershed, Northwestern Highlands of Ethiopia. *International Journal of Environmental Research and Public Health*, 15(9): 1–10, 2018. doi: 10.3390/ijerph15091903.
- R. Testoni, R. Levizzari, et M. De Salve. Coupling of unsaturated zone and saturated zone in radionuclide transport simulations. *Progress in Nuclear Energy*, 95: 84–95, 2017. doi: 10.1016/j.pnucene.2016.11.012.
- D. Thiéry. Logiciel GARDÉNIA, version v8.2. Guide d'utilisation. Rapport BRGM/RP-62797. 2014.
- R. Tolosana-Delgado, U. Mueller, et K. G. van den Boogaart. Geostatistics for Compositional Data: An Overview. *Mathematical Geosciences*, 51(4): 485–526, 2019. doi: 10.1007/s11004-018-9769-3.

- B. Tóth, M. Weynants, A. Nemes, A. Makó, G. Bilas, et G. Tóth. New generation of hydraulic pedotransfer functions for Europe. *European Journal of Soil Science*, 66(1): 226–238, 2015. doi: 10.1111/ejss.12192.
- UPDS. *Pollution concentrée - Définition, outils de caractérisation et intégration dans la méthodologie nationale de gestion des sites et sols pollués*. 2016.
- B. Usowicz et J. Lipiec. Spatial variability of soil properties and cereal yield in a cultivated field on sandy soil. *Soil and Tillage Research*, 174: 241–250, 2017. doi: 10.1016/j.still.2017.07.015.
- J. van der Lee, L. de Windt, V. Lagneau, et P. Goblet. Module-oriented modeling of reactive transport with HYTEC. *Computers & Geosciences*, 29(3): 265–275, 2003. doi: 10.1016/S0098-3004(03)00004-9.
- M. T. van Genuchten. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, 44(5): 892–898, 1980. doi: 10.2136/sssaj1980.03615995004400050002x.
- P. J. van Leeuwen. Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, 136(653): 1991–1999, 2010. doi: 10.1002/qj.699.
- K. van Looy, J. Bouma, M. Herbst, J. Koestel, B. Minasny, U. Mishra, C. Montzka, A. Nemes, Y. A. Pachepsky, J. Padarian, M. G. Schaap, B. Tóth, A. Verhoef, J. Vanderborght, M. J. van der Ploeg, L. Weihermüller, S. Zacharias, Y. Zhang, et H. Vereecken. Pedotransfer Functions in Earth System Science: Challenges and Perspectives. *Reviews of Geophysics*, 55(4): 1199–1256, 2017. doi: 10.1002/2017RG000581.
- V. van Zoest, F. B. Osei, G. Hoek, et A. Stein. Spatio-temporal regression kriging for modelling urban NO₂ concentrations. *International Journal of Geographical Information Science*, 34(5): 851–865, 2019. doi: 10.1080/13658816.2019.1667501.
- M. Varentsov, I. Esau, et T. Wolf. High-resolution temperature mapping by geostatistical kriging with external drift from large-eddy simulations. *Monthly Weather Review*, 148(3): 1029–1048, 2019. doi: 10.1175/MWR-D-19-0196.1.
- E. A. Varouchakis, P. G. Theodoridou, et G. P. Karatzas. Spatiotemporal geostatistical modeling of groundwater levels under a Bayesian framework using means of physical background. *Journal of Hydrology*, 575: 487–498, 2019. doi: 10.1016/j.jhydrol.2019.05.055.
- H. Wackernagel, C. Lajaunie, N. Blond, C. Roth, et R. Vautard. Geostatistical risk mapping with chemical transport model output and ozone station data. *Ecological Modelling*, 179(2): 177–185, 2004. doi: 10.1016/j.ecolmodel.2004.06.018.
- Y. Wang, O. V. Akeju, et T. Zhao. Interpolation of spatially varying but sparsely measured geo-data: A comparative study. *Engineering Geology*, 231: 200–217, 2017. doi: 10.1016/j.enggeo.2017.10.019.
- R. Webster et M. A. Oliver. *Geostatistics for Environmental Scientists, 2nd Edition*. Statistics in Practice. Wiley, wiley edition, 2007. ISBN 978-0-470-02858-2.
- J. H. M. Wösten, A. Lilly, A. Nemes, et C. Le Bas. Development and use of a database of hydraulic properties of European soils. *Geoderma*, 90(3-4): 169–185, 1999. doi: 10.1016/S0016-7061(98)00132-3.
- X. Yang, G. Tartakovsky, et A. M. Tartakovsky. Physics-Informed Kriging: A Physics-Informed Gaussian Process Regression Method for Data-Model Convergence. *arXiv:1809.03461v2 [stat.ML]*, pages 1–24, 2018.
- X. Yang, D. Barajas-Solano, G. Tartakovsky, et A. M. Tartakovsky. Physics-informed CoKriging: A Gaussian-process-regression-based multifidelity method for data-model convergence. *Journal of Computational Physics*, 395: 410–431, 2019. doi: 10.1016/j.jcp.2019.06.041.
- Y. Yang, J. Wu, et G. Christakos. Prediction of soil heavy metal distribution using Spatiotemporal Kriging with trend model. *Ecological Indicators*, 56: 125–133, 2015. doi: 10.1016/j.ecolind.2015.03.034.

