Université du Québec Institut national de la recherche scientifique Centre Eau Terre Environnement

CALIBRATION DE PARAMÈTRES HYDROGÉOLOGIQUES PAR ASSIMILATION DE DONNÉES ÉLECTRIQUES DANS LE SUIVI D'UNE CONTAMINATION

Par

Véronique Bouzaglou

Thèse présentée pour l'obtention du grade de *Philosophiæ doctor*, Ph.D. en sciences de la terre

Jury d'évaluation

Président du jury et examinateur interne

Examinateur externe

Examinateur externe

Directeur de recherche

Codirecteur de recherche

Claudio Paniconi INRS - ETE

Maxime Claprood Matrix Solutions

Jean-Sébastien Dubé École de technologie supérieure

Erwan Gloaguen INRS - ETE

Matteo Camporese University of Padova

©Véronique Bouzaglou, 2016

Remerciements

Tout d'abord, un grand merci à mon directeur de thèse, Erwan. Il a souvent eu à me ramener dans la bonne direction quand je commençais à errer, sans compter, que, sans lui, il y aurait probablement encore, un dernier test à faire... Je reconnais maintenant que son intuition était toujours bonne. Comme il le dit, lui même, "c'est l'expérience".

Merci à mon codirecteur, Matteo. Je pouvais toujours compter sur lui pour répondre à mes questions ou discuter de mes résultats rapidement, malgré les six heures de décalage. Je le remercie aussi, avec toute l'équipe du département d'environnement, pour l'accueil à l'université de Padoue.

Merci à mon codirecteur Fateh pour l'accueil au sein de son équipe d'hydrologie statistique. J'ai découvert plusieurs méthodes en les côtoyant.

Je tiens aussi à remercier Abderrezak d'avoir pris le temps de m'initier à ERT2D, mais aussi d'avoir toujours répondu aussi vite à l'appel lorsque j'avais besoin de plus d'explications.

Merci à mes collègues des bureaux 2401 et 2402. Marc, Pat et Emmanuelle ont été d'une précieuse aide quand j'avais des questions d'ordre hydrogéologique. J'ai pu discuter assimilation de données d'un point de vue théorique avec Gab. Lorenzo a été un soutien important pour tous mes problèmes informatiques. Marie-Amélie et Shiva ont été des consoeurs dans la folle aventure du doctorat. Avec Simon, j'ai pu questionner le pourquoi du comment de l'hydrogéophysique, ou... de la politique française. Sans Guillaume, j'aurais été trop souvent la dernière à quitter l'INRS le soir. Et, finalement, les histoires hautes en couleur de Pierre, Martin et Jean-Sébastien ont souvent égayé mes journées.

Je voudrais aussi remercier Yohann, Lauriane, Pierre, Silvia, Dikra, Carlotta, François, Sandra, Étienne et tous les autres qui ont été de passage dans notre groupe du midi. Sans vous, mon séjour à Québec aurait été moins agréable. Merci à Thérèse qui a été là pour m'écouter et me conseiller.

Je voudrais aussi remercier l'équipe de COSMO de m'avoir accueillie dans leur laboratoire quand il a été temps pour moi de retourner è Montréal.

Merci à mes parents, Guy et Hélène, de m'avoir encouragée à retourner aux études et tout le long de mon parcours. Merci aussi à mon frère, Philippe, qui a cru en moi.

Enfin, merci François! Pour tous les vendredis soirs où je rentrais de Québec à l'odeur du souper que tu avais préparé et pour tous les départs déchirants - parce que c'était incroyablement triste, même si ça faisait 100 fois qu'on vivait ce départ et que ce n'était que pour quelques jours.

Résumé

L'eau souterraine est une ressource importante pour le développement humain. Bien qu'elle soit partiellement protégée, elle est tout de même vulnérable à la contamination naturelle ou anthropique. La caractérisation hydrogéologique des aquifères est nécessaire pour leur gestion et leur protection.

L'hydrogéophysique est un domaine transdiciplinaire qui s'intéresse à la caractérisation et au suivi des processus hydrogéologiques par des méthodes géophysiques. Ces méthodes ont l'avantage d'être peu ou pas invasives et de permettre de couvrir des larges étendues à des faibles coûts. Les méthodes électriques, tel que la tomographie par resistivité électrique (ERT), sont particulièrement appropriées pour le suivi de contaminants ou de traceurs salins. Ces méthodes sont sensibles aux changements de résistivité électrique engendrés par la variation de la concentration en sel. Toutefois, l'information relevée grâce à ces méthodes est souvent utilisée de façon qualitative. Dans la présente thèse, il est proposé d'assimiler les mesures obtenues lors d'un suivi par résistivité électrique d'une contamination saline dans le temps pour la caractérisation hydrogéologique.

Dans un premier temps, un algorithme d'assimilation de données par filtres de Kalman d'ensemble pour le suivi d'un contaminant a été développé. Une première version de l'algorithme permet de modéliser la conductivité hydraulique à partir de données de concentrations de la contamination saline à des puits. La seconde version de l'algorithme est plutôt basée sur l'assimilation des données de résistance électrique obtenues par ERT. Une dernière version de l'algorithme permet l'assimilation des deux types de données.

Un cas synthétique de contamination saline dans un aquifère confiné a été mis en place pour tester les différents algorithmes. La comparaison des trois types d'assimilation a permis de déterminer que l'assimilation des données de résistance électrique permet de modéliser l'hétérogénéité de la conductivité hydraulique aussi bien que l'assimilation de données de concentration aux puits. L'assimilation des deux types de données simultanément n'a pas apporté d'amélioration significative par rapport à l'assimilation des données électriques ou de concentrations aux puits exclusivement. Ceci confirme que les deux types de données contiennent une information similaire concernant la conductivité hydraulique.

Des modifications à l'algorithme original des EnKFs ont été mises en place et comparées sur le cas synthétique. Dans le cas à l'étude, les résultats de l'estimation de la conductivité hydraulique étaient similaires que la mise à jour soit effectuée par filtres de Kalman stochastiques ou déterministes, lorsque les ensembles étaient assez grands. Des résultats similaires ont aussi été observés lorsque l'assimilation des données électriques se faisait de façon couplée ou découplée (mise à jour des concentrations ensuite assimilées pour mettre à jour la conductivité hydraulique) et que l'assimilation des données de concentration et de résistance électrique soit effectuée simultanément ou séquentiellement.

Par ailleurs, un ensemble initial des conductivités hydrauliques pour l'assimilation par EnKFs a été créé en tenant compte d'une incertitude sur les paramètres géostatistiques pour simuler cet ensemble. Les résultats de l'assimilation des concentrations aux puits et/ou des résistances électriques avec ce nouvel ensemble initial ont été comparés aux résultats de l'assimilation avec un ensemble initial de conductivités hydrauliques créé par simulations aux paramètres géostatistiques fixes. L'incertitude sur les paramètres géostatistiques utilisés pour créer l'ensemble de conductivités hydrauliques de départ a, dans la plupart des cas, empêché des mises à jour significatives par l'assimilation des données.

Dans un second temps, la méthodologie développée pour l'assimilation des résistances électriques a été testée sur une expérience de suivi d'une intrusion d'eau saline en laboratoire. Cette méthode d'assimilation des résistances électriques par EnKF a été comparée à une méthode d'assimilation par EnKF des résistivités obtenues par inversion. La conductivité hydraulique du bac de sable de l'expérience était considérée homogène et avait été préalablement mesurée en laboratoire. L'assimilation des résistances électriques (ou résistivités inversées) était utilisé pour mettre à jour la dispersivité en plus de la conductivité hydraulique. La conductivité hydraulique obtenue par assimilation des données électriques (que ce soit les résistances ou les résistivités inversées) convergeait vers la valeur obtenue en laboratoire, tandis que la dispersivité était peu sensible à la mise à jour, probablement due à la faible résolution des mesures électriques qui ne permettait pas de bien imager la zone de transition entre l'eau douce et l'eau salée.

L'étude sur le cas synthétique et sur l'expérience d'eau saline en laboratoire a permis de conclure que la conductivité hydraulique peut être modélisée par l'assimilation de données de résistance électrique lors du suivi d'une contamination par ERT.

Mots-clés: Inversion hydrogéophysique; assimilation de données; filtres de Kalman d'ensemble; tomographie par résistivité électrique; contamination des eaux souterraines; intrusion saline.

Abstract

Although partially protected, groundwater is still vulnerable to natural or anthhropic contamination. Characterisation of aquifers is essential to their good management and their protection. Hydrogeophysics is a cross-disciplinary field whose broad objective is to characterize and monitor groundwater processes through geophysical methods. These methods are most often non-invasive or minimally invasive and allow for large-scale coverage. Electrical methods, such as electrical resistivity tomography (ERT), are particularly well-suited for monitoring saline contaminations or tracer-tests, since electrical resistivity is sensitive to the concentration of salt in water. However, the information obtained through ERT is often used qualitatively. This thesis proposes to assimilate ERT measurements of a saline contamination through time to characterize groundwater parameters.

An algorithm for the update of hydraulic conductivity through the monitoring of a salt contamination by ensemble Kalman filter was first developed. A first version of the algorithm assimilated salt concentration measurements in boreholes. A second version of the algorithm assimilated the electrical resistance measurements obtained from a ERT surface survey. A final version of the algorithm assimilated both types of data to update the hydraulic conductivity.

A synthetic case study of a saline contamination in a confined aquifer with heterogeneous hydraulic conductivity was defined to test the different algorithms. The results obtained by the assimilation of electrical resistance data weres as good as when the concentration data in boreholes were assimilated. Moreover, if both types of data were simultaneously assimilated, the results were not much better than if only one type of data was assimilated. This confirms that both data types contain similar information concerning hydraulic conductivity.

A number of modifications to the original EnKF algorithm were tested on the synthetic case. In this case study, the results were similar whether stochastic or deterministic filters, i.e., the ensemble transform Kalman filter, were used, when the ensemble size was big enough. Similar results were also observed when the assimilation of electrical resistance data was done in a coupled or uncoupled fashion (update of the concentrations from resistance data followed by the update of the hydraulic conductivity through the assimilation of concentrations) and whether the assimilation of electrical resistance and concentration was done sequentially or simultaneously.

A final test case was created that took into account the uncertainty of the geostatistical parameters used to create the initial ensemble of hydraulic conductivities. The results of the assimilation of borehole concentrations and/or electrical resistance data were compared to the data assimilation using an initial ensemble of hydraulic conductivities created with fixed geostatistical parameters. In most cases, the uncertainty on the geostatistical parameters used to create the initial hydraulic conductivity ensemble prevented significant updates of the hydraulic conductivity from the data assimilation. The methodology developed to assimilate the electrical resistance data was then further tested on a real saltwater intrusion laboratory experiment. The assimilation of electrical resistance data by EnKF was compared to the assimilation of inverted resistivities also by EnKF. The hydraulic conductivity of the sandbox used for the experiment was considered homogeneous and had been previously measured in the laboratory. The assimilation of electrical resistances (or inverted resistivities) was used to update the dispersivity as well as the hydraulic conductivity. The hydraulic conductivity updated by assimilating electrical data (whether resistances or inverted resistivities) converged towards the measured value. The dispersivity, however, seemed less sensitive to the update through the assimilation of electrical data, probably due to the low resolution of the ERT data, insufficient to capture the transition zone between the salt water and freshwater.

Keywords: Hydrogeophysical inversion; data assimilation; ensemble Kalman filters; electrical resistivity tomography; groundwater transport; saltwater intrusion.

Table des matières

R	ésum	é		\mathbf{v}	
\mathbf{A}	Abstract vii				
Τŧ	able o	des ma	tières	ix	
\mathbf{Li}	ste d	les figu	res	xi	
\mathbf{Li}	ste d	les tabl	eaux x	vii	
\mathbf{Li}	ste d	les abr	éviations	xix	
1	Intr	oducti	on	1	
2	Rev 2.1 2.2 2.3	rue de Inversi 2.1.1 2.1.2 Intégra 2.2.1 Motiva	littérature on hydrogéologique Méthodes géostatistiques Filtres de Kalman d'ensemble ation de données géophysiques Inversion hydrogéophysique ation de la présente étude	5 6 7 9 13 14 15	
3	Ass : 3.1	imilatio Métho 3.1.1 3.1.2 3.1.3 3.1.4 3.1.5 3.1.6 3.1.7 3.1.8 3.1.9 3.1.10 3.1.11	on de données hydrogéophysique des d'assimilation de données séquentielles discrètes Récursion bayésienne Filtres de Kalman Filtres de Kalman d'ensemble Filtres à racine carrée (square-root filter) État augmenté Assimilation des données découplée Assimilation séquentielle de mesures Conservation de la masse Problèmes liés à la finitude du nombre d'ensemble Normes d'évaluation Assimilation de données de résistivité électrique pour la mise à jour des conductivités hydrauliques par filtres de Kalman d'ensemble	 19 20 21 22 24 26 31 33 35 37 38 40 40 	
	3.2	Modèle 3.2.1 3.2.2	e de prédiction	42 43 45	

		$3.2.3 \\ 3.2.4$	Modélisation électrique	. 46 . 49
4	Cas	synth	étique 2D	53
	4.1	Défini	tion du cas synthétique	. 53
	4.2	Ensen	nble initial	. 57
	4.3	Assim	ilation des données de concentrations aux puits	. 60
		4.3.1	Paramètres géostatistiques connus	. 60
		4.3.2	Paramètres géostatistiques variables	. 65
	4.4	Assim	ilation des données de résistance	. 69
		4.4.1	Paramètres géostatistiques connus	. 70
		4.4.2	Paramètres géostatistiques variables	. 74
	4.5	Assim	ilation des données de concentrations aux puits et de résistances électriques .	. 77
		4.5.1	Paramètres géostatistiques connus	. 79
		4.5.2	Paramètres géostatistiques variables	. 84
	4.6	Résun	né des résultats de l'assimilation de données sur un cas synthétique	. 89
5	\mathbf{Ass}	imilati	on des données de résistance électrique lors d'une expérience d'in	fil-
	trat	ion d'	eau salée	93
	5.1	Expér	ience d'intrusion d'eau saline dans un bac à sable	. 93
	5.2	Modè	e hydrogéologique	. 95
	5.3	Modè		. 102
		5.3.1	Inversion électrique	. 108
	5.4	Assim	ilation des données	. 113
		5.4.1	Méthode d'assimilation directe	. 116
		5.4.2	Methode d'assimilation par inversion	. 124
	5.5	Sensit	bilité à l'ensemble initial de l'assimilation des données	. 129
		5.5.1	Methode d'assimilation directe	. 130
	56	5.5.2 Comp	Methode d'assimilation par inversion	. 134
	5.0	Comp	araison des methode d'assimilation directe et inversee	. 130
6	Con	clusio	n	141
Ré	éfére	nces		145
\mathbf{A}	Gai	n de F	Kalman - cas synthétique	157
В	\mathbf{Esti}	matio	n de la concentration avec paramètres géostatistiques variables	159
С	Sch	éma d	étaillé de l'expérience d'eau saline	161
D	Effe	t de la	a taille de la grille - expérience d'intrusion d'eau saline	163
\mathbf{E}	Inve	ersion	électrique sur un modèle synthétique - expérience d'intrusion d'e	au
	saliı	ne		167

Liste des figures

3.1	Schéma de l'assimilation séquentielle de données	20
3.2	Schéma de l'assimilation de données découplée	35
3.3	Schéma de l'assimilation séquentielle de données de potentiel électrique pour l'esti-	
	mation des paramètres hydrogéologiques	41
3.4	Schéma de l'assimilation découplée de données de potentiel électrique pour l'estima-	
	tion des paramètres hydrogéologiques	42
3.5	Schéma du modèle de prédiction	43
3.6	Dispositif de tomographie électrique à quatre électrodes	48
3.7	Assimilation du ratio des résistances	52
4.1	Champ des conductivités hydrauliques $(log_{10}K_h)$	54
4.2	Grille d'éléments finis utilisée pour la modélisation hydrogéologique	55
4.3	Valeurs du cas synthétique (a) Concentration à 182 jours (b) Concentration à 1095	
	jours (c) Pseudo-section de la résistivité électrique à 182 jours (d) Pseudo-section de	
	la résistivité électrique à 1095 jours	57
4.4	Simulations initiales (a) Moyenne de 1000 simulations de $log_{10}(K_h)$ (b) Variance de	
	1000 simulations de $log_{10}(K_h)$ (c) Exemple de simulation de $log_{10}(K_h)$	58
4.5	Simulations initiales (a) Moyenne de 1050 simulations de $log_{10}(K_h)$ (b) Variance de	
	1050 simulations de $log_{10}(K_h)$	59
4.6	Résultats pour la simulation avec paramètres connus (a) Moyenne de déviation abso-	
	lue (AAB) de $log_{10}(K_h)$ estimé en fonction du pas de temps (b) Étendue de l'ensemble	
	(EE) de $log_{10}(K_h)$ estimé en fonction du pas de temps pour EnKF et ETKF avec	
	différents nombres d'ensembles	61
4.7	Moyenne du $log_{10}(K_h)$ (a) Cas de référence (b) Pour un ensemble de 200 réalisations	
	après un pas de temps par EnKF (c) Pour un ensemble de 200 réalisations après 6	
	pas de temps par EnKF (d) Pour un ensemble de 200 réalisations après un pas de	
	temps par ETKF (e) Pour un ensemble de 200 réalisations après 6 pas de temps par	
	ETKF (f) Pour un ensemble de 1000 réalisations après un pas de temps par EnKF	
	(g) Pour un ensemble de 1000 réalisations après 6 pas de temps par EnKF \ldots	63
4.8	Variance du $log_{10}(K_h)$ (a) Pour un ensemble de 200 réalisations après un pas de	
	temps par EnKF (b) Pour un ensemble de 200 réalisations après 6 pas de temps par	
	EnKF (c) Pour un ensemble de 200 réalisations après un pas de temps par ETKF (d)	
	Pour un ensemble de 200 réalisations après 6 pas de temps par ETKF (e) Pour un	
	ensemble de 1000 réalisations après un pas de temps par EnKF (f) Pour un ensemble	
	de 1000 réalisations après 6 pas de temps par EnKF	64

4.9	Résultats pour la simulation avec paramètres géostatistiques variables (a) Moyenne de déviation absolue (AAB) de $log_{10}(K_h)$ estimé en fonction du pas de temps (b)	
	Étendue de l'ensemble (EE) de $loq_{10}(K_h)$ estimé en fonction du pas de temps pour	
	EnKF et ETKF avec différents nombres d'ensembles	66
4.10	Moyenne du $log_{10}(K_h)$ (a) Cas de référence (b) Pour un ensemble de 75 réalisations	
	après un pas de temps par EnKF (c) Pour un ensemble de 75 réalisations après 6 pas	
	de temps par EnKF (d) Pour un ensemble de 75 réalisations après un pas de temps	
	par ETKF (e) Pour un ensemble de 75 réalisations après 6 pas de temps par ETKF $$	
	(f) Pour un ensemble de 2100 réalisations après un pas de temps par EnKF (g) Pour	
	un ensemble de 2100 réalisations après 6 pas de temps par EnKF	67
4.11	Variance du $log_{10}(K_h)$ (a) Pour un ensemble de 75 réalisations après un pas de temps	
	par EnKF (b) Pour un ensemble de 75 réalisations après 6 pas de temps par EnKF	
	(c) Pour un ensemble de 75 réalisations après un pas de temps par ETKF (d) Pour	
	un ensemble de 75 réalisations après 6 pas de temps par ETKF (e) Pour un ensemble	
	de 2100 réalisations après un pas de temps par EnKF (f) Pour un ensemble de 2100	
4 1 0	realisations apres 6 pas de temps par EnKF	68
4.12	Moyenne de deviation absolue (AAB) en fonction du pas de temps pour differentes	
	talles d'ensembles: (a) pour l'assimilation des resistances par EnKF, (b) pour l'as-	
	par ETKE	71
1 13	Étendu de l'ensemble (EE) en fonction du pas de temps pour différentes tailles d'en-	11
1.10	sembles: (a) pour l'assimilation des résistances par EnKF (b) pour l'assimilation des	
	résistances par EnKF découplé. (c) pour l'assimilation des résistances par ETKF.	71
4.14	Movenne du $loq_{10}(K_h)$ (a) Cas de référence (b) pour un ensemble de 1000 réalisations	-
	après un pas de temps par EnKF couplé (c) pour un ensemble de 1000 réalisations	
	après 6 pas de temps par EnKF couplé (d) pour un ensemble de 1000 réalisations	
	après un pas de temps par EnKF découplé (e) pour un ensemble de 1000 réalisations	
	après 6 pas de temps par EnKF découplé (f) pour un ensemble de 1000 réalisations	
	après un pas de temps par ETKF (g) pour un ensemble de 1000 réalisations après 6	
	pas de temps par ETKF	73
4.15	Variance du $log_{10}(K_h)$ (a) pour un ensemble de 1000 réalisations après un pas de	
	temps par EnKF (b) pour un ensemble de 1000 réalisations après 6 pas de temps par	
	EnKF (c) pour un ensemble de 1000 realisations apres un pas de temps par EnKF	
	decouple (d) pour un ensemble de 1000 realisations apres 6 pas de temps par EnKF	
	(f) pour un ensemble de 1000 réalisations après 6 pas de temps par ETKF	74
4 16	(1) pour un ensemble de 1000 realisations après o pas de temps par ETKF \ldots	14
4.10	tailles d'ensembles: (a) nour l'assimilation des résistances par EnKE (b) nour l'as-	
	similation des résistances par EnKF découplé (c) pour l'assimilation des résistances	
	par ETKF.	75
4.17	Étendu de l'ensemble (EE) en fonction du pas de temps pour différentes tailles d'en-	
-	sembles: (a) pour l'assimilation des résistances par EnKF, (b) pour l'assimilation des	
	résistances par EnKF découplé, (c) pour l'assimilation des résistances par ETKF.	75

4.18	Moyenne du $log_{10}(K_h)$ (a) Cas de référence (b) pour un ensemble de 1050 réalisations après un pas de temps par EnKF (c) pour un ensemble de 1050 réalisations après 6	
	pas de temps par EnKF (d) pour un ensemble de 1050 réalisations après un pas de	
	temps par EnKF découplé (e) pour un ensemble de 1050 réalisations après 6 pas de	
	temps par EnKF découplé (f) pour un ensemble de 1050 réalisations après un pas de	
	temps par ETKF (g) pour un ensemble de 1050 réalisations après 6 pas de temps par	
	ETKF	76
4.19	Variance du $log_{10}(K_h)$ (a) pour un ensemble de 1050 réalisations après un pas de	
	temps par EnKF (b) pour un ensemble de 1050 réalisations après 6 pas de temps par	
	EnKF (c) pour un ensemble de 1050 réalisations après un pas de temps par EnKF	
	découplé (d) pour un ensemble de 1050 réalisations après 6 pas de temps par EnKF	
	découplé (e) pour un ensemble de 1050 réalisations après un pas de temps par ETKF	
	(f) pour un ensemble de 1050 réalisations après 6 pas de temps par ETKF	77
4.20	Moyenne de déviation absolue (AAB) en fonction du pas de temps pour différentes	
	tailles d'ensembles: (a) pour l'assimilation par EnKF couplé, (b) pour l'assimilation	
	par EnKF découplé, (c) pour l'assimilation par ETKF.	79
4.21	Étendu de l'ensemble (EE) en fonction du pas de temps pour différentes tailles d'en-	
	sembles: (a) pour l'assimilation par EnKF couplé, (b) pour l'assimilation par EnKF	
	découplé, (c) pour l'assimilation par ETKF.	80
4.22	Moyenne du $log_{10}(K_h)$ pour un ensemble de 1000 réalisations (a) Cas de référence (b)	
	après un pas de temps pour le schéma 1 (c) après 6 pas de temps pour le schéma 1	
	(d) après un pas de temps pour le schéma 2 (e) après 6 pas de temps pour le schéma	
	2 (f) après un pas de temps pour le schéma 3 (g) après un pas de temps pour le	
	schéma 4 (h) après 6 pas de temps pour le schéma 3 (i) après un pas de temps pour	
	le schéma 4 (j) après 6 pas de temps pour le schéma 4 (k) après un pas de temps pour	
	le schéma 5 (l) après 6 pas de temps pour le schéma 5 (m) après un pas de temps	
	pour le schéma 6 (n) après 6 pas de temps pour le schéma 6 $\ldots \ldots \ldots \ldots \ldots$	82
4.23	Variance du $log_{10}(K_h)$ pour un ensemble de 1000 réalisations (a) après un pas de	
	temps pour le schéma 1 (b) après 6 pas de temps pour le schéma 1 (c) après un pas	
	de temps pour le schéma 2 (d) après 6 pas de temps pour le schéma 2 (e) après un	
	pas de temps pour le schéma 3 (f) après un pas de temps pour le schéma 4 (g) après	
	6 pas de temps pour le schéma 3 (h) après un pas de temps pour le schéma 4 (i)	
	après 6 pas de temps pour le schéma 4 (j) après un pas de temps pour le schéma 5	
	(k) après 6 pas de temps pour le schéma 5 (l) après un pas de temps pour le schéma	
	6 (m) après 6 pas de temps pour le schéma 6	83
4.24	Moyenne de déviation absolue (AAB) en fonction du pas de temps pour différentes	
	tailles d'ensembles (a) pour l'assimilation par EnKF couplé (b)pour l'assimilation	0.4
1.05	par EnKF decouple (c) Pour un pour l'assimilation par E'I'KF	84
4.25	Ltendu de l'ensemble (EE) en fonction du pas de temps pour differentes tailles d'en-	
	sembles (a) pour l'assimilation par EnKF (b) pour l'assimilation par EnKF découplé	0.4
	(c) pour l'assimilation par ETKF	84

4.26 4.27	Moyenne du $log_{10}(K_h)$ pour un ensemble de 1050 réalisations (a) Cas de référence (b) après un pas de temps pour le schéma 1 (c) après 6 pas de temps pour le schéma 1 (d) après un pas de temps pour le schéma 2 (e) après 6 pas de temps pour le schéma 2 (f) après un pas de temps pour le schéma 3 (g) après un pas de temps pour le schéma 4 (h) après 6 pas de temps pour le schéma 3 (i) après un pas de temps pour le schéma 4 (j) après 6 pas de temps pour le schéma 4 (k) après un pas de temps pour le schéma 5 (l) après 6 pas de temps pour le schéma 5 (m) après un pas de temps pour le schéma 6 (n) après 6 pas de temps pour le schéma 6	87
5.1	Schéma de l'expérience d'intrusion d'eau saline (n'est pas à l'échelle) (a) vue en coupe (b) vue en plan	94
5.2	Concentrations simulées sur la grille fine avec les paramètres calibrés (a) après 14 heures (b) après 18 heures (c) après 22 heures	102
5.3	Modèle électrique de l'expérience d'intrusion d'eau saline (légende de couleur décrite dans le paragraphe précédent)	104
5.4	Schéma du dispositif électrique utilisé pour l'expérience de Padoue (tiré de: Morrison	104
5.5	Pseudo-section des résistivités apparentes prédites à partir du modèle calibré (a) après 14 heures protocole directe (b) après 14 heures protocole inverse (c) après 18 heures protocole directe (d) après 18 heures protocole inverse (e) après 22 heures	105
5.6	protocole directe (1) après 22 heures protocole inverse	105
5.7	RMS des données de l'inversion en fonction du nombre d'itérations pour l'inversion	107
	(a) après 14 heures (b) après 18 heures (c) après 22 heures	111
$5.8 \\ 5.9$	Résultat de l'inversion (a) après 14 heures (b) après 18 heures (c) après 22 heures . (a) Pseudo-section des résistivités apparentes mesurées avant le début de l'expérience	111
0.0	(a) protocole directe (b) protocole inverse (c) résultat de l'inversion	112
5.10	Résistivités électriques prédites à partir du modèle calibré (a) après 14 heures (b)	119
5 11	Schéma de l'assimilation des données électriques pour l'estimation de la concentration	115
$5.11 \\ 5.12$	Histogramme de l'ensemble intial du logarithme en base 10 (a) de la perméabilité (b)	110
E 19	de la dispersivité longitudinale	116
9.19	Resistance mesuree par rapport a la resistance predite 14 neures après le debut de l'intrusion pour un oppocomont de (a) 81 cm (b) 108 cm (c) 150 cm	110
514	Schéme de l'assimilation des données électriques pour l'actimation de la concentration	110
$5.14 \\ 5.15$	Moyenne des concentrations (kg/kg) prédites par les EnKF (a) à 14 heures (b) à 18	119
	heures (c) à 22 heures. Les dimensions des schémas sont en mètres	120

5.16	Moyenne des concentrations mises à jour par les EnKF (a) à 14 heures (b) à 18 heures (c) à 22 heures	120
5.17	Concentrations simulées avec les paramètres calibrés (a) à 14 heures (b) à 18 heures (c) ≥ 22 l	100
5.18	(c) a 22 heures	120
	fonction du pas de temps	121
5.19	Ensemble du facteur auto-calibrant après l'assimilation pour un espacement de (a) 81 cm (b) 108 cm (c) 150 cm	121
5.20	Histogramme de l'ensemble du logarithme en base 10 de la perméabilité hydraulique (a) ensemble initial (b) ensemble mis à jour à jour après 14 heures (c) ensemble mis à jour à jour après 18 heures (d) ensemble mis à jour après 22 heures	123
5.21	Histogramme de l'ensemble du logarithme en base 10 de la dispersivité longitudinale (a) ensemble initial (b) ensemble mis à jour à jour après 14 heures (c) ensemble mis	
	à jour à jour après 18 heures (d) ensemble mis à jour après 22 heures	124
5.22	Moyenne des concentrations prédites par les EnKF (a) à 14 heures (b) à 18 heures	
ະດາ	(c) à 22 heures \dots	126
0.25	Moyenne des concentrations mises à jour par les Enter (a) à 14 neures (b) à 18 neures (c) à 22 heures	126
5.24	Concentrations simulées avec les paramètres calibrés (a) à 14 heures (b) à 18 heures	
	(c) à 22 heures $\ldots \ldots \ldots$	126
5.25	AAB des concentrations prédites et mises à jour en fonction du pas de temps	127
5.26	Histogramme de l'ensemble du logarithme en base 10 de la perméabilité hydraulique	
	(a) ensemble initial (b) ensemble mis à jour à jour après 14 heures (c) ensemble mis	1.00
F 07	à jour à jour après 18 heures (d) ensemble mis à jour après 22 heures	128
5.27	(a) ensemble initial (b) ensemble mis à jour à jour après 14 heures (c) ensemble mis	
	à jour à jour après 18 heures (d) ensemble mis à jour après 22 heures	129
5.28	Histogramme de l'ensemble initial du logarithme en base 10 (a) de la perméabilité (b)	190
5 20	de la dispersivite longitudinale \dots les EnKE (a) à 14 heures (b) à 18 heures	130
0.29	Moyenne des concentrations predites par les Emtr (a) à 14 neures (b) à 16 neures (c) à 22 heures	130
5.30	Movenne des concentrations mises à jour par les EnKF (a) à 14 heures (b) à 18 heures	100
0.00	(c) à 22 heures	131
5.31	Concentrations simulées avec les paramètres calibrés (a) à 14 heures (b) à 18 heures	
	(c) à 22 heures $\ldots \ldots \ldots$	131
5.32	AAB des concentrations prédites et mises à jour en fonction du pas de temps	131
5.33	Ensemble du facteur auto-calibrant après l'assimilation pour un espacement de (a) 81 cm (b) 108 cm (c) 150 cm	132
5.34	Histogramme de l'ensemble du logarithme en base 10 de la perméabilité hydraulique	
	(a) ensemble initial (b) ensemble mis à jour à jour après 14 heures (c) ensemble mis	
	à jour à jour après 18 heures (d) ensemble mis à jour après 22 heures $\ \ldots \ \ldots \ \ldots$	133
5.35	Histogramme de l'ensemble du logarithme en base 10 de la dispersivité longitudinale	
	(a) ensemble initial (b) ensemble mis à jour à jour après 14 heures (c) ensemble mis	
F 0.0	à jour à jour après 18 heures (d) ensemble mis à jour après 22 heures \ldots	134
5.36	Moyenne des concentrations predites par les EnKF (a) a 14 heures (b) à 18 heures (a) à 22 heures	19F
5.37	(0) a 22 neuros	199
0.01	(c) à 22 heures \ldots	135

5.38	Concentrations simulées avec les paramètres calibrés (a) à 14 heures (b) à 18 heures (a) à 22 heures	125
5.39	(c) a 22 heures	100
5.40	à jour à jour après 18 heures (d) ensemble mis à jour après 22 heures	136
$5.41 \\ 5.42$	à jour à jour après 18 heures (d) ensemble mis à jour après 22 heures	137 138
5.43	de l'ensemble	140 140
A.1	Gain de Kalman au premier pas de temps (a) pour un ensemble de 20 réalisations (b) Pour un ensemble de 100 réalisations (c) Pour un ensemble de 200 réalisations (d) Pour un ensemble de 100 réalisations	157
A.2	Gain de Kalman au premier pas de temps (a) pour un ensemble de 75 réalisations avec paramètres variables (b) Pour un ensemble de 225 réalisations avec paramètres variables (c) Pour un ensemble de 2100 réalisations avec paramètres variables réali- sations	158
B.1 B.2	Concentration synthétique après 182 jours	160 160
C.1	Schéma détaillé de l'expérience d'intrusion d'eau saline (n'est pas à l'échelle)	162
D.1	Concentrations simulées (a) Grille fine au premier pas de temps (b) Grille fine au dernier pas de temps (c) Grille grossière au premier pas de temps (d) Grille grossière au dernier pas de temps (e) Différence au premier pas de temps (f) Différence (en	
D.2	pourcentage) au dernier pas de temps	163
D.3	premier pas de temps (f) Différence (en pourcentage) au dernier pas de temps Résistivités apparentes simulées pour l'acquisition par mode arrièere (a) Grille fine au premier pas de temps (b) Grille fine au dernier pas de temps (c) Grille grossière au premier pas de temps (d) Grille grossière au dernier pas de temps (e) Différence au premier pas de temps (f) Différence au dernier pas de temps	164
E.1	Pseudo-section des résistivités apparentes simulées à partir du modèle calibré (a)	100
E.2	après 14 heures protocole directe (b) après 14 heures protocole inverse	$\begin{array}{c} 167 \\ 167 \end{array}$

Liste des tableaux

4.1	Paramètres géostatistiques de $log_{10}(K_h)$	54
4.2	Paramètres du modèle de transport	55
4.3	Simulations initiales avec paramètres géostatistiques variables	59
4.4	Variables d'état et d'observations des différents schémas	79
4.5	Normes d'évaluation sur l'ensemble des conductivités hydrauliques pour l'assimilation	
	avec paramètres géostatistiques connus	90
4.6	Normes d'évaluation sur l'ensemble des conductivités hydrauliques pour l'assimilation	
	avec paramètres géostatistiques variables	91
5.1	Pentes des régressions entre les résistances prédites et mesurées pour différents espa-	
	cements entre les électrodes	118
5.2	Différence moyenne entre les résistivités obtenues par inversion des mesures de résis-	
	tance et les résistivités prédites grâce au modèle calibré en base logarithmique $\ . \ .$	125
B.1	AAB de l'ensemble de concentrations avant (simulations intiales) et après le premier	
	pas de temps dans différents cas d'assimilation	159
B.2	Étendue de l'ensemble de concentrations avant (simulations initiales) et après le pre-	
	mier pas de temps dans différents cas d'assimilation	159
B.3	AAB avant et après assimilation à partir des résistivités (cas couplé) avec un ensemble	
	de 75 membres	160

Liste des sigles et acronymes

Acronymes

AAB	Moyenne de déviation absolue
\mathbf{EE}	Étendue de l'ensemble
EKF	Filtre de Kalman étendu
EnKF	Filtre de Kalman d'ensemble
ETKF	Filtre de Kalman d'ensemble transformés
\mathbf{KF}	Filtre de Kalman
TDS	Total des sels dissous

Alphabet Latin

\mathbf{C}	Matrice de l'ensemble des concentrations
\mathbf{d}	Vecteur des observations
D	Tenseur du coefficient de dispersion
\overline{D}_i	Coefficient de dispersion dans la direction i
D_m	Diffusivité moléculaire
f	Fonction de densité de probabilité
g	Fonction de transfert
\mathbf{G}	Matrice de la fonction de transfert
h	Fonction liant l'état aux observations
\mathbf{H}	Matrice de la fonction liant l'état aux observations
Κ	Gain de Kalman
K_h	Conductivité hydraulique
K_i	Conductivité hydraulique dans la direction i
n	porosité
n_e	Porosité effective
N_e	Taille de l'ensemble
\mathbf{P}	Matrice de covariance de l'état
\mathbf{Q}	Matrice de covariance de l'erreur du modèle
\mathbf{R}	Résistance
$\mathbf{R}\mathbf{R}$	Ratio des résistances
S_s	Coefficient d'emmagasinement
S_{op}	Coefficient d'emmagasinement de la pression
S_w	Degré de saturation
u	Vecteur de l'erreur du modèle
v_i	Vitesse de Darcy dans la direction i
\mathbf{W}	Vecteur de l'erreur d'observation
V	Potentiel electrique
\mathbf{W}	Matrice de covariance de l'erreur de mesure
\mathbf{x}	Vecteur d'état

Alphabet grec

Dispersivité longitudinale
Dispersivité transversale
Paramètres à calibrer
Perméabilité du sol
Densité
Résistivité apparente
Conductivité électrique totale du sol
Conductvité électrique de l'eau
Viscosité du fluide

Chapitre 1

Introduction

Un peu moins du tiers de l'eau potable se trouve sous la forme d'eau souterraine dans le monde (ONU, 2014). L'eau souterraine, étant protégée partiellement par le couvert de sol des polluants présents en surface, nécessite moins de traitement que l'eau de surface, ce qui en fait une ressource privilégiée d'eau potable pour les décennies à venir. Au Québec, dans plusieurs régions, l'approvisionnement en eau de surface est insuffisant ou requiert des infrastructures trop lourdes par rapport à l'approvisionnement en eau souterraine. Ainsi, l'eau souterraine constitue 21% de l'approvisionnement en eau potable du Québec (MDDEP, 2002). Cette eau souterraine est, toutefois, elle aussi sujette à la menace de la contamination, de la surconsommation pour l'approvisionnement agricole ou urbain et aussi à la montée des eaux due au réchauffement climatique.

Un cas particulier de la contamination de la nappe phréatique touche les aquifères côtiers, qui peuvent être contaminés par des intrusions d'eau salée. Les changements climatiques et l'augmentation du niveau de la mer auront un effet important sur les intrusions d'eau salée dans ces aquifères (Rasmussen *et al.*, 2013). Au Canada, ce problème se présente plus particulièrement dans les provinces de l'Atlantique (Prince Edward Island Department of Environment & Justice, 2011) et aux Îles-de-la-Madeleine (BAPE, 2013) ainsi qu'en Colombie-Britannique (Klassen & Allen, 2016).

La gestion, la protection et la restauration environnementales durables des aquifères qu'ils soient côtiers ou intérieurs, reposent, entre autres, sur la prédiction des écoulements du transport de masse de l'eau souterraine (Marsily *et al.*, 2005), lesquels sont principalement fonction de la distribution spatiale de la conductivité hydraulique dans l'aquifère (Sudicky & Huyakrn, 1991), (Wood, 2000) ou (Zheng & Gorelick, 2003). De gros efforts sont donc mis dans la caractérisation hydrogéologique des aquifères, en particulier la caractérisation de la conductivité hydraulique (Gómez-Hernández *et al.*, 2016). Une fois la caractérisation effectuée, des modèles numériques permettent de prédire l'écoulement dans le temps. Toutefois, les méthodes conventionnelles de caractérisation des aquifères comportent certaines limites qui se réfléteront ensuite dans la prédiction de l'écoulement.

Les méthodes conventionnelles de caractérisation des aquifères sont basées sur :

1. les analyses en laboratoire d'échantillons dits non-remaniés récoltés le long de forages sur la zone d'intérêt. La distribution de ces forages est souvent sub-optimal pour caractériser l'hétérogénéité de la conductivité hydraulique dans le sol.

2. les mesures hydrogéologiques consistant en une série de tests de perméabilité in-situ, qui fournissent un estimé de K aux abords des forages (par exemple, les essais entre obturateurs ou « slug tests ») ou des tests de pompage et de traceurs qui fournissent K à l'échelle de l'écartement entre les forages.

Bien qu'essentielles, ces données sont, dans le premier cas, relativement précises mais locales et dans l'autre, à plus grande échelle mais à une résolution trop grossière. De plus, ces méthodes ne donnent pas d'information sur la continuité spatiale des propriétés hydrogéologiques.

