THESE

Pour l'obtention du Grade de

DOCTEUR DE L'UNIVERSITE DE POITIERS

(Faculté des Sciences Fondamentales et Appliquées)
(Diplôme National - Arrêté du 25 mai 2016)

École Doctorale 649 :
« *Rosalind Franklin* » - Énergie – Environnement - Biosanté

Secteur de Recherche :
Structure et évolution de la terre et autres planètes

Présentée par :

Alexis GRIS

****

Optimisation des volumes de dépollution par modélisation analytique inverse

****

Directeurs de Thèse :

Laurent CANER et Jacques BODIN

****

Soutenue le 04/12/2025

devant la Commission d'Examen

****

JURY

Dr. Mariem KACEM-BOUREAU	Ecole Centrale Lyon	Rapportrice
Pr. Olivier ATTEIA	Bordeaux INP	Rapporteur
Dr. Aude DOUILLET	Université de Poitiers	Examinatrice
Dr. Jean RILLARD	BRGM	Examinateur
Pr. Nicolas FATIN-ROUGE	Université de Poitiers	Examinateur
Pr. Laurent CANER	Université de Poitiers	Directeur

RÉSUMÉ

L'évaluation des volumes de sols contaminés par des substances anthropiques représente un défi à la fois sanitaire et méthodologique. Sur le plan scientifique, l'un des principaux enjeux est d'arriver à appréhender de manière effective l'hétérogénéité intrinsèque du sol, qui rend difficile l'estimation de la distribution spatiale des substances polluantes dans les zones non-échantillonnées. À ce jour, l'évaluation s'appuie majoritairement sur des méthodes d'interpolation spatiale qui génèrent un effet de lissage ou dans le cas de la modélisation numérique, sur des approches très complexes et coûteuses à mettre en œuvre. L'objectif de ce travail est de développer une nouvelle méthodologie permettant de mieux estimer le volume total de sol impacté par une pollution à partir de données de concentration ponctuelles réparties dans l'espace. La méthodologie proposée repose sur une modélisation inverse utilisant une solution analytique du transport en zone non saturée. Le modèle est appliqué séquentiellement à des triplets de points définis selon une triangulation de Delaunay. La combinaison des résultats d'inversion permet de reconstruire une distribution spatialement hétérogène des contaminants dans le sous-sol.

La modélisation analytique inverse (IAM) est plus simple à mettre en œuvre qu'un modèle numérique spatialement distribué grâce à des hypothèses simplificatrices. La méthode a été appliquée à des cas synthétiques homogènes et hétérogènes avec différentes distributions des points d'observation. Les résultats montrent une réduction significative de l'erreur moyenne avec l'IAM en comparaison avec les méthodes d'interpolations spatiales : en fonction de la distribution des points d'observation, la méthode IAM permet une réduction de 15 à 25 % de l'erreur dans un milieu spatialement homogène et de 40 à 70 % dans un milieu hétérogène. La méthode est particulièrement utile lorsque le nombre de points est limité (moins de 10 points d'observation pour les cas testés). La méthode a ensuite été mise en application et éprouvée sur 32 cas réels.

ABSTRACT

The assessment of soil volumes contaminated by anthropogenic substances represents both a health and a methodological challenge. From a scientific perspective, one of the main issues is to effectively apprehend the intrinsic heterogeneity of the soil, which complicates the estimation of pollutants spatial distribution in unsampled areas. Currently, assessment predominantly relies on spatial interpolation methods that generate a smoothing effect, or, in the case of numerical modeling, on approaches that are highly complex and costly to implement.

The objective of this work is to develop a novel methodology to better estimate the total volume of soil impacted by pollution, utilizing spatially distributed point concentration data. The proposed methodology relies on an inverse modeling approach using an analytical solution for transport in the unsaturated zone. The model is applied sequentially to point triplets defined by Delaunay triangulation. The combination of inversion results allows for the reconstruction of a spatially heterogeneous distribution of contaminants within the subsurface.

The Inverse Analytical Modeling (IAM) is simpler to implement than a spatially distributed numerical model due to its simplifying assumptions. The method was applied to both homogeneous and heterogeneous synthetic cases, each featuring various distributions of observation points. The results demonstrate a significant reduction in average error with IAM compared to spatial interpolation methods: depending on the distribution of observation points, the IAM method provides a 15% to 25% reduction in error in a spatially homogeneous medium and a 40% to 70% reduction in a heterogeneous medium. The method is particularly useful when the number of points is limited (fewer than 10 observation points for the cases tested). The method was subsequently applied and verified on 32 real-world cases.

REMERCIEMENTS

Cette thèse est l'aboutissement d'un parcours professionnel qui m'a conduit à la recherche à la suite de plusieurs années d'expérience dans le domaine des sites et sols pollués. Je tiens à remercier les personnes qui ont participé à leur échelle à la réussite de ce projet professionnel.

Tout d'abord, je tiens à remercier les membres de mon jury pour le temps consacré à la lecture de ce manuscrit. Vos remarques constructives contribueront indéniablement à son enrichissement.

Je remercie chaleureusement mes directeurs de thèse Laurent et Jacques pour votre confiance, votre disponibilité, vos précieux conseils tout au long de la thèse. Merci de m'avoir guidé sur la voie de la recherche scientifique qui était un domaine lointain avant le démarrage de cette mission.

Je remercie Nicolas et Mathieu d'avoir accepté de faire partie du comité scientifique de ma thèse et pour leurs avis éclairés.

Je tiens à remercier également la société APAVE de m'avoir suivi dans ce projet. Je remercie sincèrement Fred, Cédric J et Cédric B pour leur soutien. Je remercie également les membres de la direction technique Nicolas et Stéphane qui ont permis d'avoir une vision différente de mes travaux. Je remercie les collègues de l'équipe (Ben, Charlotte, Val et Clément) ainsi que mes collègues de Poitiers, de Bordeaux, de Bidart ainsi que ceux qui ont participé à l'alimentation de la base de données notamment.

Je remercie les enseignants-chercheurs, le personnel permanent de l'université de Poitiers, l'ensemble de l'équipe du laboratoire HYDRASA, les thésards en cours ou jeunes docteurs du laboratoire mais aussi ceux de paléontologie et biologie et l'équipe d'ERM pour leur accueil, leur disponibilité, leur bonne humeur et les moments de partage.

Je remercie l'équipe projet FORD, en particulier Cédric, Ignacio, Louis, Simon, Xavier et Fabrice, avec qui j'ai passé beaucoup de temps ces dernières années.

Merci aux anciens du master hydrogéologie et transferts pour les moments passés ensemble avant et durant cette thèse en particulier Jim pour les soirées rugby.

Je tiens également à remercier mes proches. Merci à mes parents et ma sœur pour leur soutien. Mes deux enfants, Mathis qui m'a apporté une énergie positive chaque jour et Adélie qui est arrivée pendant la troisième année de thèse et m'a permis de prendre du recul.

Table des matières

RÉSUMÉ III

ABSTRACT IV

REMERCIEMENTS V

LISTE DES SYMBOLES XIII

INTRODUCTION 1

LA GESTION DES SITES ET SOLS POLLUES EN FRANCE 1

OBJECTIFS DE LA THESE ET STRUCTURE DU MANUSCRIT 3

CHAPITRE 1 : ETAT DE L'ART 5

1.1. ACQUISITION DES DONNEES ET INCERTITUDES GLOBALES 6

1.2. INTERPOLATIONS DETERMINISTES ET GEOSTATISTIQUES 9

1.3. MODELISATION DES PROCESSUS D'ECOULEMENT ET DE TRANSPORT 15

1.3.1. Principales équations physiques 15

1.3.2. Modélisation analytique 18

1.3.3. Modélisation numérique 21

1.3.4. Calibration 22

1.4. METHODES UTILISANT L'INTELLIGENCE ARTIFICIELLE 25

1.5. SYNTHESE DU CHAPITRE 26

CHAPITRE 2 : APPROCHE PAR MODELISATION ANALYTIQUE INVERSE 27

2.1. PRINCIPE DE LA METHODE 28

2.2. LIMITES DE LA METHODE 33

2.3. SYNTHESE DU CHAPITRE 34

CHAPITRE 3 : APPLICATION A DES CAS SYNTHETIQUES 35

3.1. CONSTRUCTION DES CAS SYNTHETIQUES 36

3.2. APPLICATION A UN CAS SYNTHETIQUE HOMOGENE 37

3.2.1. Caractérisation de la pollution du cas synthétique 37

3.2.2. Plan d'échantillonnage 39

3.2.3. Méthode de comparaison des surfaces polluées 39

3.2.4. Comparaison des surfaces estimées par les différentes méthodes 42

3.3. CAS SYNTHETIQUE HETEROGENE AVEC DIVERSES REPARTITIONS SPATIALES DES POINTS D'OBSERVATION 44

3.3.1. Implémentation de l'hétérogénéité et caractérisation de la pollution simulée 44

3.3.2. Plan d'échantillonnage 47

3.3.3. Méthodologie de comparaison des surfaces estimés par les différentes méthodes 50
3.3.4. Analyse de sensibilité du modèle analytique 50
3.3.5. Calibration du modèle analytique 51
3.3.6. Comparaison des surfaces estimées par les différentes méthodes 54
3.4. ÉVALUATION DE LA PERFORMANCE DES DIFFERENTES METHODES SUR DES CAS SYNTHETIQUES HETEROGENES AVEC ESTIMATION EN 3D 60
3.4.1. Paramètres et caractérisation de la pollution simulée 60
3.4.2. Comparaison des résultats 70
3.4.3. Plan d'échantillonnage et stratégie de la modélisation analytique inverse 70
3.4.4. Comparaison des volumes estimés par modélisation analytique inverse avec une interpolation spatiale 72
3.5. SYNTHESE DU CHAPITRE 79
CHAPITRE 4 : APPLICATION A DES CAS REELS 81
4.1. CONSTRUCTION ET DESCRIPTION DE LA BASE DE DONNEES 82
4.2. APPLICATION DE L'IAM A UN CAS REEL SPECIFIQUE ET DETAILLE 89
4.2.1. Contexte géologique 90
4.2.2. Contexte hydrogéologique 92
4.2.3. Historique du site et investigations réalisées 93
4.2.4. Plan de gestion 96
4.2.5. Application de la méthode analytique inverse 102
4.2.6. Comparaison des résultats 110
4.3. APPLICATION DE L'IAM AUX AUTRES CAS DE LA BASE DE DONNEES 115
4.3.1. Comparaison des bords de fouille estimés pollués 115
4.4. APPLICATION DE LA MODELISATION ANALYTIQUE INVERSE A UN PANACHE DE POLLUTION DES EAUX SOUTERRAINES 122
4.4.1. Description du jeu de données 122
4.4.2. Sélection d'une solution analytique adaptée 123
4.4.3. Application de la modélisation analytique inverse 124
4.4.4. Résultats 126
CHAPITRE 5 : COMPARAISON DE L'IAM AVEC LA MODELISATION NUMERIQUE 129
5.1. CONSTRUCTION ET INITIALISATION D'UN MODELE MAILLE POUR L'INVERSION 130
5.2. DISTRIBUTION SPATIALE DES POINTS D'OBSERVATION 131
5.3. STRATEGIES D'INVERSION 131
5.3.1. Inversions avec Pest IES 132
5.3.2. Inversions avec PEST 133
5.4. SYNTHESE DU CHAPITRE 141