- H. Yong, H. Wenxiang, et G. Bincheng. Combining sedimentary forward modeling with sequential Gauss simulation for fine prediction of tight sandstone reservoir. *Marine and Petroleum Geology*, 112: 104044, 2019. doi: 10.1016/j.marpetgeo.2019.104044.
- D. Zhang, L. Shi, H. Chang, et J. Yang. A comparative study of numerical approaches to risk assessment of contaminant transport. *Stochastic Environmental Research and Risk Assessment*, 24(7): 971–984, 2010. doi: 10.1007/s00477-010-0400-5.
- Y. Zhang et M. G. Schaap. Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). *Journal of Hydrology*, 547: 39–53, 2017. doi: 10.1016/j.jhydrol.2017.01.004.
- Y. Zhang et M. G. Schaap. Estimation of Saturated Hydraulic Conductivity with Pedotransfer Functions: A Review. *Journal of Hydrology*, 575: 1011–1030, 2019. doi: 10.1016/j.jhydrol.2019.05.058.

RÉSUMÉ

La caractérisation de la contamination dans les sols et les nappes autour d'installations industrielles est un enjeu majeur de l'assainissement de ces sites. Ce travail porte sur la combinaison de deux approches utilisées classiquement pour caractériser les contaminations : la géostatistique et les simulations numériques d'écoulement et de transport. L'objectif de cette combinaison est d'améliorer les estimations d'activités en radionucléides, par l'introduction d'informations de nature physico-chimique dans un modèle probabiliste. Deux méthodes sont comparées. Elles sont basées sur la génération d'un grand nombre de simulations de panaches de contaminants, prenant en compte les incertitudes sur les paramètres d'entrée. La première méthode est un krigeage avec dérive externe, où la dérive est la moyenne de la population de simulations de panaches. La seconde méthode est un krigeage avec un variogramme empirique non stationnaire, calculé à partir de la population de simulations, qui reflète la structure spatiale du phénomène étudié.

Ces méthodes sont d'abord comparées sur un cas synthétique, constitué par la simulation de la migration d'un panache de tritium dans la zone non saturée d'une formation sableuse. Deux jeux de données acquis au voisinage de la centrale de Tchernobyl sont ensuite étudiés. Les résultats montrent l'intérêt de la combinaison des simulations numériques et de la géostatistique : les estimations spatiales ou spatio-temporelles sont plus réalistes, en particulier lorsque peu de données sont disponibles.

MOTS CLÉS

Estimation, simulations d'écoulement et de transport, incertitudes, zone non saturée, paramètres hydrodynamiques, Tchernobyl

ABSTRACT

A major issue concerning the remediation of industrial plants is the characterization of contamination within soils and groundwater. Two classical approaches are combined in this work: geostatistics and numerical simulations of flow and transport. It aims at improving the estimates of radionuclide activities by taking into account physically-based information into a mathematical model. Two methods are introduced in this work. They are both based on a large number of simulations of contaminant plumes, taking into account the uncertainties on input parameters. The first method is kriging with an external drift, in which the drift is the average of the simulations. The second method is kriging with a non-stationary empirical variogram computed from the set of simulations, which models the spatial structure of the phenomenon under study.

The performances of the methods are first assessed on a synthetic test case, which deals with the migration of a tritium plume in the unsaturated zone of a sandy aquifer. Then, they are assessed on two datasets collected around Chernobyl nuclear plant. The results underline the benefit of combining geostatistics and numerical simulations, especially when only few observations are available.

KEYWORDS

Estimate, flow and transport simulations, uncertainties, unsaturated zone, hydrodynamic parameters, Chernobyl