Les méthodes géophysiques, et en particulier les méthodes de résistivité électrique, offrent une avenue intéressante pour palier aux limites des méthodes convetionnelles de caractérisation des aquifères puisqu'elles fournissent, à faible coût, une densité spatiale élevée d'informations indirectes sur les diverses propriétés physiques des sols. Des mesures directes de la structure spatiale des propriétés tels que la conductivité hydraulique à partir de mesures électriques n'est généralement pas possible (Lesmes & Friedman, 2005). Toutefois, les méthodes de tomographie par résistivité électrique (ERT) permettent un suivi de la contamination saline (par exemple, Brewster *et al.* (1995); Buselli *et al.* (2); Singha & Gorelick (2005); Robert *et al.* (2012)). Le suivi de la contamination peut ensuite être utilisé dans la caractérisation de l'aquifère.

L'objectif principal de cette thèse est donc de développer une méthodologie de l'estimation des paramètres hydrogéologiques grâce à l'assimilation de données ERT lors de la progression d'un panache salin dans le temps. Les objectifs secondaires seront donc:

- 1. Évaluer l'apport de données électriques pour estimer les paramètres hydrogéologiques par assimilation de données dans un cas de contamination saline.
- 2. Déterminer la meilleure méthode pour assimiler des données hydrogéophysiques de contamination.
- Valider l'algorithme sur un cas de contamination saline en laboratoire et le comparer avec une méthode déjà existante basée sur l'inversion des données ERT.

Le document est divisé en quatre parties. Premièrement, une revue de littérature présente un historique des méthodes de calibration des paramètres hydrogéologiques jusqu'à ce jour. Cet historique permet de mettre en contexte la motivation de la méthodologie proposée pour la calibration des paramètres hydrogéologiques, soit l'assimilation des données de résistance électrique par EnKF. Cette méthodologie est ensuite présentée avec une mise en contexte théorique des EnKF. Un cas synthétique a ensuite été étudié pour évaluer l'efficacité de la méthode proposée. Puis, la méthode a été testée sur un essai en laboratoire représentant l'intrusion d'eau marine dans un aquifère côtier.

Chapitre 2

Revue de littérature

Différents modèles numériques ont été développés afin de résoudre les équations différentielles partielles qui régissent l'écoulement et le transport (de masse ou de chaleur) dans l'eau souterraine (par exemple, FEFLOW (Diersch, 2013), MODFLOW (Harbaugh, 2005) ou SUTRA (Voss & Provost, 2010)). Toutefois, ces modèles nécessitent la définition de plusieurs paramètres. Par exemple, la conductivité hydraulique, les sources et les puits ainsi que la capacité de stockage doivent être définis pour résoudre l'équation d'écoulement. La résolution de l'équation de transport advectif-dispersif nécessite en plus la définition du coefficient de dispersivité. La définition des conditions initiales et de conditions frontières est aussi nécessaire pour la résolution numérique de chaque équation. La résolution des équations lorsque les paramètres ont été modélisés se nomme problème directe.

Les modèles conceptuels sont des descriptions du système hydrogéologique qui incluent le tracé d'unités aux propriétés hydrogéologiques spécifiques, les hydrofaciès (par exemple, Reimus *et al.* (2003) ou Tremblay *et al.* (2014)). Une fois, ces modèles créés, ils doivent tout de même être calibrés pour définir les propriétés hydrogéologiques de chaque hydrofaciès. La calibration de ces modèles peut se faire par des méthodes d'essais-erreurs comme chez Anderson & Woessner (1992). Lorsque c'est le cas les modèles doivent être recalibrés manuellement à chaque fois que des nouvelles données deviennent disponibles. De plus, les paramètres ainsi calibrés sont souvent homogènes à l'intérieur d'un hydrofaciès. Ceci n'est pas nécessairement réllement le cas. Les hétérogénéités ainsi présentes peuvent causer des problèmes dans les simulations hydrogéologiques créées avec de tels modèles (Blouin *et al.*, 2013). Les méthodes d'inversion hydrogéologique ont été développées pour faciliter la calibration (Poeter & Hill, 1997). Dans les sections suivantes, différentes méthodes d'inversion hydrogéologique sont présentées (section 2.1). Un intérêt particulier est porté à l'extension de ces méthodes pour intégrer des observations géophysiques, l'inversion hydrogéophysique (section 2.2.1). En effet, l'assimilation de données électriques pour la mise à jour des paramètres hydrogéologiques par filtres de Kalman d'ensemble est une méthode d'inversion hydrogéophysique. De plus, l'inversion hydrogéophysique est mise en contexte dans les méthodes d'intégration de données géophysiques en hydrogéologie (section 2.2).

2.1 Inversion hydrogéologique

L'inversion hydrogéologique consiste à utiliser les réponses connues du modèle (problème directe) afin d'en déterminer les paramètres. Le problème inverse peut être résolu pour des paramètres de l'équation d'écoulement à partir de données de charges hydrauliques. Le problème inverse couplé concerne l'estimation des paramètres d'un modèle de transport couplé à un modèle d'écoulement à partir de données de concentration. Dans ce dernier cas, des paramètres de la fonction de transport (tel que la dispersivité) peuvent aussi être calibrés.

Le problème inverse en hydrogéologie est souvent considéré mal-posé. Pour qu'un problème soit bien-posé il doit respecter trois conditions (Tikhonov & Arsenin, 1977):

- l'existence d'une solution
- l'unicité de cette solution
- la stabilité de la solution sous de faibles perturbations.

Étant donné l'hétérogénéité spatiale de certains paramètres, en particulier la conductivité hydraulique et l'erreur de mesure sur les données, la solution au problème inverse n'est souvent ni unique, ni stable.

La méthode des moindres carrés régularisés permet de résoudre les problèmes mal-posés. Le terme de régularisation permet de favoriser une solution en particulier. Une contrainte de lissage peut être utilisée comme régularisation. Dans ce cas, la méthode est considérée déterministe. La méthode des moindres carrés régularisés peut aussi être dérivée de manière probabiliste en introduisant de l'information *a priori* bayésienne. Dans ce cas, le résultat de l'inversion est équivalent à l'estimateur du maximum a posteriori (McLaughlin & Townley, 1996).

La méthode des moindres carrés régularisés ou des méthodes équivalentes ont souvent été utilisées pour résoudre le problème inverse en hydrogéologie (par exemple, (Cooley, 1977) ou(Carrera & Neuman, 1986)). Sun & Yeh (1990) ont transposé cette méthode pour la solution du problème inverse couplé écoulement et transport.

2.1.1 Méthodes géostatistiques

La modélisation des paramètres hydrogéologiques en tant que fonctions aléatoires permet de tenir compte de la variabilité spatiale des paramètres. La validité de l'hypothèse sous-jacente à une solution déterministe au problème inverse, c'est-à-dire qu'une solution unique peut être satisafaisante afin de représenter les paramètres hydrogéologiques a été mise en doute par les travaux de Freeze (1975). De plus, ce nouveau paradigme devrait permettre de transformer le problème inverse mal-posé en problème bien-posé à travers une paramétrisation géostatistique appropriée (Zimmerman *et al.*, 1998).

Les variables de réponse des fonctions d'écoulement et de transport (charges hydrauliques, concentrations) peuvent être utilisées comme variables secondaires dans les méthodes d'interpolation et de simulation géostatistiques. Lorsque c'est le cas, une relation linéaire entre paramètres et variables de réponse est approximée grâce aux matrices de covariance. Ces covariances peuvent être calculées numériquement comme chez Kitanidis & Vomvoris (1983) qui ont utilisé les charges hydrauliques dans l'estimation des log-perméabilités par cokrigeage, leur permettant d'obtenir des champs de log-perméabilité lisses qui décrivaient la variabilité à petite échelle. L'ajout de mesures de charges hydrauliques dans l'estimation de log-transmissivités par cokrigeage améliorerait les valeurs estimées et la variance d'estimation (variance de cokrigeage) (Ahmed & De Marsily, 1993). Gutjahr & Wilson (1989) ont eu les mêmes résultats quant à la variance d'estimation mais ont noté que le nombre de points de mesures utilisés pour effectuer l'estimation avait une influence sur l'amélioration apportée par l'ajout des charges hydrauliques: si l'estimation était effectuée avec peu de mesures, l'ajout de charges hydrauliques entraînait seulement une faible diminution de la variance de krigeage. Par ailleurs, Kitanidis (1986) note que les charges hydrauliques peuvent apporter de l'information à grande échelle qui n'est pas nécessairement présente dans les mesures ponctuelles de transmissivités. Toutes ces méthodes ne tiennent pas compte explicitement du modèle d'écoulement.

Lorsque des méthodes tel que le cokrigeage ou les cosimulations sont utilisés pour l'estimation des paramètres, la fonction d'écoulement peut être utilisée pour lier les moments statistiques des charges hydrauliques à celles des paramètres à estimer, comme chez Dagan & Rubin (1988). C'est aussi ce que font Hoeksema & Kitanidis (1984) tout en utilisant une méthode similaire d'estimation à celle de Kitanidis & Vomvoris (1983). Lorsque la fonction d'écoulement est utilisée pour calculer des moments statistiques, cette fonction est habituellement linéarisée (Hoeksema & Kitanidis, 1984), (Dagan & Rubin, 1988). Zimmerman *et al.* (1998) ont noté que cette linéarisation avait peu d'influence sur la solution au problème inverse lorsque le champ de log-transmissivités était stationnaire, mais qu'elle était peu appropriée lors de la présence d'anomalies ou de non-stationnarités. Par ailleurs, Dietrich & Newsam (1989) notent que cette linéarisation peut causer un mauvais conditionnement des matrices de covariance, en particulier lorsque le nombre de données mesurées est élevé.

Les méthodes de maximum a posteriori (par exemple, Woodbury & Ulrych (2000), pour le problème inverse d'écoulement et Valstar (2001) ou Carrera & Neuman (1986) pour le problème inverse couplé) et de vraisemblance maximale tel que défini par Medina & Carre (2003) permettent aussi de tenir compte de la variabilité spatiale tout en tenant compte du modèle d'écoulement (ou de transport). La variabilité spatiale est intégrée à la distribution a priori dans la première méthode et dans la fonction de vraisemblance dans la seconde méthode.

Des méthodes où un champ de paramètres hydrogéologiques est d'abord créé par méthodes géostatistiques avant d'être conditionné aux valeurs de charges hydrauliques ont aussi été développées. C'est d'ailleurs en générant plusieurs simulations conditionnelles de champs de transmissivités que Delhomme (1979) a été le premier à utiliser les géostatistiques dans un contexte d'inversion hydrogéologique. Un simulateur d'écoulement est ensuite utilisé conjointement aux champs de transmissivités pour obtenir des charges hydrauliques. Seuls les champs de charges simulés pour lesquels le résultat de l'écoulement respectait assez bien les observations piézométriques étaient conservés. Gómez-Hernández *et al.* (1997) proposent de simuler un champ de transmissivités par simulations stochastiques multi-gaussiennes, puis de perturber les transmissivités simulées dans une étape de minimisation de l'erreur entre les charges hydrauliques observées et les charges hydrauliques obtenues avec les champ de transmissivités observées. Cette méthode a été nommée *Sequential self-calibration*. Capilla & Llopis-Albert (2009) et Llopis-Albert & Capilla (2010) utilisent une méthode similaire, basée sur la méthode de déformation graduelle. La méthode proposée par Capilla & Llopis-Albert (2009) a l'avantage de produire des simulations de transmissivités qui ne sont pas nécessairement gaussiennes et donc de pouvoir relaxer l'hypothèse de multi-gaussianité des données.

2.1.2 Filtres de Kalman d'ensemble

Lorsque les seules données disponibles pour estimer les paramètres, telle que l'hétérogénéité de la conductivité hydraulique du sol, sont des mesures de charges hydrauliques en régime d'écoulement permanent, les méthodes présentées précédemment permettent d'incorporer toute l'information disponible. Toutefois, lorsque des données évoluant dans le temps sont disponibles, des charges hydrauliques en régime transitoire ou des données de concentration d'une contamination par exemple, il peut être intéressant d'utiliser des méthodes pouvant assimiler les données dans le temps.

Le filtre de Kalman est une méthode d'assimilation de données linéaire et récursive qui permet de propager l'information dans le temps. La récursion des filtres de Kalman se fait en deux étapes:

- Étape de prédiction: L'état est propagé dans le temps grâce à une fonction de transfert liant l'état aux différents pas de temps.
- Étape de mise à jour: L'état est mis à jour grâce aux données observées. La correction est effectuée sur l'état prédit. Cette correction est proportionnelle à la différence entre les valeurs de l'état prédites et les observations.

La fonction de transfert doit être linéaire pour l'application des KF. Mis à part certains cas particuliers, tel que l'écoulement dans un aquifère confiné, les équations d'écoulement et de transport ne sont généralement pas linéaires. Par contre, elles peuvent être linéarisées, tout comme elles l'ont été pour faire le lien entre les matrices de covariance de charges hydrauliques et de transmissivités (voir section 2.1.1). Le filtre de Kalman appliqué à des fonctions linéarisées par séries de Taylor est connu sous le nom de filtre de Kalman étendu ou *Extended Kalman Filter (EKF)*. Townley (1983) a estimé la transmissivité à partir de données charges hydrauliques par EKF. Ferraresi *et al.* (1996) et Apriliani (2011) ont aussi utilisé une méthode semblable en appliquant des filtres de Kalman sur une version linéarisée par discrétisation des équations d'écoulement dans le premier cas, et des équations d'écoulement et de transport dans le second cas. En discrétisant l'équation d'écoulement unidimensionelle, Drécourt *et al.* (2006) ont pu implémenter un filtre de Kalman coloré qui calibre automatiquement l'erreur du modèle. Les filtres de Kalman d'ensemble (EnKF) ont été développés par (Evensen, 1994) pour pallier les problèmes liés à l'approximation de premier ordre des EKF. L'approximation linéaire des EKFengendre la négligence des moments d'ordre supérieur à deux dans l'évolution de la covariance (fermeture des moments). Cette approximation peut donc engendrer une divergence de la matrice de covariance de l'erreur, soit que la covariance estimée par le filtre rerpésente mal l'erreur réelle (Sorenson, 1974). Les filtres de Kalman d'ensemble évitent les problèmes associés à la fermeture des moments puisqu'ils font évoluer la distribution de probabilité dans le temps grâce à un ensemble de réalisations.

Les filtres de Kalman d'ensemble se prêtent bien à la résolution du problème inverse en hydrogéologie, puisque le problème directe est non-linéaire. Chen & Zhang (2006) ont réussi à obtenir un champ de conductivités hydrauliques semblable au champ de référence à partir de l'assimilation par EnKF de données de charges hydrauliques dans un modèle d'écoulement en régime transitoire. L'assimilation rectifiait même certaines connaissances des formations incorrectes *a priori*. Cette méthode d'estimation d'un champ de transmissivités a été comparée par Hendricks Franssen & Kinzelbach (2009) à la méthode des *Sequential-self calibration* présentée précédemment. Les auteurs ont conclu que les deux méthodes donnaient des résultats similaires mais que les EnKF nécessitaient moins de calcul. Afin de minimiser encore plus le temps de calcul Xu *et al.* (2013a) ont proposé une méthode de parallélisation des EnKF pour l'estimation du champ de conductivités. Li *et al.* (2012c) ont aussi réduit le temps de calcul de l'assimilation des charges hydrauliques par EnKF en effectuant une mise à l'échelle supérieure des conductivités hydrauliques.

Les lisseurs d'ensembles (Ensemble Smoother ou EnS) sont une variante des EnKF développée par Van Leeuwen & Evensen (1996). Tandis que les EnKF assimilent les données séquentiellement dans le temps, les EnS assimilent toutes les données en une seule étape. La performance d'une méthode par rapport à l'autre varie selon le problème à l'étude: lorsque les deux méthodes étaient appliquées aux équations de Lorenz, les EnKF donnaient de meilleurs résultats (Evensen & Van Leeuwen, 2000), tandis que les EnS donnaient de meilleurs résults lorsqu'appliqués à un modèle de circulation océanique (Evensen & Van Leeuwen, 1996). Crestani *et al.* (2013) ont montré que les EnKF reproduisaient bien le champ de conductivité hydraulique à partir de données d'un test de traceur et que les EnS donnaient de moins bons résultats. Plusieurs facteurs, entre autres un ensemble trop petit ou une erreur du modèle trop faible ou absente, peuvent rendre la covariance échantillonnale sous-optimale (Anderson, 2012). Panzeri *et al.* (2014) ont d'ailleurs noté une divergence des EnKF pour des ensembles trop petits dans le cadre de l'assimilation de données d'écoulement. Différentes méthodes ont été développées pour palier à ces problèmes, dont la localisation de la covariance ou l'inflation. La localisation de la covariance consiste à effectuer le produit matriciel d'Hadamard de la covariance de prédiction obtenue à partir de l'ensemble par une matrice de covariance obtenue à partir d'une fonction théorique de la covariance (Hamill *et al.*, 2001). La localisation a ensuite été appliquée aux filtres à racine-carrée, une variante des EnKF par Petrie (2008). Tong *et al.* (2012) ont appliqué la localisation et un facteur d'inflation sur la covariance de l'erreur dans l'estimation d'un champ de conductivité en assimilant des charges hydrauliques. L'estimation des conductivités par leur méthode avec localisation et inflation était meilleure que la méthode classique des EnKFs. Nan & Wu (2011) arrivent à des conclusions similaires, mais notent que la définition de portées trop faibles ou trop élevées pour la fonction théorique de la covariance pourrait empêcher de modifier la covariance empirique de façon appropriée.

L'optimalité des EnKFs est conditionnelle à l'hypothèse de multi-gaussianité. Cette hypothèse a plusieurs limites. Premièrement, elle ne peut-être vérifiée pour les fonctions de densité de probabilité de plus de deux dimensions (ou deux points). Ensuite, sous le modèle multi-gaussien, les valeurs très fortes ou très faibles sont non-corrélées spatialement (Goovaerts, 1997). Cette dernière caractéristique des modèles multi-gaussiens résulterait en des temps de parcours de l'eau beaucoup plus faibles que la réalité lorsque la log-conductivité est modélisée d'après une distribution multi-gaussienne (Gómez-Hernández *et al.*, 1997).

Différentes façons de traiter de la non-gaussianité ont été développées afin d'adapter les données hydrogéologiques, dont la transformation normale (*anamorphose gaussienne*). La déformation normale est définie comme une une transformation bijective non-linéaire des variables d'un espace non-gaussien à un espace gaussien. Certaines transformations paramétriques peuvent être appliquées pour passer à un espace gaussien. La transformation de conductivités hydrauliques qui suivent une loi log-normale en log-conductivités en serait un exemple. Lorsqu'il n'y a pas une bonne raison de croire qu'une transformation paramétrique permettrait de passer à un espace gaussien, une transformation non-paramétrique peut être effectuée. Cette transformation est équivalente à la transformation par partitions normales (*Normal-score transform*). Cette transformation consiste à associer les valeurs d'une loi normale correspondants aux quantiles de la distribution empirique. Ce type de transformation pour assurer l'optimalité des filtres de Kalman d'Ensemble a d'abord été suggéré par Bertino *et al.* (2003). Cette méthode a ensuite été adoptée par Zhou *et al.* (2011) et Xu *et al.* (2013b) pour transformer les paramètres et les variables d'état à chaque localisation et chaque pas de temps dans l'assimilation de données de charges hydrauliques pour l'estimation de conductivités bimodales (tirées d'un aquifère à deux faciès) par filtres de Kalman d'ensemble. Zhou *et al.* (2012) ont évalué l'efficacité de cette transformation couplée à de la localisation pour améliorer l'estimation de la matrice de covariance. Ils ont conclu que cette méthodologie permettait de bien estimer des champs de conductivités qui ne respectent pas l'hypothèse de multigaussianité, en particulier, leur champ bimodal; malgré les limites de l'anamorphose gaussienne.

En effet, la transformation par partitions normales n'assure pas la gaussianité conjointe des données. Elle ne garantit donc pas l'optimalité de l'estimation par filtres de Kalman. De plus, cette transformation a été développée pour des distributions continues et ne serait donc possiblement pas adaptée à des distributions "pathologiques" telles que les distributions de Dirac ou bimodales (Simon & Bertino, 2009). Si Schöniger (2010) a conclu que l'anamorphose gaussienne des données de rabattement améliorait l'estimation des données de conductivité hydraulique par EnKF, l'auteure a noté que cette même transformation n'était pas recommandée lorsque des données de charges hydrauliques étaient plutôt assimilées. La transformation des données de charges serait une étape supplémentaire qui ne garantit pas des résultats stables, ni l'amélioration de ces résultats. Elle a aussi obtenu des résultats mitigés concernant l'assimilation de données de concentration dans un cadre couplé d'écoulement et de transport: la prédiction de la concentration à un pas de temps ultérieur pouvait être améliorée ou détériorée selon le champ de conductivités hydrauliques synthétiques à estimer. L'anamorphose gaussienne des concentrations a aussi été testée dans le cas de l'assimilation de données d'un test de traceur pour la mise à jour de la conductivité hydraulique par Crestani et al. (2013) par EnKF. Dans le cas du test de traceur, il a été conclu que l'anamorphose gaussienne altérait la corrélation entre la conductivité hydraulique et les concentrations. Cette altération détériorait les résultats de l'assimilation par rapport à l'assimilation sans transformation sur les concentrations.

Les EnKF ont été étudiés en profondeur dans un cadre d'inversion hydrogéologiques sur des cas synthétiques. Ils ont aussi été mis en pratique sur des études de sites réel. Vogt *et al.* (2012) ont estimé la distribution de perméabilité dans le réservoir de *Enhanced Geothermal System at Soulz-sous-Forêts* par assimilation de données de traceur par EnKF et Liu *et al.* (2008) ont estimé

la conductivité hydraulique ainsi que la dispersivité longitudinale au site MADE (macro-dispersion experiment). Ces deux cas traitent du problème d'inversion couplé écoulement et transport, tandis que la plupart des études synthétiques traitaient du cas d'inversion de la fonction d'écoulement seulement.

2.2 Intégration de données géophysiques

Il est maintenant reconnu que l'utilisation de données géophysiques permet d'améliorer l'estimation de certains paramètres hydrogéologiques, tel que la conductivité, malgré les incertitudes et les erreurs associées à l'interprétation de ces données (Copty & Rubin, 1995), (Ruggeri *et al.*, 2013).

L'estimation de paramètres hydrogéologiques à partir de données géophysiques repose entre autres sur les relations empiriques qui ont été établies entre caractéristiques géophysiques et hydrogéologiques. Les premières caractérisations hydrogéophysiques se limitaient habituellement à inférer des propriétés hydrogéologiques, comme la conductivité hydraulique, à partir de données de résistivité électrique obtenues par inversion (ou d'autres propriétés géophysiques) et de relations pétrophysiques semi-empiriques (Borner *et al.*, 1996; Kelly, 1977; Hördt *et al.*, 2007; Slater *et al.*, 2002; Kemna *et al.*, 2004). Dans ces études, les relations pétrophysiques sont utilisées pour convertir les images géophysiques en images de propriétés hydrogéologiques. Comme une seule relation pétrophysiques est utilisée pour l'ensemble du site à l'étude, ces relations négligent par le fait même l'hétérogénéité du sous-sol.

Dans d'autres cas, les corrélations entre données géophysiques et paramètres hydrogéologiques sont plutôt établies aux puits. Les données géophysiques à grande étendue peuvent ensuite être utilisées comme variables secondaires dans les méthodes d'interpolation pour l'estimation des propriétés hydrogéologiques. Par exemple, Cassiani *et al.* (1998) intègrent des données de lenteur sismique dans un cokrigeage de la conductivité hydraulique. Dubreuil-Boisclair *et al.* (2011) utilisent des simulations séquentielles avec mise à jour bayésienne pour simuler la conductivité hydraulique entre forages à partir de données de conductivité hydraulique aux puits et de données de tomographie stochastique radar. Un danger de ces méthodes est toutefois que les corrélations obtenues aux puits ne devraient toutefois pas être généralisées puisque ces corrélations varient entre les puits (Day-Lewis & Lane, 2004). Les méthodes d'intégration de données géophysiques par relations empiriques avec les paramètres hydrogélogiques ne tiennent pas compte d'un modèle hydrogéologique. Ces méthodes ont donc le désavantage d'être souvent irréalistes d'un point de vue géologique Hinnell *et al.* (2010).

2.2.1 Inversion hydrogéophysique

Les données géophysiques peuvent apporter une information complémentaire concernant certaines variables hydrogéologiques telles que le degré de saturation en eau ou la concentration en contaminant. Certains auteurs ont proposé de convertir les résultats de l'inversion géophysique en valeurs pouvant être utilisées dans l'inversion hydrogéologique, telles que la concentration, à partir des relations pétrophysiques (par exemple, Hinnell *et al.* (2010) ou Camporese *et al.* (2011)). Toutefois, les méthodes d'inversion géophysique engendrent souvent l'apparition d'artéfacts qui n'ont pas de sens hydrogéologique.

Les données géophysiques peuvent aussi être utilisées conjointement dans l'inversion hydrogéologique. Les méthodes ainsi développées sont dites "méthodes hydrogéophysiques couplées" puisqu'elles intègrent les variables géophysiques et hydrogéologiques dans une seule étape. Ces méthodes sont habituellement des variantes de méthodes de solution au problème inverse en hydrogéologie, qui ont été présentées dans la section 2.1 sur l'inversion hydrogéologique.

Parmi les méthodes d'inversion hydrogéologique présentées, la méthode du maximum d'a posteriori a été étendue aux données géophysiques. Ainsi, la mise à jour est effectuée grâce aux mesures géophysiques et à des relations empiriques entre les données hydrogéologiques et géophysiques (Rubin et al., 1992), (Copty et al., 1993), (Hubbard & Rubin, 2000). L'amélioration apportée à la distribution a posteriori par rapport à la distribution a priori dépend des données géophysiques et du modèle pétrophysique. Toutefois, Copty et al. (1993) ont noté que même lorsque les données sismiques qu'ils ont utilisées pour mettre à jour l'estimation étaient corrompues par une erreur significative, l'intégration des données sismiques améliorait l'estimation de la conductivité hydraulique. Par ailleurs, Chen et al. (2001) ont trouvé que l'amélioration de l'estimation était la plus significative aux endroits où l'information a priori était limitée.

Les méthodes de maximum de vraisemblance pour l'inversion hydrogéologique ont aussi été étendues aux observations géophysiques. Pollock & Cirpka (2010) ont utilisé la méthode d'inversion hydrogéologique développée par Kitanidis (1995) pour estimer des conductivités hydrauliques à partir de moments temporels de la résistivité électrique lors d'un test de traceur. Irving & Singha (2010) ont appliqué l'inversion par méthode de Monte-Carlo par chaînes de Markov sur des observations de résistance électrique et de concentration pour estimer la conductivité hydraulique. Hinnell *et al.* (2010) a aussi utilisé une méthode de maximum de vraisemblance pour estimer des paramètres hydrauliques à travers les valeurs de contenu en eau dans un modèle d'écoulement couplé à un modèle électrique.

Finalement, les méthodes d'assimilation de données, tel que les EnKF ont aussi été développées pour assimiler des données géophysiques. Camporese *et al.* (2015) et Crestani *et al.* (2015) ont proposé d'assimiler des données de résistance électrique obtenues par ERT lors d'un test de traceur pour estimer la conductivité électrique par EnKF.

Certaines méthodes hydrogéophysiques couplées découlent plutôt de l'inversion des données géophysiques.Fowler & Moysey (2011) présentent une méthode dans laquelle les paramètres hydrogéologiques sont convertis en paramètres géophysiques grâce à la loi d'Archie. L'erreur entre les données géophysiques estimées et les données géophysiques observées est ensuite minimisée. Hinnell *et al.* (2010) et Huisman *et al.* (2010) ont développé des variantes de cette méthode. Ces deux études concluent que l'inversion hydrogéophysique permet de calibrer efficacement les modèles hydrogéologiques, même si Hinnell *et al.* (2010) notent que si des erreurs structurelles sont présentes dans le modèle hydrogéologique, l'inversion hydrogéophysique couplée pourrait ne pas améliorer l'interprétation hydrogéologique.

Fowler & Moysey (2011) ont effectué une analyse de sensibilité des mesures de résistance électrique obtenues par ERT dans un modèle hydrogéophysique couplé. Ils ont conclu que différents panaches de contamination pouvaient donner une même réponse électrique. Une étude de l'évolution temporelle du panache serait bénéfique en inversion hydrogéophysique.

2.3 Motivation de la présente étude

La revue de littérature a permis de faire l'historique des méthodes d'inversion en hydrogéologie. Des méthodes d'inversion déterministes ont d'abord été développées, puis ont été suivies par des méthodes probabilistes qui modélisaient les distributions de probabilité des paramètres à calibrer. Les méthodes d'assimilation de données permettent de tenir compte, non seulement d'un aspect probabiliste des paramètres à calibrer, mais aussi de l'aspect temporel des variables observées lorsque le phénomène à l'étude est transitoire (par exemple, concentration, charges hydrauliques).

La plupart de ces méthodes ont pu être étendues afin de permettre la calibration des paramètres hydrogéologiques à partir d'observations géophysiques tel que les résistances obtenues par la tomographie de résistivité électrique (ERT). L'objectif principal de cette thèse est donc de développer une méthodologie de l'estimation de paramètres hydrogélogiques par assimilation de données hydrogéophysiques.

La méthode d'assimilation de données préconisée dans cette étude est les filtres de Kalman d'ensemble (EnKF) qui ont montré de bons résultats pour l'assimilation de données de concentration (Schöniger (2010)) et de résistances électriques (Camporese *et al.* (2015)).

L'utilisation des filtres de Kalman d'ensemble comporte plusieurs limites qui ont été soulevées dans la littérature et qui sont présentées à la section 3.1.9, notamment la sensibilité à la taille de l'ensemble, la représentation de l'incertitude et les problèmes de dégénerescence de la covariance.

Afin d'évaluer la meilleure méthode d'assimilation de données hydrogéophysiques pour la calibration de paramètres hydrogéologiques par EnKF, des variantes de ces filtres ont d'abord dû être développées. Notamment, une première variante des EnKF permettant de faire l'assimilation des données de résistance électrique en deux étapes, soit l'estimation des concentrations puis l'estimation des paramètres hydrogéologiques à partir de ces concentrations, a été développée et nommée EnKF découplé. Cette nouvelle méthode est présentée à la section 3.1.6 Une seconde variante des EnKF basée sur l'assimilation séquentielle des données (Evensen & Van Leeuwen (2000)) a aussi été utilisée pour faire l'assimilation conjointe de données géophysiques (levé ERT) et de données ponctuelles (concentrations aux puits). Un filtre à racine carrée a aussi été mis en place pour faire l'assimilation de données hydrogéophysiques.

En pratique, les modèles géostatistiques sont inférés d'informations qualitatives et d'un nombre limité de données. Ainsi, ils comportent une incertitude (Jafarpour & Tarrahi, 2011). Une méthodologie a donc été mise en place pour tenir compte de l'incertitude sur les paramètres géostatistiques (portée, variance, moyenne) des simulations de l'ensemble de départ utilisé pour l'assimilation de données.
Différentes mesures hydrogéologique (concentrations) ou géophysique (levé ERT) peuvent être obtenues dans un cas de contamination. En plus d'être obtenues à un coût généralement plus faible, les mesures de résistance électrique ont une plus grande couverture spatiale que les mesures de concentrations en puits. Les différentes variantes des EnKFs décrites préalablement ont été comparées entre elles sur trois cas particuliers d'un cas synthétique de contamination saline: soit, l'assimilation de données de concentrations aux puits, l'assimilation de données de résistances et l'assimilation conjointe des deux types de données. La comparaison de ces trois cas permettra de déterminer l'apport des mesures ERT et de concentration à la mise à jour de la conductivité hydraulique.

Enfin, une étude de sensibilité à la taille de l'ensemble a été effectuée dans tous les cas d'étude étant donné que la taille de l'ensemble est un paramètre déterminant sur l'efficacité des EnKFs Yin *et al.* (2015).

Les conclusions obtenues sur le cas synthétique ont permis de déterminer une méthodologie optimale à appliquer pour assimiler des données de résistance électrique sur un cas de contamination saline en laboratoire. Cette méthode a été comparée à une méthode développée par (Camporese *et al.*, 2015) où les données de résistance sont d'abord inversées par moindres carrés, puis les résultats de l'inversion sont assimilés.

Chapitre 3

Assimilation de données hydrogéophysique

Dans la présente thèse il est proposé d'assimiler des mesures obtenues par tomographie par résistivité électrique pour estimer des paramètres hydrogéologiques, tel que la conductivité hydraulique et/ou la dispersivité. L'assimilation de données est un domaine qui regroupe différentes méthodes qui permettent d'utiliser des observations conjointement à un modèle physique afin d'améliorer les prédictions d'état obtenues par le modèle. Les méthodes d'assimilation sont habituellement divisées en deux catégories, soit les méthodes séquentielles si les observations sont assimilées à chaque pas de temps ou non-séquentielles, si les observations sont toutes assimilées en une seule étape. De plus, les méthodes d'assimilation de données peuvent traiter soit de phénomènes continus ou discrets. Cette thèse s'intéresse à l'assimilation séquentielle discrète de mesures hydrogéophysiques, seules les méthodes séquentielles et discrètes seront donc abordées dans les sections suivantes.

La figure 3.1 est une représentation schématique de l'assimilation de données séquentielle où xreprésente l'état à estimer et d_t , les observations au temps t; tandis que l'indice t|t indique que l'état est mis à jour, l'indice 0|0 indique l'initialisation de l'état et l'indice t|t-1 indique l'état prédit.



Figure 3.1 – Schéma de l'assimilation séquentielle de données

Ce chapitre est divisé en deux parties. La première partie est un cadre théorique de la méthode d'assimilation qui a été sélectionnée pour l'estimation des paramètres hydrogéologiques, soit les filtres de Kalman d'ensemble (EnKF). Cette partie comporte non-seulement l'historique du développement des EnKF, mais aussi certaines variations apportées aux EnKF originaux qui ont été mis en place dans le cadre présent d'assimilation de données. La seconde partie définit les modèles hydrogéologiques et électriques nécessaires dans l'étape de prédiction pour l'assimilation de données de résistivité électrique pour estimer des paramètres hydrogéologiques.

3.1 Méthodes d'assimilation de données séquentielles discrètes

L'algorithme des filtres de Kalman est une méthode d'assimilation des données mise de l'avant par Kalman (1960) pour des fonctions de transfert linéaires. Plusieurs adaptations des KFs classiques ont suivi, dont les filtres d'information, les filtres à racine-carrée, les filtres de Kalman étendus (*extended Kalman filter*), les filtres inodores (*Unscented Kalman Filter*) et les filtres de Kalman d'ensemble (*Ensemble Kalman Filter*). Ces trois dernières adaptations du filtre de Kalman classique permettent d'appliquer le filtre à des fonctions non-linéaires. Par ailleurs, les KFs ont d'abord été développés en temps continu (Kalman & Bucy, 1961). Ce qui suit est toutefois valide pour des variables en temps discret.

3.1.1 Récursion bayésienne

Le filtre de Kalman et ses dérivés (filtre de Kalman étendu, filtre de Kalman d'ensemble, etc...) est basé sur le modèle markovien suivant:

$$\mathbf{x}_t = g(\mathbf{x}_{t-1}) + \mathbf{u}_t \tag{3.1}$$

où \mathbf{x}_t est un vecteur de dimension $(n \times 1)$ qui représente l'état au temps t, $g(\mathbf{x}_{t-1})$ représente une fonction de transfert qui peut être linéaire ou non, et \mathbf{u}_t est un vecteur $(n \times 1)$ de l'erreur du modèle

Les observations \mathbf{d}_t sont liées à l'état par la fonction d'observation, $h(\mathbf{x}_t)$:

$$\mathbf{d}_t = h(\mathbf{x}_t) + \mathbf{w}_t \tag{3.2}$$

et \mathbf{w}_t est un vecteur $m\times 1$ de l'erreur de mesure.

Comme la propagation de l'état suit un modèle markovien, alors la fonction de densité de probabilité de l'état x au temps t conditionnelle à l'état aux temps précédents dépend seulement de l'état au dernier pas de temps, c'est-à-dire x_{t-1} , ou:

$$f(\mathbf{x}_t | \mathbf{x}_0, \dots, \mathbf{x}_{t-1}) = f(\mathbf{x}_t | \mathbf{x}_{t-1})$$
(3.3)

Et, si les erreurs $\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_t$ et $\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_t$ sont toutes indépendantes les unes des autres et indépendantes de l'état initial \mathbf{x}_0 ,

$$f(\mathbf{d}_t | \mathbf{x}_0, \dots, \mathbf{x}_t) = f(\mathbf{d}_t | \mathbf{x}_t)$$
(3.4)

Ainsi, la loi de Bayes permet d'obtenir la formule récursive suivante 3.5.

$$f(\mathbf{x}_t|\mathbf{d}_0,\ldots,\mathbf{d}_t,\mathbf{x}_0) \propto f(\mathbf{x}_0) \times \prod_{i=0}^t f(\mathbf{d}_i|\mathbf{x}_i)$$
 (3.5)

Les algorithmes d'assimilation de données séquentielles se basent sur cette formulation récursive. À chaque pas de temps, la formule de Bayes permet de calculer la fonction de densité de probabilité, a posteriori $f(\mathbf{x}_t | \mathbf{d}_0, \dots, \mathbf{d}_t, \mathbf{x}_0)$ à partir de la fonction de densité de probabilité a priori $f(\mathbf{x}_t | \mathbf{d}_0, \dots, \mathbf{d}_{t-1}, \mathbf{x}_0)$ et des nouvelles données \mathbf{d}_t :

$$f(\mathbf{x}_t | \mathbf{d}_0, \dots, \mathbf{d}_t, \mathbf{x}_0) \propto f(\mathbf{d}_t | \mathbf{x}_t) \times f(\mathbf{x}_t | \mathbf{d}_0, \dots, \mathbf{d}_{t-1}, \mathbf{x}_0)$$
(3.6)

La densité de probabilité *a priori* au temps $t, f(\mathbf{x}_t | \mathbf{d}_0, \dots, \mathbf{d}_{t-1}, \mathbf{x}_0)$ est obtenue grâce à l'équation de Chapman-Kolmogrov:

$$f(\mathbf{x}_t | \mathbf{d}_0, \dots, \mathbf{d}_{t-1}, \mathbf{x}_0) = \int f(\mathbf{x}_t | \mathbf{x}_{t-1}) f(\mathbf{x}_{t-1} | \mathbf{d}_0, \dots, \mathbf{d}_{t-1}, \mathbf{x}_0) \mathrm{d}\mathbf{x}_{t-1}$$
(3.7)

3.1.2 Filtres de Kalman

Dans le cas où la fonction de densité de probabilité conditionnelle au temps t-1, $f(\mathbf{x}_{t-1}|\mathbf{d}_0, \ldots, \mathbf{d}_{t-1}, \mathbf{x}_0)$ suit une loi gaussienne et que la fonction de densité de probabilité de l'erreur du modèle $\mathbf{u}_t \sim \mathcal{N}(0, \mathbf{Q}_t)$ est aussi gaussienne, il est possible de démontrer que, si la fonction de transfert $g(\mathbf{x}_{t-1})$ est linéaire (représentée par la matrice \mathbf{G}), la fonction de densité de probabilité conditionnelle au prochain pas de temps, $f(\mathbf{x}_t|\mathbf{d}_0, \ldots, \mathbf{d}_{t-1}, \mathbf{x}_0)$ sera aussi gaussienne II est donc possible de définir les variables de moyenne et covariance de l'état *a priori*, $\bar{\mathbf{x}}_{t|t-1}$ et $\mathbf{P}_{t|t-1}$ de façon récursive.

$$\bar{\mathbf{x}}_{t|t-1} = \mathbf{G}\bar{\mathbf{x}}_{t-1|t-1} \tag{3.8}$$

$$\mathbf{P}_{t|t-1} = \mathbf{G}\mathbf{P}_{t-1|t-1}\mathbf{G}' + \mathbf{Q}_t \tag{3.9}$$

où $\bar{\mathbf{x}}_{t-1|t-1}$ et $\mathbf{P}_{t-1|t-1}$ sont la moyenne et la covariance de l'état mis à jour au pas de temps précédent. Habituellement, la moyenne ($\bar{\mathbf{x}}_{0|0}$) et la covariance ($\mathbf{P}_{0|0}$) de l'état initial sont connus et permettent d'initialiser le processus d'assimilation de données.