DISCUSSION GENERALE, CONCLUSION ET PERSPECTIVES **143**

RAPPEL DES OBJECTIFS ET METHODOLOGIE **143**

DISCUSSION GENERALE **143**

Application à des cas synthétiques 143

Application à des cas réels 144

Modélisation numérique inverse 145

LIMITES ET PERSPECTIVES **145**

REFERENCES **149**

LISTE DES ANNEXES **164**

Liste des Figures

Figure 1 : Description des différents types d'échantillonnage utilisés en site et sols pollués (Record, 2013) 6

Figure 2 : Illustration de l'application de la géostatistique avec krigeage et simulation (RECORD, 2013) 12

Figure 3 : Logigramme de la méthodologie 31

Figure 4 : Panache de pollution après deux ans pour le cas homogène 38

Figure 5 : Distributions spatiales des points d'observation pour le cas synthétique homogène 39

Figure 6 : Exemple de seuil de probabilité différent 41

Figure 7 : Distribution spatiale de la porosité entre 0,5 m et 2,5 (zone 1) et entre 3,5 et 6 m (zone 2) 44

Figure 8 : Profil Nord-Sud de la porosité 45

Figure 9 : Vue en plans du panache de pollution après deux ans pour le cas hétérogène 46

Figure 10 : Distributions spatiales régulières des points d'observation pour le cas synthétique hétérogène 48

Figure 11 : Distributions spatiales aléatoires des points d'observation 49

Figure 12 : Analyse de sensibilité - indice de Sobol 51

Figure 13 : Exemple de calibration 53

Figure 14 : Cartes présentant la modélisation analytique inverse et une interpolation RBF à une profondeur de 3 m et un seuil de 10 000 $\mu g.l^{-1}$ 55

Figure 15 : Erreurs moyennes de surfaces estimées par la méthode IAM et par les méthodes d'interpolation à partir de distributions régulières des points d'observations 55

Figure 16 : Pourcentages de surfaces faussement négatives (a) et faussement positives (b) en fonction du seuil de dépollution pour les distributions régulières des points d'observation 56

Figure 17 : Erreurs moyennes de surfaces estimées par la méthode IAM et par les méthodes d'interpolation à partir de distributions aléatoires des points d'observations 58

Figure 18 : Pourcentages de valeurs faussement négatives (a) et faussement positives (b) en fonction du seuil pour les distributions aléatoires des points d'observation 59

Figure 19 : Vue en 3D et en coupes des champs de porosité aux profondeurs 1 m, 3 m et 5 m pour le cas hétérogène n°1 62

Figure 20 : Vue 3D et en coupes des champs de porosité aux profondeurs 1 m, 3 m et 5 m pour le cas hétérogène n°2 63

Figure 21 : Vue 3D et en coupes des champs de porosité aux profondeurs 1 m, 3 m et 5 m pour le cas hétérogène n°3 64

Figure 22 : Vue 3D et en coupes des champs de porosité aux profondeurs 1 m, 3 m et 5 m pour le cas hétérogène n°4 65

Figure 23 : Vue en plans du panache de pollution après deux ans pour le cas hétérogène n°1 66

Figure 24 : Vue en plans du panache de pollution après deux ans pour le cas hétérogène n°2 67

Figure 25 : Vue en plans du panache de pollution après deux ans pour le cas hétérogène n°3 68

Figure 26 : Vue en plans du panache de pollution après deux ans pour le cas hétérogène n°4 69

Figure 27 : Carte des configurations d'échantillonnage appliquées aux 4 cas synthétiques 70

Figure 28 : Données utilisées pour estimer le volume de terres polluées 71

Figure 29 : Graphiques de comparaison des volumes pour le cas synthétique n°1 et l'ensemble des seuils 74

Figure 30 : Graphiques de comparaison des volumes pour le cas synthétique n°2 et l'ensemble des seuils 74

Figure 31 : Graphiques de comparaison des volumes pour le cas synthétique n°3 et l'ensemble des seuils 75

Figure 32 : Graphiques de comparaison des volumes pour le cas synthétique n°4 et l'ensemble des seuils 75

Figure 33 : Comparaison entre l'interpolation et la modélisation analytique des volumes faux positifs et faux négatifs 76

Figure 34 : Comparaison entre l'interpolation et la modélisation analytique inverse avec un seuil de 100 000 $\mu g.l^{-1}$ pour le cas n°1 77

Figure 35 : Comparaison entre l'interpolation et la modélisation analytique inverse avec un seuil de 100 000 $\mu g.l^{-1}$ pour le cas n°3 78

Figure 36 : La cartographie des sites inclus dans la base de données par département 84

Figure 37 : Localisation du site où l'IAM a été appliquée 89

Figure 38 : Extrait de la carte géologique de Saint Vincent de Tyrosse (BRGM, 1987) 91

Figure 39 : Carte piézométrique (SIGES, 2025) 92

Figure 40 : Log géo-hydrogéologique (base LISA, SIGES 2025) 93

Figure 41 : Cartographie des teneurs en polluants pour l'horizon 1-2 m 95

Figure 42 : Résumé de la démarche de définition de la pollution concentrée 96

Figure 43 : Interpolation avec la méthode plus proche voisin pour l'horizon 1-2 m 99

Figure 44 : Interpolation avec la méthode polygone de Voronoï pour l'horizon 1-2 m 100

Figure 45 : Cartographie du maillage et des investigations complémentaires de sol réalisées pour la caractérisation affinée de la zone source concentrée par la société de dépollution 102

Figure 46 : Cartographie représentant les triangles de Delaunay pour le cas réel 104

Figure 47 : Indice de Sobol avec l'ensemble des paramètres pour la solution analytique de Troldbord 105

Figure 48 : Vue 2D à 1,5 m de profondeur de l'estimation des concentrations en hydrocarbures C10-C40 avec un seuil de pollution à 500 mg.kg^{-1} MS avec l'IAM 109

Figure 49 : Vue en 3D de l'estimation des concentrations en hydrocarbures C10-C40 avec un seuil de pollution à 500 mg.kg^{-1} MS avec la modélisation analytique inverse 110

Figure 50 : Cartographie de la zone terrassée et des résultats en bords de fouille pour le cas réel. 112

Figure 51 : Cartographie des surfaces polluées obtenues par différentes méthodes d'estimation réalisée à 1,5 m de profondeur 114

Figure 52 : Exemple n°1/3 de résultats pour la modélisation analytique inverse sur des cas réels 119

Figure 53 : Exemple n°(2/3) de résultats pour la modélisation analytique inverse sur des cas réels 120

Figure 54 : Exemple n°(3/3) de résultats pour la modélisation analytique inverse (3/3) sur des cas réels 121

Figure 55 : Cartographie des concentrations en COHV et direction d'écoulement des eaux souterraines 123

Figure 56 : Triangulation de Delaunay appliquée au panache de contamination des eaux souterraines 125

Figure 57 : Cartographie des concentrations pour l'application de la modélisation analytique inverse à un panache de COHV (seuil à 50 µg.l^{-1}) 127

Figure 58 : Distribution spatiale des points d'observation pour l'inversion du modèle maillé 131

Figure 59 : Localisation des points pilotes de porosité 134

Figure 60 : Graphique de calibration des concentrations correspondant à l'inversion des porosités du modèle maillé avec la méthode SVD-Assist 135

Figure 61 : Coupes 3D des concentrations simulées à partir des porosités inversées 136

Figure 62 : Comparaison des volumes au-dessus des seuils critiques pour le modèle inversé et la « pseudo-réalité » (cas synthétique) 137

Figure 63 : Distribution spatiale des concentrations sources identifiées par modélisation inverse 139

Figure 64 : Coupes 3D des concentrations simulées à partir des zones sources inversées 140

Liste des tableaux

Tableau 1 : Principaux facteurs influençant le choix de la méthode d'interpolation (Li et Heap, 2014) 13

Tableau 2 : Surfaces pour lesquelles les concentrations sont supérieures aux différents seuils de dépollution considérés pour le cas synthétique homogène 38

Tableau 3 : Pourcentage d'erreur moyenne associée à chaque distribution spatiale pour la méthode IAM et pour les 3 méthodes d'interpolation 42

Tableau 4 : Surfaces de dépassement des différents seuils de pollution pour le cas synthétique hétérogène 46

Tableau 5 : Plages de paramètres possibles pour la calibration 52

Tableau 6 : Statistique concernant les paramètres inversés 53

Tableau 7 : Exemples de valeurs pour construire le variogramme du cas n°2 61

Tableau 8 : Volumes de terres en m^3 dont les concentrations sont supérieures aux différents seuils de dépollution considérés pour les 4 cas synthétiques hétérogènes 69

Tableau 9 : Récapitulatif des Erreurs Moyennes d'Estimation pour la Modélisation Analytique Inverse (borne moyenne) et l'interpolation (4 Cas, 2 Échantillonnages) 72

Tableau 10 : Présentation de la base de données (uniquement le site 1) 85

Tableau 11 : Principales statistiques concernant les travaux de dépollution 87

Tableau 12 : Pourcentage de conformité entre les bords de fouille réalisés et les seuils de dépollution pour la base de données 88

Tableau 13 : Méthodes de définition d'une pollution concentrée (extrait guide UPDS) 97

Tableau 14 : Principales statistiques pour les données acquises autour de la cuve de fioul en hydrocarbures C10-C40 98

Tableau 15 : Plages de paramètres possibles pour l'étalonnage 106

Tableau 16 : Statistique concernant les paramètres inversés pour le cas réel 108

Tableau 17 : Plages de paramètres possibles pour l'étalonnage 113

Tableau 18 : Pourcentage de conformité pour la modélisation analytique inverse par lithologie 116

Tableau 19 : Pourcentage de conformité pour la modélisation analytique inverse réalisés et les seuils de dépollution pour la base de données 118

Tableau 20 : Plages de paramètres pour l'étalonnage pour le modèle analytique en zone saturée 125

Tableau 21 : Statistiques concernant les paramètres inversés pour le cas réel en zone saturée 126

Tableau 22 : Erreur moyenne en pourcentage pour l'inversion du modèle numérique 138