Si l'erreur d'observation \mathbf{w}_t suit aussi une loi gaussienne $\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{W}_t)$ et que la fonction qui lie l'état aux observations $h(\mathbf{x}_t)$ est aussi linéaire et représentée par la matrice \mathbf{H} , alors il sera possible de calculer la moyenne et la covariance de l'état *a posteriori*:

$$\bar{\mathbf{x}}_{t|t} = \bar{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{d}_t - \mathbf{H}\mathbf{x}_{t|t-1})$$
(3.10)

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{K}_t \mathbf{H}_t \mathbf{P}_{t|t-1}$$
(3.11)

La matrice \mathbf{K}_t est le gain de Kalman calclué de la façon suivante:

$$\mathbf{K}_{t} = \mathbf{P}_{t|t-1}\mathbf{H}_{t}^{T}(\mathbf{H}_{t}\mathbf{P}_{t|t-1}\mathbf{H}_{t}^{T} + \mathbf{W}_{t})$$
(3.12)

Le gain de Kalman est un facteur linéaire minimisant l'erreur quadratique moyenne de l'état (par exemple, Brown & Hwang (1997)).

Les équations 3.10 à 3.12 sont les équations de mise à jour du filtre de Kalman classique et permettent de définir complètement l'état *a posteriori* puisque cet état suit une loi normale. De la même façon, les équations de prédictions 3.8 et 3.9 permettent de définir complètement l'état *a priori*.

Les équations 3.8 à 3.12 sont les équations récursives du filtre de Kalman classique et minimisent la covariance de l'erreur d'estimation, dans un cadre où l'état initial ainsi que les erreurs du modèle et de mesure suivent une loi gaussienne en plus d'avoir des fonctions de transfert et d'observation linéaires.

3.1.3 Filtres de Kalman d'ensemble

Comme les systèmes dynamiques dans plusieurs applications géophysiques sont souvent nonlinéaires, des variantes du filtre de Kalman ont été développées pour l'assimilation de données dans ces systèmes non-linéaires. La première variante ayant été développée est le filtre de Kalman étendu qui se base sur l'expansion en séries de Taylor des équations non-linéaires du système. Les moments d'ordre supérieurs à deux sont ignorés dans l'EKF ce qui peut avoir des effets sur la propagation de la covariance de l'erreur d'estimation. Peu de généralités peuvent être énoncées à propos de ces effets qui doivent plutôt être étudiés au cas par cas (Evensen, 1992). De plus, Evensen (1994) a montré que cette variante pouvait montrer des incohérences entre la matrice de covariance de l'erreur estimée et l'erreur réelle. Ces incohérences seraient dues à la linéarisation de la fonction de transfert. Ainsi, Evensen (1994) a proposé une nouvelle méthodologie pour les systèmes non-linéaires pouvant résoudre ces problèmes, les filtres de Kalman d'ensemble. Reichle *et al.* (2002b) ont noté que les EnKF performaient mieux que les EKF lorsque le nombre de membres de l'ensemble est assez gros. De plus, les filtres de Kalman d'ensemble ont un coût informatique moins élevé que les EKF (Bertino *et al.*, 2002).

Les filtres de Kalman d'ensemble se basent sur les méthodes Monte Carlo pour résoudre l'équation 3.7. Ainsi la distribution de probabilité est représentée par un ensemble fini de réalisations. À l'étape de prédiction, la fonction de transfert $g(\mathbf{x}_{t-1})$ de l'équation 3.1 est appliquée sur chaque réalisation, ce qui permet d'obtenir l'ensemble $\mathbf{x}_{t|t-1}^0, \ldots, \mathbf{x}_{t|t-1}^{N_e}$ où l'exposant *i* indique le membre de l'ensemble de taille N_e .

$$\mathbf{x}_{t|t-1}^{i} = g(\mathbf{x}_{t-1|t-1}^{i}) \tag{3.13}$$

L'erreur du modèle, \mathbf{u}_t peut être ajoutée à la prédiction. Dans l'utilisation des EnKF en sciences de l'atmosphère, la modélisation de cette erreur est considérée cruciale (Dee, 1995) et donc, souvent modélisée et ajoutée au modèle de prédiction (Mitchell *et al.*, 2002). Par contre, Evensen (2003) note que cette erreur est souvent mal connue et donc, est souvent négligée. C'est le cas, dans plusieurs applications géophysiques des filtres de Kalman d'ensemble (Jensen, 2007), ainsi que dans l'application des EnKF en hydrogéologie (Hendricks Franssen & Kinzelbach, 2008). Dans le cas à l'étude, comme des phénomènes hydrogéologiques et électriques terrestres sont modélisés l'erreur du modèle sera négligée.

L'ensemble a priori permet ensuite d'estimer la moyenne, $\hat{\mathbf{x}}_{t|t-1}$, et la variance, $\hat{\mathbf{P}}_{t|t-1}$, de la distribution $f(\mathbf{x}_{t|t-1})$.

$$\hat{\mathbf{x}}_{t|t-1} = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{x}_{t|t-1}^i$$
(3.14)

$$\hat{\mathbf{P}}_{t|t-1} = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} (\mathbf{x}_{t|t-1}^i - \hat{\mathbf{x}}_{t|t-1}) (\mathbf{x}_{t|t-1}^i - \hat{\mathbf{x}}_{t|t-1})'$$
(3.15)

Lorsque N_e tend vers l'infini, les équations 3.14 et 3.15 tendent vers la vraie moyenne et la vraie variance *a priori*.

Un estimateur du gain de Kalman, qui tend vers le gain de Kalman du filtre classique pour N_e tend vers l'infini pour un système linéaire, peut maintenant être calculé.

$$\hat{\mathbf{K}}_{t} = \hat{\mathbf{P}}_{t|t-1} \mathbf{H}_{t}^{T} (\mathbf{H}_{t} \hat{\mathbf{P}}_{t|t-1} \mathbf{H}_{t}^{T} + \mathbf{W}_{t})$$
(3.16)

Une relation linéaire entre l'état et les observations, est supposée à l'équation 3.16, représentée par la matrice \mathbf{H}_t . À la section 3.1.5, la mise à jour pour une fonction d'observation non-linéaire sera discutée.

La mise à jour est ensuite appliquée à chaque membre de l'ensemble pour obtenir l'ensemble *a* posteriori.

À cette étape, un bruit de moyenne nulle devra être ajouté aux valeurs observées afin que la moyenne et la covariance de la distribution après la mise à jour équivalent à la moyenne et la covariance obtenues par filtre de Kalman classique lorsque N_e est assez grand (Burgers *et al.*, 1998). Ainsi,

$$\mathbf{d}_t^i = \mathbf{d}_t^{obs} + \mathbf{w}_t^i \tag{3.17}$$

où \mathbf{d}_t^i sont les données bruitées associées au membre *i* de l'ensemble, \mathbf{d}_t^{obs} sont les données observées au temps *t* et \mathbf{w}_t^i est une erreur aléatoire qui suit une loi $N(0, \mathbf{W}_t)$

L'ensemble mis à jour devient donc:

$$\mathbf{x}_{t|t}^{i} = \mathbf{x}_{t|t-1}^{i} + \hat{\mathbf{K}}_{t}(\mathbf{d}_{t}^{i} - \mathbf{H}\mathbf{x}_{t|t-1}^{i})$$
(3.18)

La moyenne, $\hat{\mathbf{x}}_{t|t}$, et la variance, $\hat{\mathbf{P}}_{t|t}$, de la distribution *a posteriori* $f(\mathbf{x}_{t|t})$ peuvent être calculées sur ce nouvel ensemble et, dans un cas linéaire et avec un nombre d'ensemble qui tend vers l'infini, ces valeurs devraient tendre vers les vraies valeurs calculées à partir des équations 3.10 et 3.11.

D'ailleurs, la convergence des filtres de Kalman d'ensemble vers la solution obtenue par filtre de Kalman classique pour un ensemble dont la taille tend vers l'infini a été démontrée par Mandel *et al.* (2011). Par ailleurs, le KF classique est un estimateur non-biaisé et à variance minimale. Toutefois, ces résultats (KF en tant qu'estimateur optimal et convergence des EnKF vers les KF) ne sont valides que lorsque la fonction de transfert et la fonction d'observation sont linéaires et l'état initial ainsi que les erreurs d'observation et de modèle suivent chacun une loi gaussienne. Les erreurs d'observation et de modèle doivent être indépendantes l'une de l'autre et indépendantes de l'état initial.

Si les filtres de Kalman d'ensemble convergent vers la solution obtenue par les KF classiques pour un ensemble dont la taille tend vers l'infini, l'utilisation d'un ensemble de taille fini peut engendrer plusieurs problèmes. Ces problèmes seront discutés à la section 3.1.9.

3.1.4 Filtres à racine carrée (square-root filter)

Précédemment, la nécessité d'ajouter un bruit sur les observations a été soulevée. Toutefois, selon Whitaker & Hamill (2002), l'ajout de ce bruit peut diminuer l'exactitude de la matrice de covariance de l'erreur de l'état *a posteriori* en plus de sous-estimer cette matrice. Ces problèmes surviendraient particulièrement lorsque l'ensemble est petit. Des méthodes ont été proposées pour contourner l'ajout d'une erreur d'observation aléatoire et ont été nommées filtres déterministes. Les méthodes qui nécessitent l'ajout d'une erreur d'observation sont nomées filtres stochastiques. Whitaker & Hamill (2002) concluent que les filtres déterministes sont plus justes que les filtres stochastiques. Furrer & Bengtsson (2007) ont aussi noté que les filtres déterministes reproduisent les équations de récursion des KF plus facilement que les filtres stochastiques pour des ensembles de taille finie.

Les filtres à racine carrée sont une classe des filtres déterministes. Leur justification, tirée de Tippett *et al.* (2003) est présentée ci-dessous. Les indices pour représenter le pas de temps ont été éliminés pour alléger le texte et puisque, seule l'étape de mise à jour à un pas de temps donné est présentée. Par ailleurs, une fonction linéaire liant les observations à l'état est considérée. La généralisation à une fonction non-linéaire est possible de la même façon que pour les EnKF, tel que présenté à la section 3.1.5.

La matrice $\mathbf{X'}_{t|t-1}$ contient les perturbations des membres de l'ensemble, c'est-à-dire, leur différence par rapport à la moyenne de l'ensemble ($\bar{\mathbf{x}}$). Cette matrice est définie à l'étape de prédiction.

$$\mathbf{X}_{t|t-1}^{*} = \frac{1}{\sqrt{(N_e - 1)}} (\mathbf{x}_{t|t-1}^{0} - \bar{\mathbf{x}}, \dots, \mathbf{x}_{t|t-1}^{N_e} - \bar{\mathbf{x}})$$
(3.19)

La mise à l'échelle de la matrice $\mathbf{X'}_{t|t-1}$ par le facteur $\frac{1}{\sqrt{(N_e-1)}}$ permet de calculer la covariance de prédiction non-biaisée de la façon suivante:

$$\hat{\mathbf{P}}_{t|t-1} = \mathbf{X}_{t|t-1}^{*} \mathbf{X}_{t|t-1}^{T}$$
(3.20)

L'équation 3.20 équivaut à l'équation 3.15, mais formulée différemment.

Les filtres à racine carrée sont développés à partir du calcul de la covariance *a posteriori*, tel que décrit par le filtre de Kalman classique :

$$\hat{\mathbf{P}}_{t|t} = \hat{\mathbf{P}}_{t|t-1} - \mathbf{K}_t \mathbf{H}_t \hat{\mathbf{P}}_{t|t-1}$$
(3.21)

Si les perturbations, après mise à jour sont définies comme:

$$\tilde{\mathbf{X}}_{t|t} = \mathbf{X'}_{t|t-1}\mathbf{T} \tag{3.22}$$

où \mathbf{T} est une matrice arbitraire qui répond à l'équation 3.23

$$\mathbf{T}\mathbf{T}^{T} = \mathbf{I} - \mathbf{X}_{t|t-1}^{T} \mathbf{H}^{T} (\mathbf{H}\mathbf{X}_{t|t-1}^{T}\mathbf{X}_{t|t-1}^{T}\mathbf{H}^{T} + \mathbf{W}_{t})^{-1} \mathbf{H}\mathbf{X}_{t|t-1}^{*}$$
(3.23)

Les équations 3.22 et 3.23 permettent donc de définir la matrice de covariance *a posteriori*, $\mathbf{P}_{t|t}$ à partir de l'équation 3.24:

$$\hat{\mathbf{P}}_{t|t} = \tilde{\mathbf{X}}_{t|t} \tilde{\mathbf{X}}_{t|t}^{T}$$
(3.24)

Les filtres à racine carrée cherchent à trouver une valeur pour **T** qui satisfait à l'équation 3.23, d'où leur nom. Dans le domaine des mathématiques, la racine carrée d'une matrice correspond habituellement à **S** pour $\mathbf{P} = \mathbf{SS}$. Toutefois, en ingénierie, le terme de racine carrée réfère souvent à la matrice **S**, lorsque celle-ci est plutôt la solution de $\mathbf{P} = \mathbf{SS}^T$, par exemple, chez Petrie (2008), Whitaker & Hamill (2002) ou Livings *et al.* (2008). Dans le cas d'une matrice **S** symétrique, les deux définitions sont équivalentes.

La solution de l'équation 3.23 n'est pas unique ce qui a mené à différents développements qui font tous partie de la famille des filtres d'ensemble à racine carrée, dont les filtres de Kalman évolutifs interpolés (Singular evolutive interpolated Kalman filter ou SEIK) de Pham (2001) ou les filtres de Kalman d'ensemble d'ajustement (Ensemble adjustment Kalman filter ou EAKF) de Anderson (2001). Les filtres de Kalman d'ensemble transformés (Ensemble transform Kalman filter ou ETKF) sont la formulation des filtres d'ensemble à racine carrée qui seront ici utilisés. Les ETKF ont d'abord été développés par Bishop *et al.* (2001). Wang *et al.* (2004) ont apporté des modifications afin de s'assurer que le résultat était non-biaisé.

Les ETKF se basent sur l'identité 3.25 qui permettra de simplifier la solution pour trouver **T** de l'équation 3.23 (Petrie, 2008):

$$\mathbf{I} - \mathbf{X}_{t|t-1}^{T} \mathbf{H}^{T} (\mathbf{H} \mathbf{X}_{t|t-1}^{T} \mathbf{X}_{t|t-1}^{T} \mathbf{H}^{T} + \mathbf{W}_{t})^{-1} \mathbf{H} \mathbf{X}_{t|t-1}^{T} = (\mathbf{I} - \mathbf{X}_{t|t-1}^{T} \mathbf{H}^{T} \mathbf{W}_{t}^{-1} \mathbf{H} \mathbf{X}_{t|t-1}^{T})^{-1}$$
(3.25)

Ainsi, il suffira d'effectuer la décomposition de $\mathbf{X}_{t|t-1}^T \mathbf{H}^T \mathbf{W}_t^{-1} \mathbf{H} \mathbf{X}_{t|t-1}^{-1}$ en valeurs propres:

$$\mathbf{X}_{t|t-1}^{T} \mathbf{H}^{T} \mathbf{W}_{t}^{-1} \mathbf{H} \mathbf{X}_{t|t-1}^{T} = \mathbf{U} \Lambda \mathbf{U}^{T}$$
(3.26)

où U est une matrice carrée où la colonne *i* est le vecteur propre u_i et Λ est une matrice diagonale dont l'élément *i* de la diagonale est la valeur propre, λ_i du membre de gauche de l'équation 3.26.

pour calculer:

$$\mathbf{I} - \mathbf{X}_{t|t-1}^{T} \mathbf{H}^{T} (\mathbf{H} \mathbf{X}_{t|t-1}^{T} \mathbf{X}_{t|t-1}^{T} \mathbf{H}^{T} + \mathbf{W}_{t})^{-1} \mathbf{H} \mathbf{X}_{t|t-1}^{T} = \mathbf{U} (\mathbf{I} + \Lambda)^{-1} \mathbf{U}^{T}$$
(3.27)

La matrice $(\mathbf{I} + \Lambda)$ étant diagonale, il est donc facile de calculer $(\mathbf{I} + \Lambda)^{-1/2}$ pour obtenir la racine carrée du membre de gauche de l'équation 3.27 qui équivaut à la matrice **T**. Dans la version originale des ETKF proposée par Bishop *et al.* (2001), une solution de 3.23 était donc,

$$\mathbf{T} = \mathbf{U}(\mathbf{I} + \Lambda)^{-1/2} \tag{3.28}$$

Livings *et al.* (2008) a montré que cette formulation pouvait produire des estimations biaisées. Livings *et al.* (2008) a aussi montré que la solution proposée par Wang *et al.* (2004), qui sera utilisée ci-après, garantissait une estimation non-biaisée:

$$\mathbf{T} = \mathbf{U}(\mathbf{I} + \Lambda)^{-1/2} \mathbf{U}^T$$
(3.29)

Livings (2005) a proposé un algorithme équivalent aux ETKF tel que proposé par Bishop *et al.* (2001), mais plus rapide que la décomposition en valeurs propres pour la mise à jour de l'ensemble. Cet algorithme a ensuite été corrigé pour être non-biaisé par Petrie (2008).

La matrice $\hat{\mathbf{Y}}_t$ doit d'abord être définie.

$$\hat{\mathbf{Y}}_t = \mathbf{W}_t^{-1/2} \mathbf{H} \mathbf{X}_{t|t-1} \tag{3.30}$$

La décomposition en valeurs singulières de $\hat{\mathbf{Y}}_t^{\ T}$:

$$\hat{\mathbf{Y}}_t^{\ T} = \mathbf{U} \Sigma \mathbf{V}^T \tag{3.31}$$

permet d'obtenir les valeurs propres: $\Lambda = \Sigma \Sigma^T$

Les perturbations mises à jour sont donc calculées

$$\tilde{\mathbf{X}}_{t|t} = \mathbf{X}_{t|t-1}^{*} \mathbf{T} = \mathbf{X}_{t|t-1}^{*} \mathbf{U} (\mathbf{I} + \Lambda)^{-1/2} \mathbf{U}^{T}$$
(3.32)

Le gain de Kalman peut aussi être calculé

$$\begin{aligned} \mathbf{K}_{t} &= \hat{\mathbf{P}}_{t|t-1} \mathbf{H}_{t}^{T} (\mathbf{H}_{t} \hat{\mathbf{P}}_{t|t-1} \mathbf{H}_{t}^{T} + \mathbf{W}_{t})^{-1} \\ &= \mathbf{X}'_{t|t-1} \mathbf{X}'_{t|t-1}^{T} \mathbf{H}_{t}^{T} (\mathbf{H}_{t} \mathbf{X}'_{t|t-1} \mathbf{X}'_{t|t-1}^{T} \mathbf{H}_{t}^{T} + \mathbf{W}_{t})^{-1} \\ &= \mathbf{X}'_{t|t-1} \hat{\mathbf{Y}}_{t}^{T} (\hat{\mathbf{Y}}_{t} \hat{\mathbf{Y}}_{t}^{T} + \mathbf{I})^{-1/2} \\ &= \mathbf{X}'_{t|t-1} \mathbf{U} \Sigma (\Sigma^{T} \Sigma + \mathbf{I})^{-1} \mathbf{V}^{T} \mathbf{W}_{t}^{-1/2} \end{aligned}$$
(3.33)

La moyenne après assimilation peut ensuite être calculée à partir de l'équation 3.10 et du gain de Kalman ainsi obtenue.

$$\bar{\mathbf{x}}_{t|t} = \bar{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{d}_t - \mathbf{H}\mathbf{x}_{t|t-1})$$
(3.34)

Aucune erreur n'a dû être ajoutée à la variable d'observation \mathbf{d}_t .

La matrice contenant l'ensemble *a posteriori* peut finalement être calculée en rajoutant les perturbations à la moyenne *a posteriori*.

$$\mathbf{X}_{t|t} = \tilde{\mathbf{X}}_{t|t} + \bar{\mathbf{x}}_{t|t} \tag{3.35}$$

La matrice $\mathbf{X}_{t|t}$ équivaut à l'ensemble $\mathbf{x}_{t|t}^i$ de l'équation 3.18 mis à jour par EnKF.

3.1.5 État augmenté

La définition de l'état est une étape importante dans l'application des algorithmes d'assimilation de données. Dans certains cas, les variables devant être utilisées pour définir l'état iront de soi, comme, par exemple la position d'une particule dans un problème de traçage de particule. Dans ce cas, la variable à estimer (et prédire) est la même que la variable observée, soit la position de la particule. Toutefois, dans certains problèmes, la relation entre les variables que l'on désire estimer et les variables d'observations n'est pas linéaire ou des erreurs auto-corrélées peuvent être présentes et doivent donc être prises en compte. L'état devra donc comporter toutes les variables d'intérêt, dont les variables à estimer et les variables d'observation ainsi que toute variable intermédiaire qui pourrait , par exemple, permettre de tenir compte d'erreurs auto-corrélées. Ainsi, il existe principalement trois cas où l'augmentation de l'état est utilisé:

- Augmentation de l'état par les observations
- Augmentation de l'état par les paramètres
- Augmentation de l'état pour tenir compte d'erreurs auto-corrélées

Seulement les deux premiers cas seront mis en pratique pour l'assimilation des données telle que mise en place dans le présent document. Les lecteurs s'intéressant à l'intégration d'erreurs auto-corrélées dans les EnKF par état augmenté peuvent se référer à Reichle *et al.* (2002a).

L'augmentation de l'état par les observations est proposée dans un cadre plus général d'inversion par Tarantola (2005). Cette stratégie peut être utilisée en particulier lorsque la fonction $h(\mathbf{x}_t)$ de l'équation 3.2 n'est pas linéaire. Ainsi, si une variable transitoire \mathbf{y}_t est créée, il sera possible de ré-écrire les équations de prédiction 3.1 et d'observation 3.2 de la façon suivante:

$$\begin{bmatrix} \mathbf{x}_t \\ \mathbf{y}_t \end{bmatrix} = \begin{bmatrix} f(\mathbf{x}_{t-1}) \\ h(\mathbf{x}_t) \end{bmatrix} + \begin{bmatrix} \mathbf{u}_t \\ \mathbf{0} \end{bmatrix}$$
(3.36)

$$\mathbf{d}_{t} = \begin{bmatrix} \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t} \\ \mathbf{y}_{t} \end{bmatrix} + \mathbf{w}_{t}$$
(3.37)

La relation entre les observations et la variable d'état devient donc linéaire. L'ajout des observations à la variable d'état a été utilisé , entre autres, par Clark *et al.* (2008) dans un contexte d'assimilation de données par filtres de Kalman d'ensemble.

Dans le cas où des paramètres du modèle qui définit la dynamique de l'état doivent être estimés, il est aussi possible d'utiliser un état augmenté. Habituellement, les paramètres du modèle, θ seront des variables statiques dans le temps et donc, les indices de temps ont été enlevés à leur notation. L'équations 3.1 devient:

$$\begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{x}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta} \\ f(\mathbf{x}_{t-1}) \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{u}_t \end{bmatrix}$$
(3.38)

Et l'équation qui lie les observations à l'état devient:

$$\mathbf{d}_{t} = \begin{bmatrix} \mathbf{0} & \mathbf{H} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{x}_{t} \end{bmatrix} + \mathbf{w}_{t}$$
(3.39)

L'augmentation de l'état avec les paramètres pour l'estimation de ceux-ci a, entre autres, été utilisée par Hendricks Franssen & Kinzelbach (2008), Chen & Zhang (2006) et Li *et al.* (2012b) dans un cadre d'assimilation de données hydrogéologiques. Toutefois, cette méthodologie peut causer certains problèmes, en particulier, une incompatibilité entre l'état et les paramètres (Wen *et al.*, 2005). Une méthode pour résoudre cette incompatibilité sera présentée à la section 3.1.8.

Enfin, l'état pourrait aussi être augmenté par les paramètres en plus d'être augmenté par les observations. Ce sera le cas dans l'estimation des paramètes hydrogéologiques par assimilation de données de tomographie de résistivité électrique. En effet, dans ce cas, les observations, soit la résistance électrique mesurée, sont une fonction non-linéaire de la variable d'état, soit la concentration en contaminant. De plus, les paramètres hydrogéologiques (conductivité hydraulique et/ou dispersivité) doivent être estimés. L'équation qui lie l'état du temps t - 1 au temps t devient:

$$\begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{x}_t \\ \mathbf{y}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta} \\ f(\mathbf{x}_{t-1}) \\ h(\mathbf{x}_t) \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{u}_t \\ \mathbf{0} \end{bmatrix}$$
(3.40)

Et l'équation qui lie les observations à l'état au temps t + 1 devient:

$$\mathbf{d}_{t} = \begin{bmatrix} \mathbf{0} & \mathbf{H}_{x} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{x}_{t} \\ \mathbf{y}_{t} \end{bmatrix} + \mathbf{w}_{t}$$
(3.41)

 \mathbf{H}_x représente la relation linéaire qui peut exister entre la variable d'état \mathbf{x}_t et les variables d'observations. Par exemple, dans le cas où la variable d'état est la concentration d'un contaminant, la matrice \mathbf{H}_x pourrait permettre de transformer la variable de concentration en concentrations observées à des puits. Si, aucune observation n'a de lien linéaire avec la variable d'état \mathbf{x}_t , alors $\mathbf{H}_x = \mathbf{0}$.

3.1.6 Assimilation des données découplée

L'augmentation de l'état tel que présenté aux équations 3.36 à 3.41 permet d'appliquer directement l'algorithme des EnKF (ou d'autres algorithmes dérivés des KFs pour des fonctions de transfert non-linéaires) sur la nouvelle définition de l'état. Ici, une nouvelle méthode est proposée pour l'assimilation des données par EnKF lorsque l'état est augmenté par les paramètres et les observations

tel que présenté à l'équation 3.40. L'état ainsi augmenté est rappelé ici: \mathbf{x}_t

Il est proposé d'effectuer l'assimilation en deux étapes. D'abord les données observées, \mathbf{d}_t sont assimilées pour estimer la variable d'état, \mathbf{x}_t . Puis, la variable d'état estimée est elle-même assimilée pour estimer les paramètres $\boldsymbol{\theta}$. Les équations de prédiction doivent d'abord être appliquées, pour chaque variable de l'état et pour chaque membre de l'ensemble, *i*. Étant donné que les paramètres ne sont pas variables dans le temps, $\boldsymbol{\theta}_{t|t-1}^i = \boldsymbol{\theta}_{t-1|t-1}^i$. La fonction non-linéaire liant la variable d'état $\mathbf{x}_{t|t-1}^i$ aux paramètres est ensuite appliquée.

$$\mathbf{x}_{t|t-1}^{i} = f(\boldsymbol{\theta}_{t-1|t-1}^{i}, \mathbf{x}_{t-1|t-1}^{i})$$
(3.42)

Dans certains cas, dont les *Restart-EnKF* présentés plus loin à la section 3.1.8, la variable d'état $\mathbf{x}_{t|t-1}$ sera fonction des paramètres $\boldsymbol{\theta}_{t-1|t-1}$ seulement et non, de la variable d'état $\mathbf{x}_{t-1|t-1}$ au pas de temps précédent. Comme l'état est augmenté avec la variable d'observation $\mathbf{y}_{t|t-1}$, cette variable doit aussi être prédite.

$$\mathbf{y}_{t|t-1}^{i} = h(\mathbf{x}_{t|t-1}^{i}) \tag{3.43}$$

La première étape consite à assimiler les observations afin d'estimer la variable $\mathbf{x}_{t|t}$. À cette étape l'état prédit est défini par: $\begin{bmatrix} \mathbf{x}_{t|t-1} \\ \mathbf{y}_{t|t-1} \end{bmatrix}$

L'équation de mise à jour de l'ensemble devient donc:

$$\begin{bmatrix} \mathbf{x}_{t|t}^{i} \\ \mathbf{y}_{t|t}^{i} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{t|t-1}^{i} \\ \mathbf{y}_{t|t-1}^{i} \end{bmatrix} + \hat{\mathbf{K}}_{t}^{x} (\mathbf{d}_{t}^{i} - \begin{bmatrix} \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t|t-1}^{i} \\ \mathbf{y}_{t|t-1}^{i} \end{bmatrix})$$
(3.44)

où \mathbf{d}_t^i sont les observations auxquelles une erreur a été ajoutée et le $\hat{\mathbf{K}}_t^x$ est le gain de Kalman calculé pour l'état tel que défini précédemment.

Une fois, la variable d'état $\mathbf{x}_{t|t}$ mise à jour, cette variable devient les données d'observation. Comme cette variable est déjà constituée d'un ensemble, il n'est pas nécessaire d'ajouter un bruit aux observations. L'équation de mise à jour de l'ensemble à la deuxième étape permet d'estimer les paramètres, $\boldsymbol{\theta}_{t|t}$.

$$\begin{bmatrix} \boldsymbol{\theta}_{t|t}^{i} \\ \mathbf{x}_{t|t}^{i} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}_{t|t-1}^{i} \\ \mathbf{x}_{t|t-1}^{i} \end{bmatrix} + \hat{\mathbf{K}}_{t}^{\theta} (\mathbf{x}_{t|t}^{i} - \begin{bmatrix} \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t|t-1}^{i} \\ \mathbf{y}_{t|t-1}^{i} \end{bmatrix})$$
(3.45)

où $\mathbf{x}_{t|t}^i$ est l'ensemble de la variable d'état calculé à l'étape précédente et $\hat{\mathbf{K}}_t^{\theta}$ est le gain de Kalman calculé pour l'état contenant les paramètres et la variable d'état $\mathbf{x}_{t|t-1}$.

Aux deux étapes, le gain de Kalman est calculé à partir des équations 3.15 et 3.16 présentées en 3.1.

Cette méthode d'assimilation permet d'estimer les paramètres $\boldsymbol{\theta}$ à partir de la variable d'observation \mathbf{y}_t à en passant par l'estimation de la variable d'état \mathbf{x}_t qui fait le lien entre les paramètres et la variable d'observation.

Dans le cas, de l'assimilation de données électriques pour l'estimation de paramètres hydrogéologiques, la méthode d'assimilation découplée est représentée par le schéma 3.2.



Figure 3.2 – Schéma de l'assimilation de données découplée

3.1.7 Assimilation séquentielle de mesures

Lorsqu'on est appelé à assimiler des données qui ont des erreurs de mesure à des échelles différentes, il peut être nécessaire d'effectuer une mise à l'échelle de ces mesures, afin d'éviter des erreurs numériques dues à la troncation des valeurs propres les plus faibles lors de l'utilisation d'une pseudo-inverse dans le calcul du gain de Kalman (Evensen, 2003). Afin de contourner cette difficulté, il est possible d'effectuer une assimilation séquentielle des mesures si leurs erreurs sont non-corrélées. La preuve de la validité de cette assimilation séquentielle peut être trouvée dans Evensen & Van Leeuwen (2000) et est résumée ci-dessous:

Nous cherchons à connaître la distribution *a posteriori* de la variable d'état \mathbf{x}_t connaissant des données au temps t, $\mathbf{d}_t^1 \dots \mathbf{d}_t^k$ c'est-à-dire que nous cherchons à connaître $f(\mathbf{x}_t | \mathbf{d}_t^1 \dots \mathbf{d}_t^k)$, si k données différentes sont observées au temps t

$$f(\mathbf{x}_t | \mathbf{d}_t^1 ... \mathbf{d}_t^k) \propto f(\mathbf{x}_t) \times f(\mathbf{d}_t^1 ... \mathbf{d}_t^k | \mathbf{x}_t)$$
(3.46)

Si, les erreurs de mesure, \mathbf{w}_t^k sont indépendantes les unes des autres, alors $f(\mathbf{d}_t^1|\mathbf{x}_t), f(\mathbf{d}_t^2|\mathbf{x}_t) \dots, f(\mathbf{d}_t^k|\mathbf{x}_t)$ seront indépendants, puisque

$$\mathbf{d}_t^n = \mathbf{H}_k \times \mathbf{x}_t^n + \mathbf{w}_t^n \tag{3.47}$$

et donc,

$$f(\mathbf{d}_t^n | \mathbf{x}_t^n) = f(\mathbf{d}_t^n - \mathbf{H}_n \times \mathbf{x}_t^n) = f(\mathbf{w}_t^n)$$
(3.48)

Donc, si les erreurs de mesure \mathbf{w}_t^n sont indépendantes pour $n = 1 \dots k$, alors,

$$f(\mathbf{d}_t^1 \dots \mathbf{d}_t^k | \mathbf{x}_t) = f(\mathbf{d}_t^1 | \mathbf{x}_t) \times \dots \times f(\mathbf{d}_t^k | \mathbf{x}_t)$$
(3.49)

Ainsi, il sera possible d'écrire

$$f(\mathbf{x}_t | \mathbf{d}_t^1 \dots \mathbf{d}_t^k) \propto f(\mathbf{x}_t) \times f(\mathbf{d}_t^1 | \mathbf{x}_t) \times \dots \times f(\mathbf{d}_t^k | \mathbf{x}_t)$$
(3.50)

ou,

$$f(\mathbf{x}_t | \mathbf{d}_t^1 \dots \mathbf{d}_t^k) \propto f(\mathbf{x}_t | \mathbf{d}_t^1 \dots \mathbf{d}_t^{k-1}) \times f(\mathbf{d}_t^k | \mathbf{x}_t)$$
(3.51)

En effectuant l'assimilation de différentes variables mesurées séquentiellement, les fausses corrélations pouvant apparaître entre ces variables seront aussi évitées. Toutefois, Houtekamer & Mitchell (2001) notent que cette façon d'assimiler pourrait causer d'autres problèmes, soit des des problèmes d'écrasement de la covariance dont il sera question à la section 3.1.9.

3.1.8 Conservation de la masse

Dans certains cas, il peut être désiré que les variables d'état respectent des contraintes physiques comme la conservation de la masse totale. Janjic et al. (2014) montrent que les filtres de Kalman d'ensemble et les filtres à racine carrée permettent la conservation de la masse lors de l'étape de mise à jour si l'état est défini par une variable qui conserve la masse à l'étape de prédiction. Toutefois, des modifications, telles que la reparamétrisation de l'état par anamorphose gaussienne ou l'intégration d'un processus de localisation rendront l'assimilation non-conservative. Par ailleurs, la mise en place d'un état augmenté avec les paramètres du modèle et/ou avec toute autre variable qui ne fait pas partie de l'état à masse fixe ne garantira plus la conservation de la masse dans l'étape d'assimilation. L'ajout de contraintes aux filtres de Kalman d'ensemble (ou aux filtres à racine carrée) est une façon d'imposer la conservation de la masse, lorsque celle-ci n'est pas garantie. Dans leur formulation de base, les filtres de Kalman ne peuvent pas tenir compte de contraintes sur les états; que ces contraintes soient linéaires ou non, d'égalité ou d'inégalité. Il existe plusieurs développements pour intégrer ces différentes contraintes dans la version classique des filtres de Kalman, par exemple, en projetant l'état prédit sur un espace contraint, en tronquant la distribution de probabilité obtenue par l'assimilation ou en utilisant les contraintes d'égalité en tant que mesures sans erreur (voir Simon (2010)). Cette dernière méthode a été étendue aux filtres de Kalman d'ensemble afin d'imposer la conservation de la masse ((Phale & Oliver, 2011; Pan & Wood, 2006; Li et al., 2012a)).

L'intégration de contraintes n'est pas la seule méthode qui permet d'assurer la conservation de la masse. Les filtres de Kalman d'ensemble itératifs ont été développés pour assurer la compatibilité entre les paramètres et les états lorsque les EnKF sont appliqués sur des états augmentés. En d'autres termes, les EnKF itératifs permettent de s'assurer que le lien entre les états et les paramètres respecte toujours l'équation, habituellement non-linéaire, qui devrait les relier, malgré une mise à jour linéaire. Ainsi, lorsque les équations d'écoulement et de transport sont respectées, la masse sera aussi conservée. L'idée des EnKF itératifs est d'estimer les paramètres dans un premier temps, puis d'utiliser la modélisation directe pour estimer les variables d'état à partir des nouveaux paramètres mis à jour.

Deux types d'EnKF itératifs ont été développés. Dans les *Restart-EnKF* (Gu & Oliver, 2007), les variables d'état sont mises à jour en résolvant le problème directe à partir du temps **0**, tandis que dans les *Confirming-EnKF* (Wen *et al.*, 2005; Moradkhani *et al.*, 2005), les variables d'état sont mises à jour à partir de leur estimation au dernier pas de temps. Toutefois, Song *et al.* (2014) ont

montré que le *Confirming-Enkf* ne réussissait pas toujours à résoudre les incompatibilités entre paramètres et variables d'état en particulier, lors de l'assimilation de données dans la zone vadose suivant l'équation de Richards.

L'assimilation de données se fera donc par *Restart-EnKF*, comme chez Camporese *et al.* (2015), Hendricks Franssen & Kinzelbach (2008) ou même Crestani *et al.* (2015).

3.1.9 Problèmes liés à la finitude du nombre d'ensemble

Les méthodes de type Monte Carlo d'assimilation de données sont basées sur un échantillonnage aléatoire pour représenter une distribution probabiliste. En effet, elles se basent sur la prémisse que, si un échantillon de taille infini était tiré de la distribution, les statistiques calculées sur cet échantillon représenteraient parfaitement la distribution. Dans le cas des filtres de Kalman d'ensemble, comme la distribution de l'état est supposée gaussienne, cette distribution peut être définie exclusivement par les deux premiers moments, c'est-à-dire la moyenne et la covariance. Ainsi, le calcul de la moyenne et de la covariance échantillonnale sur l'ensemble devraient être représentatifs de la distribution réelle. Le calcul de la covariance échantillonnale est particulièrement importante, étant donné que cette covariance est utilisée pour calculer le gain de Kalman. Hamill *et al.* (2001) ont montré que, plus le nombre d'ensembles diminuait, plus l'erreur relative sur la covariance échantillonnale augmentait.

Les deux problèmes principaux qui surviennent lorsque l'ensemble est trop petit pour être représentatif de la covariance réelle sont l'apparition de fausses corrélations et l'écrasement de la covariance. Les fausses corrélations peuvent causer une correction de variables de l'état a priori qui ne sont pas corrélées aux observations à l'étape de mise à jour. Cette correction sera non-fondée et arbitraire. Lorsqu'il y a écrasement de la covariance et que l'étendu de l'ensemble devient trop faible, les observations n'auront presque plus de poids dans l'étape de mise à jour. Ainsi, l'ensemble peut dévier de l'état réel, un phénomène nommé divergence (Anderson & Anderson, 1999). L'absence d'erreur du modèle peut entraînter cette divergence (Mitchell *et al.*, 2002) ainsi qu'exacerber l'écrasement de la covariance (Lorenc, 2003).

Même si les filtres sans perturbation des observations, tel que les ETKF permettent d'obtenir une matrice de covariance a posteriori plus exact (Whitaker & Hamill, 2002), les ETKF sont aussi susceptibles de produire des matrices de covariance non-représentatives de la réalité que les EnKF dû à la finitude de l'ensemble (Petrie, 2008). La grosseur de l'ensemble est donc un paramètre crucial dans l'assimilation des données par des méthodes de type Monte Carlo. Ainsi, si l'ensemble est trop petit, l'assimilation de nouvelles données pourrait aller jusqu'à détériorer l'estimation des variables d'état (Mandel *et al.*, 2011). Il sera donc important de déterminer un ensemble dont le nombre de membres est assez grand pour éviter les problèmes liés à la finitude de l'ensemble tout en permettant une assimilation des données assez rapide.

Des méthodes ont aussi été développées pour contrer les problèmes de fausses corrélations et d'écrasement de la covariance. Anderson & Anderson (1999) ont proposé de grossir la matrice de covariance de l'ensemble prédit grâce à un facteur d'inflation. Les perturbations de l'état seront donc multipliées par ce facteur qui sera un peu plus élevé que 1. Cette méthode a aussi été utilisée par Hamill *et al.* (2001) et Whitaker & Hamill (2002). Ce facteur est défini de façon heuristique.

Comme la méthode du facteur d'inflation ne contre pas l'apparition de fausses corrélations, une méthode de localisation a été développée (Houtekamer & Mitchell, 2001; Whitaker & Hamill, 2002; Hamill *et al.*, 2001). Cette méthode consiste à faire le produit matriciel d'Hadamard (aussi connu sous le nom de produit de Schur) de la matrice de covariance de l'ensemble prédit et d'une matrice de corrélations. Cette matrice de corrélations sera calculée à partir d'une fonction de corrélation avec une portée donnée. Toutes les corrélations au-delà de la portée auront une valeur nulle, permettant ainsi d'éliminer les fausses corrélations qui apparaissent à de grandes distances. La portée est aussi déterminée de façon heuristique. Si la portée choisie est trop grande, des fausse corrélations continueront d'apparaître dans la matrice de covariance localisée. Si la portée choisie est trop faible, des covariances importantes seront perdues (Petrie, 2008).

Les méthodes pour palier les problèmes liés à un nombre trop petit de réalisations de l'ensemble développés jusqu'à présent sont limitées par une sélection arbitraire de paramètres: le facteur d'inflation ou la portée de la fonction de corrélation dans le cas de la localisation. Ainsi, il a été décidé de plutôt effectuer une analyse de sensibilité au nombre de réalisations de l'ensemble afin de déterminer une taille de l'ensemble assez petite pour ne pas présenter un trop gros effort de calcul mais permettant tout de même d'améliorer l'estimation des variables d'état à partir des données observées.

3.1.10 Normes d'évaluation

Afin d'évaluer si l'assimilation des données observées permet d'améliorer l'estimation des variables d'état, deux normes seront utilisées: la moyenne de déviation absolue (Average absolute bias ou AAB) qui mesure l'exactitude de l'estimation et l'étendue moyen de l'ensemble (EE) qui mesure, non seulement l'incertitude mais permet aussi de juger s'il y a présence d'écrasement de la covariance. Ces normes sont souvent utilisées dans les méthodes utilisant des ensembles (Fortin *et al.*, 2014).

La moyenne de déviation absolue et l'étendue moyen de l'ensemble sont donnés par les équations suivantes:

$$AAB = \frac{1}{N_b} \sum_{r=1}^{N_b} \frac{1}{N_e} \sum_{i=1}^{N_e} |x_{r,i,t} - x_{r,ref}|$$
(3.52)

$$EE = \left(\frac{1}{N_b} \sum_{r=1}^{N_b} \frac{1}{N_e} \sigma_{x_{r,t}}^2\right)^{1/2}$$
(3.53)

Dans le cas à l'étude, ces statistiques sont calculées pour l'estimation du paramètre de conductivité hydraulique. Ainsi, N_b représente le nombre de points du paramètre (ou de l'état) et équivaut au nombre d'éléments du modèle puisque la conductivité hydraulique est définie aux éléments, N_e représente le nombre d'ensembles et $x_{r,i,t}$ représente la conductivité hydraulique estimée à l'élément r à l'itération t de l'ensemble i tandis que $x_{r,ref}$ représente la conductivité hydraulique de référence à l'élément r. Par ailleurs, $\sigma_{x_{t,r}}^2$ représente la variance de la variable d'état (la conductivité hydraulique dans ce cas) à l'élément r à l'itération t.

3.1.11 Assimilation de données de résistivité électrique pour la mise à jour des conductivités hydrauliques par filtres de Kalman d'ensemble

La théorie de l'assimilation de données par filtres de Kalman d'ensemble ainsi que les limites des normes d'évaluation de cette méthode ont été présentés. De plus, différentes variantes des EnKFs ont été proposés soit:

- Les filtres de Kalman d'ensemble classiques (EnKF)
- Les filtres de Kalman à racine carrée (ETKF)
- L'assimilation de données par EnKF ou ETKF découplée
- L'assimilation de données séquentielle par EnKF ou ETKF

Ces différentes méthodes seront utilisées pour estimer des paramètres hydrogéologiques (soit la conductivité hydraulique et/ou la dispersivité) à partir de données hydrogéophysiques (soit des concentrations d'un contaminant salin en puits et/ou des résistances électriques obtenues par tomographie de résistivité électrique). Les figures 3.3 et 3.4 présentent l'assimilation de données classique ou découplée respectivement avec les variables d'intérêt pour l'estimation des paramètres hydrogéologique, θ^{HG} à partir de données de concentration, C, et/ou de résistance électrique, R. Sur les deux figures, l'étape de mise à jour peut se faire, soit grâce aux équations de mise à jour des EnKFs classiques ou aux équations de mise à jour des ETKFs. Enfin, l'étape de prédiction est décrite à la section 3.2. Plus précisément, la section 3.2.1 présente le modèle de prédiction pour obtenir les concentrations, tandis que les sections 3.2.2 à 3.2.4 présentent le modèle de prédiction pour obtenir les résistances électriques, une fois les concentrations connues.



Figure 3.3 – Schéma de l'assimilation séquentielle de données de potentiel électrique pour l'estimation des paramètres hydrogéologiques



Figure 3.4 – Schéma de l'assimilation découplée de données de potentiel électrique pour l'estimation des paramètres hydrogéologiques

3.2 Modèle de prédiction

Dans la section précédente, présentant la méthode d'assimilation de données par filtres de Kalman d'ensemble, il a été mentionné que l'assimilation se faisait en deux étapes: l'étape de prédiction et l'étape d'analyse. Le schéma d'assimilation implémenté nécessite la création d'un état contenant des paramètres (la conductivité hydraulique), la variable d'état (la concentration) ainsi que les observations (les résistances mesurées), tel que présenté à la section 3.1.5. De plus, le schéma d'assimilation est basé sur le calcul de la concentration et des résistances à partir du temps 0, ce qui a été nommé *Restart-EnKF*. Ainsi, à chaque pas de temps, l'étape de prédiction a deux composantes: d'abord un modèle hydrogéologique d'écoulement et de transport est utilisé pour obtenir des concentrations à partir des conductivités hydrauliques et ensuite des lois pétrophysiques permettent de convertir les concentrations en résistivités électriques du sol et un modèle de tomographie électrique permet de calculer les valeurs de potentiel mesurées. Les différentes étapes du modèle de prédiction sont présentées à la figure 3.5.



Figure 3.5 – Schéma du modèle de prédiction

3.2.1 Modélisation hydrogéologique

La modélisation de la contamination dans l'eau souterraine se fait en deux étapes. Dans un premier temps, l'écoulement de l'eau doit être modélisé.

La loi de Darcy énonce que l'écoulement de l'eau dans les milieux poreux saturés est régi par le gradient hydraulique. Si les directions principales d'anisotropie de la conductivité hydraulique coincident avec les trois axes, x, y et z, alors la vitesse de Darcy dans la direction x, v_x , est donnée par:

$$v_x = -K_x \frac{dh}{dx} \tag{3.54}$$

où K_x est la conductivité hydraulique dans la direction x, et h est la charge hydraulique, et il en va de même pour la vitesse dans les deux autres directions.

Un bilan massique sur un volume de contrôle pour une masse conservatrice dans un milieu saturé permet d'obtenir l'équation d'écoulement:

$$\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} = S_s \frac{\partial h}{\partial t}$$
(3.55)

où S_s est le coefficient d'emmagasinement.

En utilisant la définition de la vitesse de Darcy dans chaque direction, l'équation 3.55 peut être ré-écrite en fonction des charges hydrauliques:

$$\frac{\partial}{\partial x}(K_x\frac{\partial h}{\partial x}) + \frac{\partial}{\partial y}(K_y\frac{\partial h}{\partial y}) + \frac{\partial}{\partial z}(K_z\frac{\partial h}{\partial z}) = S_s\frac{\partial h}{\partial t}$$
(3.56)

Dans le cas, d'un écoulement en régime permanent, $S_s \frac{\partial h}{\partial t} = 0.$

La résolution de l'équation d'écoulement permet de calculer la vitesse interstitielle moyenne qui régit la portion advective de l'équation de transport. La vitesse interstitielle moyenne, \bar{v} , est liée à la vitesse de Darcy, v, grâce à la porosité effective, n_e .

$$\bar{v} = \frac{v}{n_e} \tag{3.57}$$

L'advection est un des trois mécanismes entrant en jeu dans le transport de contaminants.Le deuxième mécanisme est la dispersion hydrodynamique qui regroupe la diffusion moléculaire et la dispersion mécanique. Finalement, les contaminants réactifs peuvent subir une dégradation biochimique ou radioactive ou réagir avec le sol ou l'eau. Pour les contaminants conservatifs (ou nonréactifs), ce dernier mécanisme n'intervient pas. Dans le cas où le contaminant est conservatif, le taux de variation de la masse du contaminant à l'intérieur d'un volume de contrôle est donc égal à la différence entre le flux de contaminant entrant dans le volume moins le flux sortant. Ce bilan de masse et la définition des mécanismes d'advection et de dispersion permet d'obtenir l'équation d'advection-dispersion:

$$\left[\frac{\partial}{\partial x}\left(D_x\frac{\partial C}{\partial x}\right) + \frac{\partial}{\partial y}\left(D_y\frac{\partial C}{\partial y}\right) + \frac{\partial}{\partial z}\left(D_z\frac{\partial C}{\partial z}\right)\right] - \left[\frac{\partial}{\partial x}(\bar{v}_xC) + \frac{\partial}{\partial y}(\bar{v}_yC) + \frac{\partial}{\partial z}(\bar{v}_zC)\right] = \frac{\partial C}{\partial t} \quad (3.58)$$

où C représente la concentration. La première partie du membre de gauche de l'équation représente la dispersion et la seconde partie représente l'advection. Le coefficient de dispersion dans la direction x, D_x est donné par:

$$D_x = \alpha \bar{v}_x + D * \tag{3.59}$$

où $\alpha_x \bar{v}_x$ représente la dispersion mécanique avec la dispoersivité dynamique, α_x , et D^* , le coefficient de diffusion effectif.

Lorsque la conductivité hydraulique du milieu est hétérogène, les équations différentielles 3.56 et 3.58 ne peuvent pas être résolues analytiquement. Des modèles numériques par différences finies, par volumes finis ou par éléments finis ont donc été développés. Le logiciel de modélisation **FEFLOW** (DHI-WASY, 2012) permet de résoudre les équations d'écoulement 3.56 et de transport 3.58 par éléments finis, plus particulièrement par la méthode de Galerkin. Ce logiciel a été utilisé dans les études de cas présentées au chapitre suivant. Une discussion approfondie sur la modélisation numérique des équations d'écoulement et de transport avec le logiciel **FEFLOW** peut être trouvée dans Diersch (2014).

3.2.2 Modèle pétrophysique qui lie la contamination à la résistivité électrique

Il a été proposé de faire un suivi de la contamination, et par le biais de ce suivi, une estimation des propriétés hydrogéologiques à partir de mesures électriques. Dans la section 2.2 plusieurs études hydrogéologiques utilisant des mesures électriques avec succès ont été présentées. Ces études sont justifiées par la relation existante entre la conductivité électrique de l'eau et la conductivité du sol. Lorsque la conductivité de surface est négligeable comme dans le cas des aquifères de sable ou de gravier sans argile, cette relation est donnée par la loi d'Archie (Archie, 1942).

$$\sigma_{sol} = \frac{1}{a} \sigma_w n^m S_w^n \tag{3.60}$$

où σ_{sol} est la conductivité électrique d'un sol saturé, a est le facteur de tortuosité, σ_w est la conductivité du fluide dans le sol qui peut varier selon la concentration en contaminant, n est la

porosité du sol, m est un exposant exprimant le degré de cimentation du sol, S_w est le degré de saturation du sol et n est un exposant lié à la saturation. Le ratio $\frac{a}{\phi^m}$ est défini comme le facteur de formation, F.

De plus, une relation empirique permet de lier la conductivité du fluide dans le sol au niveau de contamination (Lesmes & Friedman, 2005):

$$\sigma_w = k_e \times TDS \tag{3.61}$$

où TDS est l'acronyme de total dissolved solids ou total des solides dissous en mg/L et k_e est une constante qui varie entre $1, 2 \times 10^{-4}$ et $2, 0 \times 10^{-4}$ selon la température et la composition ionique.

Plusieurs auteurs ont d'ailleurs noté des relations linéaires entre la concentration en ions dissous et la conductivité de l'eau sur des sites spécifiques e.g. Kemna *et al.* (2002) (ions de bromure) ou Singha & Gorelick (2005) (NaCl). Ainsi, si la concentration en ions dissous et le degré de saturation d'un sol sont connus, il est possible de calculer la conductivité électrique du sol sous les conditions de respect de la loi d'Archie.

3.2.3 Modélisation électrique

Habituellement, la résistivité du sol ne sera pas mesurée directement. Les mesures de tomographie par résistivité électrique permettent plutôt d'obtenir des séries de mesures de différence de potentiel en réponse à un courant électrique émis dans le sol.

La distribution du potentiel électrique dans un milieu conducteur à travers lequel passe un courant constant est décrit par l'équation de Poisson:

$$\nabla \cdot (\sigma_{sol} \nabla V) = -I(\delta(\vec{r} - \vec{r}_{+}) - \delta(\vec{r} - \vec{r}_{-}))$$
(3.62)

où I est la courant du dipôle formé par l'électrode positive positionnée en \vec{r}_+ et l'électrode négative positionnée en \vec{r}_- , V est le potentiel électrique et σ_{sol} représente la conductivité électrique du sol qui peut être hétérogène. Si la conductivité électrique du sol est hétérogène, la solution de l'équation 3.62 devra être obtenue numériquement, soit par éléments, différences, ou volumes finis. Les conditions frontières des modèles numériques sont définies par défaut pour représenter le sol. Le modèle résultant est communément appelé demi-espace et est limité par une condition de Neumann de flux-nul à la surface pour représenter l'interface sol-air, qui est un isolant. Les conditions aux autres frontières sont plus arbitraires. En effet, du remplissage est habituellement effectué afin d'éloigner ces frontières le plus possible de la zone d'étude et de limiter les effets de bords dû à un grille finie qui tente de représenter un milieu infini. Tout en utilisant du remplissage, Pidlisecky & Knight (2008) définissent des conditions de Neumann à toutes les frontières tandis que Gunther (2004) définit plutôt des conditions de Robin à ces frontières. Dey & Morrison (1979) ont montré que ces conditions de Robin permettent de limiter les effets de bords dû à un grille finie qui tente de représenter un milieu infini. Ce sont ces dernières conditions de Robin qui seront définies aux trois frontières du modèle numérique qui ne représentent pas l'interface sol-air.

Lorsque l'équation 3.62 doit être résolue en deux dimensions, la conductivité est considérée constante et infinie dans la troisième dimension, ce qui permet de réduire la conductivité à deux dimensions. Toutefois, la source et le potentiel restent des fonctions en trois dimensions. Une transformation dans le domaine de Fourier permet de réduire le potentiel électrique à deux dimensions (Dey & Morrison, 1979). Le \tilde{V} est le potentiel transformé grâce à l'équation 3.63

$$\tilde{V}(x,y,k_z) = 2\int_0^\infty (x,y,z)\cos(k_z z)dz$$
(3.63)

où k_z est le nombre d'onde.

Dans le domaine de Fourier, l'équation 3.62 devient:

$$\frac{\partial}{\partial x} \left(\sigma_{sol} \frac{\partial \tilde{V}}{\partial x}\right) + \frac{\partial}{\partial y} \left(\sigma_{sol} \frac{\partial \tilde{V}}{\partial y}\right) - k_z^2 \sigma_{sol} \tilde{V} = -I(\delta(\vec{r} - \vec{r}_+) - \delta(\vec{r} - \vec{r}_-))$$
(3.64)

Les conditions frontières sont aussi transformées dans le domaine de Fourier.

L'équation 3.64 est ensuite résolue par différences finies. Une discussion sur la discrétisation en différences finies peut être trouvée dans Gunther (2004). Une fois, l'équation 3.64 résolue pour \tilde{V} , la transformée de Fourier inverse est utilisée pour retrouver le potentiel, V. Cette façon de résoudre l'équation de Poisson en supposant la conductivité uniforme et infinie dans la troisième dimension et en passant par le domaine de Fourier est communément appelée, modèle 2.5D.

La résolution de l'équation de Poisson permet d'obtenir des valeurs de potentiel électrique qui sont mesurées lors d'un levé de tomographie électrique. La résistance, R, peut être calculée à partir du potentiel, V et du courant émis, I.

$$R = \frac{\Delta V}{I} \tag{3.65}$$

où ΔV est la différence de potentiel mesuré entre les électrodes.

Ces résistances sont souvent transformées en résistivité apparente (ρ_a) qui équivaut à la résistivité qui serait observée pour un même dispositif dans un sol homogène.

$$\rho_a = R \times k_g \tag{3.66}$$

où k_g est un facteur géométrique.

La figure 3.6 présente une configuration possible d'un dispositif à quatre électrodes avec un courant I injecté à l'aide du dipôle AB et un potentiel de courant, δV , mesuré entre les électrodes M et N. Le facteur géométrique, k_g est donné par l'équation 3.67.

$$k_g = \frac{2\pi}{\frac{1}{AM} - \frac{1}{AN} - \frac{1}{BM} + \frac{1}{BN}}$$
(3.67)



Figure 3.6 – Dispositif de tomographie électrique à quatre électrodes

La configuration du dispositif présenté à la figure 3.6 est de type Wenner où a représente l'espacement entre les dipôles. Plusieurs autres configurations existent. La sensibilité des mesures varie selon les configurations. Une discussion sur la sensibilité des différences configurations peut être trouvé dans M.H.Loke (1999).

3.2.4 Levé de résistivité électrique dans le temps

Un cas particulier des mesures de résistivité électrique est le levé de résistivité électrique dans le temps (time-lapse ERT). Ces levés servent à observer une variation dans la conductivité électrique totale du sol et sont donc très utiles en hydrogéologie afin de suivre l'évolution d'un traceur ou d'un contaminant par exemple. Comme il est rare que des propriétés physique du sol telles que sa porosité ou sa cimentation varient, les changements en conductivité électrique observés par tomographie dans le temps seront habituellement dus soit à un changement de degré de saturation, soit à un changement de la conductivité du fluide. Dans la présente étude, il sera supposé que les variations de conductivité électrique dans le temps sont dues à des variations de concentration d'un contaminant puisque le degré de saturation des sols à l'étude sera considéré constant. Ainsi, il sera possible de lier la conductivité électrique du sol au temps t à la conductivité électrique du sol au moment initial (avant contamination) et la concentration du contaminant à ce moment. Étant donné que les propriétés physiques du sol sont considérées constantes, la conductivité totale du sol au temps t est résumé par la formule:

$$\sigma_{sol}^t = F \times \sigma_w^t \tag{3.68}$$

où le facteur de formation F est une constante pour un pixel donné dans la modélisation d'un sol donné. Par ailleurs, comme la conductivité de l'eau peut être décrite comme la somme des concentrations des ions dans l'eau multipliés par les valeurs de conductivité (nombre de valence, degré de dissociation, mobilité de Valence), associés à chacun de ces ions (Lesmes & Friedman, 2005), la conductivité de l'eau au temps t est calculée à partir de la somme (Vanderborght *et al.*, 2005):

$$\sigma_w^t = \sigma_w^0 + k_e \times TDS \tag{3.69}$$

Cette équation est équivalente à l'équation $\sigma_{bulk}^t = \sigma_{bulk}^0 + constante * concentration$ de Vanderborght *et al.* (2005) ou Pollock & Cirpka (2008), pour un facteur de formation constant.

Il est aussi à noter que la température peut faire varier la conductivité de l'eau. Celle-ci est donc supposée stable dans le temps. Des valeurs typiques de conductivité de l'eau douce sont d'environ $10^{-2}S/m$ à $25^{\circ}C$.

Étant donné que l'intérêt des levés de résistivité électrique dans le temps réside plus dans la variation de la résistivité électrique que dans la valeur absolue de résistivité. Différentes façons d'évaluer cette variation ont été proposées. Certains auteurs commencent par effectuer l'inversion sur les valeurs de résistivité apparentes observées, puis faire la différence entre les résistivités inversées au temps t, et les résistivités inversées au temps 0, par exemple, Ramirez *et al.* (1993), LaBrecque *et al.* (1996), Singha & Gorelick (2005) ou Hayley *et al.* (2009). Toutefois, cette méthode engendre une diminution de la qualité due à l'addition des erreurs liées à l'inversion.

Une solution permettant d'éviter ce genre de propagation de l'erreur est d'appliquer la méthode proposée par Daily & Owen (1991), soit effectuer l'inversion sur le ratio des résistivités. Cette méthode présente plusieurs avantages, en particulier, d'atténuer, si ce n'est d'éliminer les effets dus à des garnissages de contact trop conducteurs ou trop résistants entourant les électrodes ou de faire ressortir des variations subtiles selon Daily *et al.* (2004). Cette méthode permet aussi d'éviter d'avoir à tenir compte des effets de conditions frontières tridimensionnelles dans le cas d'une inversion bidimensionnelle (Ramirez *et al.*, 1996) ou de l'utilisation des grilles d'élément finies différentes pour le modèle directe et l'inversion (Daily & Owen, 1991). Cette méthode a donc été utilisée dans l'analyse de données de résistivité électrique pour monitorer un essai de traceur (par exemple, Slater *et al.* (2000), Cassiani *et al.* (2006), Cassiani *et al.* (2009) ou Perri *et al.* (2012)) ou pour monitorer la saturation d'un sol (Binley *et al.*, 2002; Ramirez & Daily, 2001).

Il est proposé d'évaluer le ratio de conductivités électriques dans le modèle directe. Ainsi, la valeur de conductivité électrique utilisée dans le modèle directe est donné par:

$$\sigma = \frac{\sigma_{sol}^t}{\sigma_{sol}^0} \times \sigma_{homo} \tag{3.70}$$

$$= (1 + k_e \times TDS) \times \sigma_{homo} \tag{3.71}$$

Si le ratio $\frac{\sigma_{sol}^t}{\sigma_{sol}^0}$ est faible ou si la résistivité initiale du sol est plutôt homogène, alors il est possible de faire l'approximation suivante où l est la fonction de modélisation directe

$$l(\frac{\sigma_{sol}^{t}}{\sigma_{sol}^{0}} \times \sigma_{homo}) \cong \frac{l(\sigma_{sol}^{t})}{l(\sigma_{sol}^{0})} \times l(\sigma_{homo})$$
(3.72)

Et si le sol est un demi-espace isotrope et homogène ($\sigma_{sol}^0 = constante$), cette approximation devient une égalité. Si le ratio est trop élevé pour que cette approximation soit valide, Daily *et al.* (2004) fait la supposition que le ratio permettra tout de même de révéler les changements dans la résistivité électrique dus à la propagation du contaminant.

Dans le cadre de l'assimilation de données, cette approximation sera utilisée pour comparer les données prédites aux données mesurées. Ainsi, le ratio de résistances prédites au temps t sera donné par:

$$RR_{predit} = l(\sigma) = l((1 + k_e \times TDS_t) \times \sigma_{homo})$$
(3.73)

ou, selon la notation des filtres de Kalman d'ensemble utilisée jusqu'à présente.

$$RR_{t|t-1} = l(\sigma) = l((1 + k_e \times C_{t|t-1}) \times \sigma_{homo})$$
(3.74)

où $l(\mathbf{x})$ est le modèle numérique qui permet de calculer le potentiel aux électrodes à partir d'un modèle de conductivité électrique du sol, σ et ensuite, de transformer ce potentiel en résistance, RR_{predit} ou $RR_{t|t-1}$. La concentration en sel dissous au temps t, TDS_t ou $C_{t|t-1}$, est obtenue à partir du modèle hydrogéologique et σ_{homo} est une valeur de résistivité du sol qui peut être choisie de façon arbitraire.

Ce ratio de résistance prédit, $RR_{t|t-1}$ est comparé au ratio des résistances mesurées, RR_{obs} .

$$RR_{obs} = \frac{R_t^{obs}}{R_0^{obs}} * R_{homo}$$
(3.75)

où R_t^{obs} est la résistance observée aux électrodes au temps, t, R_0^{obs} est la résistance mesurée avant le début de l'infiltration du contaminant et R_{homo} est la valeur de résistance théorique d'un sol avec une conductivité homogène, σ_{homo} .

Ainsi, l'assimilation des données de résistance électrique par EnKF (ou ETKF) peut être résumée par la figure suivante:

Figure 3.7 – Assimilation du ratio des résistances


Chapitre 4

Cas synthétique 2D

La méthode proposée d'assimilation de données de résistances électriques et de concentrations en NaCl mesurées sur quelques forages pour estimer la conductivité hydraulique a été testée sur un cas synthétique bidimensionnel. Plusieurs scénarios de l'assimilation de données sont proposés et validés en les comparant au modèle de référence et en les comparant entre eux. L'assimilation des données de résistances électriques et de concentrations mesurées en forages a donc été testée séparément et conjointement pour évaluer l'apport de chaque type de donnée dans la calibration de la conductivité hydraulique. De plus l'algorithme de mise à jour des ETKF présenté à la section 3.1.4 a été comparée à l'algorithme classique des EnKFs. Enfin, la méthodologie découplée (section 3.1.6) a été comparée à la méthodologie classique d'augmentation de l'état (section 3.1.5) pour l'assimilation des données de résistance. L'assimilation séquentielle présentée à la section 3.1.7 a aussi été testée pour l'assimilation conjointe des données de résistance et de concentration en forage.

4.1 Définition du cas synthétique

Le modèle synthétique sur lequel les tests ont été effectuées est un modèle bidimensionnel de 20 mètres de profondeur et 100 mètres de longueur dans lequel un panache de NaCl se propage. La figure 4.1 présente le champ de conductivités hydrauliques du cas synthétique. Ce champ a été généré par simulation par bandes tournantes avec le logiciel *Isatis* (Isatis (2015)). Comme, il est généralement accepté qu'à l'intérieur d'un hydrofaciès, les conductivités hydrauliques suivent une distribution de type log-normale (Meerschaert *et al.*, 2013), les simulations ont été effectuées sur le logarithme de la conductivité hydraulique. Les propriétés géostatistiques du champ simulé sont présentées au tableau 4.1.

Moyenne (m/s)	-5.6 $(2, 51 \times 10^{-6} \text{ m/s})$		
Variogramme	Sphérique		
Variance	0,1848		
Portée horizontale (m)	22		
Portée verticale (m)	5		
Effet de pépite	0,01848		

Tableau 4.1 – Paramètres géostatistiques de $log_{10}(K_h)$



Figure 4.1 – Champ des conductivités hydrauliques $(log_{10}K_h)$

La modélisation d'écoulement et de transport a été calculée par éléments finis avec le simulateur hydrogéologique *Feflow*. Le modèle représente un écoulement en régime permanent bidimensionnel en coupe de 5800 éléments. La grille est formée d'éléments quadrilatéraux (figure 4.2). La taille des éléments est fixée à 0,50 mètres dans la direction horizontale. Dans la direction verticale la taille des éléments varie de 0,25 mètres en surface à 1 mètre en profondeur. L'aquifère modélisé est confiné et saturé. La recharge a été fixée à 0,00108 m/j grâce à une condition de Neumann à la frontière horizontale supérieure. La frontière horizontale inférieure ainsi que la frontière en amont (à 0 mètres) sont des frontières imperméables, donc avec une condition de Neumann de flux nul. Une condition frontière de Dirichlet a été fixée le long de la frontière verticale en aval (une frontière de flux sortant) avec une charge hydraulique fixe de 41 mètres. Enfin, la conductivité hydraulique verticale est fixée à 1 % de la conductivité hydraulique horizontale.

Paramètre	Valeur
Porosité	0,30
D*	$10^{-9}m^2/s$
$lpha_L$	$5 \mathrm{m}$
$lpha_t$	$0,5 \mathrm{m}$
Nombre de Péclet maximal	0,23
Pas de temps initial	0,001 jour

Tableau 4.2 – Paramètres du modèle de transport



Figure 4.2 – Grille d'éléments finis utilisée pour la modélisation hydrogéologique

La contamination est représentée par une frontière de Dirichlet de concentration constante à 1000 mg/L à la surface du modèle entre 20 et 30 mètres. Les éléments ont été raffinés autour de cette frontière. Toutes les autres frontières dans la modélisation de transport sont des frontières de flux nul. Toutefois des contraintes sont appliquées à la frontière verticale en aval afin de permettre au contaminant de sortir du modèle. La discrétisation temporelle est basée sur un schéma automatique de contrôle des pas de temps par prédiction-correction. Cette méthodologie permet de déterminer la longueur du pas de temps en se basant sur la variation de la concentration en fonction du pas de temps. Elle est expliquée plus en profondeur dans Diersch (2009). Les paramètres du modèle de transport sont résumés au tableau 4.2.

La résistivité électrique est ensuite modélisée sur un domaine bidimensionnel par différences finies grâce au logiciel *ERT2d* (Bouchedda (2010)). La conductivité électrique est calculée à partir des données de concentration grâce à la loi d'Archie en supposant un facteur de formation homogène ainsi que l'absence d'argile afin que la loi d'Archie soit applicable sans avoir à tenir compte de la conductivité de surface. La conductivité totale du sol saturé avec de l'eau non-contaminée est fixée à $3,9811 \times 10^{-3}$ S/m et la conductivité de surface est considérée négligeable. La conductivité de l'eau non-contaminée est fixée à $1 \times 10^{-2}S/m$, ce qui implique un facteur de formation d'environ 2,5. La conductivité de l'eau étant associée aux ions présents dans l'eau, elle est proportionelle au total des sels dissous avec un facteur de proportionalité de $1, 5 \times 10^{-4}$ si les TDS sont en mg/L et la conductivité est en S/m (Butler, 2005). La conductivité électrique étant aussi fonction de la température, la température est présumée corrigée et constante.

Des puits synthétiques d'observation de la concentration ont été placés à 30 mètres et 70 mètres sur toute la hauteur du modèle. Les points d'observation à ces puits sont représentés par les traits pointillés aux figures 4.3. Chaque point correspond à un point d'échantillonnage pour un total de 60 points d'échantillonnage (30 par puits). La variance de l'erreur de mesure pour la concentration d'un ion dissous dans l'eau de puits est habituellement fixé à la limite de détection. Par contre, lorsque la concentration de l'ion dans l'eau est trop élevée, l'échantillon doit être dilué. Le cas échéant, l'erreur de mesure augmente dû au processus de dilution (Bordeleau, communication personelle, 2014). Afin de modéliser ce type d'erreur, la variance de l'erreur est calculée à partir de la relation linéaire W = 0, 2 + 0, 1 * concentration, puisque la limite de détection est fixée à 0,2 mg/L.

Le levé électrique est un protocole d'acquisition dipôle-dipôle qui couvre toute la section, c'est-àdire de 0 à 100 mètres. L'espacement entre les deux électrodes de courant ou les deux électrodes de potentiel, *a*, varie entre 1 et 7 mètres; tandis que le multiple, *n* d'espacement entre les dipôles varie entre 1 et 6, afin d'obtenir une profondeur d'investigation médiane d'environ 12 mètres (Loke, 2004). Un total de 2973 mesures de différence de potentiel sont obtenues avec ce protocole d'acquisition. Les pseudo-sections ainsi obtenues sont présentées aux figures 4.3c, 182 jours après le début de la contamination et 4.3d après 1095 jours. L'échelle de la résistivité apparente est différente sur les deux pseudo-sections puisque la résistivité apparente diminue de façon significative avec l'avancée du panache. L'écart-type de l'erreur de mesure est de 5% de la valeur de la résistance mesurée.

Comme présenté à la section 3.2.4, la méthode des ratios de résistances est utilisée pour assimiler les données électriques dans le temps. Dans le cas synthétique présenté, la résistivité électrique initiale du sol (avant la contamination) est considérée homogène. La méthodologie des ratios de résistance devrait donc être appropriée. Le ratio des résistances observées au temps, t, est donné par:

$$RR_{obs} = \frac{R_t^{obs}}{R_0^{obs}} * R_{homogène}$$
(4.1)

où R_t^{obs} est la résistance observée aux électrodes au temps, t, R_0^{obs} est la résistance mesurée avant le début de l'infiltration du contaminant et $R_{homogène}$ est la valeur de résistance théorique d'un sol avec une conductivité homogène, $\sigma_{homogène}$.

Une résistivité du sol homogène de 100 $\Omega \cdot m$ a été choisie arbitrairement. Comme un ratio est utilisé, l'erreur de mesure du ratio doit être caclulé en conséquence. Cette nouvelle erreur de mesure est estimée à partir de la formule suivante, en supposant que les deux erreurs de mesure sont non-corrélées

$$\frac{Var(RR_{obs})}{(E[RR_{obs}])^2} = \frac{Var(R_t)}{(E[R_t])^2} + \frac{Var(R_0)}{E[R_0]^2}$$
(4.2)

où E[X] représente l'espérance mathématique de X.



Figure 4.3 – Valeurs du cas synthétique (a) Concentration à 182 jours (b) Concentration à 1095 jours (c) Pseudo-section de la résistivité électrique à 182 jours (d) Pseudo-section de la résistivité électrique à 1095 jours

4.2 Ensemble initial

Idéalement, l'ensemble initial de conductivités hydrauliques devrait être créé afin de représenter les caractéristiques de l'erreur de l'estimé initial (Evensen, 2003). Dans les applications en hydrogéologie, ceci implique généralement de créer un ensemble de simulations géostatistiques qui représentent les deux premiers moments du paramètre à l'étude à partir du variogramme et de la moyenne de ce paramètre. Lorsque plusieurs hydrofaciès sont présents, il est possible d'intégrer des statistiques d'ordre supérieur, à partir de simulations multi-points pour recréer les hydrofaciès. Il sera ensuite possible de simuler la conductivité hydraulique à partir de simulations gaussiennes à l'intérieur de chaque hydrofaciès (Blouin *et al.*, 2013). Comme le cas à l'étude ne représente qu'un seul hydrofaciès, l'ensemble initial du logarithme de la conductivité hydraulique a été créé à partir de simulations géostatistiques non-conditionnelles par bandes tournantes, permettant de recréer le variogramme sphérique du cas de référence décrit précédemment (voir tableau 4.1). Les simulations des conductivités hydrauliques ont aussi la même moyenne que le cas de référence. La moyenne et la variance d'un échantillon de 1000 simulations ainsi qu'une des réalisations sont présentées à la figure suivante (figure 4.4). Tel que prévu la moyenne des 1000 simulations montre très peu de variation de conductivité hydraulique. Les variations spatiales attendues sont présentes lorsqu'une seule réalisation est présentée, comme à la figure 4.4c.



Figure 4.4 – Simulations initiales (a) Moyenne de 1000 simulations de $log_{10}(K_h)$ (b) Variance de 1000 simulations de $log_{10}(K_h)$ (c) Exemple de simulation de $log_{10}(K_h)$

Étant donné que les paramètres géostatistiques (notamment la portée) de la distribution de la conductivité hydraulique sont difficilement connus dans un cas réel, un deuxième cas de figure a été mis à l'étude. Comme, une mauvaise représentation des paramètres géostatistiques, et en particulier de la portée, dans l'ensemble initial est difficile à corriger (Camporese *et al.*, 2011; Jafarpour & Tarrahi, 2011), une représentation probabiliste de ces paramètres peut améliorer l'estimation (Jafarpour & Tarrahi, 2011). Donc, dans ce deuxième cas, les simulations initiales sont créées avec une portée, une variance et une moyenne qui varient pour représenter l'incertitude sur ces paramètres. À partir des cinq valeurs de moyenne, des trois valeurs de variance et des cinq valeurs de portée horizontale présentées au tableau 4.3, les 75 combinaisons possibles d'ensembles comprenant une moyenne, une variance et une portée ont été mis en place. La portée verticale est souvent bien connue grâce aux données de forage (Paradis *et al.*, 2014). Elle reste donc fixée à 4 mètres. L'effet de pépite est aussi fixe, à 10% de la variance.

Paramètres géostatisiques pour simuler $log_{10}(K_h)$	Valeurs				
Moyenne (m/s)	-6,6	-6,1	-5,6	-5,1	-4,6
Variance	0,1183	0,1848	0,2661		
Portée horizontale (m)	18	20	22	24	26

Tableau 4.3 – Simulations initiales avec paramètres géostatistiques variables

La moyenne et la variance de 1050 simulations effectuées par bandes tournantes avec ces paramètres sont présentées à la figure 4.5. Ces simulations ont été choisies afin d'avoir un nombre égal de chaque combinaison de moyenne, variance et portée horizontale dans l'ensemble des simulations. Le nombre de simulations devait donc être un multiple de 75 (nombre de combinaisons possibles à partir des valeurs de moyennes, variance et portée décrites au tableau 4.3).Tel qu'attendu, la variance de ces 1050 simulations est beaucoup plus élevée que la variance d'un nombre semblable de simulations (1000) avec paramètres géostatistiques fixes.



Figure 4.5 – Simulations initiales (a) Moyenne de 1050 simulations de $log_{10}(K_h)$ (b) Variance de 1050 simulations de $log_{10}(K_h)$

Par ailleurs, les mêmes modèles d'écoulement et de transport ainsi que de tomographie électrique sont utilisés pour propager l'ensemble initial dans le temps.

4.3 Assimilation des données de concentrations aux puits

Comme il a été mentionné à la section 2.1.2, les filtres de Kalman d'ensemble ont déjà été utilisés pour mettre à jour différentes propriétés hydrogéologiques des sols, telles que la transmissivité, la conductivité hydraulique, la porosité ou même le taux de recharge, habituellement à partir soit de la concentration, lorsqu'un contaminant était présent, soit des charges hydrauliques en présence d'écoulement transitoire, ou des deux (par exemple, Schöniger *et al.* (2012), Zhou *et al.* (2011), Li *et al.* (2012b)). Dans cette section, les données de concentration aux puits seront assimilées par la méthode d'assimilation conjointe des états et des paramètres présenté à la section 3.1.5 afin de mettre à jour la distribution spatiale du logarithme de la conductivité hydraulique. Ainsi, le vecteur d'état à mettre à jour devient: $\mathbf{x} = \begin{bmatrix} log_{10}(K_h) \\ Concentration \end{bmatrix}$.

La représentativité de la distribution par l'ensemble s'améliore lorsque le nombre de membres de l'ensemble augmente. Toutefois, augmenter le nombre de membre de l'ensembles implique un coût computationnel élevé étant donné que la modélisation de transport et d'écoulement doit être effectuée sur chaque membre de l'ensemble, tandis qu'un ensemble trop petit peut engendrer les problèmes énumérés à la section 3.1.9. L'assimilation de données a donc été testée sur des ensembles de tailles différentes, et ceci pour le cas de figure où les paramètres géostatistiques de la conductivité hydraulique sont considérés connus, ainsi que pour le cas de figure où ces paramètres géostatistiques varient.

4.3.1 Paramètres géostatistiques connus

Les figures 4.6a à 4.6b présentent les normes d'évaluation pour l'assimilation de données de concentrations aux puits par EnKF et par ETKF, lorsque l'ensemble initial de conductivités hydrauliques est simulé avec les mêmes paramètres géostatiques que le cas synthétique présentés au tableau 4.1.



Figure 4.6 – Résultats pour la simulation avec paramètres connus (a) Moyenne de déviation absolue (AAB) de $log_{10}(K_h)$ estimé en fonction du pas de temps (b) Étendue de l'ensemble (EE) de $log_{10}(K_h)$ estimé en fonction du pas de temps pour EnKF et ETKF avec différents nombres d'ensembles

Dans ce premier cas d'étude, l'estimation est meilleure(AAB plus faible) pour des ensembles de taille plus grande (200 ou 1000 membres) que pour des ensembles de taille plus faible (20 ou 100 membres). Plusieurs auteurs ont observé une amélioration de l'estimation avec une augmentation de la taille de l'ensemble, dont Reichle *et al.* (2002a) et Yin *et al.* (2015) pour l'assimilation de données d'humidité du sol, Mitchell *et al.* (2002) pour l'assimilation de données atmosphériques et Hendricks Franssen & Kinzelbach (2008) pour l'assimilation de charges hydrauliques. Dans tous les cas, les auteurs réussissent tout de même à obtenir des bons résultats avec des ensembles de taille assez faible (12 pour Yin *et al.* (2015) et 64 pour Houtekamer & Mitchell (1998)) malgré des états de dimension élevée (6000 pour l'assimilation de données d'humidité du sol de Yin *et al.* (2015) et 660 000 pour l'assimilation de données atmosphériques à partir de concentrations aux puits, un ensemble ayant un minimum de 100 membres semble nécessaire pour améliorer l'estimation. En effet, une détérioration de la solution est notée pour un ensemble de 20 membres.

La détérioration de la solution pour un ensemble de 20 membres réfléte des fausses corrélations entre les concentrations aux puits et les conductivités hydrauliques. À partir du troisième pas de temps, l'assimilation avec un ensemble de 100 réalisations semble aussi diverger (AAB augmente) ce qui semble aussi être dû à l'apparition de fausses corrélations. Ces fausses corrélations ont été observées en comparant les matrices de covariance avec des ensembles de tailles différentes. Plus la taille de l'ensemble augmente, plus la covariance est considérée représentative.

Ainsi, contrairement à la mesure d'erreur de l'estimation, l'étendue de l'ensemble augmente avec la taille afin de mieux représenter la variance. Cette augmentation de l'étendue de l'ensemble en fonction de sa taille peut être observée à la figure 4.6a Une étendue plus grande est favorable. En effet, si la covariance de l'ensemble est sous-estimée (l'étendue de l'ensemble est plus faible qu'elle ne devrait l'être en réalité), le gain de Kalman mettra plus de poids sur l'a priori par rapport aux observations. La correction de l'a priori par les observations sera donc plus faible qu'elle ne devrait l'être. Ce phénomène a été nommé dégénérescence du filtre (ou *Filter Inbreeding*) par Houtekamer & Mitchell (1998).