Liste des annexes

ANNEXE 1 : CARTOGRAPHIES DES RESULTATS DE LA MODELISATION ANALYTIQUE INVERSE POUR LES 32 SITES DE LA BASE DE DONNEES

LISTE DES SYMBOLES

α : paramètre de van Genuchten [L^{-1}]

α_L : dispersivité longitudinale [L]

α_T : dispersivité transverse [L]

b : épaisseur de l'aquifère [L]

β : paramètre de van Genuchten [sans unité]

C_{crit} : valeur seuil de concentration [$M.L^{-3}$]

C_s : concentration du composant adsorbé sur la matière organique du sol [$M.L^{-3}$]

C_t : concentration totale [$M.L^{-3}$]

Cw_0 : concentration d'injection du polluant dans l'eau [$M.L^{-3}$]

Cw : concentration du soluté dans la phase aqueuse [$M.L^{-3}$]

D^e_w: coefficient de diffusion de l'eau dans les pores [$L^2.T^{-1}$]

D_w : coefficient de diffusion dans l'eau [$L^2.T^{-1}$]

D_a: coefficient de diffusion dans la phase gazeuse [$L^2.T^{-1}$]

D^e_a : coefficient de diffusion effectif dans la phase gazeuse du milieu poreux [$L^2.T^{-1}$]

D_{lh} : coefficient de dispersion hydrodynamique [L^2T^{-1}]

D_e : coefficient de dispersion effectif [L^2T^{-1}]

foc : teneur en carbone total [sans unité]

h : potentiel de l'eau [L]

H : constante de Henry du polluant [sans unité]

i : nombre total de jeux de paramètres optimisés [sans unité]

i' : nombre de jeux de paramètres conduisant à une valeur $C_w > C_{crit}$ [sans unité]

$Ieff$: vitesse d'infiltration [$L.T^{-1}$]

J_a : flux de soluté dans la phase gazeuse [$M.L^{-2}.T^{-1}$]

J_t : flux de soluté total [$M.L^{-2}.T^{-1}$]

J_w : flux de soluté dans la phase liquide [$M.L^{-2}T^{-1}$]

K : conductivité hydraulique en zone non saturée [$L.T^{-1}$]

Koc : coefficient de partage carbone organique/eau [$L^3.M^{-1}$]

Kd : coefficient de partage entre l'eau et le sol [$L^3.M^{-1}$]

Kr : conductivité hydraulique relative [sans unité]

λ : constante de dégradation de premier ordre [T^{-1}]

n : exposant de Brooks-Corey [sans unité]

θw : teneur en eau volumique [sans unité]

θa : teneur en air volumique [sans unité]

θr : teneur en eau volumique résiduelle [sans unité]

θs : teneur en eau volumique à saturation [sans unité]

θso : fraction volumique de sol [sans unité]

θ : porosité [sans unité]

ψ : variable d'intégration [sans unité]

ρb : masse volumique apparente du sol [$M.L^{-3}$]

Q : débit d'injection du polluant [$L.T^{-1}$]

qc : vitesse d'écoulement de l'eau [LT^{-1}]

R : coefficient de retard [sans unité]

S : terme source de l'équation de Richards [T^{-1}]

S_e : saturation effective du sol [sans unité]

t : temps [T]

V : vitesse d'écoulement dans la direction x [$L.T^{-1}$]

ω : terme source de l'équation de conservation de la masse du polluant [$M.L^{-3}.T^{-1}$]

$xobs$: Coordonnée spatiale horizontale [L]

X : longueur de la source [L]

Xs : coordonnée du centre de la source [L]

$yobs$: coordonnée spatiale horizontale [L]

Y : largeur de la source [L]

Ys : coordonnée du centre de la source [L]

z : coordonnée spatiale verticale [L]

INTRODUCTION

LA GESTION DES SITES ET SOLS POLLUES EN FRANCE

Depuis 1990, l'étalement urbain en France a confronté les professionnels de l'urbanisme et de la construction à des problématiques de pollution des sols et des eaux souterraines liées à la présence d'anciennes activités industrielles qui se sont succédé pendant des décennies.

Différentes bases de données permettent de recenser en partie ces sites telles que la Carte des Anciens Sites Industriels et Activités de Service (CASIAS) qui comporte près de 323 000 sites (Géorisques, 2025a) ou la base Information de l'administration concernant une pollution suspectée ou avérée (ancienne dénomination : BASOL) qui inclut plus de 11 200 sites (Géorisques, 2025b). La réhabilitation de ces sites représente des coûts financiers importants mais constitue un enjeu majeur pour limiter l'artificialisation des sols.

Pellet et Laville-Timsit (1993) définissent le sol comme « *l'ensemble de la zone non-saturée comprise entre la surface topographique et le niveau de battement de la nappe. C'est donc une entité d'épaisseur et de composition très variables, constituée par ailleurs de plusieurs phases : solide, liquide et gaz. La matière organique vivante, en décomposition ou déjà transformée, y joue également un rôle important, sur les niveaux les plus superficiels. Les interactions entre ces phases relèvent et de processus divers, tels que l'infiltration, l'exfiltration, la volatilisation, la diffusion, le lessivage, l'adsorption, la désorption, …*

Il s'agit donc d'un milieu complexe ouvert, vivant et… éminemment hétérogène. »

Le site internet https://www.georisques.gouv.fr/, réalisé en partenariat entre le Ministère de la Transition Écologique, de la Biodiversité, de la Forêt, de la Mer et de la Pêche, et le BRGM, donne la définition suivante d'un site pollué : « *un site qui, du fait d'anciens dépôts de déchets ou d'infiltration de substances polluantes, présente une pollution susceptible de provoquer une nuisance ou un risque pérenne pour les personnes ou l'environnement* ».

Les principaux polluants correspondent aux polluants organiques (hydrocarbures, hydrocarbures aromatiques polycycliques (HAP)…) et aux polluants inorganiques comme les métaux lourds et les métalloïdes (Santépubliquefrance, 2025).

Dans la méthodologie nationale sites et sols pollués d'avril 2017 (Ministère de la transition écologique et solidaire, 2017), le terme « pollution concentrée » a été explicité (il avait été évoqué dans la circulaire du 8 février 2007 relative aux sites et sols pollués). La pollution concentrée correspond au « *volume de milieu souterrain à traiter, délimité dans l'espace, au sein duquel les concentrations en une ou plusieurs substances sont significativement supérieures aux concentrations de ces mêmes substances à proximité immédiate de ce volume* ». La suppression de cette pollution concentrée est qualifiée de prioritaire lorsqu'elle est identifiée. Différents « outils » sont proposés permettant de définir un seuil de concentration à partir duquel une pollution concentrée est identifiée. Ces méthodes font notamment appel à l'estimation des volumes de sols pollués dépassant un seuil de pollution.

Il est primordial de pouvoir estimer au mieux les volumes de terres à traiter afin de limiter les impacts financiers (Dakins et al., 1994), sanitaires (Oliver et al., 2015, Carre et al., 2015) et environnementaux (Perez-Lucas et al., 2019). L'évaluation des volumes de sols contaminés est une tâche complexe (Tao et al., 2022) notamment au regard de l'hétérogénéité du sol. Elle implique l'utilisation de méthodes appropriées pour estimer la concentration des substances polluantes dans les zones non-échantillonnées.

OBJECTIFS DE LA THESE ET STRUCTURE DU MANUSCRIT

L'objectif de cette thèse est de développer une approche permettant de mieux estimer les panaches de pollution des sols en zone non saturée notamment en milieu hétérogène.

La méthodologie nationale des sites et sols pollués précise que « *lorsque des pollutions concentrées sont identifiées (flottants sur les eaux souterraines, terres fortement imprégnées de produits, produits purs…), la priorité consiste d'abord à déterminer les modalités de suppression des pollutions concentrées, plutôt que d'engager des études pour justifier leur maintien en l'état, en s'appuyant sur la qualité déjà dégradée des milieux ou sur l'absence d'usage de la nappe.* ». Cette notion a été renforcée avec la mise en place du décret n°2024-742 du 6 juillet 2024 où, en cas de non-suppression d'une pollution concentrée, l'exploitant doit justifier de la coupure des voies de transfert de la pollution concentrée et établir un bilan environnemental global avec des résultats favorables pour le maintien de la source de pollution par rapport à la suppression de celle-ci.

Le manuscrit est séparé en 5 chapitres. La première partie de ce mémoire présente les méthodologies déjà existantes pour estimer les volumes de sols pollués. Le chapitre 2 présente en détail la nouvelle méthodologie développée correspondant à la modélisation analytique inverse.

Pour comparer cette nouvelle méthode avec les méthodes existantes en 2D et en 3D, plusieurs cas synthétiques homogènes et hétérogènes ont été créés et présentés dans le chapitre 3.

Dans le chapitre 4, la méthode est mise en œuvre sur un grand nombre de cas réels et comparée aux résultats obtenus sur le terrain. Enfin, la modélisation analytique inverse est appliquée à un panache de polluant dans les eaux souterraines.

Dans une approche méthodologique alternative, une méthode de modélisation numérique inverse est explorée dans le chapitre 5.

CHAPITRE 1 : ETAT DE L'ART

Ce chapitre a pour objectif de faire un état de l'art des différentes approches pour dimensionner les volumes de terres impactées par une pollution. Les deux approches les plus fréquemment utilisées dans la littérature scientifique et/ou par les professionnels sont les suivantes :

- Les interpolations spatiales déterministes ;
- La géostatistique par krigeage et les simulations géostatistiques.

D'autres méthodes moins utilisées telles que la modélisation numérique et analytique seront également présentées. Ce chapitre se limite aux méthodes exploitant des données ponctuelles de concentrations en polluants dans les sols.

1.1. ACQUISITION DES DONNEES ET INCERTITUDES GLOBALES

L'étape initiale d'un diagnostic de pollution des sols correspond à la réalisation d'une étude historique et documentaire visant à identifier les potentielles sources de pollution. Suite à cette étape, des investigations ciblées seront mises en œuvre aux abords des sources ainsi définies. La stratégie d'échantillonnage choisie la plus fréquente est dite préférentielle : elle consiste à réaliser des sondages et des prélèvements ponctuels de sol à proximité ou au droit des sources de pollution identifiées ou supposées. Toutefois, si la connaissance de l'historique du site est trop incomplète pour définir ces cibles, un échantillonnage aléatoire ou systématique est mis en œuvre (Fig.1) (Record, 2013).

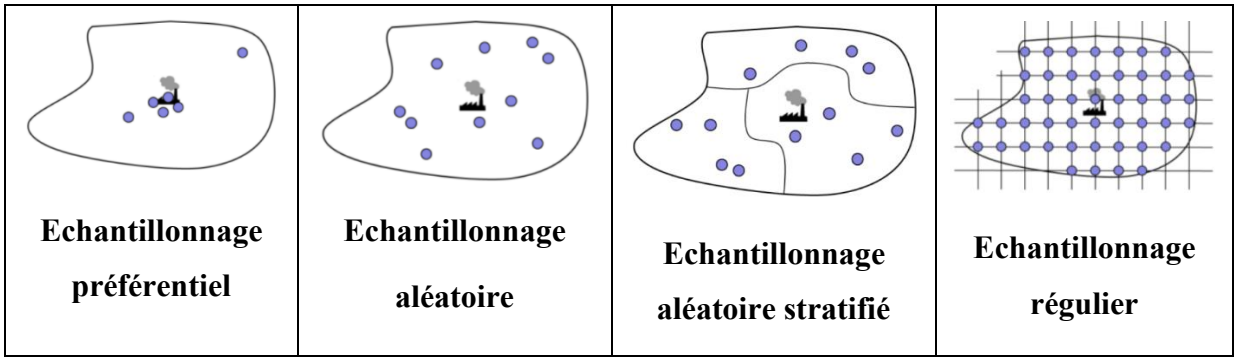


Figure 1 : Description des différents types d'échantillonnage utilisés en site et sols pollués (Record, 2013)

Les concentrations en polluant dans les sols, mesurées et utilisées, sont exprimées en $mg.kg^{-1}$ sur une base de matière sèche (MS). Le coût des investigations pouvant être élevé, le nombre de données disponibles à l'issue du diagnostic de pollution est souvent limité. Les données acquises lors d'un ou plusieurs diagnostics permettent la réalisation d'un plan de

gestion (Colombano, 2010). Ce document vise à déterminer le seuil de pollution concentrée à atteindre ainsi que les méthodes de traitement permettant d'atteindre cet objectif. Les mesures ponctuelles de concentration jouent donc un rôle majeur dans l'estimation des volumes de terres polluées (Environnemental Protection Agency (EPA), 1996, IAEA, 2004) et doivent être le plus représentative possible (Ramsey et Hewitt, 2005). Les incertitudes associées aux investigations en sites et sols pollués sont nombreuses et ne se limitent pas à la méthode d'échantillonnage (INERIS, 2020, Guyonnet et al., 2019).

Guridi (2024) distingue trois principales sources d'incertitudes :

i) l'absence de compréhension exacte du milieu et de la pollution. Cette notion est principalement liée à un échantillonnage incomplet du site (nombre limité de données). Cette source d'incertitudes peut être reliée directement à l'hétérogénéité du sol de chaque site. Celle-ci constitue un facteur structurant majeur des distributions spatiales des polluants, en particulier des contaminants organiques et peut conduire à des variations de concentrations de plusieurs ordres de grandeur à de très faibles distances. Cette hétérogénéité résulte de la variabilité conjointe des propriétés physiques, chimiques et biologiques du sol, mais également de l'hétérogénéité géologique et lithologique sous-jacente, qui conditionne la texture, la minéralogie, le pH et la capacité de stabilisation de la matière organique (Beven, 1996 ; Goovaerts, 1997). Les contrastes lithologiques, notamment entre formations argileuses, sableuses ou carbonatées, induisent des discontinuités spatiales marquées des propriétés pédologiques, se traduisant par des différences importantes de rétention et d'accumulation des polluants organiques. À l'échelle locale, la distribution hétérogène des fractions organiques stables, fortement influencée par la nature de la roche mère et les processus pédogénétiques associés, crée des zones de rétention préférentielle où les contaminants peuvent s'accumuler de manière ponctuelle, générant des gradients de concentration très marqués (Huang et al., 2003 ; Ukalska-Jaruga et al., 2023). À des échelles plus larges, l'interaction entre hétérogénéité lithologique, propriétés du sol et modalités de dépôt des sources anthropiques conduit à des structures spatiales complexes et souvent discontinues, caractérisées par une forte variabilité intra-site des concentrations (Froger et al., 2021). Cette organisation spatiale rend l'interprétation des niveaux de contamination particulièrement sensible à la stratégie d'échantillonnage et souligne

la nécessité d'intégrer explicitement les contrôles géologiques et pédologiques dans les analyses spatiales et l'évaluation des risques environnementaux.

ii) l'existence de divergences et de fluctuations dans les mesures. Cette notion fait référence à l'échantillonnage avec notamment la méthode de forage, la méthode de prélèvements, le type de flaconnage, la localisation, le changement de protocole entre intervenants (Favereaux et al., 2019) mais aussi les analyses en laboratoire. Les rapports d'analyses sont souvent fournis avec une estimation de l'incertitude de la mesure (généralement de l'ordre de 40%). Cependant, les incertitudes liées à la préparation de l'échantillon, à la contamination potentielle par un opérateur ou l'atmosphère et ainsi qu'au sous échantillonnage ne sont pas prises en compte (Amairic et al., 2021).

iii) les procédures d'estimation de la pollution : il est possible de quantifier ces incertitudes pour certaines techniques d'interpolations spatiales, notamment pour les méthodes géostatistiques (Chiles et Delfiner, 2012, Boudreault et al., 2016).

Pour vérifier le succès des opérations de dépollution par excavation, des prélèvements de contrôle sont réalisés en bords et fonds de fouille. Cette procédure vise à confirmer que les objectifs de dépollution définis dans le plan de gestion ont été atteints. Ces prélèvements peuvent être de trois types, en fonction des objectifs de l'étude (Le Corfec, 2011) :

- Echantillons ponctuels, effectués à des localisations et profondeurs précises pour caractériser un point spécifique ;
- Échantillons composites, obtenus par le mélange de plusieurs prélèvements ponctuels issus d'une même zone homogène, afin d'obtenir une représentativité globale d'un secteur ou d'un horizon donné ;
- Echantillon sélectif : échantillon choisi délibérément en utilisant un plan d'échantillonnage qui élimine des matériaux ou en sélectionne selon des caractéristiques pertinentes.

En cas de dépassement des seuils de dépollution fixés, une phase de terrassement supplémentaire, couramment appelée reprise de travaux, peut être engagée générant des coûts supplémentaires non prévus.

1.2. INTERPOLATIONS DETERMINISTES ET GEOSTATISTIQUES

Les interpolations spatiales font partie des méthodes les plus utilisées pour déterminer un volume de terres polluées. Ces méthodes permettent d'estimer les concentrations de polluants dans des zones non échantillonnées en se basant sur des modèles mathématiques appliqués aux données existantes à l'intérieur du domaine spatial délimité par les échantillons existants. À l'inverse, l'extrapolation désigne l'estimation de valeurs pour des points localisés en dehors de cette zone d'observation initiale (Burrough et McDonnell, 2015). De manière générale, le terme interpolation est tout de même utilisé. On distingue classiquement deux catégories principales de méthodes d'interpolation : i) les approches déterministes et ii) les approches géostatistiques. Les estimations basées sur des méthodes d'interpolation spatiales déterministes ou géostatistiques s'appuient sur des maillages 2D ou 3D.

La littérature recense une multitude de méthodes d'interpolation. Li et Heap (2008) dans une étude comparative portant sur 26 méthodes (incluant des approches déterministes et géostatistiques), ont souligné qu'aucune méthode en particulier ne permettait de minimiser systématiquement l'incertitude associée à l'estimation car leur efficacité dépend fortement des conditions spécifiques à chaque site et des caractéristiques des données (Isaaks et Srivastava, 1989). Leur revue de 51 articles a néanmoins indiqué que le krigeage ordinaire (KRG) était en général la méthode la plus fréquemment recommandée. Zhang et al. (2024) soulignent que la meilleure interpolation peut différer suivant le composé recherché pour un même site : la méthode des fonctions de base radiale (RBF) s'est avérée plus performante pour le cadmium et le molybdène, la pondération par l'inverse de la distance (IDW) pour l'arsenic et le mercure, tandis que le krigeage ordinaire (OK) donnait les meilleurs résultats pour le plomb. L'hétérogénéité des propriétés des différents composés, leurs modes de transport et de transformation, les caractéristiques de leurs sources et la structure de leur corrélation spatiale, justifient qu'une méthode d'interpolation unique ne soit pas universellement optimale pour tous les composés.

Li et al (2020), ont comparé 4 méthodes pour caractériser une contamination du sol par l'antimoine, concluant que le krigeage ordinaire 3D surpassait en termes de performance la méthode des fonctions de base radiale (RBF), la méthode du plus proche voisin et la méthode de Shepard quadratique modifiée (MQS). La méthode RBF estime la valeur en un point non échantillonné par une combinaison linéaire de fonctions radiales centrées sur les points de données existants. Plusieurs fonctions radiales sont utilisables, la plus utilisée est la

multiquadratique. La méthode des plus proches voisins identifie les K points d'observation les plus proches du point où l'on souhaite estimer une valeur. La valeur estimée est ensuite calculée comme la moyenne des valeurs de ces K voisins. La méthode MQS est une technique d'interpolation basée sur la pondération des points voisins en fonction de leur distance au point d'estimation. Contrairement à la méthode classique de Shepard (IDW), la version MQS utilise une fonction quadratique pour un ajustement plus lisse des poids.

La classification spatiale basée sur les polygones de Voronoï est également considérée comme une technique d'interpolation **(**Deng et al., 2024). Cependant, son application pour l'estimation de volumes pollués requiert une délimitation précise de la source afin d'éviter une surestimation des zones contaminées due à des polygones de forte concentration disproportionnément grands (Wang et al., 2023). Les méthodes d'interpolation déterministes présentent l'avantage d'une mise en œuvre relativement simple, nécessitant peu de données d'entrée (uniquement les coordonnées et les concentrations pour certaines méthodes), ce qui explique leur utilisation fréquente. Un inconvénient majeur réside toutefois dans leur incapacité à fournir une quantification de l'incertitude associée aux estimations (Ersoy, 2021; Xie et al., 2011). Les techniques d'interpolation ont toutes un effet de lissage qui sous-estime les valeurs locales élevées et surestime les valeurs locales faibles aux emplacements non échantillonnés (Journel et al., 2000). Ces effets de lissage induisent des erreurs dans l'estimation des volumes de sols pollués compromettant ainsi l'efficacité des mesures de dépollution mises en œuvre (Goovaerts, 2000).