Figure 4.7 – Moyenne du $log_{10}(K_h)$ (a) Cas de référence (b) Pour un ensemble de 200 réalisations après un pas de temps par EnKF (c) Pour un ensemble de 200 réalisations après 6 pas de temps par EnKF (d) Pour un ensemble de 200 réalisations après un pas de temps par ETKF (e) Pour un ensemble de 200 réalisations après 6 pas de temps par ETKF (f) Pour un ensemble de 1000 réalisations après un pas de temps par EnKF (g) Pour un ensemble de 1000 réalisations après 6 pas de temps par EnKF

Comme noté préalablement, la mesure d'erreur d'estimation (AAB) pour l'assimilation avec un ensemble de 200 membres est plus faible que pour l'assimilation avec 1000 membres. Malgré cette détérioration, il est possible d'observer aux figures 4.7b, 4.7c, 4.7f et 4.7g que les mêmes structures apparaissent tout au long de l'assimilation de données pour des ensembles de 200 ou 1000 membres. Plusieurs facteurs peuvent rentrer en jeu pour expliquer une meilleure estimation avec un ensemble plus petit. Dans le cas présenté ici, le gain de Kalman calculé à partir de l'ensemble avec 200 membres permet des poids plus élevés sur les observations que celui calculé à partir de l'ensemble avec 1000 membres (les gains de Kalman au premier pas de temps sont présentés à l'annexe A). Si la covariance calculée à partir de l'ensemble de 200 membres ne présente pas trop de fausses corrélations, il peut être avantageux d'appliquer une plus grande correction à partir des observations, ce qui semble être le cas ici.



Figure 4.8 – Variance du $log_{10}(K_h)$ (a) Pour un ensemble de 200 réalisations après un pas de temps par EnKF (b) Pour un ensemble de 200 réalisations après 6 pas de temps par EnKF (c) Pour un ensemble de 200 réalisations après un pas de temps par ETKF (d) Pour un ensemble de 200 réalisations après 6 pas de temps par ETKF (e) Pour un ensemble de 1000 réalisations après un pas de temps par EnKF (f) Pour un ensemble de 1000 réalisations après 6 pas de temps par EnKF

Les figures 4.8a à 4.8f montrent que la variance de l'ensemble diminue avec l'ajout de pas de temps à l'assimilation et pour un nombre d'ensemble plus faible. Ceci avait aussi été observé pour l'étendue de l'ensemble. Au premier pas de temps, la variance pour des ensembles de 200 ou 1000 réalisations est plus élevée dans le bas et aux bords du domaine puisque la contamination est plutôt circonscrite dans la portion supérieure et loin des bords. Cette variance plus élevée en périphérie du modèle disparaît au dernier pas d'assimilation avec un ensemble de 200 membres, mais est toujours présente au dernier pas d'assimilation avec 1000 membres. Étant donné que la contamination semble atteindre l'ensemble du domaine au dernier pas de temps (figure 4.3b), cette variance élevée n'est possiblement pas représentative de l'information présente dans les données, puisque, même si les données de concentration sont connues seulement aux puits à 30 et 70 mètres, la concentration montre une forte continuité spatiale. D'ailleurs, la variance au puits de 30 mètres est faible peu importe le nombre d'ensemble et pour tous les pas de temps, indiquant une bonne corrélation entre la conductivité hydraulique et les données de concentration à ce puits. Toutefois, cette faible variance est moins apparente pour le puits à 70 mètres.

Les résultats pour EnKF et ETKF sont très similaires pour des ensembles de taille de plus de 100. La similitude peut être notée, non seulement dans les statistiques d'étendu de l'ensemble et de l'erreur d'estimation, mais aussi aux figures 4.7b à 4.7e et 4.8a à 4.8d où les moyennes et variances des ensembles de 200 réalisations assimilées par EnKF et ETKF sont presque identiques. Ceci indique que la perturbation aléatoire de l'erreur utilisée dans les EnKF est représentative de la covariance de l'erreur à partir d'une centaine de réalisations.

Il est aussi intéressant de noter qu'aucune mise à l'échelle n'a été effectuée sur les observations des concentrations aux puits avant que ces données ne soient assimilées. La mise à l'échelle des mesures, en normalisant les mesures par leur écart-type, afin qu'elles aient une variance similaire est préconisé par Evensen (2003). Ceci est nécessaire afin d'éviter la troncation des valeurs propres associées à des mesures dont l'ordre de grandeur serait faible lors de l'inversion nécessaire au calcul du gain de Kalman (équation 3.12). Dans le cas à l'étude, les concentrations observées varient d'environ 0, 1 à 1000 mg/L. La différence de 5 ordres de grandeur entre les mesures les plus faibles et les plus élevées ne semble pas affecter l'inversion. Plus le temps avance, plus cette différence faiblit puisque la concentration minimale du champ augmente tandis que la concentration maximale varie peu. La ressemblance entre les résultats des ETKF et des EnKF confirme que la mise à l'échelle n'est pas nécessaire dans ce cas-ci puisqu'une mise à l'échelle est intégrée, par défaut, dans l'algorithme des ETKF (équation 3.30).

4.3.2 Paramètres géostatistiques variables

L'assimilation de données de concentrations aux puits a aussi été testée sur des ensembles de taille variable pour un ensemble initial de conductivités hydrauliques produit à partir des paramètres géostatistiques variables présentés au tableau 4.3. Des ensembles de taille variant entre 75 et 2100 membres ont été testés. L'assimilation avec un ensemble de taille plus élevée (2100 membres) que pour l'assimilation des données avec un ensemble initial aux paramètres géostatistiques fixes a été testée puisqu'il a été supposé que l'assimilation avec un ensemble initial aux paramètres variables serait plus sensible à la taille de l'ensemble. Les normes d'évaluation pour différentes tailles d'ensemble en fonction du pas de temps sont présentées à la figure 4.9.



Figure 4.9 – Résultats pour la simulation avec paramètres géostatistiques variables (a) Moyenne de déviation absolue (AAB) de $log_{10}(K_h)$ estimé en fonction du pas de temps (b) Étendue de l'ensemble (EE) de $log_{10}(K_h)$ estimé en fonction du pas de temps pour EnKF et ETKF avec différents nombres d'ensembles

Premièrement, l'étendue de l'ensemble initial ne varie pas beaucoup en fonction du nombre de membres de l'ensemble, tout comme pour l'ensemble initial formé avec les paramètres géostatiques connus. Par contre, cet étendu initial est beaucoup plus élevé que pour les ensembles formés avec les bons paramètres, ce qui est attendu puisque la portée, la variance et la moyenne varient d'une simulation à l'autre. Comme pour le cas précédent, l'étendue de l'ensemble est plus élevé pour des ensembles de taille plus élevée après assimilation de données. Toutefois, dans le cas présent, l'assimilation ne semble pas converger pour des ensembles de plus de 500 membres. Dans ces ensembles de taille élevée, si l'assimilation des données améliore faiblement l'estimation de la conductivité hydraulique pour les premiers pas de temps, l'ajout de nouvelles données détériore la solution après le troisième pas de temps. Ceci n'est toutefois pas le cas pour les ensembles de 75 et 225 réalisations, du moins pour l'assimilation par EnKF. Il est aussi intéressant de noter que pour les ensembles de 75 et 225 membres, les EnKF et les ETKF donnent des solutions dont la différence n'est pas négligeable.



Figure 4.10 – Moyenne du $log_{10}(K_h)$ (a) Cas de référence (b) Pour un ensemble de 75 réalisations après un pas de temps par EnKF (c) Pour un ensemble de 75 réalisations après 6 pas de temps par EnKF (d) Pour un ensemble de 75 réalisations après un pas de temps par ETKF (e) Pour un ensemble de 75 réalisations après 6 pas de temps par ETKF (f) Pour un ensemble de 2100 réalisations après un pas de temps par EnKF (g) Pour un ensemble de 2100 réalisations après 6 pas de temps par EnKF



Figure 4.11 – Variance du $log_{10}(K_h)$ (a) Pour un ensemble de 75 réalisations après un pas de temps par EnKF (b) Pour un ensemble de 75 réalisations après 6 pas de temps par EnKF (c) Pour un ensemble de 75 réalisations après un pas de temps par ETKF (d) Pour un ensemble de 75 réalisations après 6 pas de temps par ETKF (e) Pour un ensemble de 2100 réalisations après un pas de temps par EnKF (f) Pour un ensemble de 2100 réalisations après 6 pas de temps par EnKF

Les moyennes des ensembles assimilés par EnKF et ETKF pour des ensembles de 75 réalisations sont similaires (4.10c et 4.10d). Toutefois, la variance est différente au dernier pas de temps (figures 4.11c et 4.11d) Il est à noter que les échelles de couleur varient entre les premiers et derniers pas de temps pour faire ressortir la différence entre la variance obtenue par EnKF et par ETKF au dernier pas de temps. L'ajout de perturbations aléatoires pourrait améliorer la solution ce qui expliquerait les meilleurs résultats obtenus par EnKF que par ETKF pour un ensemble de 75 réalisations. En effet, Lei *et al.* (2010) notent que les filtres de Kalman d'ensemble stochastiques sont plus stables que les filtres de Kalman d'ensemble déterministes, en particulier lors de la présence d'ensembles contenant des données éloignées de la moyenne. Lawson & Hansen (2004) avaient aussi noté que les filtres de Kalman d'ensemble avec perturbation stochastique des observations arrivaient mieux à représenter une variable d'état gaussienne, ce qui est le cas des états représentant la conductivité hydraulique dans le cas à l'étude.

Le comportement de l'assimilation assez erratique pousse à mettre un bémol sur l'assimilation avec un ensemble initial aux paramètres géostatistiques variables. En effet, dans ce cas, lorsque l'ensemble initial a été créé avec des paramètres géostatistiques variables, des ensembles de plus petite taille (75 ou 225 membres) donnaient des meilleurs résultats que des ensembles de plus grande taille. Ceci est contraire à ce qui a été observé dans la littérature et dans le cas précédent où l'ensemble initial était créé avec des paramètres géostatistiques fixes et pourrait s'expliquer par l'étendue élevée de l'ensemble initial. Même si Erdal & Cirpka (2016) et Jafarpour & Tarrahi (2011) concluent qu'une représentation probabiliste des paramètres géostatistiques peut mener à une solution acceptable de l'état par assimilation de données par filtres de Kalman d'ensemble, plusieurs facteurs semblent jouer un rôle important dans la convergence de l'assimilation, tel que l'ajout de perturbations, le nombre d'ensembles, et très probablement la dimension et la complexité du problème.

4.4 Assimilation des données de résistance

Dans la section précédente, il a été montré qu'il est possible d'estimer avec précision la conductivité hydraulique par assimilation de données de concentrations par filtres de Kalman d'ensemble. Les relations pétrophysiques présentées à la section 3.2.2, justifient le suivi des concentrations à partir de méthodes électriques. Cette section vise donc à présenter les résultats de l'assimilation des données obtenues par le levé dipôle-dipôle décrit à la section 4.1 pour l'estimation de la conductivité hydraulique. Le ratio des résistances a été assimilé, tel que présenté à la section 3.2.4.

En plus de l'assimilation par EnKF et ETKF, l'assimilation a aussi été effectuée par la méthode présentée à la section 3.1.6 et qui a été nommée, EnKF découplé. Il est à noter qu'une méthode équivalente n'a pas été développée dans le cas des ETKF. En effet, la méthode découplée implique que les observations pour la mise à jour des conductivités hydrauliques soient obtenues de l'ensemble de concentrations estimées précédemment à partir des données de résistances. Les observations sont donc représentées par la moyenne de l'ensemble et la covariance associée peut aussi être calculée empiriquement. Toutefois, cette covariance empirique n'est pas nécessairement définie positive. Ceci est problématique dans les cas des ETKF puisque la racine carrée doit être calculée sur cette matrice. Certaines méthodes, telle que la régularisation, peuvent être utilisées pour rendre une matrice défini positive. Ces méthodes ont été explorées brièvement sans grand succès. Dans le cas présent, les ETKFs n'ont finalement pas été testés avec la méthodologie découplée. Les deux différents cas pour la génération de l'ensemble initial de conductivités hydrauliques considérés à la section précédente, soit la génération d'un ensemble à partir des mêmes paramètres géostatiques que le cas synthétique et la génération d'un ensemble initial à partir de paramètres géostatistiques variables ont aussi été étudiés dans l'assimilation de données de résistance.

Comme la relation entre les données de résistance et les concentrations est non-linéaire, l'état a dû être augmenté comme présenté à la section 3.1.5. L'état pour l'assimilation par EnKF ou ETKF devient donc:

$$\mathbf{x} = \begin{bmatrix} log_{10}(K_h) \\ C \\ \frac{R_t}{R_0} \end{bmatrix}.$$

où C est la concentration, $\frac{R_t}{R_0}$ est le ratio de la résistance électrique mesurée au pas de temps t par rapport à la résistance électrique mesurée au temps 0, c'est-à-dire avant le début de l'infiltration, comme présenté à la section 3.2.4.

Toutefois, pour la méthode des EnKF découplé, l'état à la première itération, qui permet d'estimer les concentrations sur tout le champ à partir des résistances électriques est:

$$\mathbf{x} = \begin{bmatrix} C_{interm\acute{e}diaire} \\ \frac{R_t}{R_0} \end{bmatrix}.$$

À la seconde itération, pour estimer les conductivités hydrauliques à partir des concentrations estimées, $C_{intermédiaire}$, à l'étape précédente, l'état devient: $\mathbf{x} = \begin{bmatrix} log_{10}(K_h) \\ C \end{bmatrix}$.

4.4.1 Paramètres géostatistiques connus

L'assimilation a été effectuée sur des ensembles de tailles variables lorsque les réalisations de ces ensembles étaient générées à partir des paramètres connus, afin de déterminer l'effet de la taille de l'ensemble sur l'assimilation des données de résistance. La statistique d'erreur, AAB et la statistique d'étendue de l'ensemble sont présentées aux figures 4.12 et 4.13.



Figure 4.12 – Moyenne de déviation absolue (AAB) en fonction du pas de temps pour différentes tailles d'ensembles: (a) pour l'assimilation des résistances par EnKF, (b) pour l'assimilation des résistances par EnKF découplé, (c) pour l'assimilation des résistances par ETKF.



Figure 4.13 – Étendu de l'ensemble (EE) en fonction du pas de temps pour différentes tailles d'ensembles: (a) pour l'assimilation des résistances par EnKF, (b) pour l'assimilation des résistances par EnKF découplé, (c) pour l'assimilation des résistances par ETKF.

Encore une fois, plus la taille de l'ensemble augmente, plus l'étendu de l'ensemble est élevé après assimilation. Et, comme pour l'assimilation de la concentration, un minimum de 200 membres de l'ensemble semble être nécessaire pour que l'ajout d'information à chaque pas de temps améliore l'estimation. Par ailleurs la statistique d'erreur lorsque l'assimilation se fait par EnKF est meilleure que l'assimilation par ETKF et ce, de façon plus marquée que lorsque les données de concentration étaient assimilées, même pour un ensemble ayant 1000 membres. Une différence entre l'assimilation de données de résistance par rapport à l'assimilation de données de concentration est la non-linéarité de la fonction liant l'état (les concentrations, qui sont elles-mêmes une fonction non-linéarie du paramètre à estimer) aux observations (les résistances). En effet, la résistance observée est sensible à la concentration sur une partie du domaine, tandis que l'observation ponctuelle de concentrations n'est sensible qu'à la concentration en un point. Ainsi, la résistance observée présente une couverture latérale, tandis que les observations de concentration aux puits sont ponctuelles. Un ou plusieurs de ces facteurs pourrait expliquer que la façon de traiter les observations, déterministe vs stochastique, ait un effet plus prononcé que lorsque les données de concentrations sont utilisées directement.

Toutefois, malgré la différence qui peut être soulevée dans la statistique d'erreur, la moyenne et la variance de la conductivité hydraulique assimilée par EnKF ou ETKF sont très similaires. Les figures 4.14b et 4.14c représentent la conductivité hydraulique mise à jour par assimilation des données de résistance par EnKF. La ligne d'électrodes du protocole d'ERT est représentée à la surface des figures. Les structures pouvant être observées sur ces figures sont aussi présentes aux figures 4.14f et 4.14g qui montrent l'assimilation par ETKF. Les variances sont aussi très similaires en plus d'être cohérentes avec la profondeur d'investigation du dispositif utilisé, puisque la variance est beaucoup plus faible en surface qu'en profondeur après l'assimilation au premier pas de temps. Si la démarcation entre les variances faibles et fortes est plus profonde au sixième pas de temps, c'est que le panache s'est répandu plus profondément. Malgré la présence du panache en profondeur, la variance reste plus élevée que lorsque des données de concentration étaient assimilées puisque les concentrations étaient assimilées sur toute la profondeur du domaine.

Enfin, si les EnKF couplés et découplés montrent des résultats très similaires sous tous les aspects (AAB, étendue de l'ensemble, moyenne et variance de l'ensemble des conductivités hydrauliques) lorsque l'ensemble contient 1000 membres, les EnKF découplés semblent présenter un petit avantage pour des ensembles plus faibles (de 200 membres). Cet avantage pour des plus petits ensembles serait cohérent avec l'élimination de corrélations fortuites entre la résistance et la conductivité hydraulique de la méthode découplée.



Figure 4.14 – Moyenne du $log_{10}(K_h)$ (a) Cas de référence (b) pour un ensemble de 1000 réalisations après un pas de temps par EnKF couplé (c) pour un ensemble de 1000 réalisations après 6 pas de temps par EnKF couplé (d) pour un ensemble de 1000 réalisations après un pas de temps par EnKF découplé (e) pour un ensemble de 1000 réalisations après 6 pas de temps par EnKF découplé (f) pour un ensemble de 1000 réalisations après un pas de temps par EnKF découplé (f) pour un ensemble de 1000 réalisations après un pas de temps par ETKF (g) pour un ensemble de 1000 réalisations après 6 pas de temps par ETKF



Figure 4.15 – Variance du $log_{10}(K_h)$ (a) pour un ensemble de 1000 réalisations après un pas de temps par EnKF (b) pour un ensemble de 1000 réalisations après 6 pas de temps par EnKF (c) pour un ensemble de 1000 réalisations après un pas de temps par EnKF découplé (d) pour un ensemble de 1000 réalisations après 6 pas de temps par EnKF découplé (e) pour un ensemble de 1000 réalisations après un pas de temps par EnKF (f) pour un ensemble de 1000 réalisations après 6 pas de temps par ETKF

4.4.2 Paramètres géostatistiques variables

Dans cette section, l'objectif est toujours l'estimation de la conductivité hydraulique à partir de données de résistances électriques, toutefois, les simulations de l'ensemble initiales ont été générées avec les paramètres variables décrits au tableau 4.3. Ainsi, cette section vise à déterminer s'il est possible d'estimer les conductivités hydrauliques à partir de données de résistance électrique lorsque les paramètres géostatistiques de la variable à estimer sont plus ou moins bien connus. Dans la section précédente, il a été conclu que les résistances électriques permettaient d'améliorer l'estimation de la conductivité hydraulique si les simulations initiales de cette conductivité étaient générées à partir de paramètres connus. Les figures 4.16a à 4.16c montrent que l'erreur d'estimation (AAB) diminue en effet avec l'assimilation de données pour les 3 schémas d'assimilation lorsque la taille de l'ensemble est de 1050 réalisations ou plus. Toutefois, l'assimilation avec un ensemble de 1050 réalisations semble un peu plus stable que l'assimilation avec un plus grand ensemble de 2100 réalisations, puisque, par exemple, la statistique d'erreur atteint un minimum au troisième pas pour ensuite augmenter pour l'assimilation des résistances par EnKF découplé et au quatrième pas de temps pour l'assimilation par EnKF couplé. L'étendue de l'ensemble reste tout de même plus élevée pour tous les pas de temps lorsque l'ensemble comporte 2100 réalisations que lorsque l'ensemble n'en comporte que 1050.



Figure 4.16 – Moyenne de déviation absolue (AAB) en fonction du pas de temps pour différentes tailles d'ensembles: (a) pour l'assimilation des résistances par EnKF, (b) pour l'assimilation des résistances par EnKF découplé, (c) pour l'assimilation des résistances par ETKF.



Figure 4.17 – Étendu de l'ensemble (EE) en fonction du pas de temps pour différentes tailles d'ensembles: (a) pour l'assimilation des résistances par EnKF, (b) pour l'assimilation des résistances par EnKF découplé, (c) pour l'assimilation des résistances par ETKF.

Les structures qui apparaissent dans les cartes de conductivité hydraulique sont très similaires que l'ensemble initial de conductivités hydrauliques soit généré à partir des bons paramètres géostatistiques ou à partir d'une gamme de paramètres (figures 4.14 et 4.18). Les conductivités hydrauliques estimées à partir de l'ensemble aux paramètres variables sont toutefois plus faibles en moyenne. En effet, la moyenne du logarithme des conductivités hydrauliques estimées à partir de l'ensemble aux paramètres variables est d'environ -5,4, tandis qu'elle est de -5,6 lorsque l'ensemble était généré grâce aux paramètres connus. Ce biais a été introduit par l'assimilation des données puisque la moyenne de l'ensemble initial était aussi de -5,6 (les moyennes des simulations variaient mais étaient centrées sur la vraie moyenne).

Les variances montrent aussi le même comportement que lorsque l'assimilation des données de résistance électrique se faisait à partir d'un ensemble généré à partir de paramètres géostatistiques fixes et connus; c'est-à-dire une variance qui diminue avec l'ajout de données, et ceci, particulièrement en profondeur.



Figure 4.18 – Moyenne du $log_{10}(K_h)$ (a) Cas de référence (b) pour un ensemble de 1050 réalisations après un pas de temps par EnKF (c) pour un ensemble de 1050 réalisations après 6 pas de temps par EnKF (d) pour un ensemble de 1050 réalisations après un pas de temps par EnKF découplé (e) pour un ensemble de 1050 réalisations après 6 pas de temps par EnKF découplé (f) pour un ensemble de 1050 réalisations après un pas de temps par ETKF (g) pour un ensemble de 1050 réalisations après 6 pas de temps par ETKF



Figure 4.19 – Variance du $log_{10}(K_h)$ (a) pour un ensemble de 1050 réalisations après un pas de temps par EnKF (b) pour un ensemble de 1050 réalisations après 6 pas de temps par EnKF (c) pour un ensemble de 1050 réalisations après un pas de temps par EnKF découplé (d) pour un ensemble de 1050 réalisations après 6 pas de temps par EnKF découplé (e) pour un ensemble de 1050 réalisations après un pas de temps par ETKF (f) pour un ensemble de 1050 réalisations après 6 pas de temps par ETKF

Si l'assimilation des données de résistance électrique ne semble pas appropriée pour des ensembles trop petits, les figures de l'annexe B permettent toutefois de remarquer que, malgré un petit ensemble, les données de résistance électrique permettent d'améliorer l'estimation des concentrations par rapport à l'ensemble de concentrations prédites (AAB plus faible sur l'ensemble de concentrations mises à jour que sur l'ensemble de concentrations prédites). Même si la mise à jour des concentrations à partir des données de résistances est bénéfique, l'assimilation de ces concentrations mises à jour (par la méthode découplée) pour estimer la conductivité hydraulique ne l'est pas nécessairement.

4.5 Assimilation des données de concentrations aux puits et de résistances électriques

Dans les sections précédentes, le champ de conductivité hydraulique a pu être mis à jour à partir d'observations de concentrations à des puits, ainsi qu'à partir de données de résistance électrique obtenues par tomographie. La présente section vise à évaluer si l'assimilation des deux types de données est bénéfique par rapport à l'assimilation de seulement un des deux types de données. Plusieurs schémas d'assimilation de ces données sont possibles. Les différents schémas étudiés sur ce cas sont présentés ci-dessous:

- 1 EnKF couplé simultané. Les données de concentration et de résistance électrique sont assimilées en même temps pour estimer les concentrations ainsi que la conductivité hydraulique sur tout le domaine en une seule étape.
- 2 EnKF couplé séquentiel. Les données de résistance électrique sont d'abord assimilées, suivies des données de concentrations au puits. Les concentrations et la conductivité hydraulique sur tout le domaine sont estimées dans la même étape.
- 3 EnKF découplé simultané. Les données de concentration et de résistance électrique sont assimilées en même temps pour estimer les concentrations sur tout le domaine. Les concentrations ainsi obtenues sont utilisées pour estimer la conductivité hydraulique.
- 4 EnKF découplé séquentiel. Les données de résistance électrique sont d'abord assimilées suivies des données de concentrations aux puits pour estimer les cocentrations sur tout le domaine. Les concentrations ainsi obtenues sont utilisées pour estimer la conductivité hydraulique.
- 5 ETKF couplé simultané. Le premier schéma est ici répété, mais en utilisant l'algorithme des ETKF plutôt que l'algorithme des EnKF pour l'étape de mise à jour.
- 6 ETKF couplé séquentiel. Le second schéma est ici répété, mais en utilisant l'algorithme des ETKF plutôt que l'algorithme des EnKF pour l'étape de mise à jour.

La définition de l'état et des données assimilées dans les différentes étapes de ces schémas sont résumés au tableau 4.4 dans lequel C représente la concentration sur l'ensemble du domaine, C_{puits} représente la concentration observée aux puits et $C_{estimé}$ représente la concentration estimée sur tout le domaine à partir des observations, soit les concentrations aux puits et les résistances électriques.

		Schémas 1 et 5	Schémas 2 et 6	Schéma 3	Schéma 4
		EnKF et ETKF	EnKF et ETKF	EnKF découplé	EnKF découplé
		couplé simultané	couplé séquentiel	$\operatorname{simultan\acute{e}}$	séquentiel
Étape 1	État	$log_{10}(K_h), \mathrm{C}, \frac{R_t}{R_0}$	$log_{10}(K_h), \mathrm{C}, \frac{R_t}{R_0}$	C, $\frac{R_t}{R_0}$	C, $\frac{R_t}{R_0}$
	Données	$C_{puits}, \ \frac{R_t}{R_0}$	$\frac{R_t}{R_0}$	$C_{puits}, \frac{R_t}{R_0}$	$\frac{R_t}{R_0}$
Étape 2	État	n/a	$log_{10}(K_h), \mathcal{C}, \frac{R_t}{R_0}$	$log_{10}(K_h), C$	C, $\frac{R_t}{R_0}$
	Données	n/a	C_{puits}	$C_{estim\acute{ ext{e}}}$	C_{puits}
Étape 3	État	n/a	n/a	n/a	$log_{10}(K_h), \mathcal{C}$
	Données	n/a	n/a	n/a	$C_{estim\acute{ ext{e}}}$

Tableau 4.4 – Variables d'état et d'observations des différents schémas

4.5.1 Paramètres géostatistiques connus

Les statistiques d'erreur de l'assimilation ainsi que d'étendue de l'ensemble selon les six schémas sont résumés dans les figures 4.20 et 4.21 pour des ensembles ayant 100 et 1000 membres.



Figure 4.20 – Moyenne de déviation absolue (AAB) en fonction du pas de temps pour différentes tailles d'ensembles: (a) pour l'assimilation par EnKF couplé, (b) pour l'assimilation par EnKF découplé, (c) pour l'assimilation par ETKF.



Figure 4.21 – Étendu de l'ensemble (EE) en fonction du pas de temps pour différentes tailles d'ensembles: (a) pour l'assimilation par EnKF couplé, (b) pour l'assimilation par EnKF découplé, (c) pour l'assimilation par ETKF.

D'une part, l'assimilation des concentrations et des résistances électriques permet d'améliorer l'estimation des conductivités hydrauliques pour un ensemble de 1000 réalisations avec des résultats très similaires peu importe le schéma d'assimilation. La statistique d'erreur la plus faible (AAB) pour un ensemble de 100 réalisation est atteinte lors de l'assimilation simultanée des concentrations et des résistances électriques par EnKF couplé. Ces résultats sont aussi très similaires, quoique sensiblement meilleurs par rapport à la statistique d'erreur, que les résultats de l'assimilation par résistances électriques exclusivement. Du point de vue de l'étendue de l'ensemble, celle-ci est sensiblement plus faible que pour l'assimilation de la résistance électrique seulement, qui était elle-même plus faible que pour l'assimilation des concentrations. Ces résultats sont attendus puisque l'ajout d'information devrait améliorer l'estimation.

D'autre part, l'assimilation des concentrations et des résistances électriques détériore l'estimation dans tous les schémas d'assimilation lorsque l'ensemble n'est composé que d'une centaine de réalisations et ce, même si l'assimilation des données de concentration permettait d'améliorer l'estimation des conductivités hydrauliques pour un ensemble de 100 réalisations. Ceci laisse croire que les erreurs induites par un ensemble trop faible de réalisations pour représenter les mesures électriques ne peuvent être compensées par l'assimilation des données de concentration.

Un point important à soulever est la similitude entre l'assimilation par ETKF séquentiel et simultané. Au premier pas de temps, ces deux schémas d'assimilation donnent en fait des valeurs de AAB et EE égales. Tippett *et al.* (2003) a noté que les deux premiers moments statistiques des ETKF et ETKF séquentiels devraient être égaux pour une même covariance de l'erreur prédite. Ce-

81

pendant les statistiques d'ordre supérieurs ne seraient pas nécessairement égales. Pour cette raison, les résultats des ETKF séquentiels et simultanés divergent aux pas de temps suivants. Ceci confirme d'ailleurs que l'état ne suit pas une loi multigaussienne, puisque, le cas échéant, les deux premiers moments décriraient complètement l'ensemble et donc, il ne devrait pas y avoir de divergence au pas subséquent.



Figure 4.22 – Moyenne du $log_{10}(K_h)$ pour un ensemble de 1000 réalisations (a) Cas de référence (b) après un pas de temps pour le schéma 1 (c) après 6 pas de temps pour le schéma 1 (d) après un pas de temps pour le schéma 2 (e) après 6 pas de temps pour le schéma 2 (f) après un pas de temps pour le schéma 3 (g) après un pas de temps pour le schéma 4 (h) après 6 pas de temps pour le schéma 3 (i) après un pas de temps pour le schéma 4 (j) après 6 pas de temps pour le schéma 4 (k) après un pas de temps pour le schéma 5 (l) après 6 pas de temps pour le schéma 5 (m) après un pas de temps pour le schéma 6 (n) après 6 pas de temps pour le schéma 6



Figure 4.23 – Variance du $log_{10}(K_h)$ pour un ensemble de 1000 réalisations (a) après un pas de temps pour le schéma 1 (b) après 6 pas de temps pour le schéma 1 (c) après un pas de temps pour le schéma 2 (d) après 6 pas de temps pour le schéma 2 (e) après un pas de temps pour le schéma 3 (f) après un pas de temps pour le schéma 4 (g) après 6 pas de temps pour le schéma 3 (h) après un pas de temps pour le schéma 4 (i) après 6 pas de temps pour le schéma 4 (j) après un pas de temps pour le schéma 5 (k) après 6 pas de temps pour le schéma 5 (l) après un pas de temps pour le schéma 6 (m) après 6 pas de temps pour le schéma 6

Si les champs du $log_{10}(K_h)$ estimés (figure 4.22)sont très similaires aux champs assimilés à partir des résistances électriques exclusivement (figure 4.14), la variance (figure 4.23) montre, non seulement le même comportement que pour l'assimilation des données de résistances électrique, mais

aussi la baisse de variance autour du puits à 30 mètres qui peut être observé lors de l'assimilation des données de concentration seulement.

4.5.2 Paramètres géostatistiques variables

Dans la section suivante, l'assimilation des données de concentrations aux puits et de résistances électriques ont été appliquées au cas synthétique pour un ensemble initial de conductivités hydrauliques généré avec les paramètres géostatistiques variables décrits au tableau 4.3. Les différents schémas d'assimilation présentés à la section précédente ont aussi été appliqués ici. Les figures 4.24 et 4.25 présentent la statistique d'erreur et l'étendue de l'ensemble pour l'assimilation selon les différents schémas mis en place.



Figure 4.24 – Moyenne de déviation absolue (AAB) en fonction du pas de temps pour différentes tailles d'ensembles (a) pour l'assimilation par EnKF couplé (b)pour l'assimilation par EnKF découplé (c) Pour un pour l'assimilation par ETKF



Figure 4.25 – Étendu de l'ensemble (EE) en fonction du pas de temps pour différentes tailles d'ensembles (a) pour l'assimilation par EnKF (b) pour l'assimilation par EnKF découplé (c) pour l'assimilation par ETKF

L'assimilation des données de concentrations aux puits ainsi que des données de résistivité améliore l'estimation pour des ensembles de 1050 membres, mais pas pour des ensembles de seulement 75 membres pour tous les schémas d'assimilation. Tel qu'observé aux sections 4.3.2 et 4.4.2, un ensemble de 75 membres était suffisant pour l'assimilation des données de concentrations aux puits, mais pas pour l'assimilation des données de résistance électrique. Le même phénomène était observé lorsque l'assimilation des concentrations aux puits et des résistances était effectuée sur un ensemble initial avec les paramètres géostatistiques connus: un ensemble de 100 membres était suffisant pour assimiler les données de concentration, mais pas pour assimiler les données de résistances électriques, ni lorsque les deux types de données étaient assimilées. Ceci laisse croire que l'assimilation des deux types de données conjointement ne convergera pas pour des ensembles dont la taille n'aurait pas permis que l'un ou l'autre type de donnée ne converge individuellement. En effet, les fausses corrélations entre les données et les variables d'état à estimer, pouvant apparaître pour de trop petits ensembles seraient toujours présentes malgré l'ajout de données de nature différente.

Par ailleurs, l'assimilation des donnés de résistance électrique en plus des données de concentration permet d'obtenir une statistique d'erreur plus faible que lorsqu'une seule de ces variables d'observation n'était assimilée pour un ensemble de 1050 réalisations. Toutefois, l'assimilation des concentrations aux puits seulement pour un ensemble de 75 réalisations permettait d'obtenir de meilleurs résultats que l'assimilation des concentrations aux puits et des résistances électriques pour un ensemble de 1050 réalisations. La faible erreur d'estimation pour l'assimilation des données de concentration seulement avec un ensemble de 75 réalisations semblait toutefois une anomalie. De plus, l'étendue de l'ensemble de 75 membres assimilé à partir des concentrations aux puits est aussi plus faible que lorsque les deux types de données sont assimilées pour un ensemble de 1050 réalisations. De plus le ratio de l'étendue de l'ensemble par rapport à la mesure de biais moyen (AAB) est aussi plus faible pour l'ensemble de 75 réalisations en assimilant seulement les concentrations que pour l'ensemble de 1050 réalisations en assimilant les deux types de données. Conserver une étendue assez élevée par rapport à la mesure d'erreur est important pour bien représenter l'incertitude (Fortin *et al.*, 2014). En effet, si l'étendue de l'ensemble est beaucoup plus faible que le biais moyen de l'estimation; alors l'ensemble assimilé va surévaluer la précision de l'estimation.

Pour un ensemble assez grand, c'est-à-dire pour l'ensemble de 1050 réalisations, les 6 schémas d'assimilation donnent des résultats satisfaisants. Par contre, lorsque l'assimilation se fait de façon couplée que ce soit par ETKF ou EnKF, un biais apparaît dans la moyenne du champ de conductivité hydraulique. Les figures 4.26c, 4.26c, 4.26k et 4.26m montrent des conductivités hydrauliques systématiquement plus élevées que la conductivité hydraulique de référence (figure 4.26a). Ceci apparaît en particulier pour l'assimilation par EnKF couplé et séquentiel. Ce biais pouvait aussi être observé pour l'assimilation des données de résistance électrique seulement sur un ensemble initial aux paramètres géostatiques variables. Lorsque seules les données de résistance électrique étaient assimilées, l'assimilation par EnKF découplé montrait aussi un biais; ce qui n'est pas le cas ici. Les EnKF découplé montrent ici un avantage sur les autres méthodes, tandis que les EnKF couplés séquentiels montrent des résultats sensiblement moins bons que les autres méthodes. Au second pas de temps, la première étape des EnKF couplés séquentiels, soit l'assimilation des résistances électriques fait même diverger la solution. Toutefois, l'assimilation des concentrations, qui suit l'assimilation des résistances électriques permet de ramener l'estimation à une solution plus réaliste. La solution s'améliore au pas de temps suivant.



Figure 4.26 – Moyenne du $log_{10}(K_h)$ pour un ensemble de 1050 réalisations (a) Cas de référence (b) après un pas de temps pour le schéma 1 (c) après 6 pas de temps pour le schéma 1 (d) après un pas de temps pour le schéma 2 (e) après 6 pas de temps pour le schéma 2 (f) après un pas de temps pour le schéma 3 (g) après un pas de temps pour le schéma 4 (h) après 6 pas de temps pour le schéma 3 (i) après un pas de temps pour le schéma 4 (j) après 6 pas de temps pour le schéma 4 (k) après un pas de temps pour le schéma 5 (l) après 6 pas de temps pour le schéma 5 (m) après un pas de temps pour le schéma 6 (n) après 6 pas de temps pour le schéma 6



Figure 4.27 – Variance du $log_{10}(K_h)$ pour un ensemble de 1050 réalisations (a) après un pas de temps pour le schéma 1 (b) après 6 pas de temps pour le schéma 1 (c) après un pas de temps pour le schéma 2 (d) après 6 pas de temps pour le schéma 2 (e) après un pas de temps pour le schéma 3 (f) après un pas de temps pour le schéma 4 (g) après 6 pas de temps pour le schéma 3 (h) après un pas de temps pour le schéma 4 (i) après 6 pas de temps pour le schéma 4 (j) après un pas de temps pour le schéma 5 (k) après 6 pas de temps pour le schéma 5 (l) après un pas de temps pour le schéma 6 (m) après 6 pas de temps pour le schéma 6
4.6 Résumé des résultats de l'assimilation de données sur un cas synthétique

Un cas synthétique de contamination saline dans un aquifère à conductivité hydraulique hétérogène a été développé afin d'évaluer différentes méthodes d'assimilation de données pour l'estimation de la conductivité hydraulique. L'assimilation des données de concentration, de résistance électrique obtenues par ERT et des deux types de données conjointement a été comparée afin d'évaluer l'apport de chacun de ces types de données. De plus, différentes variantes des filtres de Kalman d'ensemble ont été mises en place pour chacun des types de données afin de déterminer la meilleure méthode d'assimilation de chaque type de donnée. Étant donné l'importance de la taille de l'ensemble sur l'efficacité de l'assimilation, une analyse de sensibilité au nombre de réalisations de l'ensemble a aussi été effectué pour chaque cas d'assimilation.

Parmi les variantes des EnKFs comparées, les filtres à racine carrée ont été comparés aux EnKFs classiques. En général, pour de grands ensembles, ces deux méthodes montraient peu de différences. Les ETKFs évitent la perturbation stochastique des observations, en intégrant la matrice de covariance de l'erreur directement dans le calcul de la mise à jour. Ainsi, lorsque l'ensemble est assez grand, la perturbation stochastique des observations tel qu'appliquée dans les EnKFs devrait bien représenter la matrice de covariance de l'erreur des observations et il ne devrait pas y avoir de différence importante entre les deux méthodes. Toutefois, contrairement à ce qui serait attendu, c'est-à-dire que les ETKFs performent mieux pour de petits ensembles puisqu'ils représentent mieux la matrice de covariance de l'erreur d'observation, la mise à jour par EnKFs donnait des meilleurs résultats pour de petits ensembles dans le cas présenté. Tel que discuté précédemment (section 3.1.9), les ETKFs sont sujets aux mêmes problèmes (écrasement de la covariance et divergence)que les EnKFs lorsque les ensembles sont trop petits. L'ajout de perturbations aurait possiblement un effet bénéfique dans le cas des petits ensembles pour contrer ces problèmes liés à la taille.

Une seconde variante des EnKFs dite EnKF découplé a été comparé aux EnKFs classiques pour l'assimilation des résistances et l'assimilation des résistances et des concentrations. Ces deux méthodes donnaient des résultats similaires avec de très faibles variations dans les normes d'évaluation (biais moyen et étendu de l'ensemble) lorsque la taille de l'ensemble était assez élevée. Les méthodes simultanées et séquentielles pour l'assimilation conjointe des résistances et des concentrations donnaient aussi des résultats similaires lorsque les ensembles étaient assez grands. Même si des différences notables entre ces différentes méthodes ont pu être observées pour des petits ensembles, ces ensembles étaient trop petits pour que l'assimilation soit bénéfique.

Le tableau 4.5 présente les normes d'évaluation au dernier pas de temps pour la méthodologie et le nombre d'ensemble avec le biais moyen (AAB) final sur les conductivités hydrauliques le plus faible pour chaque type de variable assimilé (concentration, résistances électriques, concentrations et résistances électriques). Dans le cas de l'assimilation avec un ensemble initial de conductivités hydrauliques généré à partir de paramètres géostatistiques connus; l'assimilation des données de résistance permet d'obtenir un biais moyen de la conductivité hydraulique plus faible que l'assimilation des concentrations et que l'assimilation des deux types de données permet d'obtenir un biais moyen encore plus faible que l'une ou l'autre variable séparément. Ceci était attendu puisque les données de résistances électriques ont une plus grande couverture spatiale que les données de concentration qui sont des données ponctuelles à deux puits. De plus, il était aussi attendu que l'assimilation des deux types de données améliore l'estimation des conductivités hydrauliques par rapport à un seul type de donnée puisque les données de concentration et de résistance apporteraient des informations complémentaires. Toutefois, il est à noter que la différence dans le biais moyen entre l'assimilation de chaque type de données ne varie pas beaucoup. La faiblesse de ces variations est possiblement due à la faible hétérogénéité de la conductivité hydraulique. Ainsi, un cas avec de plus grandes variations de conductivité hydraulique pourrait montrer plus de bénéfices apportés par l'assimilation des résistances par rapport à l'assimilation des concentrations.

Tableau 4.5 – Normes d'évaluation	n sur l'ensemble d	les conductivités	hydrauliques	pour	l'assimilation
avec paramètres géostatistiques co	nnus				

Donnée assimilée	Méthode	Taille de l'ensemble	Biais moyen final	Étendue final
				de l'ensemble
Concentrations	EnKF	200	0,4268	0,3027
Résistances électriques	EnKF découplé	1000	0,3963	0,3154
Concentrations et	EnKF découplé			
résistances électriques	et séquentiel	1000	$0,\!3795$	0,2909

Le tableau 4.6 présente les normes d'évaluation au dernier pas de temps pour chaque type de variable assimilé (concentration, résistances électriques, concentrations et résistances électriques) lorsque les paramètres géostatistiques utilisés pour créer l'ensemble de départ étaient variables. Un ensemble de faible taille (75 membres) était suffisant pour assimiler les données de concentration. Toutefois, lorsque la taille de l'ensemble augmentait, l'algorithme d'assimilation semblait avoir de la difficulté à appliquer des corrections par rapport à l'ensemble de départ. Ceci était aussi vrai pour des ensembles de grande taille dans le cas de l'assimilation des résistances ou de l'assimilation conjointe des résistances et des concentrations. Partir d'un ensemble généré à partir de paramètres géostatistiques variables pour l'assimilation créait donc un cadre où il était difficile d'effectuer des mises à jour significatives.

Il a déjà été noté que faire varier la portée pour générer l'ensemble initial de conductivités hydrauliques est possiblement problématique. En effet, Bailey & Baú (2012) ont conclu qu'il était difficile d'estimer la portée de la conductivité hydraulique en assimilant une donnée indirecte de la conductivité, soit le niveau de la nappe, dans leur cas d'étude; et ce, même s'ils étaient capables de bien estimer la moyenne et la variance de la conductivité hydraulique à partir de cette donnée. Un solution possible serait de partir de plusieurs ensembles ayant des paramètres géostatistiques homogènes et de faire l'assimilation séparément pour chacun des ensembles.

Donnée assimilée	Méthode	Taille de l'ensemble	Biais moyen final	Étendue final
				de l'ensemble
Concentration	EnKF	75	0,4875	0,3528
Résistances électriques	EnKF couplé	1050	0,6662	0,6835
Concentrations et	EnKF découplé			
résistances électriques	et séquentiel	1050	0,6006	0,6164

Chapitre 5

Assimilation des données de résistance électrique lors d'une expérience d'infiltration d'eau salée

5.1 Expérience d'intrusion d'eau saline dans un bac à sable

Du 13 au 14 mars 2015, une expérience d'intrusion d'eau saline dans un bac de sable a été effectuée au laboratoire d'hydraulique de l'université de Padoue. Le sable était composé de billes de silice avec une taille nominale de 200 à 800 μm . Des mesures de tomographie électrique de surface ont été acquises tout au long de l'expérience afin de déterminer si ce type de mesures permettait de faire la surveillance d'une intrusion d'eau salée qui peuvent être problématiques dans les aquifères côtiers. La figure 5.1 présente des vue en plan et en couple de l'expérience. Un schéma plus détaillé de l'expérience, incluant la disposition des électrodes peut être trouvé à l'annexe C.



Figure 5.1 – Schéma de l'expérience d'intrusion d'eau saline (n'est pas à l'échelle) (a) vue en coupe (b) vue en plan

Dans la présente thèse, ces mesures électriques ont été utilisées afin d'estimer deux paramètres hydrogéologiques par filtre de Kalman d'ensemble: la perméabilité et la dispersivité longitudinale. Les concentrations ont aussi été prédites puis mises à jour par EnKF à chaque pas de temps. Tout comme pour le cas synthétique, des modèles hydrogéologiques et électriques ont d'abord dû être définis afin de prédire l'état dans le temps. Dans le cadre de ce projet, une méthode d'assimilation des résistivités électriques obtenues par inversion des mesures de résistance a aussi été développée à des fins de comparaison avec les études préalablement effectuées par le groupe Camporese *et al.* (2015).

Plusieurs études ont été effectuées sur cette expérience d'eau saline préalablement. Des mesures des paramètres hydrogéologiques du sable ont été effectuées en laboratoire, dont des mesures de conductivité hydraulique qui ont permis de calculer la perméabilité (Bertorelle, 2014; Constantini, 2015). Par ailleurs, la dispersivité longitudinale a été calibrée lors d'une étude numérique. Ces valeurs seront considérées comme les valeurs de référence pour ces paramètres.

Même si l'expérience a duré 24 heures et que des données de résistivité électrique étaient disponibles à des intervalles de 20 minutes, seuls 5 pas de temps ont été conservés. Le premier pas de temps est à 14 heures après le début de l'expérience. À ce pas de temps l'intrusion d'eau saline apparaît nettement sur les données de résistivité électrique. Les pas de temps subséquents sont pris à un intervalle de 2 heures afin d'observer une assez grande différence dans les mesures entre deux pas de temps.

5.2 Modèle hydrogéologique

L'expérience d'intrusion d'eau saline est représentée par un modèle d'écoulement en régime transitoire dans un milieu saturé - non-saturé couplé à un modèle de transport aussi en régime transitoire. Le modèle tient compte de la variation de densité du fluide, due à la concentration en sel qui varie. Le solveur *SUTRA* (Voss & Provost, 2010) a été utilisé pour résoudre les équations d'écoulement et de transport par éléments finis.

Tout comme pour le développement des équations d'écoulement et de transport pour un milieu saturé et à densité fixe, l'établissement d'un bilan massique permet d'obtenir les équations d'écoulement et de transport dans un milieu saturé-non-saturé et avec un liquide à densité variable. L'équation d'écoulement est donnée par:

$$\left(S_w \rho S_{op} + n\rho \frac{\partial S_w}{\partial p}\right) + \left(nS_w \frac{\partial \rho}{\partial C}\right) \frac{\partial C}{\partial t} - \nabla \cdot \left[\left(\frac{\kappa \kappa_r \rho}{\mu} \cdot (\nabla p - \rho \mathbf{g})\right)\right] = Q_p \tag{5.1}$$

où S_w représente le degré de saturation, ρ représente la densité totale, S_{op} est le coefficient d'emmagasinement de la pression, n représente la porosité, p représente la pression, κ est un tenseur de perméabilité tandis que κ_r est la perméabilité relative pour l'écoulement non-saturé qui est considérée isotrope, μ est la viscosité du fluide, \mathbf{g} est un vecteur qui représente la gravité. Q_p représente une source ou un puits.

Le coefficient d'emmagasinement de la pression S_{op} est estimé par la relation linéaire:

$$S_{op} = a(1-n) + bn$$
 (5.2)

où *a* et *b* sont la compressibilité de la matrice solide et du fluide, respectivement. Le coefficient d'emmagasinement commun, S_s , qu'on retrouve dans l'équation d'écoulement à densité fixe (équation 3.55), peut être lié au coefficient d'emmagasinement de la pression par la relation $S_s = \rho g S_{op}$.

Une relation similaire lie la perméabilité à la conductivité hydraulique, K:

$$\kappa = \frac{K\mu}{\rho g} \tag{5.3}$$

La densité du fluide ρ est fonction de la concentration, C. Une approximation de premier ordre est utilisée pour définir cette relation dans le solveur SUTRA

$$\rho = \rho_0 + \frac{\partial \rho}{\partial C} (C - C_0) \tag{5.4}$$

où $\frac{\partial \rho}{\partial C}$ est présumée constante.

Chapitre 5. Assimilation des données de résistance électrique lors d'une expérience d'infiltration d'eau salée97

L'équation de transport est, pour une situation où l'adsorption est considérée négligeable, telle qu'implémentée dans le solveur numérique, *SUTRA*:

$$\frac{\partial (nS_w\rho C)}{\partial t} = -\nabla \cdot (nS_w\rho \mathbf{v}C) + \nabla \cdot [nS_w\rho(D_mI + \underline{\underline{D}} \cdot \nabla C] + nS_w\rho\Gamma_w + Q_pC *$$
(5.5)

Où D_m représente la diffusivité moléculaire et I est la matrice identité, tandis que \underline{D} représente un tenseur de la dispersion qui peut donc être anisotrope. Le terme $nS_w\rho\Gamma_w$ est associé à la production de contaminant dissous par des réactions chimiques, biologiques ou radioactives et Γ_w est le taux de production par unité de masse de fluide. Le terme Q_pC^* représente l'entrée de contaminant dans le système par des sources du fluide dans lesquelles ce contaminant serait dissout. Donc, C^* est la concentration massique du contaminant dans le fluide à une source de fluide.

Lorsque l'écoulement est bidimensionnel, le tenseur de dispersion est défini de la façon suivante:

$$\underline{\underline{D}} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$
(5.6)

avec

$$D_{xx} = \frac{1}{v^2} (d_L v_x^2 + d_T v_y^2) \tag{5.7}$$

$$D_{yy} = \frac{1}{v^2} (d_T v_x^2 + d_L v_y^2) \tag{5.8}$$

et,

$$D_{xy} = D_{yx} = \frac{1}{v^2} (d_L - d_T) (v_x v_y)$$
(5.9)

où v, v_x et v_y sont la magnitude totale et dans les directions, x et y du vecteur de vitesse, \mathbf{v} , respectivement.

Les coefficients de dispersion longitudinale et transversale, d_L et d_T sont calculés à partir des dispersivités longitudinales et transversales, α_L et α_T et de la magnitude totale de la vitesse, v:

$$d_L = \alpha_L v \tag{5.10}$$

$$d_T = \alpha_T v \tag{5.11}$$

Le détail de ces équations d'écoulement et de transport peut être trouvé dans Voss & Provost (2010) ou dans Voss (1999).

La configuration de l'expérience permet de modéliser l'écoulement et le transport en deux dimensions puisque l'écoulement se fait selon un plan parallèle à la longueur du bac de sable. Les dimensions du modèle équivalent aux dimensions du bac de sable de l'expérience, c'est-à-dire 5 mètres en longueur et 50 cm en profondeur. Ce modèle exclut donc les deux bacs d'eau présents dans l'expérience, qui sont plutôt représentés par des conditions frontières.

Comme mentionné précédemment, la simulation d'écoulement se fait en régime transitoire et il en va de même pour la simulation de transport. Par ailleurs, une petite zone en surface n'est pas saturée, la simulation se fait donc dans un mode saturé-non-saturé. Étant donné que l'écoulement est principalement horizontal et qu'il se situe presqu'exclusivement dans la zone saturée, la zone non-saturée n'a que peu d'influence sur le résultat de la simulation.

Les paramètres du fluide et de la matrice solide doivent d'abord être spécifiés:

- La viscosité du fluide est considérée constante à 0,001 $kg/(m \cdot s)$
- La densité du fluide est considérée variable et suit l'équation 5.4 avec les propriétés suivante:
 - Concentration en sel de l'eau pure: $(C_0) = 0 kg/kg$
 - Densité de l'eau pure: $(\rho_0) = 1000 \ kg/m^3$
 - Coefficient de variation de la densité du fluide: $(\frac{\partial\rho}{\partial C})=722,15~(kg\cdot kg)/(kg\cdot m^3)$
- La valeur moyenne de la porosité déterminée en laboratoire est de 0,367. Cette valeur est utilisée dans le modèle hydrogéologique.
- Le coefficient d'emmagasinement de la pression est calculé à partir de l'équation 5.2 et des paramètres suivants, en plus de la porosité:
 - La compressibilité du fluide est de $4,47\times 10^{-10}~[kg/(m\cdot s^2)]^{-1}$
 - La compressibilité de la matrice solide est de $1 \times 10^{-8} [kg/(m \cdot s^2)]^{-1}$

Les valeurs de diffusion moléculaire et de dispersivité sont nécessaires pour la simulation de transport.

- La valeur de diffusion moléculaire généralement acceptée de $10^{-9} m^2/s$ pour le sel (NaCl) à 20 °C est utilisé.
- Il est présumé que la dispersivité transversale est d'un ordre de grandeur plus faible que la dispersivité longitudinale, et donc $\alpha_T = \frac{\alpha_L}{10}$

Les valeurs de perméabilité du fluide et de dispersivité longitudinale seront estimées par assimilation de données électriques. Toutefois, la valeur de perméabilité du fluide calculée à partir des mesures de conductivité hydraulique en laboratoire est de $1, 3 \times 10^{-10} m^2$. Cette valeur sera utilisée comme valeur de référence lors de l'estimation de la perméabilité.

La dispersivité longitudinale a, quant à elle, été calibrée à partir d'un modèle numérique à $\alpha_L = 0,001m$ (Constantini, 2015). Dans cette étude, les résultats du modèle numérique ont été comparés à des photographies du panache d'eau saline dans le temps. La dispersivité longitudinale calibrée lors de l'étude numérique devait aussi tenir compte des instabilités numériques pouvant être générées par une dispersivité trop faible pour une maille trop grossière (la finesse de la maille était limitée par des considérations de temps de calcul). La dispersivité ainsi obtenue était un peu plus élevée que la dispersivité réelle puisque la zone de transition était un peu plus épaisse sur le modèle numérique que sur les photographies à un temps donné. Cette valeur de dispersivité longitudinale a tout de même été considérée comme la valeur de référence.

Des conditions frontières de Dirichlet (charge hydraulique constante) permettent de tenir compte des deux bacs d'eau présents aux extrémités du modèle. Lorsque la modélisation se fait avec SUTRAdes pressions plutôt que des charges doivent être spécifiées aux frontières de Dirichlet. Ces pressions sont calculées à partir de la densité du fluide, qui, peut-être variable, et de la définition de la charge hydraulique, h (équation 5.12).

$$h = \frac{P}{\rho g} + z \tag{5.12}$$

où z représente l'élévation à partir d'un datum arbitraire. À la frontière correspondant au bassin d'eau pure (frontière amont), la charge hydraulique est fixée à 0,726 mètres et la densité est celle de l'eau douce, soit 1000 kg/m^3 . La charge hydraulique à la frontière correspondant au bassin d'eau salée (frontière aval) est fixée à 0,706 mètres. Les quatre premiers centimères (à partir du dessus du bassin) sont une frontière d'eau douce, dont la densité est de 1000 kg/m^3 . Le reste de la frontière est une frontière d'eau salée dont la densité est de 1033 kg/m^3 et la concentration est de 0,0456 kg/kg. La salinité de l'eau a été déterminée grâce à une sonde qui mesure la conductivité électrique et la température de l'eau et qui, grâce à une calibration préalable, calcule la concentration en sel de l'eau avec une précision de 10% à partir de ces mesures (Constantini, 2015). La portion d'eau douce de la frontière aval a été fixée pour tenir compte de la poussée de l'eau douce qui force l'eau salée à s'éloigner de cette frontière. Les frontières supérieure et inférieure du bassin sont considérées des frontières à flux nul. De plus, les frontières en amont et en aval du bac de sable qui dépassent les bassins d'eau salée et douce, respectivement, sont aussi considérées des frontières à flux nul, soit environ 7 cm au-dessus du bassin d'eau salée et 5 cm au-dessus du bassin d'eau-douce.

La concentration en contaminant (sel, dans ce cas) du fluide est définie sur toute frontière où le fluide peut intégrer le modèle, soit aux frontières de Dirichlet décrites précédemment. Pour la frontière amont (bassin d'eau douce), la concentration était fixée à 0 kg/kg. La concentration a aussi été fixée à 0 kg/kg pour les quatre premiers centimètres à partir du dessus du bassin d'eau salée qui représentent la frontière aval. La concentration de l'eau salée a été fixée à 0,0456 kg/kg pour les 36 cm inférieurs de la frontière aval représentant le bassin d'eau salée. Cette façon de spécifier la concentration du fluide entrant dans le modèle implique un mélange entre le fluide existant dans le modèle et le fluide entrant et serait donc plus réaliste que de fixer des conditions frontières de concentration (Voss, 1999).

Des conditions initiales de pression et de concentration doivent être spécifiées avant d'exécuter le modèle. Les valeurs de pression initiales ont été obtenues en exécutant le modèle d'écoulement une première fois en régime permanent. Les concentrations initiales étaient nulles sur l'ensemble du modèle.

Une grille d'éléments finis doit être définie. La discrétisation d'un modèle de transport de contaminant est habituellement dirigé par des considérations numériques. Dans le cas de l'assimilation de données par EnKF, le temps de calcul des simulations peut être très restrictif puisque plusieurs simulations doivent être roulées pour obtenir un ensemble de réalisations. Un compromis doit donc être effectué entre le temps de calcul et la précision de la solution. Le modèle calibré a d'abord été roulé sur une grille fine avec une discrétisation de $1mm \times 2,5mm$ qui prenait plusieurs heures. Le modèle avec les mêmes paramètres du fluide, de la matrice solide et de transport, ainsi que les mêmes conditions frontières et une même discrétisation temporelle a ensuite été roulé sur une grille plus grossière avec une discrétisation de $1cm \times 1cm$ qui ne prenait plus que quelques minutes. Les résultats des simulations de transport des deux modèles sont présentés à l'annexe C et permettent d'observer que la différence entre les concentrations calculées sur les deux grilles se retrouvent principalement autour de la zone de transition. Ces différences sont plus importantes après une simulation de 22 heures (le dernier pas de temps de l'assimilation) qu'après une simulation de 14 heures (le premier pas de temps de l'assimilation). Les données de résistance électrique simulées à partir des résultats des simulations hydrogéologiques sur les deux grilles différentes sont très semblables (moins de 5 % de différence). La résolution des données de résistance est fixée par l'espacement entre les électrodes qui est de 3 cm. Ainsi, l'utilisation de la grille fine n'est pas justifiée étant donné son coût computationnel élevé et la faible résolution du protocole électrique.

La discrétisation temporelle a été fixée à des pas de temps de 60 secondes. Des discrétisations temporelles plus fines avaient été testées et n'avaient pas apporté d'amélioration sur la grille fine (Constantini, 2015).

Les figures suivantes présentent la concentration en sel à différents pas de temps simulés sur la grille fine à partir des paramètres de perméabilité hydraulique et de dispersivité longitudinale de référence. Aux figures 5.2a à 5.2c, les concentrations ont été transférées de la grille fine à la grille grossière. Ces valeurs seront utilisées comme référence pour évaluer l'efficacité de l'assimilation de données.



Figure 5.2 – Concentrations simulées sur la grille fine avec les paramètres calibrés (a) après 14 heures (b) après 18 heures (c) après 22 heures

La prédiction des concentrations dans le cadre de l'assimilation par EnKF, se fait en roulant les modèles de transport et d'écoulement à partir du temps 0, comme présenté à la section 3.1.8.

5.3 Modèle électrique

Deux méthodes peuvent être utilisées pour assimiler des données de tomographie électrique dans le but d'estimer des concentrations et/ou des paramètres hydrogéologiques. La première méthode consiste à comparer les données de résistance prédites aux données de résistance observées et sera ici nommée "méthode d'assimilation directe". La seconde méthode consiste à comparer les données de résistivité du sol prédites aux données de résistivité du sol obtenues par l'inversion des résistances observées. Cette seconde méthode sera nommée "méthode d'assimilation par inversion". Dans les deux cas, la relation pétrophysique entre les concentrations et la résistivité électrique ainsi qu'un modèle électrique doivent être définis.

La loi d'Archie a été désignée précédemment pour faire le lien entre les concentrations et la résistivité électrique. Elle sera aussi utilisée dans le contexte d'assimilation de données électriques

pour l'expérience d'intrusion d'eau saline. Le facteur de formation est un coefficient linéaire qui lie la résistivité du fluide à la résistivité totale du sol. La formule de Humble permet de déterminer le facteur de formation, F, pour un sable non-consolidé (Ward, 1991):

$$F = \frac{0.62}{n^{2,5}} \tag{5.13}$$

Dans cette expérience d'intrusion d'eau saline, la configuration de l'expérience avec des billes de verre peut être assimilé à un sable non-consolidé dont la porosité, n a été déterminée en laboratoire.

Le calcul de la résistivité totale du sol par la loi d'Archie nécessite de connaître la conductivité du fluide. La conductivité du fluide est une fonction linéaire de la concentration en sel, tel que présenté à l'équation 3.61. Les conductivités de l'eau salée et de l'eau douce de l'expérience on été mesurées en laboratoire.

- Conductivité de l'eau douce (concentration nulle de NaCl): 0,0478 S/m
- Conductivité de l'eau salée à une concentration de 0,0456kg/kg de NaCl: 6,7295 S/m

La relation entre la conductivité de l'eau, σ_w et la concentration en sel de l'eau, C, devient donc:

$$\sigma_w = \sigma_{w0} + 146,5279 \times C \tag{5.14}$$

où σ_{w0} est la conductivité de l'eau douce.

La loi d'Archie et la relation 5.14 permettent de calculer la conductivité totale du sol, lorsque la concentration en sel, C, est connue. Toutefois, la résistivité du sol n'est pas mesurée directement. Un dispositif pôle-dipôle permet plutôt de mesurer une différence de potentiel à un dipôle en surface pour un courant injecté à un pôle. Cette différence de potentiel est donnée par la loi de Poisson, comme présenté à l'équation 3.62.

La modélisation de ce protocole d'acquisition de données électriques se fait par différences finies tel que présenté en 3.2.3. Pour des raisons de simplicité, la grille électrique a été conçue afin que les noeuds de la grille électrique correspondent aux noeuds de la grille d'éléments finis du modèle hydrogéologique. Les valeurs de conductivité électriques aux noeuds de la grille électrique peuvent donc être calculées à partir des concentrations en sel de l'eau obtenues aux noeuds de la grille du modèle hydrogéologique et de la loi d'Archie et de l'équation 5.14.

Comme mentionné à la section 3.2.3, les conditions frontières définies par défaut par le modèle sont une condition de Neumann à l'interface sol-air et des conditions de Robin à toutes les autres frontières, afin de représenter un demi-espace infini. Toutefois, dans le cas de l'expérience d'intrusion d'eau saline, le milieu n'est pas infini et des conditions de flux-nul sont présentes aux quatre frontières. Idéalement, des conditions de Neumann seraient imposées aux quatre frontières, mais, pour des raisons de simplicité, ces conditions de Neumann sont remplacées par des sections de la grille avec une résistivité très élevée. Dans la figure 5.3, la section bourgogne représente une résistivité de $2 \times 10^{20} \ \Omega \cdot m$ qui remplace les conditions frontières de Neumann, la section jaune représente le bac de sable, la section bleu marine représente le bassin d'eau salée avec une résistivité de 0, 1486 $\Omega \cdot m$ et la section bleue représente le bassin d'eau douce avec une résistivité de 20, 92 $\Omega \cdot m$ (ces valeurs ont été déterminées en laboratoire). Les valeurs de résistivité dans les bassins d'eau douce et d'eau salée et dans la section bourgogne sont constantes tout au long de l'expérience. La section de résistivité très élevée (section bourgogne) dans le bas du modèle électrique permet, entre autres, de représenter la surélévation du bac de sable par rapport aux bassins d'eau douce et salée.



Figure 5.3 – Modèle électrique de l'expérience d'intrusion d'eau saline (légende de couleur décrite dans le paragraphe précédent)

Dans la portion représentant le bac de sable, la discrétisation de la grille de différences finies est la plus fine avec une distance entre les cellules de 3 cm dans les deux directions (longueur et hauteur). La discrétisation de la grille devient plus grossière dans les bassins d'eau douce et salée (6 cm en longueur \times 3 cm en profondeur dans le bassin d'eau salée et 19,2 cm \times 3 cm dans le bassin d'eau douce) et encore plus grossière à l'extérieur, pour les marges représentant la résistivité très élevée. La distance entre les électrodes du protocole d'ERT et le bassin d'eau douce justifie une discrétisation aussi grossière. Par ailleurs, une discrétisation plus fine pour la portion représentant le bac de sable, 1 cm dans la longueur, a aussi été testée. La différence entre les potentiels prédits avec les deux discrétisations ne dépassait jamais 3 %.

La figure 5.4 est une schématisation d'un dispositif pôle-dipôle tel qu'utilisé dans l'expérience d'eau saline. Ce dispositif est consitué d'un pôle d'injection de courant (le pôle A) et d'un dipôle où le potentiel électrique est observé (dipôle M-N). Le pôle d'injection de courant est en fait un dipôle dont le second pôle est situé à l'infini, ou très loin du pôle d'injection. Dans l'expérience de Padoue, le pôle à l'infini est situé à l'extrémité amont du bac de sable.



Figure 5.4 – Schéma du dispositif électrique utilisé pour l'expérience de Padoue (tiré de: Morrison & Gasperikova (2016))

Les électrodes sont disposées sur une ligne suivant l'axe de la longueur du bac de sable et débutant à 2,79 mètres de la frontière amont du bac de sable jusqu'à 4,92 mètres. L'espacement entre les électrodes est de 3 cm. L'électrode de courant à distance est placée à 6 cm de la frontière amont, soit à 2,73 mètres de la première électrode en amont de la ligne pôle-dipôle. Le protocole d'échantillonnage contient deux parties, soit une première partie pour laquelle l'électrode de courant est placée en amont de la ligne pôle-dipôle et une deuxième partie pour laquelle l'électrode de courant est placée en aval de la ligne pôle-dipôle. La distance entre les électrodes du dipôle de potentiel, variable a dans la figure 5.4, varie entre 3 et 18 cm (par incréments de 3 cm) et la distance entre le pôle de courant et le dipôle de potentiel, variable n dans la figure 5.4, varie entre 1 et 12 fois la distance entre les électrodes du dipôle pour le protocole d'acquisition des données. La prise de mesure électrique se fait à un intervalle de 20 minutes et le temps 0 correspond au moment du début de l'expérience d'intrusion d'eau salée.

Les mesures électriques de l'expérience sont représentées par des pseudo-sections. Deux pseudosections séparées sont présentées, soit l'une pour le protocole avec l'électrode de courant en amont du dipôle de potentiel, dit protocole directe, et une pour le protocole avec l'électrode de courant en aval du dipôle de potentiel, dit protocole inverse. Les figures 5.5a à 5.5f représentent les résistivités apparentes prédites à partir du modèle hydrogéologique défini précédemment et du modèle électrique présenté à la figure 5.3, tandis que les figures 5.6a à 5.6f représentent les résistivités apparentes mesurées.



Figure 5.5 – Pseudo-section des résistivités apparentes prédites à partir du modèle calibré (a) après 14 heures protocole directe (b) après 14 heures protocole inverse (c) après 18 heures protocole directe (d) après 18 heures protocole inverse (e) après 22 heures protocole directe (f) après 22 heures protocole inverse



Figure 5.6 – Pseudo-section des résistivités apparentes mesurées (a) après 14 heures protocole directe (b) après 14 heures protocole inverse (c) après 18 heures protocole directe (d) après 18 heures protocole inverse (e) après 22 heures protocole directe (f) après 22 heures protocole inverse

Si des structures similaires sont apparentes sur les pseudo-sections mesurées et simulées, les résistivités apparentes observées sont beaucoup plus élevées que les résistivités apparentes prédites. La configuration de l'expérience semble expliquer ce phénomène. En effet, le modèle électrique qui permet d'obtenir le potentiel aux électrodes du protocole de tomographie est un modèle 2.5 D, c'est-à-dire que le modèle suppose que la résistivité électrique est constante dans la troisième dimension jusqu'à l'infini tel que mentionné à la section 3.2.3. Par contre, les parois du bac de sable dans le sens de la longueur (figure 5.1) impliqueraient plutôt des frontières de flux-nul dans cette troisième dimension. Ces parois de flux-nul pourraient expliquer les résistivités apparentes observées plus élevées que les résistivités apparentes simulées. Par ailleurs, l'espacement entre les électrodes affecte la sensibilité des résistivités observées. Plus cet espacement est grand, plus les mesures seront sensibles à des variations de résistivité dans le sol, éloignées de la ligne d'acquistion des données.

De plus, pour un même espacement des électrodes du dipôle de potentiel et entre le dipôle de potentiel et le pôle de courant, des différences importantes peuvent être observées entre les mesures obtenues pour le protocole directe et pour le protocole inverse. Ces différences peuvent s'expliquer par le fait que le deuxième pôle de courant, à l'infini, est placé à l'extrémité amont du bac de sable. Ainsi, lorsque les mesures sont prises pour le protocole directe, les électrodes ont une configuration ressemblant à un protocole dipôle-dipôle, tandis que le protocole inverse ressemble plus à un protocole Wenner-Schlumberger, c'est-à-dire, les électrodes de potentiel sont placées entre les électrodes de courant.

Les contraintes de temps n'ont pas permis de répéter des mesures afin d'obtenir une estimation de la variance de l'erreur de mesure. L'écart-type a donc été modélisée par une relation linéaire, en supposant que l'écart-type est proportionnel à la donnée mesurée. La relation suivante a été fixée:

$$\sigma_e = 0,5735 \times |R| + 0,022 \tag{5.15}$$

où σ_e est l'écart-type de l'erreur, |R| est la valeur absolue de la résistance mesurée et l'ordonnée à l'origine, 0,022, représente la résistance minimale qui peut-être mesurée par l'appareil. Un coefficient très élevé, soit 57,35% a été choisi. Ce coefficient devrait permettre d'assimiler les données mesurées efficacement sans sur-évaluer l'importance des données mesurées dans le processus d'assimilation.

5.3.1 Inversion électrique

La modélisation directe permet de calculer le potentiel électrique dû à une source grâce à l'équation de Poisson (équation 3.62) lorsque la conductivité électrique du sol (σ_{sol}) est connue. Toutefois, lorsque le sol est sondé par tomographie électrique, il est plutôt question de retrouver la conductivité du sol à partir de mesures de potentiel à la surface ou en forage. L'inversion électrique permet d'estimer la valeur de la conductivité du sol à partir de ces mesures. Ce qui suit présente l'inversion électrique telle qu'elle a été implémentée pour estimer les valeurs de conductivité du sol dans l'expérience d'intrusion d'eau saline. Une discussion plus complète sur l'inversion électrique peut être trouvée dans Gunther (2004).

En général, le processus d'inversion cherche à reconstituer les paramètres, m, à partir des données, d, lorsque ces données sont reliées aux paramètres par une fonction: d = f(m) + e où ereprésente une erreur de mesure. Ainsi, il s'agira de minimiser les résidus au carré, soit, la différence entre les données observées et les valeurs prédites à partir de la fonction et des paramètres. Comme l'amplitude de l'erreur sur les données peut varier d'une donnée à l'autre, il sera possible de mettre des poids variables sur les données afin de tenir compte de ces erreurs. Il faudra donc minimiser la norme L_2 des résidus pondérés:

$$\Phi_d = ||D(d - f(m))||_2^2 \tag{5.16}$$

où D représente une matrice de poids associés aux données. Dans le cas de l'inversion de données de résistivité électrique, l'hypothèse de résidus gaussiens est posée ce qui justifie l'utilisation de la norme L_2 , mais d'autres normes pourraient être utilisées.

Comme dans la plupart des cas d'inversion géophysique, il y a plus de paramètres pour représenter le sol à l'étude qu'il n'y a de données, le problème est mal posé. Ajouter un terme de régularisation à la fonction à minimiser permettra de résoudre ce problème mal-posé. Le terme de régularisation, donné en 5.17, permettra aussi de limiter des forts contrastes dans les paramètres inversés.

$$\Phi_m = ||C(m - m_0)||_2^2 \tag{5.17}$$

où C représente une matrice de contrainte et m_0 est le modèle initial suggéré.

Un poids de régularisation, λ est aussi intégré à la fonction finale à minimiser:

$$\Phi = \Phi_d + \lambda \Phi_m \tag{5.18}$$

La minimisation de cette équation se fait par méthode de Gauss-Newton, ce qui permet d'obtenir l'équation 5.19 pour l'itération, k:

$$m^{k+1} = m^k + (J^T D^T D J + \lambda C^T C)^{-1} [J^T D^T D (d - f(m^k) - \lambda C^T C (m^k - m^0)]$$
(5.19)

110

À partir de cette équation et du développement de d = f(m) + e en série de Taylor de premier ordre, il est possible de développer une approximation des paramètres estimés (équation 5.22).

D'abord l'inverse généralisé, J^{\dagger} , est défini:

$$J^{\dagger} = (J^T D^T D J + \lambda C^T C)^{-1} J^T D^T$$
(5.20)

Ainsi que la matrice de résolution, R

$$R = J^{\dagger} D J \tag{5.21}$$

$$m^{est} = Rm^{vrai} + (I - R)m^0 + J^{\dagger}De$$
(5.22)

À partir de l'équation 5.22, il est possible de définir la covariance d'estimation des paramètres, en notant que la covariance des vrais paramètres est nulle $(Cov(m^{vrai}) = 0)$ et que la covariance des résidus est égale à la covariance des données Cov(e) = Cov(d). De plus, dans l'implémentation de l'inversion des données électriques de l'intrusion d'eau saline, la matrice de poids, D est une matrice diagonale de l'inverse des écarts-types des données, soit, $Cov(e) = Cov(d) = D^{-2}$

$$Cov(m^{est}) = (I - R)Cov(m^0)(I - R)^T + J^{\dagger}DCov(d)(J^{\dagger}D)^T$$
(5.23)

La première partie de l'équation 5.23 concernant la covariance du modèle initial est habituellement ignorée, puisque cette covariance n'est pas connue (Gunther, 2004).

Par ailleurs, l'inversion est ici effectuée sur le logarithme de la résistivité afin d'éviter des valeurs négatives. Dans ce cas, les données observées sont aussi transformées pour obtenir leur logarithme dans la procédure d'inversion.

Un autre paramètre à considérer lors de l'inversion est la valeur du poids de régularisation, λ . Ce poids de régularisation a été fixé à 0,1 pour la première itération et diminue de moitié pour les itérations subséquentes. Enfin, un maximum de 15 itérations ont été effectuées pour chaque inversion. Les figures 5.7a,5.7b et 5.7c montrent que des itérations subséquentes n'apporteraient probablement que de très faibles améliorations.



Figure 5.7 – RMS des données de l'inversion en fonction du nombre d'itérations pour l'inversion (a) après 14 heures (b) après 18 heures (c) après 22 heures



Figure 5.8 – Résultat de l'inversion (a) après 14 heures (b) après 18 heures (c) après 22 heures

Les figures 5.8a à 5.8c présentent les résultats de l'inversion des résistances mesurées en laboratoire à différents pas de temps lors de l'expérience d'intrusion d'eau saline. Les résultats de l'inversion présentent des faibles résistivités aux faibles profondeurs à tous les pas de temps. Cette couche peu résistante peut aussi être observée sur les pseudo-sections à faible profondeur, en particulier sur les pseudo-sections représentant le protocole directe (figures 5.6b, 5.6d, 5.6f). Cette couche est aussi présente sur le résultat de l'inversion des résistances mesurées avant le début de l'intrusion d'eau saline (figure 5.9).

Par ailleurs, ces artéfacts d'inversion horizontaux à faible profondeur peuvent aussi être observés sur les résultats de l'inversion sur un modèle synthétique d'intrusion d'eau saline (annexe D). Toutefois, les résistivités inversées ne sont pas aussi faibles en surface sur le modèle synthétique. Ainsi, ces très faibles résistances doivent être dues à un phénomène qui n'est pas pris en compte dans le modèle électrique. La présence de frontières imperméables dans la troisième dimension n'est pas prise en compte dans l'inversion et pourrait influencer les valeurs des résistances. Plus l'espacement entre les électrodes est grand, plus la mesure sera sensible à des variations de résistance éloignées des électrodes. Ainsi, les mesures prises avec des électrodes très espacés seraient plus influencées par ces frontières imperméables que les mesures prises avec des électrodes à faible espacement. De plus, des grilles métalliques séparent le bassin d'eau salée du bac de sable et pourraient aussi causer une couche de faible resistivité en surface. La présence de frontières imperméables dans la direction perpendiculaire au plan du modèle électrique et de grilles métalliques couplée au protocole d'acquisition pourrait expliquer la présence de ces faibles résistivités en surface lorsque les mesures électriques sont inversées.



Figure 5.9 – (a) Pseudo-section des résistivités apparentes mesurées avant le début de l'expérience (a) protocole directe (b) protocole inverse (c) résultat de l'inversion

De plus, comme des différences importantes ont été notées entre les résistivités apparentes mesurées et prédites par le modèle hydrogéologique calibré, il est probable que ces mêmes différences se retrouvent dans les résistivités inversées par rapport aux résistivités prédites à partir du modèle hydrogéologique calibré. L'inversion des mesures de résistance nécessite le calcul d'un modèle électrique directe qui est, encore une fois, un modèle 2,5D qui ne tient donc pas compte des frontières de flux-nul dans la troisième dimension. Les figures 5.10a à 5.10c montrent les résistivités électriques calculées à partir des concentrations en sel du modèle de contamination calibré et de la loi d'Archie avec le facteur de formation défini en 5.13. Les résistivités sont environ 10 fois plus faibles que les résistivités calculées grâce à l'inversion des données mesurées. L'assimilation des résistivités inversées pour la mise à jour des paramètres hydrogéologiques devra tenir compte de cette différence de magnitude.



Figure 5.10 – Résistivités électriques prédites à partir du modèle calibré (a) après 14 heures (b) après 18 heures (c) après 22 heures

5.4 Assimilation des données

Une procédure d'assimilation des données par filtre de Kalman d'ensemble a été mise en place pour estimer, en plus de la concentration à chaque pas de temps, la dispersivité longitudinale ainsi que la perméabilité hydraulique. Deux différentes méthodes d'assimilation des résistances électriques observées ont été comparées. La première méthode consiste à calculer des résistances électriques prédites à partir du modèle électrique décrit en 5.3 et des concentrations en sel transformées en résistivités électriques grâce à la loi d'Archie, nommée méthode d'assimilation directe. La deuxième méthode, nommée méthode d'assimilation par inversion, consiste à transformer les concentrations en sel prédites en résistivités électriques prédites, puis à comparer ces résistivités à celles obtenues par l'inversion des données de résistance mesurées par la tomographie électrique.

Dans les deux cas, l'assimilation des données électriques, qu'elles soient les mesures directes ou les résistivités inversées, permettra, dans un premier temps, de mettre à jour les concentrations. Les valeurs des concentrations incluses dans l'état se limitent aux concentrations qui se trouvent à plus de trois mètres de la frontière amont. Pour la durée de l'expérience, la contamination est restée circonscrite dans cet espace. De plus, tout comme pour le cas synthétique, la prédiction des concentrations se fait à partir du temps 0, c'est-à-dire avant le début de l'intrusion d'eau saline pour chaque pas de temps (méthode des *Restart-EnKF*).

Ce nouvel ensemble de concentrations mises à jour sera ensuite assimilé pour mettre à jour l'ensemble des paramètres, c'est-à-dire, la perméabilité hydraulique et la dispersivité longitudinale. Cette méthode a été nommée EnKF découplé et a été présentée à la section 3.1.6.

L'assimilation se fait donc en deux étapes. L'état à la première étape est donné par $\mathbf{x} = \begin{bmatrix} C \\ R \end{bmatrix}$ où R est soit la résistance observée ou la résistivité inversée.

À la deuxième étape, l'état est donné par
$$\mathbf{x} = \begin{bmatrix} log_{10}(\kappa) \\ log_{10}(\alpha) \\ C \end{bmatrix}$$
.

À la deuxième étape, seul un sous-ensemble de la variable intermédiaire, C, est utilisée pour la mise à jour des paramètres κ et α : la concentration est donnée aux noeuds de la grille d'éléments finis. Seuls les noeuds auxquels aucun membre de l'ensemble de concentrations prédites ne présentait une concentration nulle ont été conservés pour la mise à jour des paramètres. Utiliser seulement un sous-ensemble des concentrations devrait permettre de stabiliser la solution, puisque moins de données seront utilisées pour estimer les paramètres.

La figure 5.11 représente cette méthodologie dans le contexte de l'expérience d'intrusion d'eau saline en laboratoire où α représente la dispersivité longitudinale, κ représente la perméabilité hydraulique, C représente la concentration en sel, R représente la résistance électrique observée et ρ représente la résistivité électrique obtenue par inversion. Deux différentes méthodes ont été développées pour l'étape de mise à jour de la variable intermédiaire selon que la donnée assimilée soit les résistances, R, ou les résistivités inversées, ρ . Ces méthodes seront explicitées dans les sections appropriées.