Les méthodes géostatistiques sont utilisées dans le domaine des sites et sols pollués depuis les années 1980 (Moore et McLaughlin, 1980). Ces méthodes s'appuient sur la théorie des variables régionalisées, qui repose sur l'hypothèse que les points de mesure situés à proximité les uns des autres ont une probabilité plus élevée de présenter des valeurs similaires que ceux qui sont plus éloignés. L'application de la géostatistique à l'étude des sites contaminés implique typiquement une séquence d'étapes (Géosipol, 2005) :

- Il est crucial de disposer d'un nombre suffisant de données ; un minimum de 25 à 30 observations pour la variable d'intérêt est souvent recommandé (Oliver et Webster, 2014). Dans les cas où une proportion significative de données est inférieure à la limite de quantification, des approches spécifiques, comme l'utilisation de variables indicatrices, peuvent être intégrées (D'or et al., 2009) ;
- Une analyse exploratoire des données est conduite pour appréhender la distribution statistique des échantillons prélevés. Une transformation des données peut s'avérer

nécessaire si la distribution brute ne correspond pas aux hypothèses sous-jacentes des méthodes utilisées. En effet, les concentrations de contaminants présentent fréquemment des distributions fortement dissymétriques, caractérisées par une abondance de valeurs faibles et la présence de quelques valeurs extrêmes élevées. De telles distributions ne satisfont généralement pas l'hypothèse multigaussienne requise par de nombreuses techniques géostatistiques standards. Diverses transformations, telles que les transformations gaussiennes anamorphique ou logarithmique, peuvent être appliquées pour normaliser ou symétriser les données (RECORD, 2013) ;

- L'étape suivante consiste en la modélisation de la structure spatiale des données via l'établissement d'un variogramme expérimental et l'ajustement d'un modèle théorique (par exemple, sphérique, exponentiel, gaussien). Le variogramme quantifie la variance (ou semi-variance $\gamma(h)$) en fonction de la distance de séparation (h) entre les points d'observation ;
- Une fois la structure spatiale caractérisée, une cartographie de la variable peut être réalisée à l'aide d'un estimateur géostatistique, le krigeage. L'estimation par krigeage permet d'obtenir la meilleure estimation linéaire au sens de la variance d'erreur minimale. Les méthodes les plus utilisées sont le krigeage simple et le krigeage ordinaire. La distinction entre le krigeage ordinaire et le krigeage simple réside dans l'hypothèse faite sur la moyenne de la variable étudiée : cette dernière est considérée comme inconnue et localement constante pour le krigeage ordinaire, alors qu'elle doit être connue et globalement constante pour le krigeage simple. D'autres méthodes de krigeage peuvent être utilisées suivant la problématique (co-krigeage, krigeage avec dérive externe…) (Cavalcante Fraga, 2019).

Comme pour les méthodes déterministes, un inconvénient des estimations par krigeage est leur tendance à produire des résultats lissés, qui ne représentent pas fidèlement la variabilité locale réelle (Demougeot-Renard, 2004). Ce phénomène de lissage atténue la continuité spatiale en moyennant les valeurs estimées. Pour surmonter cet inconvénient, notamment pour l'estimation de volumes de sol à traiter, la réalisation de simulations conditionnelles est souvent jugée plus appropriée (Géosipol, 2005). Une simulation conditionnelle génère de multiples réalisations équiprobables du champ de concentration du polluant, honorant les données aux points d'échantillonnage tout en reproduisant la variabilité spatiale définie par le variogramme. Contrairement au krigeage qui produit une seule carte estimée, la simulation fournit un ensemble de cartes possibles (Fig.2), permettant ainsi de quantifier l'incertitude et d'obtenir la

distribution de probabilité des valeurs possibles de la teneur moyenne de la maille (Demougeot-Renard et al., 2004).

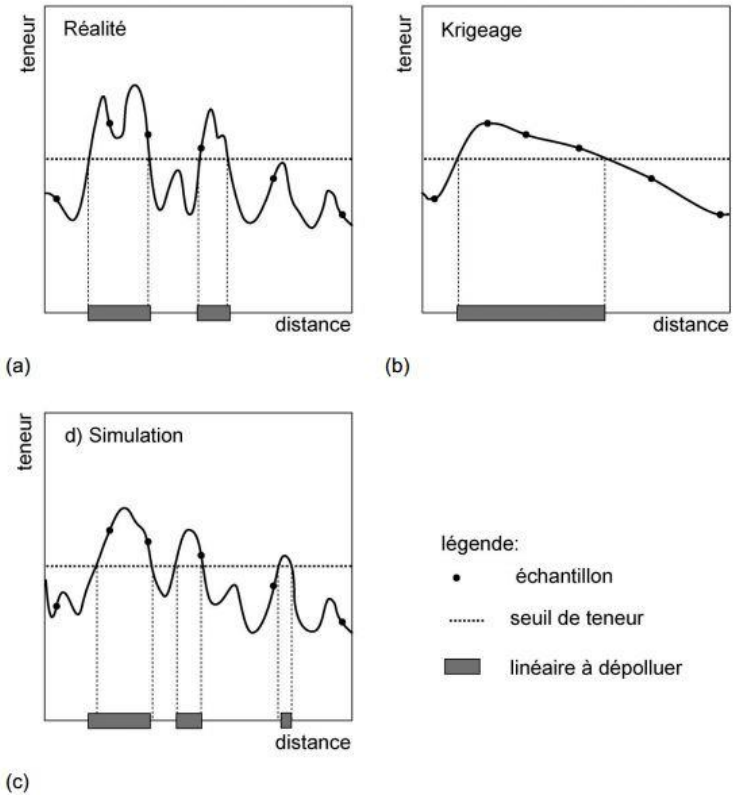


Figure 2 : Illustration de l'application de la géostatistique avec krigeage et simulation (RECORD, 2013)

Un grand nombre de simulations géostatistiques permet de représenter la variabilité locale. Une carte de probabilité de dépassement d'un seuil peut être réalisée à partir des simulations. Celle-ci est calculée avec la proportion de cartes simulées dépassant une teneur en contaminant fixée comme seuil de contamination. Les résultats obtenus permettent d'obtenir l'incertitude pour chaque maille (Boudreault et al., 2016). Deux types de méthodes sont couramment utilisées : les simulations séquentielles et les simulations à bandes tournantes (Chilès et al., 2012). A partir d'une première valeur tirée aléatoirement, la simulation séquentielle consiste en la génération progressive des valeurs à chaque nœud d'une grille de simulation, en s'appuyant sur les valeurs simulées lors des précédentes itérations, de manière à respecter les contraintes de conditionnement (variogramme et histogramme). Les méthodes courantes incluent la simulation séquentielle gaussienne et la simulation séquentielle par indicatrices, qui permet d'intégrer des informations incomplètes sur l'état de contamination des sols. La méthode des bandes tournantes simplifie une simulation 2D (ou 3D) en effectuant des simulations 1D le long de lignes générées aléatoirement. La simulation 2D (ou 3D) est ensuite reconstruite en

moyennant les valeurs projetées des simulations 1D. Cet algorithme génère des simulations reproduisant l'histogramme et le variogramme des données, mais sans respecter les données mesurées.

Les méthodes géostatistiques présentées ci-avant peuvent être utilisées pour estimer le volume d'un sol pollué. Néanmoins, l'estimation par krigeage présente des limitations notables, principalement liées à son effet de lissage intrinsèque qui tend à masquer la variabilité locale, et peut ne pas fournir des résultats optimaux pour la délimitation fine des zones contaminées. En revanche, les simulations conditionnelles offrent une représentation plus réaliste de cette variabilité et permettent une quantification robuste de l'incertitude spatiale. Cependant, la mise en œuvre des simulations est plus exigeante et plus complexe qu'une simple estimation par krigeage (Xie et al., 2011). Le nombre de données nécessaires (plusieurs dizaines d'échantillons) ainsi que la répartition spatialement homogène des points de mesure sont importants pour ces méthodes, ce qui, en pratique, limite souvent leur usage dans le domaine des sites et sols pollués.

Les méthodes géostatistiques peuvent aussi être combinées à d'autres méthodes. Des équations de transport (voir chapitre 1.3) peuvent par exemple être ajoutées en dérivée externe (cas du krigeage calculé à partir de valeurs simulées) (Le Coz et al., 2020, Pannecoucke, 2020). La dérivée externe permet d'estimer une variable peu échantillonnée en s'aidant d'une autre variable connue en tout point. Li et al. (2022), ont mis en œuvre un krigeage bayésien 3D empirique en prenant en compte les variations verticales de lithologie. Les données sont séparées en sous-ensembles afin de calculer un variogramme par lithologie. Tao et al. (2019) ont associé des réseaux de neurones à rétropropagation à un krigeage tridimensionnel pour estimer les délimitations dans les sols de plusieurs polluants organiques et inorganiques. Les résultats montrent une diminution significative de l'erreur d'estimation.

Plusieurs facteurs influencent la sélection de la méthode la plus appropriée, les principaux facteurs sont détaillés dans le Tableau 1 (Li et Heap, 2014). Les auteurs insistent sur la nécessité d'une approche contextuelle pour la sélection de la méthode la plus appropriée, celle-ci devant être déterminée au cas par cas.

Tableau 1 : Principaux facteurs influençant le choix de la méthode d'interpolation (Li et Heap, 2014)

Facteur	Détail
Densité d'échantillonnage	Une densité plus élevée améliore généralement la performance de la plupart des méthodes, en particulier pour les variables à forte variabilité spatiale. Cependant, les bénéfices apportés par chaque nouveau prélèvement diminuent significativement une fois qu'une certaine quantité d'informations est déjà acquise. Lorsque les données sont rares, les hypothèses sous-jacentes sur la variation entre les échantillons peuvent différer, et le choix d'une méthode d'interpolation et de ses paramètres devient critique.
Répartition spatiale des données	La distribution spatiale des points (régulière, aléatoire ou préférentielle) a un impact significatif. L'échantillonnage préférentiel est souvent moins performant que l'échantillonnage régulier ou aléatoire stratifié. L'échantillonnage régulier est souvent préféré pour capturer la structure spatiale globale.
Nombre d'échantillons	La taille de l'échantillon et la conception de l'échantillonnage affectent la fiabilité du variogramme. Généralement, si la taille de l'échantillon est inférieure à 50 (d'autres sources évoquées précédemment indiquent 25 à 30 échantillons), les variogrammes dérivés sont souvent erratiques avec peu ou pas de structure spatiale évidente. Plus la taille de l'échantillon est grande, plus le variogramme est estimé avec précision.
Isotropie ou tendance	La présence de tendances directionnelles dans les données (non-stationnarité de la moyenne) ou d'anisotropie (variation de l'autocorrélation spatiale selon la direction) doit être prise en compte. Des méthodes comme le krigeage universel ou le krigeage avec dérivée externe peuvent gérer les tendances, et les modèles de variogramme peuvent être ajustés pour l'anisotropie.

1.3. MODELISATION DES PROCESSUS D'ECOULEMENT ET DE TRANSPORT

La modélisation consiste en l'élaboration d'une représentation simplifiée de processus réels à partir d'équations mathématiques et physiques. L'objectif est d'analyser et de mieux appréhender le comportement de systèmes complexes par l'intermédiaire d'un modèle qui en capture les caractéristiques essentielles. En fonction du degré de simplification apporté à la description des conditions aux limites et de l'hétérogénéité spatiale du milieu, ces équations peuvent être résolues par des méthodes analytiques ou numériques.

1.3.1. Principales équations physiques

1.3.1.1. Equations décrivant l'écoulement

L'écoulement de l'eau en zone non saturée est communément décrit par l'équation de Richards (1931). L'équation 1.1 présente la formule avec un écoulement vertical unidimensionnel selon l'axe z (Eq.1.1). L'écriture de cette équation en une dimension (1D) implique la négligence des composantes horizontales de l'écoulement (en x et y) par rapport à la composante verticale (en z), une hypothèse couramment admise dans la littérature relative à la zone non saturée (ZNS) (Simunek et van Genuchten, 1999, Varado et al., 2006, Li et al., 2025). Il est cependant possible de recourir à une formulation tridimensionnelle (3D) si une description des composantes horizontales de l'écoulement s'avère nécessaire pour une analyse plus exhaustive.

$$\frac{\partial\, \theta_w(h)}{\partial t} = \frac{\partial}{\partial z}\left(K(\theta_w)\frac{\partial h}{\partial z}\right) + S \qquad (1.1)$$

où θ_w est la teneur en eau volumique [-], t est le temps [T], K est la conductivité hydraulique en zone non saturée [$L.T^{-1}$], h est le potentiel de l'eau [L], S est le terme source [T^{-1}].

Cette équation décrit les variations de la teneur en eau volumique (θ_w), du potentiel matriciel (h) et de la conductivité hydraulique (K) dans le temps et l'espace. Elle doit donc être résolue en utilisant des modèles de capillarité associant h et θ_w. Les approches couramment utilisées permettant de résoudre cette équation sont celles de Brooks et Corey (1966) (Eq.1.2-1.4) et de Van Genuchten (1980) (Eq.1.5-1.7), décrivant la courbe de conductivité hydraulique et la courbe de rétention d'eau dans les sols et en conditions non saturées :

$$K_r = S_e^n \tag{1.2}$$

Où

$$K_r = \frac{K_{\theta w}}{K_s} \tag{1.3}$$

Où K_r est la conductivité hydraulique relative [-], n [-] est l'exposant de Brooks et Corey et S_e [-] est la saturation effective du sol définie par l'équation suivante :

$$S_e = \frac{\theta_w - \theta_r}{\theta_s - \theta_r} \tag{1.4}$$

Où θr [-] est la teneur en eau volumique résiduelle, et θ_s la teneur en eau volumique à saturation [-].

Les paramètres θ_r et θ_s varient en fonction du type de sol, notamment de leur granulométrie (texture), structure et distribution des tailles de pores. Ils peuvent être mesurés en laboratoire ou calculés à partir du code ROSETTA (Schaap et al., 2001) qui utilise notamment la granulométrie du sol (pourcentage de sable, limon et argile), la texture, la fraction de carbone organique, la densité apparente et la teneur en eau volumique mesurée à la capacité de rétention et au point de flétrissement. Ce code de calcul est notamment intégré dans des logiciels de simulation d'écoulement en zone non saturée tel que HYDRUS (Šimůnek et al., 2005).

$$K_r = S_e^{1/2}[1 - \left(1 - S_e^{\frac{1}{\Upsilon}}\right)^{\Upsilon}]^2 \tag{1.5}$$

$$S_e = [1 + \alpha(h)^{\beta}]^{-\Upsilon} \tag{1.6}$$

Où

$$\Upsilon = 1 - \frac{1}{\beta} \tag{1.7}$$

Où α [L^{-1}] et β [-] sont les paramètres de Van Genuchten. Le paramètre α est l'inverse de la succion d'entrée d'air. Le paramètre β contrôle la pente de la courbe de rétention d'eau dans la zone de transition entre saturation et drainage ; plus β est grand, plus cette transition est rapide. Une valeur α élevée est corrélée à une plus faible capacité de rétention d'eau matricielle du sol, traduisant une désaturation plus rapide du sol. Le paramètre β caractérise la pente de la courbe

de rétention d'eau et est lié à la distribution de la taille des pores du sol. Plus β est grand, plus la distribution de la taille des pores est étroite.

1.3.1.2. Equations décrivant le transport

Les processus de transports généralement pris en compte sont : i) l'advection, ii) la dispersion, iii) la diffusion, iv) la sorption, v) la volatilisation, et vi) la dégradation (Zhang, 2019).

Les équations de transport sont basées sur la conservation de la masse du polluant (Eq. 1.8 à 1.10) :

$$\frac{\partial C_t}{\partial t} = \frac{\partial J_t}{\partial z} - \omega \tag{1.8}$$

Avec (en considérant l'absence de phase organique dans l'échantillon) :

$$C_t = \rho_b C_s \theta_{so} + \theta_w C_w + C_a \theta_a \tag{1.9}$$

Où :

$$J_t = J_w + J_a \tag{1.10}$$

Où C_t correspond à la concentration totale du polluant [$M.L^{-3}$], J_t est le flux de soluté total [$M.L^{-2}.T^{-1}$] , ω est un terme source représentant un gain ou une perte du polluant et pouvant représenter diverses réactions chimiques ou biologiques [$M.L^{-3}.T^{-1}$], ρb est la masse volumique apparente du sol [$M.L^{-3}$], C_s est la concentration du composant adsorbé sur la matière organique du sol [$M.L^{-3}$], C_w est la concentration du soluté dans l'eau [$M.L^{-3}$], C_a est la concentration du composant dans l'air du sol [$M.L^{-3}$], θ_{so} est la fraction volumique de sol [-], θ_a est la teneur volumique en air [-], J_w est le flux de soluté dans la phase liquide [$ML^{-2}T^{-1}$], J_a est le flux de soluté dans la phase gazeuse [$ML^{-2}T^{-1}$].

Une grande majorité de modèles ne considèrent que le transport dans la phase aqueuse (Karapanagioti et al., 2003). Les modèles d'écoulement de fluides multiphasiques nécessitent généralement des équations pour chaque phase fluide (eau, air, phase organique). La phase organique est donc majoritairement négligée ou intégrée dans la phase aqueuse afin de limiter les calculs (Cushman et Tartakovsky, 2016).

Les termes J_w et J_a correspondent à la somme des 3 principaux processus actifs : l'advection, la dispersion hydrodynamique et la diffusion moléculaire. A noter que les termes

de transports advectif et dispersif dans la phase gazeuse sont généralement négligés par rapport à la diffusion gazeuse, ces phénomènes ne sont donc pas inclus dans les équations suivantes :

$$J_w = qc - \theta_w D_w \frac{\partial c_w}{\partial z} - \theta_w D_{lh} \frac{\partial c_w}{\partial z} = qc - \theta_w D_e \frac{\partial c_w}{\partial z} \quad (1.11)$$

$$J_a = -\theta_a D_a \frac{\partial c_a}{\partial z} \quad (1.12)$$

où qc est le terme d'advection [LT^{-1}], D_w correspond au coefficient de diffusion dans l'eau [$L^2.T^{-1}$], D_{lh} est le coefficient de dispersion hydrodynamique [$L^2.T^{-1}$], D_e est le coefficient de dispersion effectif [$L^2.T^{-1}$], D_a est le coefficient de diffusion dans la phase gazeuse.

La combinaison des équations 1.10, 1.11 et 1.12 conduit à l'équation d'advection-dispersion. Dans un système unidimensionnel en l'absence de phase volatile et en considérant des réactions de sorption linéaires, réversibles et à cinétique instantanée, l'équation d'advection-dispersion peut s'écrire (Eq 1.13 à 1.15) :

$$\frac{\partial(\rho_b C_s + \theta_w C_w)}{\partial t} = \frac{\partial(\theta_w R C_w)}{\partial t} = \frac{\partial}{\partial z}(\theta_w D_e \frac{\partial c_w}{\partial z}) - \frac{\partial qc}{\partial z} - \omega \quad (1.13)$$

Où :

$$R = 1 + \frac{\rho b}{\theta_w} K_d \quad (1.14)$$

Avec :

$$Kd = foc * K_{oc} \quad (1.15)$$

où foc est la fraction en carbone total [-] et K_{oc} est le coefficient de partage carbone organique/eau [$L^3.M^{-1}$].

1.3.2. Modélisation analytique

En formulant l'hypothèse d'un milieu spatialement homogène avec des conditions aux limites simples, il est possible de résoudre les équations de transport (comme l'équation d'advection-dispersion) par des méthodes analytiques. L'équation de Richards, qui décrit l'écoulement en milieu non saturé, ne possède généralement pas de solution analytique

(Cushman et Tartakovsky, 2016). Cet aspect met en évidence les limites de l'approche analytique pour l'écoulement. Lu et Zhang (2004) ont, par exemple, développé une version très simplifiée de l'équation de Richards. Cette simplification leur a permis d'obtenir une solution analytique pour un système multicouche (c'est-à-dire prenant en compte une stratification verticale), dans le cas d'un écoulement unidimensionnel en milieu non saturé. Il convient donc de noter que la résolution analytique est principalement applicable à l'équation de transport, où l'écoulement est implicitement pris en compte via le terme advectif. Les solutions analytiques (ou modèles analytiques) sont des équations mathématiques qui sont simples et rapides à utiliser. Elles possèdent l'avantage d'être exactes et continues dans le temps et dans l'espace. Les hypothèses de simplification impliquent généralement :

- une géométrie simple ;
- un milieu homogène ;
- un écoulement uniforme en régime permanent ;
- une injection instantanée ou continue d'un seul polluant.

Par conséquent, les solutions analytiques sont naturellement inadaptées aux problèmes de pollution impliquant un milieu hétérogène et/ou un écoulement transitoire et/ou des réactions de transport non linéaires (Šimůnek, 2006). L'écoulement de l'eau et le transport de soluté sont fortement influencés par la présence de macropores et d'éléments structurés, mais aussi par la variabilité spatiale des propriétés du sol. Certaines solutions analytiques prennent en compte des sols homogènes stratifiés (Srivastava et al., 1991).

La majorité des solutions analytiques publiées dans la littérature ont été dérivées en appliquant des transformations (ex. Laplace ou Fourier) aux équations d'écoulement et de transport. Ces transformations permettent de simplifier les équations différentielles partielles en équations différentielles ordinaires. La transformation de Kirchhoff a été largement utilisée pour convertir des équations non linéaires en équations différentielles linéaires qui sont résolues analytiquement.