Figure 5.11 – Schéma de l'assimilation des données électriques pour l'estimation de la concentration

L'ensemble initial a été créé de la même façon pour l'assimilation des données de résistances et des données de résistivité inversée. Un ensemble de paramètres a d'abord été créé. Les deux paramètres qui doivent être mis à jour, soit la perméabilité et la dispersivité longitudinale, sont considérés homogènes dans le bac de sable étant donné la façon dont cette expérience a été mise en place. Les deux paramètres sont aussi considérés indépendants. Chaque paramètre est transformé en logarithme en base 10. Il a été décidé de représenter l'ensemble initial du logarithme de chacun des paramètres par une loi uniforme, centrée sur la valeur calibrée du paramètre. L'ensemble initial s'étend sur deux ordres de grandeur dans le cas de la dispersivité et quatre ordres de grandeur dans le cas de la perméabilité. La loi uniforme est la densité de probabilité qui maximise l'entropie (ou de façon heuristique, qui correspond à une connaissance minimale (Conrad, 2005)) pour une distribution bornée. La borne minimale de la distribution de la dispersivité est donc définie par le logarithme en base 10 de la valeur calibrée moins un et la borne maximale est définie par le logarithme de la valeur calibrée plus un, pour représenter deux ordres de grandeur. Les bornes minimale et maximale de la distribution de perméabilité sont définies par le logarithme en base 10 de la valeur calibrée moins 2 et plus 2, respectivement, pour représenter un total de quatre ordres de grandeur.

Afin d'assurer que les deux paramètres ne soient pas corrélés, l'ensemble initial est créé à partir de toutes les permutations possibles des valeurs de chaque paramètre. Ainsi, l'ensemble initial contient 15 différentes valeurs de perméabilité et 15 valeurs de dispersivité. En faisant toutes les permutations entre des valeurs de perméabilité et de dispersivité, l'ensemble initial final contient $225 (15 \times 15)$ membres et la covariance empirique entre les deux paramèters est nulle. La taille de l'ensemble a été fixée à 225 membres. Même si aucune analyse de sensibilité n'a été effectuée pour déterminer la taille optimale de l'ensemble optimal, des critères peuvent permettre de diagnostiquer un ensemble trop petit, tel qu'une étendue de l'ensemble qui diminue de façon importante. De plus, le cas synthétique a pu donner une idée de la taille de l'ensemble à générer. En effet, l'assimilation de résistances par la méthodologie découplée améliorait l'estimation de la conductivité hydraulique pour un ensemble de 200 membres lorsque les paramètres géostatistiques étaient connus (figure 4.12). Dans le cas présent, seuls deux paramètres (plutôt que 5800 points de conductivité hydraulique) sont à calibrer et ces paramètres sont considérés non-corrélés. Un ensemble d'un peu plus de 200 membres a donc été considéré suffisant pour tester l'assimilation.

La figure 5.12 présente l'histogramme de l'ensemble initial pour chacun des paramètres. La ligne noire indique la valeur qui avait été calibrée manuellement. Tous les autres paramètres du modèle hydrogéologique sont restés invariables. Ce modèle a permis d'obtenir l'ensemble initial des concentrations grâce au modèle hydrogéologique SUTRA, présenté préalablement.



Figure 5.12 – Histogramme de l'ensemble intial du logarithme en base 10 (a) de la perméabilité (b) de la dispersivité longitudinale

5.4.1 Méthode d'assimilation directe

À la section 5.3, la différence entre les résistances mesurées et les résistances prédites à partir du modèle calibré a été discutée. Des différences entre les mesures du protocole directe et du protocole inverse ont aussi été présentées. Afin de tenir compte de ces deux phénomènes, il a été décidé de faire l'assimilation des données de résistance électrique mesurées séquentiellement (voir la section 3.1.7)en fonction de l'espacement entre l'électrode de courant et le dipôle de potentiel. Seules les mesures du protocole inverse ont été utilisées puisqu'elles étaient plus rapprochées de la frontière aval et donc, du panache d'eau salée. Trois espacements du protocole inverse ont donc été sélectionnés: 0,81 mètres, 1,08 mètres et 1,50 mètres.

L'utilisation d'un ratio de résistances (résistance au pas de temps, t par rapport à la résistance au temps t_0 , c'est-à-dire, avant le début de l'intrusion d'eau saline), comme pour le cas synthétique, aurait pu être considéré pour s'affranchir du facteur linéaire entre les résistances mesurées et prédites. Toutefois, l'utilisation de cette méthode nécessite le respect de certaines conditions. En particulier, la résistivité électrique du milieu au temps t_0 doit être minimalement homogène. Dans le cas de l'expérience d'intrusion d'eau saline, les bassins d'eau douce et d'eau salée aux extrémités du bac de sable ne permet pas de considérer la résistivité électrique du milieu initial comme étant homogène. De plus, ce facteur doit être constant dans le temps ce qui pourrait ne pas être le cas, en particulier pour un espacement de 0,81 mètres (voir le tableau 5.1). Donc, la résistance au pas de temps t a été utilisée comme donnée observée et un facteur linéaire a été intégré à l'assimilation pour tenir compte de la relation entre les résistances mesurées et prédites.

Des graphiques de corrélation, ainsi que des régressions linéaires, entre les résistances mesurées et les résistances prédites (à partir du modèle calibré) pour chacun de ces espacements ont confirmé qu'un lien linéaire existait entre les deux, avec des coefficients de détermination dépassant 0,8. Le tableau 5.1 présente la pente obtenue par régression linéaire pour chaque espacement aux différents pas de temps. Les figures 5.13a à 5.13c montrent les diagrammes croisés des valeurs de résistance mesurée en fonction de la résistance prédite qui ont mené à ces régressions linéaires au premier pas de temps. L'ordonnée à l'origine a été fixée à zéro dans toutes les régressions, afin de s'assurer qu'une résistance prédite nulle correspondait à une résistance observée nulle. Dans tous les cas, les diagrammes croisés montrent une structure non-linéaire. Ceci est particulièrement évident pour un espacement de 108 cm, puisque les points de résistances élevées s'éloignent de la courbe de régression, mais peut aussi être observé pour les autres espacements. Ces divergences de la linéarité sont faibles et n'auront possiblement qu'un effet négligeable sur l'assimilation.

Espacement	Temps écoulé				
(cm)	(h)				
	14	16	18	20	22
81	3,31	3,37	3,44	$3,\!50$	3,53
108	3,37	3,38	3,40	3,42	3,41
150	3,28	3,29	$3,\!29$	$3,\!28$	3,25

Tableau 5.1 – Pentes des régressions entre les résistances prédites et mesurées pour différents espacements entre les électrodes



Figure 5.13 – Résistance mesurée par rapport à la résistance prédite 14 heures après le début de l'intrusion pour un espacement de (a) 81 cm (b) 108 cm (c) 150 cm

Le facteur de proportionnalité varie d'un pas de temps à l'autre, en particulier pour un espacement de 81 cm. De plus, ce facteur varie selon l'espacement entre les électrodes. Toutefois, ces variations sont assez faibles. Il a été décidé d'ajuster ce facteur de proportionnalité automatiquement dans l'assimilation dans l'espoir de tenir compte de ces variations. Ce facteur de proportionnalité devient donc un facteur auto-calibrant. L'assimilation s'est faite en quatre étapes pour chaque pas de temps. À la première étape, l'état était donné par:

 $\mathbf{x} = \begin{bmatrix} F_1 \\ R_1 \\ C_1 \end{bmatrix}$ où F_1 est le facteur de proportionalité (ou facteur auto-calibrant) liant la résistance

mesurée et la résistance prédite pour l'espacement 1, R_1 est la résistance et C_1 est la concentration.

Dans le modèle de prédiction, C_1 est obtenu par le modèle hydrogéologique et les paramèteres de perméabilité et de dispersivité estimés au temps précédent. Si R_1^* est la résistance prédite à partir des concentrations prédites C_1 , de la loi d'Archie, et du modèle électrique, alors $R_1 = R_1^* \times F_1$. Le modèle électrique est ensuite calculé sur les concentration mises à jour à l'étape 1, ce qui permet d'obtenir, une résistivité prédite, R_2 à partir du facteur de proportionnalité (ou auto-calibrant) F_2 . Et, ainsi de suite, jusqu'à ce que toutes les résistances aient été assimilées pour obtenir une concentration mise à jour finale. Cette méthode d'assimilation des résistances en plusieurs étapes est une méthode d'assimilation itérative, qui s'apparente aux méthodes présentées en 3.1.8 pour assurer la conservation de la masse. Dans le cas présent, l'assimilation itérative a été mise en place pour assurer l'indépendance des facteurs de proportionnalité d'un espacement à l'autre. Un schéma de la méthodologie proposée pour mettre à jour les concentrations à partir de données de résistance observées en tenant compte d'un facteur de proportionnalité entre les résistances observées et les résistances prédites est présenté à la figure 5.14.



Figure 5.14 – Schéma de l'assimilation des données électriques pour l'estimation de la concentration

L'estimé final de la concentration est assimilée à son tour pour mettre à jour les paramètres de perméabilité et dispersivité hydrauliques comme dans la méthode nommée découplée présentée en 3.1.6 et aussi mise en place das le cas synthétique.

Un ensemble initial (a priori) est créé pour chaque facteur de proportionnalité auto-calibrant, F_1 , F_2 et F_3 . Comme ce facteur ne varie pas beaucoup en fonction de l'espacement, l'ensemble initial pour F_1 , F_2 et F_3 est initialisé à partir d'une même loi uniforme sur l'intervalle [1, 9; 4, 9]. Cet intervalle a été choisi en se basant sur les pentes des régressions linéaires entre résistances mesurées et prédites à partir du modèle calibré. De plus, comme ce facteur n'est pas un paramètre d'intérêt (il sert seulement à améliorer l'estimation des concentrations) et que son évolution dans le temps n'est pas connue, il est ré-initialisé à chaque pas de temps.

Les figures 5.15a à 5.15c et 5.16a à 5.16c montrent la moyenne de l'ensemble des concentrations avant et après l'assimilation des mesures de résistance électrique pour 3 pas de temps. Les figures 5.17a à 5.17c montrent les concentrations obtenues par le modèle hydrogéologique calibré pour fins de comparaison.



Figure 5.15 – Moyenne des concentrations (kg/kg) prédites par les EnKF (a) à 14 heures (b) à 18 heures (c) à 22 heures. Les dimensions des schémas sont en mètres.



Figure 5.16 – Moyenne des concentrations mises à jour par les EnKF (a) à 14 heures (b) à 18 heures (c) à 22 heures



Figure 5.17 – Concentrations simulées avec les paramètres calibrés (a) à 14 heures (b) à 18 heures (c) à 22 heures



Figure 5.18 – Moyenne de déviation absolue (AAB) des concentrations prédites et mises à jour en fonction du pas de temps



Figure 5.19 – Ensemble du facteur auto-calibrant après l'assimilation pour un espacement de (a) 81 cm (b) 108 cm (c) 150 cm

Au premier pas de temps, il est possible d'observer une différence notable entre la moyenne des concentrations prédites et mises à jour par filtres de Kalman d'ensemble. Avant l'assimilation, le panache de contamination moyen se rendait à 3 mètres. L'ensemble initial de perméabilités hydrauliques contenait des perméabilités 2 ordres de grandeur plus élevées que le modèle calibré. La contamination associée à ces perméabilités se rendait donc beaucoup plus loin que la contamination du modèle calibré. L'assimilation de données électriques permet de contraindre la contamination à une distance qui ressemble plus au panache obtenu avec le modèle calibré. Toutefois des artéfacts de contaminant apparaissent au-dessus du panache qui ne semblent pas réalistes d'un point de vue hydrogéologique. Au troisième pas de temps, l'assimilation permet d'allonger le panache pour qu'il représente mieux la contamination réelle par rapport au panache prédit. Même si le panache *a posteriori* est plus long, la concentration diminue rapidement après 4,5 mètres ce qui n'est pas observé, ni sur le panache prédit, nis sur le panache obtenu à partir du modèle calibré.

Au dernier pas de temps, la moyenne des concentrations ne varie pas beaucoup après l'assimilation par rapport à la prédiction. La moyenne des concentrations prédites est déjà très proche des concentrations simulées avec le modèle calibré. Il y a donc peu d'ajustement à effectuer aux concentrations. La figure 5.18 qui présente la moyenne de déviation absolue (AAB) pour les concentrations avant et après l'assimilation pour chaque pas de temps, confirme qu'au dernier pas de temps, l'assimilation des données électriques n'est plus utile pour corriger les concentrations.

Les diagrammes de quartiles de la figure 5.19 montrent que les facteurs auto-calibrants restent généralement centrés autour de la valeur obtenue par la régression. Plus l'assimilation progresse (plus le nombre de pas de temps augmente), plus la variance de l'ensemble du facteur auto-calibrant après la mise à jour est faible même si l'ensemble *a priori* est toujours le même. Ceci peut s'expliquer par l'ensemble de résistance prédites qui est moins variable après plusieurs pas d'assimilation.

Les figures 5.20a à 5.20d et 5.21a à 5.21d présentent les histogrammes des ensembles de perméabilité hydraulique et de dispersivité respectivement tel qu'ils évoluent dans le processus d'assimilation des données de résistance électrique. Même si une loi uniforme est utilisée pour représenter l'ensemble initial de ces paramètres, il est attendu que les ensembles mis à jour représentent plutôt des lois gaussiennes. Plus le nombre de pas de l'assimilation augmente, plus les histogrammes de ces deux paramètres tendent vers des lois normales et plus, la variance de l'ensemble de chaque paramètre diminue. Ces deux phénomènes sont beaucoup plus prononcés pour la perméabilité, tandis que la dispersivité semble beaucoup moins sensible à l'assimilation des données de résistance électrique. En effet, une perméabilité plus élevée engendre un panache de concentration qui se rend plus loin, ce qui pourra se réfléter sur les mesures de résistance électrique. La dispersivité, quant à elle, aura surtout un effet sur la zone de transition entre l'eau salée et l'eau douce. Dans le cas à l'étude, il y a lieu de croire que la dispersivité réelle était plus faible que la valeur calibrée de 0,001 mètres. Avec une dispersivité aussi faible, la zone de transition est assez restreinte. Les mesures de résistance intègrent la résistivité électrique sur une portion du sol. Comme mentionné précédemment, la résolution de ces mesures est fonction de l'espacement entre les électrodes. Il semblerait donc que cet espacement n'est pas assez faible pour avoir une résolution permettant de bien délimiter une zone de transition aussi restreinte.



Figure 5.20 – Histogramme de l'ensemble du logarithme en base 10 de la perméabilité hydraulique (a) ensemble initial (b) ensemble mis à jour à jour après 14 heures (c) ensemble mis à jour à jour après 18 heures (d) ensemble mis à jour après 22 heures



Figure 5.21 – Histogramme de l'ensemble du logarithme en base 10 de la dispersivité longitudinale (a) ensemble initial (b) ensemble mis à jour à jour après 14 heures (c) ensemble mis à jour à jour après 18 heures (d) ensemble mis à jour après 22 heures

5.4.2 Méthode d'assimilation par inversion

L'assimilation a ensuite été effectuée avec les données de résistivité électrique inversée. Le logarithme en base 10 des données inversées a été utilisé pour l'inversion. Ainsi, les résistivités prédites à partir des concentrations prédites ont aussi été transformées en base logarithmique. Comme observé à la section 5.3.1, des différences importantes apparaissent entre les résistivités inversées et les résistivités prédites avec le modèle calibré. Ces différences sont probablement dues aux frontières de flux-nul dans la troisième dimension qui sont négligées dans le modèle électrique bi-dimensionnel. Un facteur auto-calibrant, F_a , ajouté à la prédiction des résistivités permettra de tenir compte de
	Temps écoulé (h)						
	14	16	18	20	22		
Différence	3,30	3,05	2,88	2,75	$2,\!65$		

Tableau 5.2 – Différence moyenne entre les résistivités obtenues par inversion des mesures de résistance et les résistivités prédites grâce au modèle calibré en base logarithmique

cette différence. Ce facteur auto-calibrant, F_a est similaire au facteur auto-calibrant présenté à la section précédente. Toutefois, dans la méthode d'assimilation par inversion, le facteur est additif plutôt que multiplicatif comme c'était le cas pour l'assimilation directe. Ce facteur est ajouté à l'état qui devient:

$$\mathbf{x} = \begin{bmatrix} F_a \\ log_{10}(\rho) \\ C \end{bmatrix}$$
. Le logarithme de la résistivité prédite est donc donné par $log_{10}(\rho) = log_{10}(\rho_{archie}) + C$

 F_a où ρ_{archie} est la résistivité prédite à partir de la loi d'Archie, des concentrations et du facteur de formation défini en 5.13.

La différence moyenne a d'abord été calculée entre le logarithme des résistivités prédites et obtenues par inversion (tableau 5.2) pour déterminer l'ensemble initial du facteur F_a .

Le facteur F_a est initialisé par une loi uniforme sur l'intervalle [0,4]. Comme l'évolution de ce facteur dans le temps n'est pas connu, il est réinitialisé à chaque pas de temps.

La covarariance des données assimilées, soit les résistivités inversées a été calculé à partir de l'équation 5.23. Toutefois, comme cette matrice de covariance n'était pas définie-positive, il a été décidé de considérer l'erreur de mesure d'un point de résistivité à un autre nul. Ceci permettait d'obtenir une matrice diagonale qui aiderait à stabiliser l'assimilation.

Les figures 5.22a à 5.22c et 5.23a à 5.23c montrent la moyenne de l'ensemble des concentrations avant et après l'assimilation des résistivités électriques inversées pour 3 pas de temps. Les concentrations mises à jour après le premier pas de temps ont un comportement similaire aux concentrations mises à jour par assimilation des mesures de résistance. La mise à jour permet de contraindre la contamination à 1,5 mètres de la frontière aval tout en créant des artéfacts au-dessus du panache de contamination. Toutefois, dans ce cas-ci, une diminution de la concentration apparaît aussi dans le panache en se rapprochant de la frontière aval. Cette diminution de concentration est peu réaliste d'un point de vue hydrogéologique et, est probablement dû à la forme de la résistivité inversée (plus forte au centre et qui diminue en s'éloignant du centre). La figure 5.25 indique que la mise à jour de la concentration aux quatrième et cinquième pas de temps ne sont plus très bénéfiques par rapport à la contamination prédite. La même chose était observée pour l'assimilation des données de résistance mesurée.



Figure 5.22 – Moyenne des concentrations prédites par les EnKF (a) à 14 heures (b) à 18 heures (c) à 22 heures



Figure 5.23 – Moyenne des concentrations mises à jour par les EnKF (a) à 14 heures (b) à 18 heures (c) à 22 heures



Figure 5.24 – Concentrations simulées avec les paramètres calibrés (a) à 14 heures (b) à 18 heures (c) à 22 heures



Figure 5.25 – AAB des concentrations prédites et mises à jour en fonction du pas de temps

Les figures 5.26a à 5.26d et 5.27a à 5.27d présentent les histogrammes de l'évolution des ensembles de perméabilité hydraulique et de dispersivité dans l'assimilation des données de résistivité inversée. Encore une fois, les ensembles de chacun des paramètres hydrauliques tendent vers une loi gaussienne après l'assimilation des données électriques. La variance de chaque ensemble diminue aussi au fur de l'assimilation. Même si les ensembles du logarithme de la perméabilité hydraulique ne sont pas centrés autour de la valeur calibrée après l'assimilation, la différence entre la moyenne et la valeur calibrée n'est d'environ que de 0,3; indiquant une estimation qui devrait permettre une prédiction de la concentration proche de la réalité. La variance de l'ensemble de perméabilités diminue de façon plus importante avec l'assimilation avec données inversées qu'avec l'assimilation avec données directes.

Ceci peut aussi être observé sur l'ensemble de la dispersivité longitudinale. D'ailleurs la dispersivité longitudinale mise à jour s'éloigne de la valeur calibrée plus l'assimilation progresse. Cette correction biaisée de la dispersivité pourrait être due à de fausses corrélations dû à l'écrasement de l'ensemble beaucoup plus prononcé que pour l'assimilation avec les données de résistances directes. Dans le cas à l'étude, l'assimilation des résistivités inversées est plus encline à l'écrasement de l'ensemble que l'assimilation directe des résistances, étant donné que plus de données sont utilisées dans l'assimilation des résistivités inversées que dans l'assimilations des résistances. De plus, seule la diagonale de la matrice d'erreur des résistivités inversées est utilisée dans l'assimilation. Ainsi, toutes les covariances hors-diagonal sont ignorées. Ceci pourrait aussi affecter l'écrasement de l'ensemble.



Figure 5.26 – Histogramme de l'ensemble du logarithme en base 10 de la perméabilité hydraulique (a) ensemble initial (b) ensemble mis à jour à jour après 14 heures (c) ensemble mis à jour à jour après 18 heures (d) ensemble mis à jour après 22 heures



Chapitre 5. Assimilation des données de résistance électrique lors d'une expérience d'infiltration d'eau salée 129



Figure 5.27 – Histogramme de l'ensemble du logarithme en base 10 de la dispersivité longitudinale (a) ensemble initial (b) ensemble mis à jour à jour après 14 heures (c) ensemble mis à jour à jour après 18 heures (d) ensemble mis à jour après 22 heures

5.5 Sensibilité à l'ensemble initial de l'assimilation des données

L'assimilation des données a été effectuée précédemment sur un ensemble initial centré sur des valeurs de perméabilité hydraulique et de dispersivité préalablement calibrés. Il peut être intéressant de déterminer comment l'algorithme d'assimilation réagirait si l'ensemble initial n'était pas centré sur les bonnes valeurs. Étant donné que la perméabilité hydraulique semble être le paramètre le plus sensible à l'assimilation de données électriques, seul l'ensemble initial de ce paramètre a été modifié. La moyenne de l'ensemble initial de la perméabilité hydraulique était plus faible que la valeur calibrée. Les figures 5.28a et 5.28b montrent les histogrammes des ensembles initiaux du

logarithme de la perméabilité hydraulique et de la dispersivité longitudinale. La valeur calibrée est indiquée par la ligne noire sur l'histogramme.



Figure 5.28 – Histogramme de l'ensemble intial du logarithme en base 10 (a) de la perméabilité (b) de la dispersivité longitudinale

5.5.1 Méthode d'assimilation directe

Les figures 5.29a à 5.29c et 5.30a à 5.30c montrent la moyenne de l'ensemble des concentrations avant et après l'assimilation des résistivités électriques inversées pour 3 pas de temps. Aux trois pas de temps, l'assimilation des données de résistance semble améliorer l'estimation de la concentration par rapport aux concentrations prédites. Ceci est confirmé par la statistique d'erreur AAB sur les concentrations (figure 5.32). Par ailleurs, les facteurs auto-calibrants semblent aussi généralement centrés autour des valeurs obtenues par régression linéaire. Le facteur auto-calibrant pour l'espacement de 108 cm est légèrement sous la valeur obtenue par régression linéaire, mais ceci n'affecte pas l'estimation des concentrations de façon importante.



Figure 5.29 – Moyenne des concentrations prédites par les EnKF (a) à 14 heures (b) à 18 heures (c) à 22 heures



Figure 5.30 – Moyenne des concentrations mises à jour par les EnKF (a) à 14 heures (b) à 18 heures (c) à 22 heures



Figure 5.31 – Concentrations simulées avec les paramètres calibrés (a) à 14 heures (b) à 18 heures (c) à 22 heures



Figure 5.32 – AAB des concentrations prédites et mises à jour en fonction du pas de temps



Figure 5.33 – Ensemble du facteur auto-calibrant après l'assimilation pour un espacement de (a) 81 cm (b) 108 cm (c) 150 cm

Les figures 5.34a à 5.34d et 5.35a à 5.35d présentent les histogrammes de l'évolution des ensembles de perméabilité hydraulique et de dispersivité dans l'assimilation des données de résistivité inversée. La dispersivité semble, encore une fois, peu sensible à l'assimilation des données de résistances. La distribution de la perméabilité, quant à elle, se rapproche de la valeur calibrée avec la progression de l'assimilation des données de résistance. De plus, la variance de la distribution de la perméabilité est systématiquement plus élevée que, lorsque l'assimilation se faisait sur un ensemble intial centré sur la valeur calibrée. Ceci laisse croire que l'addition de pas subséquents à l'assimilation pourrait encore améliorer l'estimation de la perméabilité.



Figure 5.34 – Histogramme de l'ensemble du logarithme en base 10 de la perméabilité hydraulique (a) ensemble initial (b) ensemble mis à jour à jour après 14 heures (c) ensemble mis à jour à jour après 18 heures (d) ensemble mis à jour après 22 heures



Figure 5.35 – Histogramme de l'ensemble du logarithme en base 10 de la dispersivité longitudinale (a) ensemble initial (b) ensemble mis à jour à jour après 14 heures (c) ensemble mis à jour à jour après 18 heures (d) ensemble mis à jour après 22 heures

5.5.2 Méthode d'assimilation par inversion

Les figures 5.36a à 5.36c et 5.37a à 5.37c montrent la moyenne de l'ensemble des concentrations avant et après l'assimilation des résistivités électriques inversées. L'assimilation des données de résistivités électriques inversées améliore l'estimation de la concentration au premier pas de temps. Au troisième pas de temps, l'ensemble des perméabilités hydrauliques n'englobe pas la valeur déterminée expérimentalement (figure 5.39c). Comme un grand nombre de données sont assimilées (par rapport à l'assimilation directe), la mise à jour des perméabilités est plus susceptible de souffrir d'écrasement de la covariance. La diminution de la variance de l'ensemble de perméabilités mis à jour est donc excessive. Ce nouvel ensemble se concentre autour d'une valeur plus faible que la valeur expérimentale. La longueur de la contamination prédite (figure 5.36b) est donc plus restreinte que la contamination obtenue avec le modèle calibré (figure 5.38b). Ceci expliquerait que l'assimilation des données de résistivité électrique inversées ne permet pas de corriger la contamination de façon significative au troisième pas de temps. En effet, il serait difficile de corriger la concentration en un noeud pour lequel tous les ensembles affichent une concentration nulle, puisqu'ainsi, la variance en ce point devrait aussi être nulle.



Figure 5.36 – Moyenne des concentrations prédites par les EnKF (a) à 14 heures (b) à 18 heures (c) à 22 heures



Figure 5.37 – Moyenne des concentrations mises à jour par les EnKF (a) à 14 heures (b) à 18 heures (c) à 22 heures



Figure 5.38 – Concentrations simulées avec les paramètres calibrés (a) à 14 heures (b) à 18 heures (c) à 22 heures



Figure 5.39 – Histogramme de l'ensemble du logarithme en base 10 de la perméabilité hydraulique (a) ensemble initial (b) ensemble mis à jour à jour après 14 heures (c) ensemble mis à jour à jour après 18 heures (d) ensemble mis à jour après 22 heures





Figure 5.40 – Histogramme de l'ensemble du logarithme en base 10 de la dispersivité longitudinale (a) ensemble initial (b) ensemble mis à jour à jour après 14 heures (c) ensemble mis à jour à jour après 18 heures (d) ensemble mis à jour après 22 heures



Figure 5.41 - AAB des concentrations prédites et mises à jour en fonction du pas de temps

Les figures 5.39a à 5.39d et 5.40a à 5.40d présentent les histogrammes de l'évolution des ensembles de perméabilité hydraulique et de dispersivité dans l'assimilation des données de résistivité inversée. L'assimilation n'ayant pas permis d'améliorer l'estimation de la concentration après le premier pas de temps, elle ne permet pas non plus de rapprocher l'estimation de la perméabilité ni de la dispersivité des valeurs calibrées de façon significative. Comme dans le cas où les ensembles initiaux de perméabilité et de dispersivité étaient centrés sur leurs valeurs préalablement calibrées, la dispersivité diminue plus l'assimilation progresse. La perméabilité moyenne, quant à elle, augmente pour se rapprocher faiblement de la valeur calibrée plus l'assimilation progresse. Toutefois, la variance des distributions des deux paramètres diminue de façon très importante. Avec des variances aussi faibles, des pas d'assimilation subséquents pourraient n'avoir aucun effet bénéfique sur l'estimation des paramètres.

5.6 Comparaison des méthode d'assimilation directe et inversée

Les figures 5.42a et 5.43a montrent la moyenne de déviation absolue pour les deux paramètres estimés par l'assimilation de données électriques par EnKF, soit la perméabilité hydraulique et la dispersivité longitudinale. Les figures 5.42b et 5.43b présentent l'étendue de l'ensemble de ces paramètres. Ces statistiques sont présentées pour les deux méthodes utilisées ainsi que pour les deux ensembles initiaux. Les deux méthodes permettent de diminuer la moyenne de déviation absolue de la perméabilité hydraulique que l'ensemble initial soit centré sur la vraie valeur ou non. Toutefois, la méthode directe semble présenter de meilleurs résultats, avec un ensemble plus étendu pour les deux paramètres et une movenne de déviation absolue plus faible pour la perméabilité hydraulique. Ces résultats semblent contraires aux résultats obtenus par Camporese et al. (2015) qui concluaient que la méthode d'assimilation à partir des données de résistances directes montraient plus d'écrasement de la covariance que la méthode d'assimilation à patir des données de résistivités inversées. Dans la présente étude, la comparaison entre les deux méthodes est biaisée puisque l'assimilation des résistances se fait sur un sous-ensemble de données. La diminution du nombre de données peut être bénéfique pour l'assimilation par EnKFs surtout si l'information contenue dans les données est redondante. De plus, la matrice de covariance de l'erreur des résistivités inversées a été tronquée pour assimiler ces données. Cette troncation crée une perte d'information. Enfin, l'artéfact à la surface des résistivités inversées pourrait aussi influencer négativement l'assimilation à partir de ces données. D'ailleurs, Dannowski & Yaramanci (1999) ont noté que les propriétés hydrogéologiques inférées de propriétés géophysiques comporteraient toujours une certaine ambiguïté. Cette ambiguïté serait associée non seulement aux erreurs dans l'acquisition, mais aussi à l'inversion des données géophysiques (Linde et al., 2006).

Comme discuté précédemment, la dispersivité est beaucoup plus difficile à ajuster par l'assimilation de données électriques peu importe la méthode utilisée.

Enfin, il est intéressant de noter que, malgré la discrétisation grossière de la grille d'éléments finis, l'assimilation permet d'améliorer l'estimation de la conductivité hydraulique. En effet, pour certains membres de l'ensemble (conductivité hydraulique élevée et/ou dispersivité longitudinale faible), la modélisation de la concentration présentait des oscillations puisque la grille était trop grossière. Malgré l'erreur numérique importante sur la modélisation de la concentration pour certains membres de l'ensemble, l'assimilation des résistances électriques était bénéfique pour l'estimation des concentrations ou de la conductivité hydraulique. Ceci est possiblement, au moins partiellement dû à la faible résolution des données de résistances électriques. En effet, les oscillations présentes dans le modèle hydrogéologique n'étaient pas nécessairement présentes sur le résultat de la modélisation électrique.



Figure 5.42 – Statistiques d'erreur pour le logarithme en base 10 de la perméabilité hydraulique en fonction du pas de temps pour les différents cas d'assimilation (a) AAB (b) Étendue de l'ensemble



Figure 5.43 – Statistiques d'erreur pour le logarithme en base 10 de la dispersivité longitudinale en fonction du pas de temps pour les différents cas d'assimilation (a) AAB (b) Étendue de l'ensemble

Chapitre 6

Conclusion

L'objectif principal de cette thèse était d'évaluer la calibration des paramètres hydrauliques à partir d'un suivi par résistance électrique dans des cas de contamination saline de la nappe par filtres de Kalman d'ensemble. Un algorithme d'assimilation de données de résistance électrique par EnKF dans un cas de contamination saline a donc été développé.

L'algorithme a d'abord été testé sur un cas synthétique bidimensionnel à conductivité hydraulique hétérogène, avec une contamination de surface. L'application des filtres de Kalman d'ensemble pour assimiler des données de résistance provenant d'un sondage en configuration dipôle-dipôle et des données de concentrations en puits a permis de conclure que l'assimilation des deux types de données donnaient des résultats similaires quant à la calibration du champ de conductivités hydrauliques lorsque les paramètres géostatistiques (movenne, variance, portée) du champ étaient connus et utilisés pour créer l'ensemble initial du champ. Comme ces deux types de données apportaient une information similaire, leur assimilation conjointe n'apportait pas une amélioration significative de la calibration du champ de conductivités par rapport à l'assimilation des données de concentration ou de résistances exclusivement. Ainsi, les paramètres hydrogéologiques peuvent être calibrés par assimilation de données électriques dans un cas de contamination saline. Les observations électriques montrant une plus grande continuité spatiale que les observations des concentrations en puits, il était attendu que l'assimilation des observations électriques permette de mieux calibrer le champ de conductivités hydrauliques que l'assimilation de données de concentrations ponctuelles (aux puits). L'étude sur un cas synthétique n'a toutefois pas montré cet avantage attendu. Il est possible qu'une plus forte hétérogénéité du champ de conductivités hydrauliques soit nécessaire pour

mieux déterminer l'apport de l'assimilation des données électriques dans la calibration des conductivités hydrauliques. Il serait donc intéressant d'effectuer une étude de sensibilité à l'hétérogénéité du champ de charges hydrauliques sur l'assimilation des données électriques et de concentrations aux puits.

Comme il existe différentes variantes des filtres de Kalman d'ensemble, une étude de ces variantes sur le cas synthétique a permis de comparer ces méthodes. Certaines variantes des EnKFs ont aussi été développées et comparées sur le cas synthétique.

La première variante des EnKFs développée est une méthodologie tenant compte d'une certaine variabilité des paramètres géostatistiques. Il a été noté que, dans ce cas, il est beaucoup plus difficile de calibrer le champ de conductivités peu importe le type de données utilisées, à l'exception des données de concentration aux puits pour un petit ensemble. Il est donc proposé de créer différents ensembles avec des paramètres géostatistiques fixes pour chaque ensemble et d'effectuer l'assimilation sur chaque ensemble indépendamment.

Des filtres de Kalman d'ensemble transformés (ETKF) sont des filtres dits déterministes, c'està-dire qu'une erreur de mesure aléatoire n'a pas à être ajoutée à l'ensemble prédit. Tel qu'attendu, les résultats de l'assimilation par ETKFs et EnKFs étaient similaires pour des grands ensembles, mais les EnKFs performaient mieux lorsque les ensembles étaient petits.

Une méthologie heuristique, dite découplée, pour l'assimilation des données de résistance, a été développée. Cette méthodologie a été comparée à la méthodologie classique, soit la méthodologie couplée. Les résultats des deux méthodologies étaient très similaires. Cette nouvelle méthodologie a une certaine flexibilité, en permettant de conserver seulement un sous-ensemble de la variable intermédiaire, soit les concentrations dans ce cas d'étude.

Enfin, une étude de sensibilité a été effectuée pour évaluer la taille optimale de l'ensemble dans chaque cas d'assimilation (type de données assimilé et méthode d'assimilation). Lorsque seules les concentrations aux puits étaient assimilées, des ensembles de petite taille (200 membres pour l'assimilation avec paramètres géostatistiques connus et 75 membres pour l'assimilation avec paramètres géostatistiques variables) permettaient d'obtenir des champs de conductivité hydraulique avec les plus faibles déviation par rapport au champ de référence. Augmenter la taille de l'ensemble semblait plus avantageux pour la calibration des paramètres hydrogéologiques par assimilation des résistances électriques que par assimilation des concentrations en puits. Ainsi, il est recommandé d'effectuer une étude de sensibilité à la taille de l'ensemble pour tout problème d'assimilation de données par EnKF tel que préalablement soulevé par Houtekamer & Mitchell (2001).

La seconde partie de cette thèse visait à valider l'algorithme sur une expérience de contamination saline en laboratoire. Cette expérience constituait en une intrusion saline dans une boîte de sable quasi-saturée de perméabilité homogène. L'assimilation des données de résistance servait à calibrer cette perméabilité ainsi que la dispersivité longitudinale, aussi considérée homogène. La méthode d'assimilation des résistances qui avait été testée sur le cas synthétique a été comparée à une méthode d'assimilation des résistivtés inversées. Dans cette deuxième méthode, une inversion était d'abord effectuée sur les résistances pour obtenir un champ de résistivités électriques. Ces résistivités étaient ensuite assimilées pour calibrer les paramètres. Deux cas ont aussi été testés: dans le premier cas, l'ensemble intial de la perméabilité était centré sur la vraie valeur de la perméabilité (qui avait été déterminée en laboratoire) et dans le second cas, l'ensemble initial était biaisé. Dans les deux cas, la dispersivité longitudinale était centrée sur une valeur calibrée manuellement au préalable. Dans les deux cas, l'assimilation permettait d'améliorer l'estimation de la perméabilité. Et dans les deux cas, l'assimilation par la méthode directe surpassait la méthode par inversion. Toutefois, la comparaison entre les deux méthodes n'est pas équitable puisqu'une méthodologie pour diminuer le nombre de données nécessaires à l'assimilation a été utilisée dans le cas de la méthode directe. Une diminution du nombre de données assimilées pourrait avantager la méthode directe en diminuant les risques d'écrasement de la covariance, par exemple.

L'assimilation des résistances n'améliorait pas significativement l'estimation de la dispersivité et l'assimilation des résistivités inversées détériorait l'estimation de la dispersivité. La dispersivité réelle était possiblement très faible. Une dispersivité aussi faible serait difficile à calibrer à partir de mesures électriques de faible résolution. Il est donc probable que l'assimilation de données électriques permette de calibrer une dispersivité plus élevée, soit à l'échelle d'un levé sur le terrain.

Le cas synthétique et le cas d'intrusion saline en laboratoire avaient quelques limites quant à leur représentativité de cas réels. Premièrement, ces deux cas ont été traités en deux dimensions. Cette méthode pourrait être généralisée à des cas tridimensionnels en utilisant des modèles hydrogéologiques et électriques en trois dimensions. Toutefois, l'utilisation de cette méthode sur un cas réel tridimensionnel sera influencé de façon importante par la disposition des électrodes. En effet, les levés électriques en trois dimensions sont sensibles à l'orientation des dispositifs (Chambers *et al.*, 2002). Par exemple, des variations électriques dues à la contamination parallèles aux lignes de levés seraient peu apparentes sur ces levés.

Les deux cas représentaient aussi des aquifères saturés (ou presque). La loi d'Archie permet de tenir compte du degré de saturation du sol. L'algorithme pourrait donc être utilsé en présence d'une zone non-saturée. Étant donné que le degré de saturation varie aussi en fonction de la conductivité hydraulique, la performance de l'algorithme dans un milieu avec une zone non-saturée importante devra être étudiée.

Comme mentionné préalablement, la variance de la conductivité hydraulique du cas synthétique était faible. Les paramètes hydrogéologiques calibrés dans l'expérience d'intrusion d'eau saline étaient, quant à eux, homogènes. Il serait donc intéressant d'évaluer l'assimilaition de données électriques pour calibrer des champs de conductivité hydraulique avec des variances plus élevées qui pourraient être représentatives de cas réel. De plus, l'algorithme a été testé sur des cas ne présentant qu'un seul hydrofaciès. Étant donné que plusieurs hydrofaciès sont souvent présents dans un aquifère, l'algorithme devrait être testé sur un cas présentant une multitude d'hydrofaciès. L'utilisation de simulations multipoints permettrait de tenir compte de ces hydrofaciès.

Références

- Ahmed S & De Marsily G (1993). Cokriged estimation of aquifer transmissivity as an indirect solution of the inverse problem: A practical approach. Water Resources Research, 29(2):521–530. DOI:10.1029/92WR00226.
- Anderson JL (2001). An ensemble adjustment kalman filter for data assimilation. Monthly Weather Review, 129(12):2884–2903. DOI:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.
- Anderson JL (2012). Localization and sampling error correction in ensemble kalman filter data assimilation. Monthly Weather Review, 140(7):2359–2371.
- Anderson JL & Anderson SL (1999). A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 126:2741– 2758.
- Anderson MP & Woessner WW (1992). Applied groundwater modeling: simulation of flow and advective transport. Academic press.
- Apriliani E (2011 2011). The groundwater pollution estimation by the ensemble kalman filter. Canadian Journal on Science and Engineering Mathematics, 2(2):60–63.
- Archie GE (1942). The electrical resistivity log as an aid in determining some reservoir characteristics. *Petroleum Transactions*, 146:54–62.
- Bailey RT & Baú D (2012). Estimating geostatistical parameters and spatially-variable hydraulic conductivity within a catchment system using an ensemble smoother. *Hydrology and Earth System Sciences*, 16(2):287–304. DOI:10.5194/hess-16-287-2012.
- BAPE (2013). TOTO. http://www.bape.gouv.qc.ca/sections/rapports/publications/bape297.pdf. electronic.
- Bertino L, Evensen G & Wackernagel H (2002). Combining geostatistics and kalman filtering for data assimilation in an estuarine system. *Inverse Problems*, 18(1):23.
- Bertino L, Evensen G & Wackernagel H (2003). Sequential data assimilation techniques in oceanography. *International Statistical Review*, 71(2):223–241. DOI:10.1111/j.1751-5823.2003.tb00194.x.
- Bertorelle E (2014). Laboratory experiments on the saltwater intrusion process. Mémoire de maîtrise, University of Padova.
- Binley A, Cassiani G, Middleton R & Winship P (2002). Vadose zone flow model parameterisation using cross-borehole radar and resistivity imaging. *Journal of Hydrology*, 267(3–4):147 – 159. DOI:http://dx.doi.org/10.1016/S0022-1694(02)00146-4.