Les premiers modèles analytiques unidimensionnels de transport en zone non saturée datent des années 1980 (Van Genuchten, 1981, Jury et al., 1983, Parker et van Genuchten, 1984, Ünlü et al., 1992, Lahvis et Baehr 1997). Des modèles multidimensionnels de transport en zone non-saturée ont été proposés par Leij et al., (1991), Šimůnek et al., (1999a), Šimůnek, (2006), Troldbord et al., (2009). Certains de ces modèles ont été implémentés dans des logiciels tels

que ASTM (ASTM, 2000) et RISC4 (Spence et al., 2001) et peuvent être utilisés pour réaliser des calculs de risques sanitaires en estimant notamment la concentration inhalée en polluant à partir de la volatilisation des polluants présents dans les sols (Diallo et al., 2015, Liu et al., 2021, Pu et al., 2024). Une comparaison de diverses solutions analytiques de transport a été réalisée par Stoppiello et al., (2020) en appliquant les modèles à des jeux de données similaires. Différents polluants ont été considérés (benzène, benzo(a)pyrène, chlorure de vinyle, trichloroéthylène et aldrine), ainsi que différents types de textures de sols (sables, limons et argiles). L'étude met en évidence que les modèles les plus simplifiés conduisent aux concentrations les plus importantes avec parfois plusieurs ordres de grandeur d'écart.

Dans la zone saturée, il existe des solutions analytiques aux équations d'écoulement. La méthode des éléments analytiques permet de superposer plusieurs solutions analytiques dans le temps et dans l'espace. Cette méthode possède les avantages des solutions analytiques à savoir, l'absence de maillage et pas de conditions aux limites à fixer. Elle a été développée principalement pour l'écoulement en zone saturée (Strack, 2003, Fitts, 2013) où les champs d'écoulement associés à chaque élément analytique (un puits de pompage, une zone de recharge…) sont ensuite combinés (ou superposés). Bien que cette méthode soit capable de représenter des cas complexes (présences de puits, sources, variation de conductivité hydraulique, présence de rivières ou limites imperméables), la géométrie et les conditions aux limites restent simples. L'utilisation de cette méthode en zone non saturée est limitée. Une autre méthode appelée méthode analytique finie s'appuie sur une discrétisation spatiale (maillage) de la zone d'étude et applique la transformation de Kirchhoff pour résoudre analytiquement l'équation de l'écoulement non saturé à l'échelle de chaque maille. L'objectif est d'obtenir une meilleure stabilité et précision de calcul qu'avec une méthode numérique, notamment dans les zones de transitions hétérogènes (Zhang et al., 2016, Zhang, 2018). Cette méthode est très dépendante de la taille des mailles et peut nécessiter des temps de calcul importants.

Les solutions analytiques sont fréquemment employées pour caractériser les panaches de polluants en zone saturée (Domenico et Robbins,1985, Paladino et al., 2018, Siregar et al., 2024). Cependant, leur application est moins répandue en zone non saturée pour la détermination des volumes de terres polluées. Cette limitation tient principalement à la complexité inhérente des processus d'écoulement et de transport en milieu non saturé (variabilité spatiale des propriétés hydriques, non-linéarité de la relation pression-humidité), qui rendent les hypothèses simplificatrices requises pour une résolution analytique moins

représentatives ou difficiles à satisfaire, comparativement aux conditions plus stables souvent rencontrées en zone saturée.

1.3.3. Modélisation numérique

La modélisation numérique s'appuie sur une discrétisation de l'espace et du temps pour résoudre les équations d'écoulement et/ou de transport. Cette approche offre l'avantage de pouvoir intégrer des géométries et des conditions aux limites complexes, reflétant plus fidèlement les conditions géologiques et hydrogéologiques naturelles. Bien que les modèles numériques permettent ainsi un plus grand réalisme, leur implémentation peut s'avérer complexe et longue à mettre en œuvre (Schaap et al., 2004). La modélisation numérique de l'écoulement en zone non saturée demeure complexe, notamment en raison de la non-linéarité de l'équation de Richards, du nombre de paramètres à acquérir et des hypothèses de simplification (Rajabi Naraki, 2024).

Le développement de la modélisation numérique en hydrogéologie a débuté dans les années 1960 et a initialement été axé sur la simulation des écoulements en zone saturée (Tyson et Weber, 1964). L'application à la zone non saturée a suivi dans les années 1980 avec l'émergence de codes comme UNSAT (David et Neuman, 1983) et SWMII (Vogel, 1987). Les modèles numériques capables de simuler simultanément l'écoulement et le transport de solutés en milieu non saturé sont apparus dans les années 1990, incluant des logiciels tels que VS2DI (Healy, 2010), HYDRUS-1D (Šimůnek et al., 1998) et MODFLOW-SURFACT (HydroGeoLogic, 1996). Des développements plus récents ont conduit à des codes plus avancés dotés de capacités améliorées de gestion des maillages non structurés, de modélisation 3D intégrée écoulement/transport, de simulation de phénomènes complexes tels que le transport dépendant de la densité… Ces codes incluent MODFLOW USG (Panday, 2013), HydroGeoSphere (Brunner et al., 2012), SUTRA (Voss et Provost, 2010) et FEFLOW (FEFLOW, 2015).

Pour intégrer des processus hydrogéochimiques complexes, la modélisation du transport réactif peut être réalisée à l'aide de logiciels spécialisés, tels que PHREEQC (Parkhurst et Appelo, 2013). La modélisation numérique en zone non saturée est également employée pour confronter les résultats d'essais en laboratoire (expériences en colonne) à différents modèles conceptuels d'écoulement et de transport (Viotti et al., 2005 ; Yang et al., 2020, Shu et al., 2023).

Les simulations numériques de panaches de pollutions sont majoritairement réalisées en zone saturée (Burnett et Frind, 1987, Beyer et al., 2006, He et al., 2024), mais plusieurs études ont également cherché à quantifier le temps nécessaire à une pollution pour traverser la zone non saturée et contaminer la nappe sous-jacente (Sousa et al., 2013, Potrykus et al., Szymkiewicz et al., 2019). Huan et al. (2020) et El-Aassar et al. (2023) ont utilisé HYDRUS 1D pour estimer le risque de contamination en nitrates et en métaux lourds des nappes d'eaux souterraines par rapport à la localisation des sources potentielles. Ciampi et al. (2021) ont appliqué la modélisation hydrogéochimique à grande échelle pour prédire l'atténuation naturelle de contaminations en COHV dans les eaux souterraines issues de sources localisées dans les sols.

1.3.4. Calibration

Indépendamment de l'approche de modélisation adoptée, qu'elle soit analytique ou numérique, la calibration du modèle représente une étape fondamentale (Anderson et al., 2015). L'objectif principal de la calibration est d'identifier les valeurs optimales de paramètres qui permettent de minimiser l'écart (ou la fonction objectif) entre les mesures réalisées sur le terrain (concentrations par exemple) et les valeurs correspondantes simulées par le modèle. On désignera par « inversion » l'ensemble des méthodes permettant d'automatiser la calibration d'un modèle à l'aide d'algorithmes.

Les paramètres de modèles peuvent être déterminés par des méthodes d'inversion. À titre d'exemple, des paramètres liés à l'équation de Richards, tels que α, n et Ks du modèle de van Genuchten, peuvent être inversés pour caractériser les propriétés hydrauliques des sols en zone non saturée. De manière plus générale, cette inversion est réalisée à l'aide de méthodes d'optimisation ou d'algorithmes d'inversion variés, tels que le Global Multilevel Coordinate Search (GMCS) combiné séquentiellement avec l'algorithme classique Nelder-Mead Simplex (NMS) (Lambot et al., 2002), le PIMLI (Vrugt et al, 2001), ou PEST (Parameter ESTimation) (Dohertyb, 2018). Ces algorithmes ne sont pas spécifiques à l'équation de Richards et peuvent être appliqués à diverses problématiques d'inversion, incluant l'estimation de paramètres pour des modèles de transport.

Concernant les modèles de transport de polluants, qu'il s'agisse de modèles analytiques ou numériques, de nombreux paramètres peuvent faire l'objet d'une inversion. Cela inclut non seulement les paramètres définissant une source de pollution (tels que sa localisation, ses dimensions, sa concentration initiale et sa durée d'émission), qui sont souvent inconnus ou peu

contraints par les données, mais également les paramètres régissant les processus de transport et de réactivité. Ces derniers peuvent comprendre les coefficients d'advection, de dispersion, ainsi que les paramètres de sorption ou de dégradation du polluant. L'inversion de ces différents types de paramètres permet de mieux caractériser les phénomènes de propagation et d'atténuation des contaminants dans les milieux poreux, y compris dans les eaux souterraines (Essouayed, 2019). Seul l'outil PEST est détaillé dans cette synthèse, il est aujourd'hui l'un des logiciels d'optimisation les plus performants et flexibles (Espinet et al, 2013, Anderson et al. 2015, Winston, 2024).

PEST est un logiciel libre qui permet d'optimiser les paramètres d'un modèle non linéaire et d'analyser l'incertitude associée aux paramètres optimisés (Doherty et al, 2010_a). Il peut être couplé à de nombreux modèles dont les modèles analytiques et numériques précédemment évoqués. PEST utilise la méthode de Gauss-Marquardt-Levenberg qui permet d'optimiser les paramètres en suivant un processus itératif. Un problème d'optimisation pouvant avoir plusieurs solutions possibles, une incertitude significative en découle. Dans ce contexte, PEST intègre des outils dédiés à l'analyse de ces incertitudes, ce qui est crucial pour évaluer la robustesse de la calibration du modèle.

Une fonction SVD (Singular Value Decomposition) peut être utilisée avec PEST, permettant d'analyser la structure et la qualité des solutions possibles compte tenu de la sensibilité de certains paramètres (Kalman, 1996). Une décomposition de la matrice jacobienne (ou matrice de sensibilité) est réalisée, permettant de déterminer les valeurs singulières pour optimiser la convergence du processus d'inversion des paramètres vers une solution optimale.

Lorsque le nombre d'observations disponibles est limité ou que le modèle est particulièrement complexe, des techniques de régularisation peuvent être employées durant le processus d'inversion. La régularisation introduit des contraintes supplémentaires dans le problème inverse, visant à réduire l'instabilité de la solution et à établir un compromis entre l'ajustement aux données mesurées et la plausibilité ou la régularité du modèle. Parmi les diverses formes de régularisation, celle de Tikhonov est l'une des plus connues (Tikhonov et al., 1995) ; elle permet par exemple de privilégier les solutions qui présentent une homogénéité spatiale maximale (Alcolea et al., 2008) ou bien les solutions s'écartant le moins possible de valeurs prédéfinies par l'utilisateur.

Vis-à-vis du processus d'inversion, deux points de vigilance découlent de la non-linéarité des modèles de transport et du nombre de paramètres impliqués. Il s'agit d'une part de la

sensibilité aux valeurs initiales des paramètres et d'autre part de l'existence éventuelle d'une multiplicité de solutions optimales équivalentes dans l'espace des paramètres :

- La première notion traduit une convergence possible du processus d'optimisation vers un minima local de la fonction objectif au lieu du minima global recherché. Pour essayer de contourner ce problème, il est possible de recourir à une procédure dite « multi-start » (Martínez-Estudillo, 2006, Yin et Sykes, 2006, Piotrowski et Napiorkowski, 2011) qui consiste en la répétition du processus d'optimisation en partant de valeurs initiales différentes afin d'augmenter les chances de converger vers la solution optimale globale. Les valeurs initiales peuvent être tirées aléatoirement dans des gammes de valeurs identifiées dans la littérature ou bien basées sur une connaissance à priori du site investigué ;
- Le second point, communément appelé « non-unicité de la solution » dans la littérature, concerne les cas où le processus d'inversion peut converger vers des jeux de paramètres de valeurs différentes mais équivalentes en termes de calibration du modèle.