- Bishop CH, Etherton BJ & Majumdar SJ (2001). Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. *Monthly Weather Review*, 129(3):420–436. DOI:10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.
- Blouin M, Martel R & Gloaguen E (2013). Accounting for aquifer heterogeneity from geological data to management tools. *Groundwater*, 51(3):421–431. DOI:10.1111/j.1745-6584.2012.00982.x.
- Borner FD, Schopper JR & Weller A (1996). Evaluation of transport and storage properties in the soil and groundwater zone from induced polarization measurements. *Geophysical prospecting*, 44:583–601.
- Bouchedda A (2010). Inversion conjointe des données électriques et de radar en forage. Thèse de doctorat, Ecole Polytechnique de Montréal.
- Brewster M, Annan A, Greenhouse J, Kueper B, Olhoeft G, Redman J & Sander K (1995). Observed migration of a controlled dnapl release by geophysical methods. *Ground Water*, 33:977–987.
- Brown RG & Hwang PYC (1997). Introduction to Random Signals and Applied Kalman Filtering with Matlab Exercises and Solutions. John Wiley and Sons.
- Burgers G, Jan van Leeuwen P & Evensen G (1998). Analysis scheme in the ensemble kalman filter. Monthly Weather Review, 126:1719–1724.
- Buselli G, Barber C, Davis G & Salama R (2). Detection of groundwater contamination near waste disposal sites with transient electromagnetic and electrical methods. *Geotechnical and environmental geophysics*, 1990:27–39.
- Butler DK, éditeur (2005). Near-surface Geophysics. Butler DK, éditeur. volume 13. SEG Books.
- Camporese M, Cassiani G, Deiana R & Salandin P (2011). Assessment of local hydraulic properties from electrical resistivity tomography monitoring of a three-dimensional synthetic tracer test experiment. *Water Resources Research*, 47(12):n/a–n/a. DOI:10.1029/2011WR010528.
- Camporese M, Cassiani G, Deiana R, Salandin P & Binley A (2015). Coupled and uncoupled hydrogeophysical inversions using ensemble kalman filter assimilation of ert-monitored tracer test data. *Water Resources Research*, 51(5):n/a.
- Capilla JE & Llopis-Albert C (2009). Gradual conditioning of non-gaussian transmissivity fields to flow and mass transport data: 1. theory. *Journal of Hydrology*, 371(1–4):66 74. DOI:http://dx.doi.org/10.1016/j.jhydrol.2009.03.015.
- Carrera J & Neuman SP (1986). Estimation of aquifer parameters under transient and steady state conditions: 1. maximum likelihood method incorporating prior information. Water Resources Research, 22(2):199–210. DOI:10.1029/WR022i002p00199.
- Cassiani G, Bruno V, Villa A, Fusi N & Binley AM (2006). A saline tracer test monitored via time-lapse surface electrical resistivity tomography. *Journal of Applied Geophysics*, 59(3):244 – 259. DOI:http://dx.doi.org/10.1016/j.jappgeo.2005.10.007.
- Cassiani G, Böhm G, Vesnaver A & Nicolich R (1998). A geostatistical framework for incorporating seismic tomography auxiliary data into hydraulic conductivity estimation. *Journal of Hydrology*, 206(1–2):58 – 74. DOI:http://dx.doi.org/10.1016/S0022-1694(98)00084-5.

- Cassiani G, Godio A, Stocco S, Villa A, Deiana R, Frattini P & Rossi M (2009). Monitoring the hydrologic behaviour of steep slopes via time-lapse electrical resistivity tomography. *Near Surface Geophysics, special issue on Hydrogeophysics*, 7(5-6):475–486.
- Chambers J, Ogilvy R, Kuras O, Cripps J & Meldrum P (2002). 3d electrical imaging of known targets at a controlled environmental test site. *Environmental Geology*, 41(6):690–704. DOI:10.1007/s00254-001-0452-4.
- Chen J, Hubbard S & Rubin Y (2001). Estimating the hydraulic conductivity at the south oyster site from geophysical tomographic data using bayesian techniques based on the normal linear regression model. *Water Resources Research*, 37(6):1603–1613. DOI:10.1029/2000WR900392.
- Chen Y & Zhang D (2006). Data assimilation for transient flow in geologic formations via ensemble kalman filter. Advances in W, 29:1107–1122.
- Clark MP, Rupp DE, Woods RA, Zheng X, Ibbitt RP, Slater AG, Schmidt J & Uddstrom MJ (2008). Hydrological data assimilation with the ensemble kalman filter: Use of streamflow observations to update states in a distributed hydrological model. Advances in Water Resources, 31(10):1309 – 1324. DOI:http://dx.doi.org/10.1016/j.advwatres.2008.06.005.
- Conrad K (2005). *PROBABILITY DISTRIBUTIONS AND MAXIMUM ENTROPY*. http://www.math.uconn.edu/ kconrad/blurbs/analysis/entropypost.pdf. electronic.
- Constantini FM (2015). Un caso test per l'intrusione saline: evidenze di laboratorio e numeriche. Mémoire de maîtrise, University of Padova.
- Cooley RL (1977). A method of estimating parameters and assessing reliability for models of steady state groundwater flow: 1. theory and numerical properties. *Water Resources Research*, 13(2):318–324. DOI:10.1029/WR013i002p00318.
- Copty N & Rubin Y (1995). A stochastic approach to the characterization of lithofacies from surface seismic and well data. *Water Resources Research*, 31(7):1673–1686. DOI:10.1029/95WR00947.
- Copty N, Rubin Y & Mavko G (1993). Geophysical-hydrological identification of field permeabilities through bayesian updating. *Water Resources Research*, 29(8):2813–2825. DOI:10.1029/93WR00745.
- Crestani E, Camporese M, Bau D & Salandin P (2013). Ensemble kalman filter versus ensemble smoother for assessing hydraulic conductivity via tracer test data assimilation. *Hydrology and Earth System Sciences*, 17:1517–1531.
- Crestani E, Camporese M & Salandin P (2015). Assessment of hydraulic conductivity distributions through assimilation of travel time data from ert-monitored tracer tests. *Advances in Water Resources*, 84:23–36.
- Dagan G & Rubin Y (1988). Stochastic identification of recharge, transmissivity, and storativity in aquifer transient flow: A quasi-steady approach. *Water Resources Research*, 24(10):1698–1710. DOI:10.1029/WR024i010p01698.
- Daily W & Owen E (1991). Cross-borehole resistivity tomography. *GEOPHYSICS*, 56(8):1228–1235. DOI:10.1190/1.1443142.

- Daily W, Ramirez A, Binley A & LaBrecque D (2004). Electrical resistance tomography. *The Leading Edge*, 23(5):438–442.
- Dannowski G & Yaramanci U (1999). Estimation of water content and porosity using combined radar and geoelectrical measurements. *Eur. J. Environ. Eng. Geophys*, 4(1):71–85.
- Day-Lewis FD & Lane JW (2004). Assessing the resolution-dependent utility of tomograms for geostatistics. *Geophysical Research Letters*, 31(7):n/a–n/a. DOI:10.1029/2004GL019617.
- Dee DP (1995). On-line estimation of error covariance parameters for atmospheric data assimilation. *Monthly Weather Review*, 123(4):1128–1145. DOI:10.1175/1520-0493(1995)123<1128:OLEOEC>2.0.CO;2.
- Delhomme JP (1979). Spatial variability and uncertainty in groundwater flow parameters: A geostatistical approach. *Water Resources Research*, 15(2):269–280. DOI:10.1029/WR015i002p00269.
- Dey A & Morrison HF (1979). Resistivity modelling for arbitrarily shaped two-dimensional structures. *Geophysical Prospecting*, 27(1):106–136.
- DHI-WASY S (2012). FEFLOW 6.1 Finite Element Subsurface Flow and Transport Simulation System: User Manual. DHI-WASY, éditeur. DHI-WASY GmbH.
- Diersch H (2009). Feflow-white papers vol. i. DHI-WASY, GmbH.
- Diersch HJ (2013). FEFLOW: finite element modeling of flow, mass and heat transport in porous and fractured media. Springer Science & Business Media.
- Diersch HJ (2014). FEFLOW: Finite Element Modeling of Flow, Mass and Heat Transport in Porous and Fractured Media. Springer, éditeur. Springer.
- Dietrich CR & Newsam GN (1989). A stability analysis of the geostatistical approach to aquifer transmissivity identification. *Stochastic Hydrology and Hydraulics*, 3(4):293–316.
- Drécourt JP, Madsen H & Rosbjerg D (2006). Bias aware kalman filters: Comparison and improvements. Advances in Water Resources, 29:707–718.
- Dubreuil-Boisclair C, Gloaguen E, Marcotte D & Giroux B (2011). Heterogeneous aquifer characterization from ground-penetrating radar tomography and borehole hydrogeophysical data using nonlinear bayesian simulations. *Geophysics*, 76(4):13–25.
- Erdal D & Cirpka O (2016). Joint inference of groundwater-recharge and hydraulic-conductivity fields from head data using the ensemble kalman filter. *Hydrology and Earth System Sciences*, 20:555–569.
- Evensen G (1992). Using the extended kalman filter with a multilayer quasi-geostrophic ocean model. Journal of Geophysical Research: Oceans, 97(C11):17905–17924. DOI:10.1029/92JC01972.
- Evensen G (1994). Sequential data assimilation with a nonlinear quasigeostrophic model using monte carlo methods to forecast error statistics. *Journal of geophysical research*, 99(10):143–162.
- Evensen G (2003). The ensemble kalman filter: theoretical formulation and practical implementation. Ocean Dynamics, 53(4):343–367. DOI:10.1007/s10236-003-0036-9.

- Evensen G & Van Leeuwen PJ (1996). Assimilation of geosat altimeter data for the agulhas current using the ensemble kalman filter with a quasigeostrophic model. *Monthly Weather Review*, 124(1): 85–96.
- Evensen G & Van Leeuwen PJ (2000). An ensemble kalman smoother for nonlinear dynamics. Monthly Weather Review, 128(6):1852–1867.
- Ferraresi M, Todini E & Vignoli R (1996). A solution to the inverse problem in groundwater hydrology based on kalman filtering. *Journal of Hydrology*, 175(1–4):567 – 581. DOI:http://dx.doi.org/10.1016/S0022-1694(96)80025-4.
- Fortin V, Abaza M, Anctil F & Turcotte R (2014). Why should ensemble spread match the rmse of the ensemble mean? *Journal of Hydrometeorology*, 15(4):1708–1713. DOI:10.1175/JHM-D-14-0008.1.
- Fowler DE & Moysey SM (2011). Estimation of aquifer transport parameters from resistivity monitoring data within a coupled inversion framework. *Journal of Hydrology*, 409(1–2):545 554. DOI:http://dx.doi.org/10.1016/j.jhydrol.2011.08.063.
- Freeze RA (1975). A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media. *Water Resources Research*, 11(5):725–741. DOI:10.1029/WR011i005p00725.
- Furrer R & Bengtsson T (2007). Estimation of high-dimensional prior and posterior covariance matrices in kalman filter variants. *Journal of Multivariate Analysis*, 98:227–255.
- Gómez-Hernández JJ, Jr. JJB & Fiori A (2016). Groundwater Transport in Highly Heterogeneous Aquifers. https://eos.org/meeting-reports/groundwater-transport-in-highlyheterogeneous-aquifers. Meeting Report.
- Gómez-Hernández JJ, Sahuquillo A & Capilla JE (1997). Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data–1. theory. *Journal of Hydrology(Amsterdam)*, 203(1):167–174.
- Goovaerts P (1997). Geostatistics for Natural Resources Evaluation. Oxford University Press.
- Gu Y & Oliver D (2007). An iterative ensemble kalman filter for multiphase fluid flow data assimilation. *SPE Journal*, na:438–446.
- Gunther T (2004). Inversion Methods and Resolution Analysis for the 2D/3D Reconstruction of Resistivity Structures from DC measurements. Thèse de doctorat, TU Bergakademie Freiberg.
- Gutjahr A & Wilson J (1989). Co-kriging for stochastic flow models. *Transport in Porous Media*, 4(6):585–598. DOI:10.1007/BF00223629.
- Hamill T, Whitaker J & Snyder C (2001). Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter. *Monthly W*, 129(11):2776–2790. DOI:10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.
- Harbaugh AW (2005). MODFLOW-2005, the US Geological Survey modular ground-water model: the ground-water flow process. US Department of the Interior, US Geological Survey Reston, VA, USA.

- Hayley K, Bentley L & Gharibi M (2009). Time-lapse electrical resistivity monitoring of salt-affected soil and groundwater. *Water Resources Research*, 45(7):n/a.
- Hendricks Franssen H & Kinzelbach W (2008). Real-time groundwater flow modeling with the ensemble kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research*, 44(9):n/a.
- Hendricks Franssen HJ & Kinzelbach W (2009). Ensemble kalman filtering versus sequential selfcalibration for inverse modelling of dynamic groundwater flow systems. *Journal of Hydrology*, 365(3–4):261 – 274. DOI:http://dx.doi.org/10.1016/j.jhydrol.2008.11.033.
- Hinnell AC, Ferré TPA, Vrugt JA, Huisman JA, Moysey S, Rings J & Kowalsky MB (2010). Improved extraction of hydrologic information from geophysical data through coupled hydrogeophysical inversion. Water Resources Research, 46(4):n/a–n/a. DOI:10.1029/2008WR007060.
- Hoeksema RJ & Kitanidis PK (1984). An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling. *Water Resources Research*, 20(7):1003–1020.
- Hördt A, Blaschek R, Kemna A & Zisser N (2007). Hydraulic conductivity estimation from induced polarisation data at the field scale the krauthausen case history. *Journal of Applied Geophysics*, 62(1):33–46.
- Houtekamer PL & Mitchell HL (1998). Data assimilation using an ensemble kalman filter technique. *Monthly Weather Review*, 126(3):796–811. DOI:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.
- Houtekamer PL & Mitchell HL (2001). A sequential ensemble kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129(1):123–137. DOI:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.
- Hubbard SS & Rubin Y (2000). Hydrogeological parameter estimation using geophysical data: a review of selected techniques. *Journal of Contaminant Hydrology*, 45(1–2):3 34. DOI:http://dx.doi.org/10.1016/S0169-7722(00)00117-0.
- Huisman J, Rings J, Vrugt J, Sorg J & Vereecken H (2010). Hydraulic properties of a model dike from coupled bayesian and multi-criteria hydrogeophysical inversion. *Journal of Hydrology*, 380(1–2):62 – 73. DOI:http://dx.doi.org/10.1016/j.jhydrol.2009.10.023.
- Irving J & Singha K (2010). Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities. *Water Resources Research*, 46(11):n/a–n/a. DOI:10.1029/2009WR008340. W11514.
- Isatis (2015). version 2015. Geovariances.
- Jafarpour B & Tarrahi M (2011). Assessing the performance of the ensemble kalman filter for subsurface flow data integration under variogram uncertainty. *Water Resources Research*, 47(5): n/a–n/a. DOI:10.1029/2010WR009090. W05537.
- Janjic T, McLaughlin D, Cohn SE & Verlaan M (2014). Conservation of mass and preservation of positivity with ensemble-type kalman filter algorithms. *Monthly Weather Review*, 142(2):755–773. DOI:10.1175/MWR-D-13-00056.1.

- Jensen JP (2007). Ensemble kalman filtering for state and parameter estimation on a reservoir model. Mémoire de maîtrise, Norwegian University of Science and Technology.
- Kalman R (1960). A new approach to linear filtering and prediction problems. Transactions of the ASME- Journal of Basic Engineering, 82:35–45.
- Kalman RE & Bucy RS (1961). New results in linear filtering and prediction theory. Journal of Basic Engineering, 83:95–108.
- Kelly WE (1977). Geoelectric sounding for estimating aquifer hydraulic conductivity. *Ground* Water, 15(6):420–425. DOI:10.1111/j.1745-6584.1977.tb03189.x.
- Kemna A, Binley A & Slater L (2004). Crosshole ip imaging for engineering and environmental applications. *Geophysics*, 69(1):97–07.
- Kemna A, Kulessa B & Vereecken H (2002). Imaging and characterisation of subsurface solute transport using electrical resistivity tomography (ert) and equivalent transport models. *Journal* of Hydrology, 267(3–4):125 – 146. DOI:http://dx.doi.org/10.1016/S0022-1694(02)00145-2.
- Kitanidis PK (1986). Comment on "stochastic modeling of groundwater flow by unconditional and conditional probabilities: The inverse problem" by gedeon dagan. Water Resources Research, 22(6):984–986.
- Kitanidis PK (1995). Quasi-linear geostatistical theory for inversing. *Water Resources Research*, 31(10):2411–2419.
- Kitanidis PK & Vomvoris EG (1983). A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. Water Resources Research, 19(3):677–690. DOI:10.1029/WR019i003p00677.
- Klassen J & Allen DM (2016). Risk of saltwater intrusion in coastal bedrock aquifers: Gulf islands, bc. Department of Earth Sciences, Simon Fraser University.
- LaBrecque DJ, Ramirez AL, Daily WD, Binley AM & Schima SA (1996). Ert monitoring of environmental remediation processes. *Measurement Science and Technology*, 7:375–383.
- Lawson WG & Hansen JA (2004). Implications of stochastic and deterministic filters as ensemblebased data assimilation methods in varying regimes of error growth. *Monthly weather review*, 132(8):1966–1981.
- Lei J, Bickel P & Snyder C (2010). Comparison of ensemble kalman filters under non-gaussianity. Monthly Weather Review, 138(4):1293–1306. DOI:10.1175/2009MWR3133.1.
- Lesmes DP & Friedman SP (2005). *Hydrogeophysics*, chapitre 4, pages 87–128. Springer.
- Li B, Toll D, Zhan X & Cosgrove B (2012a). Improving estimated soil moisture fields through assimilation of amsr-e soil moisture retrievals with an ensemble kalman filter and a mass conservation constraint. *Hydrology and Earth System Sciences*, 16:105–119.
- Li L, Zhou H, Gomez-Hernandez JJ & Franssen HJH (2012b). Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via ensemble kalman filter. Journal of Hydrology, 429(0):152 – 169. DOI:http://dx.doi.org/10.1016/j.jhydrol.2012.01.037.

- Li L, Zhou H, Hendricks Franssen HJ & Gómez-Hernández JJ (2012c). Modeling transient groundwater flow by coupling ensemble kalman filtering and upscaling. *Water Resources Research*, 48(1):19.
- Linde N, Chen J, Kowalsky MB & Hubbard S (2006). *Applied Hydrogeophysics*, chapitre 22, pages 9–44. Spring.
- Liu G, Chen Y & Zhang D (2008). Investigation of flow and transport processes at the {MADE} site using ensemble kalman filter. *Advances in Water Resources*, 31(7):975 986. DOI:http://dx.doi.org/10.1016/j.advwatres.2008.03.006.
- Livings D (2005). Aspects of the ensemble kalman filter. Mémoire de maîtrise, University of Reading.
- Livings DM, Dance SL & Nichols NK (2008). Unbiased ensemble square root filters. Physica D: Nonlinear Phenomena, 237(8):1021–1028.
- Llopis-Albert C & Capilla J (2010). Stochastic inverse modelling of hydraulic conductivity fields taking into account independent stochastic structures: A 3d case study. *Journal of Hydrology*, 391(3–4):277 288. DOI:http://dx.doi.org/10.1016/j.jhydrol.2010.07.028.
- Loke MH (2004). Tutorial: 2-D and 3-D electrical imaging surveys. electronic. cf p. 27.
- Lorenc AC (2003). The potential of the ensemble kalman filter for nwp a comparison with 4d var. Quarterly Journal of the Royal Meteorological Society, 129(595):3183–3203. DOI:10.1256/qj.02.132.
- Mandel J, Cobb L & Beezley JD (2011). On the convergence of the ensemble kalman filter. *Applications of Mathematics*, 56(6):533–541. DOI:10.1007/s10492-011-0031-2.
- Marsily G, Delay F, Goncalves J, Renard P, Teles V & Violette S (2005). Dealing with spatial heterogeneity. *Hydrogeology Journal*, 13(1):161–183. DOI:10.1007/s10040-004-0432-3.
- McLaughlin D & Townley LR (1996). A reassessment of the groundwater inverse problem. Water Resources Research, 32(5):1131–1161.
- MDDEP GdQ (2002). Le puits: Demystifier l'eau souterraine. http://www.mddep.gouv.qc.ca/eau/souterraines/puits/demystifier.htm.
- Medina A & Carre J (2003). Geostatistical inversion of coupled problems: dealing with computational burden and different types of data. *Journal of Hydrology*, 281:251–264.
- Meerschaert MM, Dogan M, Dam RLV, Hyndman DW & Benson DA (2013). Hydraulic conductivity fields: Gaussian or not? Water Resources Research, 49:1–8.
- M.H.Loke (1999). Electrical imaging surveys for environmental and engineering studies : A practical guide to 2-d and 3-d surveys. Michigan Tech.
- Mitchell HL, Houtekamer PL & Pellerin G (2002). Ensemble size, balance, and model-error representation in an ensemble kalman filter. *Monthly Weather Review*, 130(11):2791–2808. DOI:10.1175/1520-0493(2002)130<2791:ESBAME>2.0.CO;2.
- Moradkhani H, Sorooshian S, Gupta HV & Houser PR (2005). Dual state-parameter estimation for hydrological models using ensemble kalman filter. Advances in Water Resources, 28:135–147.

- Morrison F & Gasperikova E (2016). The Berkeley Course in Applied Geophysics. http://appliedgeophysics.berkeley.edu. Web course.
- Nan T & Wu J (2011). Groundwater parameter estimation using the ensemble kalman filter with localization. *Hydrogeology Journal*, 19(3):547–561.
- ONU (2014). UN Water statistics. http://www.unwater.org/statistics/en/.
- Pan M & Wood EF (2006). Data assimilation for estimating the terrestrial water budget using a constrained ensemble kalman filter. *Journal of Hydrometeorology*, 7:534–546. DOI:10.1175/JHM495.1.
- Panzeri M, Riva M, Guadagnini A & S.P.Neuman (2014). Comparison of ensemble kalman filter groundwater-data assimilation methods based on stochastic moment equations and monte carlo simulation. Advances in Water Resources, 66:8–18.
- Paradis D, Tremblay L, Lefebvre R, Gloaguen E, Rivera A, Parent M, Ballard JM, Michaud Y & Brunet P (2014). Field characterization and data integration to define the hydraulic heterogeneity of a shallow granular aquifer at a sub-watershed scale. *Environmental Earth Sciences*, 72(5):1325– 1348.
- Perri M, Cassiani G, Gervasio I, Deiana R & Binley A (2012). A saline tracer test monitored via both surface and cross-borehole electrical resistivity tomography: comparison of time-lapse results. *Journal of Applied Geophysics*, 79:6–16.
- Petrie RE (2008). Localization in the ensemble kalman filter. Mémoire de maîtrise, University of Reading.
- Phale HA & Oliver DS (2011). Data assimilation using the constrained ensemble kalman filter. SPE Journal, 16:n/a.
- Pham DT (2001). Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review*, 129(5):1194–1207. DOI:10.1175/1520-0493(2001)129<1194:SMFSDA>2.0.CO;2.
- Pidlisecky A & Knight R (2008). Fw2_5d: A matlab 2.5-d electrical resistivity modeling code. Computers & Geosciences, 34(12):1645–1654.
- Poeter EP & Hill MC (1997). Inverse models: A necessary next step in ground-water modeling. Ground water, 35(2):250–260.
- Pollock D & Cirpka OA (2008). Temporal moments in geoelectrical monitoring of salt tracer experiments. *Water Resources Research*, 44(12):n/a–n/a. DOI:10.1029/2008WR007014.
- Pollock D & Cirpka OA (2010). Fully coupled hydrogeophysical inversion of synthetic salt tracer experiments. *Water Resources Research*, 46(7):n/a–n/a. DOI:10.1029/2009WR008575. W07501.
- Prince Edward Island Department of Environment L & Justice (2011). Saltwater intrusion and climate change: a primer for local and provincial decision-makers. Atlantic Climate Adaptation Solutions Association.
- Ramirez A & Daily W (2001). Electrical imaging at the large block test yucca mountain, nevada. Journal of Applied Geophysics, 46:85–100.

- Ramirez A, Daily W, Binley A, LaBrecque D & Roelant D (1996). Detection of leaks in underground storage tanks using electrical resistance methods. *Journal of Environmental and Engineering Geophysics*, 1(3):189–203.
- Ramirez A, Daily W, Labrecque D, Owen E & Chestnut D (1993). Monitoring an underground steam injection process using electrical resistance tomography. *Water Resources Research*, 29:73–87.
- Rasmussen P, Sonnenborg TO, Goncear G & Hinsby K (2013). Assessing impacts of climate change, sea level rise, and drainage canals on saltwater intrusion to coastal aquifer. *Hydrology and Earth System Sciences*, 17:421–443.
- Reichle RH, McLaughlin DB & Entekhabi D (2002a). Hydrologic data assimilation with the ensemble kalman filter. *Monthly Weather Review*, 130(1):103–114. DOI:10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2.
- Reichle RH, Walker JP, Koster RD & Houser PR (2002b). Extended versus ensemble kalman filtering for land data assimilation. *Journal of Hydrometeorology*, 3:728–740.
- Reimus P, Pohll G, Mihevc T, Chapman J, Haga M, Lyles B, Kosinski S, Niswonger R & Sanders P (2003). Testing and parameterizing a conceptual model for solute transport in a fractured granite using multiple tracers in a forced-gradient test. Water Resources Research, 39(12):n/a– n/a. DOI:10.1029/2002WR001597. 1356.
- Robert T, Caterina D, Deceuster J, Kaufmann O & Nguyen F (2012). A salt tracer test monitored with surface ert to detect preferential flow and transport paths in fractured/karstified limestones. *Geophysics*, 77(2):B55–B67.
- Rubin Y, Mavko G & Harris J (1992). Mapping permeability in heterogeneous aquifers using hydrologic and seismic data. Water Resources Research, 28(7):1809–1816. DOI:10.1029/92WR00154.
- Ruggeri P, Irving J, Holliger K, Gloaguen E & Lefebvre R (2013). Hydrogeophysical data integration at larger scales: Application of bayesian sequential simulation for the characterization of heterogeneous alluvial aquifers. *The Leading Edge*, 32(7):766–774. DOI:10.1190/tle32070766.1.
- Schöniger A (2010). Parameter estimation by ensemble kalman filters with transformed data. Mémoire de maîtrise, Universität Stuttgart.
- Schöniger A, Nowak W & Hendricks Franssen HJ (2012). Parameter estimation by ensemble kalman filters with transformed data: Approach and application to hydraulic tomography. *Water Resources Research*, 48(4):n/a.
- Simon D (2010). Kalman filtering with state constraints: a survey of linear and non-linear algorithms. *IET Control Theory and Applications*, 4(8):1303–1318. DOI:10.1049/iet-cta.2009.0032.
- Simon E & Bertino L (2009). Application of the gaussian anamorphosis to assimilation in a 3-d coupled physical-ecosystem model of the north atlantic with the enkf: a twin experiment. Ocean Science, 5:495–510.
- Singha K & Gorelick SM (2005). Saline tracer visualized with three-dimensional electrical resistivity tomography: Field-scale spatial moment analysis. *Water Resources Research*, 41(5):n/a–n/a. DOI:10.1029/2004WR003460.

- Slater L, Binley A, Daily W & Johnson R (2000). Cross-hole electrical imaging of a controlled saline tracer injection. *Journal of Applied Geophysics*, 44(2–3):85 102. DOI:http://dx.doi.org/10.1016/S0926-9851(00)00002-1.
- Slater L, Binley A, Versteeg R, Cassiani G, Birken R & Sandberg S (2002). A 3d ert study of solute transport in a large experimental tank. *Journal of Applied Geophysics*, 49(4):211 229.
- Song X, Shi L, Ye M, Yang J & Navon IM (2014). Numerical comparison of iterative ensemble kalman filters for unsaturated flow inverse modeling. Vadose Zone Journal, 13(2):n/a. DOI:10.2136/vzj2013.05.0083.
- Sorenson HW (1974). On the development of practical nonlinear filters. *Information Sciences*, 7:253–270.
- Sudicky EA & Huyakrn PS (1991). Contaminant migration in imperfectly known heterogeneous groundwater systems. *Review of Geophysics*, 29:240–253.
- Sun NZ & Yeh WWG (1990). Coupled inverse problems in groundwater modeling: 1. sensitivity analysis and parameter identification. Water Resources Research, 26(10):2507–2525. DOI:10.1029/WR026i010p02507.
- Tarantola A (2005). Inverse problem theory and methods for model parameter estimation. siam.
- Tikhonov A & Arsenin V (1977). Solutions of ill-posed problems. Wiley.
- Tippett M, Anderson J, Bishop C, Hamill T & Whitaker J (2003). Ensemble square root filters. Monthly Weather Review, 131:1485–1490.
- Tong J, Hu BX & Yang J (2012). Assimilating transient groundwater flow data via a localized ensemble kalman filter to calibrate a heterogeneous conductivity field. *Stochastic Environmental Research and Risk Assessment*, 26(3):467–478.
- Townley LR (1983). Numerical Models of Groundwater Flow: Prediction and Parameter Estimation in the Presence of Uncertainty. Thèse de doctorat, Massachusets Institute of Technology.
- Tremblay L, Lefebvre R, Paradis D & Gloaguen E (2014). Conceptual model of leachate migration in a granular aquifer derived from the integration of multi-source characterization data (st-lambert, canada). *Hydrogeology Journal*, 22(3):587–608. DOI:10.1007/s10040-013-1065-1.
- Valstar J (2001). Inverse modelling of groundwater and groundwater mass transport. Impacts of Human Activity on Groundwater Dynamics, IAHS.
- Van Leeuwen PJ & Evensen G (1996). Data assimilation and inverse methods in terms of a probabilistic formulation. MONTHLY WEATHER REVIEW-USA, 124:2898–2913.
- Vanderborght J, Kemna A, Hardelauf H & Vereecken H (2005). Potential of electrical resistivity tomography to infer aquifer transport characteristics from tracer studies: A synthetic case study. *Water Resources Research*, 41(6):23.
- Vogt C, Kosack C & Marquart G (2012). Stochastic inversion of the tracer experiment of the enhanced geothermal system demonstration reservoir in soultz-sous-forêts revealing pathways and estimating permeability distribution. *Geothermics*, 42:1–12.

- Voss CI (1999). Seawater Intrusion in Coastal Aquifers Concepts, Methods and Practices, volume 14 de Theory and appications of transport in porous media, chapitre 9 USGS SUTRA Code History, Pratical Use and Application in Hawaii, pages 249 313. Kluwer Academic Publishers.
- Voss CI & Provost AM (2010). SUTRA: A Model for Saturated-Unsaturated, Variable-Density Ground-Water Flow with Solute or Energy Transport. U.S. Geological Survey.
- Wang X, Bishop CH & Julier SJ (2004). Which is better an ensemble of positive negative pairs or a centered spherical simplex ensemble. *Monthly Weather Review*, 132(7):1590–1605. DOI:10.1175/1520-0493(2004)132<1590:WIBAEO>2.0.CO;2.
- Ward S (1991). Geotechnical and Environmental Geophysics, Volume II : Environmental and Groundwater, chapitre Resistivity and induced polarization methods, pages 147–189. Society of Exploration Geophysicists.
- Wen XH, Chen W et al. (2005). Real-time reservoir model updating using ensemble kalman filter. SPE reservoir simulation symposium, Society of Petroleum Engineers.
- Whitaker J & Hamill T (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130:1913–1924.
- Wood WW (2000). It's the heterogeneity! Ground Water, 38(1):1–1.
- Woodbury AD & Ulrych TJ (2000). A full-bayesian approach to the groundwater inverse problem for steady state flow. *Water Resources Research*, 36(8):2081–2093.
- Xu T, Gomez-Hernandez JJ, Li L & Zhou H (2013a). Parallelized ensemble kalman filter for hydraulic conductivity characterization. *Computers and Geosciences*, 52:42–49.
- Xu T, Gomez-Hernandez JJ, Zhou H & Li L (2013b). The power of transient piezometric head data in inverse modeling: An application of the localized normal-score enkf with covariance inflation in a heterogenous bimodal hydraulic conductivity field. *Advances in Water Resources*, 54:100–118.
- Yin J, Zhan X, Zheng Y, Hain CR, Liu J & Fang L (2015). Optimal ensemble size of ensemble kalman filter in sequential soil moisture data assimilation. *Geophysical Research Letters*, 42(16):6710– 6715. DOI:10.1002/2015GL063366. 2015GL063366.
- Zheng C & Gorelick SM (2003). Analysis of solute transport in flow fields influenced by preferential flowpaths at the decimeter scale. *Ground Water*, 41(2):142–155.
- Zhou H, Gómez-Hernández JJ, Franssen HJH & Li L (2011). An approach to handling nongaussianity of parameters and state variables in ensemble kalman filtering. Advances in Water Resources, 34(7):844 – 864. DOI:http://dx.doi.org/10.1016/j.advwatres.2011.04.014.
- Zhou H, Li L, Hendricks Franssen HJ & Gómez-Hernández J (2012). Pattern recognition in a bimodal aquifer using the normal-score ensemble kalman filter. *Mathematical Geosciences*, 44(2): 169–185. DOI:10.1007/s11004-011-9372-3.
- Zimmerman D, Marsily Gd, Gotway C, Marietta M, Axness C, Beauheim R, Bras R, Carrera J, Dagan G, Davies P et al. (1998). A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. Water Resources Research, 34(6):1373–1413.

Annexe A

Gain de Kalman - cas synthétique

Les figures suivantes présentent le gain de Kalman calculé pour la correction du logarithme des conductivités hydrauliques par les données de concentration aux puits observées au premier pas de temps pour différents nombres d'ensembles lorsque les paramètres géostatistiques sont connus.



Figure A.1 – Gain de Kalman au premier pas de temps (a) pour un ensemble de 20 réalisations (b) Pour un ensemble de 100 réalisations (c) Pour un ensemble de 200 réalisations (d) Pour un ensemble de 100 réalisations

Les figures suivantes présentent le gain de Kalman calculé pour la correction du logarithme des conductivités hydrauliques par les données de concentration aux puits observées au premier pas de temps pour différents nombres d'ensembles lorsque les paramètres géostatistiques sont variables.



Figure A.2 – Gain de Kalman au premier pas de temps (a) pour un ensemble de 75 réalisations avec paramètres variables (b) Pour un ensemble de 225 réalisations avec paramètres variables (c) Pour un ensemble de 2100 réalisations avec paramètres variables réalisations

Annexe B

Estimation de la concentration avec paramètres géostatistiques variables

Cet annexe présente des résultats pour l'estimation de la concentration lorsque les paramètres géostatistiques utilisés pour créer les simulations initiales varient.

Les tableaux B.1 et B.2 présentent les statistiques d'erreur et d'étendue de l'ensemble pour la mise à jour des des concentrations par filtres de Kalman d'ensemble par assimilation des données de concentrations aux puits ou des données de résistances électriques (cas couplé).

	AAB
Simulations initiales (75 ensembles)	
Simulations initiales (1050 ensembles)	70,3
Assimilation par concentrations (75 ensembles)	
Assimilation par concentrations (1050 ensembles)	
Assimilation par résistances (75 ensembles)	18,5
Assimilation par résistances(1050 ensembles)	18,7

Tableau B.1 – Moyenne de déviation absolue (AAB) de l'ensemble de concentrations avant (simulations intiales) et après le premier pas de temps dans différents cas d'assimilation

	EE
Simulations initiales (75 ensembles)	
Simulations initiales (1050 ensembles)	90,4
Assimilation par concentrations (75 ensembles)	
Assimilation par concentrations (1050 ensembles)	
Assimilation par résistances (75 ensembles)	
Assimilation par résistances (1050 ensembles)	21,1

Tableau B.2 – Étendue de l'ensemble de concentrations avant (simulations initiales) et après le premier pas de temps dans différents cas d'assimilation



Figure B.1 – Concentration synthétique après 182 jours



Figure B.2 – Concentration mise à jour après un pas de temps par EnKF (b) à partir de données de concentration (75 membres) (c) à partir de données de résistance (75 membres) (c) à partir de données de concentration (1050 membres) (c) à partir de données de résistance (1050 membres)

Les tableau B.3 montre que l'assimilation des données de résistivité avec seulement 75 ensembles permet d'améliorer l'estimation de la concentration à chaque pas de temps.

Pas de temps	1	2	3	4	5	6
A priori	62,6	112,3	132,0	116,7	$113,\!5$	102,4
A posteriori	18,5	58,4	72,7	67,2	56,0	51,7

Tableau B.3 – Moyenne de déviation absolue(AAB) de la concentration avant et après assimilation à partir des résistivités (cas couplé) avec un ensemble de 75 membres
Annexe C

Schéma détaillé de l'expérience d'eau saline



Figure C.1 – Schéma détaillé de l'expérience d'intrusion d'eau saline (n'est pas à l'échelle)

Annexe D

Effet de la taille de la grille expérience d'intrusion d'eau saline

Le modèle hydrogéologique a été appliqué sur deux grilles différentes. La discrétisation de la grille fine est de $1mm \times 2,5mm$ et celle de la grille grossière est de $1cm \times 1cm$ (longeur x profondeur). Tous les autres paramètres, incluant les conditions frontières, sont restés les mêmes d'un modèle à l'autre.

Le premier pas de temps de l'assimilation correspond à la quatorzième heure après le début de l'intrusion d'eau saline. L'assimilation se déroule jusqu'à la vingt-deuxième heure après le début de l'intrusion d'eau saline. Les figures D.1a à D.1b représentent la concentration en sel de l'eau au premier et au dernier pas de temps pour les deux grilles ainsi que la différence entre les deux.



Figure D.1 – Concentrations simulées (a) Grille fine au premier pas de temps (b) Grille fine au dernier pas de temps (c) Grille grossière au premier pas de temps (d) Grille grossière au dernier pas de temps (e) Différence au premier pas de temps (f) Différence (en pourcentage) au dernier pas de temps

Le modèle électrique bidimensionnel a aussi été appliqué à ces deux pas de temps. Les figures D.2 et D.3 représentent les résistivités apparentes obtenues par modélisation directe de la tomographie électrique des concentrations modélisées avec les deux grilles ainsi que la différence entre les deux. Comme le dispositif de tomographie électrique est un dispositif pôle-dipôle et que l'acquisition s'est fait, premièrement avec le pôle en aval du modèle, ce qui sera considéré comme le "mode avant", puis avec le pôle en amont du modèle, ce qui sera nommé "mode arrière", ci-après, les deux types d'acquisition sont présentés séparément dans les figures.



Figure D.2 – Résistivités apparentes simulées pour l'acquisition par mode avant (a) Grille fine au premier pas de temps (b) Grille fine au dernier pas de temps (c) Grille grossière au premier pas de temps (d) Grille grossière au dernier pas de temps (e) Différence au premier pas de temps (f) Différence (en pourcentage) au dernier pas de temps



Figure D.3 – Résistivités apparentes simulées pour l'acquisition par mode arrièere (a) Grille fine au premier pas de temps (b) Grille fine au dernier pas de temps (c) Grille grossière au premier pas de temps (d) Grille grossière au dernier pas de temps (e) Différence au premier pas de temps (f) Différence au dernier pas de temps

Annexe E

Inversion électrique sur un modèle synthétique - expérience d'intrusion d'eau saline

L'inversion électrique a aussi été effectuée sur les données de résistance obtenues à partir du modèle hydrogéologique calibré et du modèle électrique. La figure ?? présente la pseudo-section des données obtenues à partir du modèle calibré 14 heures après le début d'intrusion d'eau saline. Et la figure E.2 représente le résultat de l'inversion des données de cette pseudo-section.



Figure E.1 – Pseudo-section des résistivités apparentes simulées à partir du modèle calibré (a) après 14 heures protocole directe (b) après 14 heures protocole inverse



Figure E.2 – Résultat de l'inversion du cas de référence après 14 heures