Une fois le modèle calibré, une analyse d'incertitude peut être réalisée. Deux grandes sources d'incertitudes peuvent être considérées (Anderson et al., 2015) :

- les incertitudes liées à la construction du modèle, comme les hypothèses de départ et/ou les erreurs de mesures associées aux observations ayant servi à calibrer le modèle (Moore et Doherty, 2005) ;
- les incertitudes liées aux prévisions, comme les précipitations dans le futur, les calendriers de pompage, les emplacements de futures sources de pollution ou les actions politiques, économiques et sociétales qui pourraient affecter l'hydrogéologie (Hunt, 2005).

L'analyse des incertitudes peut être conduite subséquemment au processus d'inversion des paramètres. Les résultats recherchés sont l'incertitude sur la variance ou un intervalle de confiance. Les méthodes d'analyse d'incertitude linéaires, telles que celles basées sur la propagation de la variance de premier ordre sont adaptées aux problématiques d'inversion. Celles-ci s'appuient sur la matrice jacobienne et permettent de calculer rapidement les incertitudes. Elles peuvent intégrer les incertitudes liées aux erreurs de mesure. La variance est calculée en comparant la valeur réelle à celle prévue par le modèle (Moore et Doherty, 2005). D'autres méthodes existent pour calculer les incertitudes. Celles-ci permettent d'aboutir aux mêmes résultats (variance ou intervalle de confiance à 95%). Les calculs sont obtenus là aussi

à l'aide de logiciels d'estimation des paramètres. Parmi ces approches, des variantes de la méthode de Monte Carlo, telles que le Null Space Monte Carlo, sont couramment utilisées pour l'analyse d'incertitude et la caractérisation des paramètres dans des problématiques d'inversion (Hunt et al., 2001, Tonkin et al., 2009). Cette approche permet d'explorer l'espace des paramètres qui produisent des solutions acceptables, fournissant ainsi une estimation robuste de leurs incertitudes, à la différence de la méthode de Monte Carlo standard qui est principalement employée pour la simulation directe ou l'intégration numérique. Celle-ci utilise un grand nombre de simulations (à minima 100 jusqu'à plusieurs milliers) du modèle. L'objectif est d'explorer l'espace des paramètres afin d'obtenir une caractérisation de la distribution des paramètres qui produisent des solutions acceptables.

1.4. METHODES UTILISANT L'INTELLIGENCE ARTIFICIELLE

Les modèles utilisant l'intelligence artificielle sont de plus en plus utilisés pour établir des cartes de pollution des sols. Les modèles les plus utilisés sont ceux à base de réseaux de neurones artificiels.

Les données acquises sur le terrain, issues de sites pollués, révèlent majoritairement des relations ou processus non-linéaires, ce qui complexifie significativement leur modélisation. Les réseaux de neurones artificiels sont adaptés à cette problématique (Chukwunonso et al., 2022). Ils sont capables de modéliser efficacement des processus non-linéaires **(**Gautam, 2023). Ils comportent un modèle à trois types de couches (des couches d'entrée, des couches cachées et des couches de sortie) avec de nombreux liens entre elles. Zhou (2015) a utilisé une approche par réseaux de neurones artificiels (modèle ANN) couplée à un algorithme génétique adaptif (AGA). Ce modèle a permis d'identifier des sources de pollution potentielles sur une zone de superficie conséquente (400 km²). Trois cent quinze échantillons de sol de surface ont été prélevés puis classés selon 5 catégories (zone résidentielle, zone industrielle, région montagneuse, zone de circulation, parc ou espace vert). Des cartographies de concentrations pour les 8 métaux/ métalloïdes (arsenic, cadmium, chrome, cuivre, mercure, nickel, plomb et zinc) ont été réalisées à l'aide du modèle ANN rétropropagation. Le modèle AGA a permis d'identifier les coordonnées des sources de pollution.

Une méthode de Machine Learning appelée Fuzzy Classification a été utilisée par Jia et al. (2021) pour déterminer des concentrations en Arsenic dans les sols. Les données ont été obtenues à partir d'une imagerie aérienne haute résolution. Quatre algorithmes d'apprentissage

automatiques ont été développés pour prédire les résultats. L'algorithme ERF (Extreme Random Forest) a permis d'obtenir les meilleurs résultats. Guridi et al. (2024), propose une approche associant les techniques d'interpolation avec du Deep Learning. Celle-ci est appelée "Evol", elle génère un modèle synthétique du champ spatial du contaminant, en extrait des caractéristiques d'échantillons et entraîne un réseau neuronal profond pour estimer le volume de sol contaminé et son incertitude. Les résultats montrent une réduction significative (jusqu'à 89%) de la taille des intervalles de confiance en volume.

1.5. SYNTHESE DU CHAPITRE

La quantification des volumes de terres affectées par des polluants représente un défi scientifique complexe. Les méthodes les plus utilisées aujourd'hui sont les interpolations déterministes et la géostatistique. Ces méthodes présentent des résultats très différents suivants les différents contextes et les polluants rencontrés.

La modélisation numérique est peu utilisée en zone non saturée compte tenu du travail important à mettre en œuvre pour réaliser la calibration. Bien que la modélisation analytique offre une alternative d'application plus aisée, elle n'est naturellement pas adaptée à des cas complexes présentant une hétérogénéité marquée.

CHAPITRE 2 : APPROCHE PAR MODELISATION ANALYTIQUE INVERSE

Ce chapitre introduit une méthode d'estimation des volumes de terres polluées qui s'appuie sur l'utilisation combinée d'une solution analytique et de techniques d'inversion. Il est souligné que cette approche ne constitue pas un développement théorique fondamental nouveau, mais plutôt une méthodologie innovante qui intègre et adapte des outils et concepts existants (solutions analytiques et méthodes d'inversion) pour aborder le problème de l'estimation des volumes de sols pollués en milieu hétérogène.

2.1. PRINCIPE DE LA METHODE

La plupart des milieux naturels étant spatialement hétérogènes, l'hypothèse d'homogénéité inhérente aux modèles analytiques limite leur applicabilité théorique aux cas réels. La méthodologie novatrice proposée dans la thèse vise à surmonter cette limitation et permettre d'appréhender l'hétérogénéité spatiale du milieu. Au lieu d'utiliser de manière synchrone l'ensemble des points d'observations disponibles, l'idée est d'appliquer séquentiellement le modèle à différents sous-ensembles de points d'observations. Cette démarche génère de multiples modèles locaux, chacun représentant un modèle calibré sur une partie des données. La combinaison de ces modèles multiples permet ensuite de reconstruire une représentation globale décrivant l'hétérogénéité du milieu. Plusieurs approches pourraient être envisagées pour définir les sous-ensembles de points ; celle retenue dans la thèse considère des triplets de points d'observation définis selon une triangulation de Delaunay. Le modèle analytique est donc séquentiellement inversé sur chaque triplet de point. L'ensemble des paramètres de la solution analytique présentée ci-après sont inversés. Les gammes de paramètres choisies pour l'inversion peuvent provenir de la bibliographie, de valeurs empiriques, ou en cohérence avec les caractéristiques du site.

Deux points de vigilance découlant de la non-linéarité des modèles de transport et du nombre de paramètres impliqués ont été évoqués dans la section (1.3.4). Il s'agit d'une part de la sensibilité aux valeurs initiales des paramètres et d'autre part de l'existence éventuelle d'une multiplicité de solutions optimales équivalentes dans l'espace des paramètres.

Pour répondre à ces points de vigilance :

- pour chaque triangle de Delaunay, si le processus d'inversion ne parvient pas à atteindre la valeur cible de la fonction objectif (c'est-à-dire en cas d'échec de la calibration), une nouvelle tentative d'optimisation est initiée. Cette nouvelle tentative implique un nouveau tirage aléatoire des paramètres initiaux. Ce processus

de re-échantillonnage des paramètres suivi d'une nouvelle inversion est répété jusqu'à l'obtention d'une solution satisfaisante ou jusqu'à un maximum de dix tentatives par triangle ;

- l'estimation d'un volume de sol pollué à partir de données de concentrations ponctuelles doit intégrer autant que possible la multiplicité de solutions et il est donc nécessaire de multiplier les inversions afin de pouvoir identifier différents jeux de paramètres optimaux associés à chaque triplet de points d'observation. Dix inversions sont réalisées par triangle de Delaunay dans le cas présent.

Une fois toutes les inversions réalisées, une multitude de jeux de paramètres est disponible, chacun permettant de calculer la concentration en tout point du milieu (x, y, z) à l'aide d'un modèle analytique tridimensionnel de transport.

Il est possible d'estimer, à partir de l'ensemble des résultats, la probabilité de dépasser en chaque point un seuil de concentration donné :

$$p(C_w(x,y,z,t) > C_{crit}) = \frac{i'}{i} \tag{2.1}$$

où C_{crit} est la valeur seuil de concentration[M.L^{-3}], i est le nombre total de jeux de paramètres optimisés [-], et i' est le nombre de jeux de paramètres conduisant à une valeur $C_w(x, y, z, t) > C_{crit}$.

Les calculs de probabilité peuvent être pondérés afin de donner plus d'importance aux valeurs de concentrations calculées à partir des jeux de paramètres optimisés sur le triangle de Delaunay contenant le point x, y, z :

$$p(C_w(x,y,z,t) > C_{crit}) = \frac{na'}{na} * \upsilon + \frac{nb'}{nb} * \sigma \tag{2.2}$$

Avec na' + nb' = i' ; na + nb = i ; σ+ υ = 1

où na'/na est la proportion de modèles optimisés sur le triangle de Delaunay contenant le point x, y, z et conduisant à une valeur $C_w(x, y, z, t) > C_{crit}$, nb'/nb est la proportion de modèles optimisés sur les triangles de Delaunay « externes », i.e., ne contenant pas le point x, y, z, et conduisant eux aussi à une valeur $C_w(x, y, z, t) > C_{crit}$, υ et σ sont les coefficients de pondérations respectivement attribués aux deux proportions.

Il est ainsi possible d'établir une carte quantifiant l'incertitude spatiale de dépassement d'un seuil de pollution sur une zone d'étude. Le calcul a été réalisé à l'aide d'un code développé en R (R Core Team, 2024).

Le logigramme de la Figure 3 résume la méthodologie. Dans les prochaines sections, cette méthodologie est appelée modélisation analytique inverse ou IAM (Inverse Analytical Modeling).

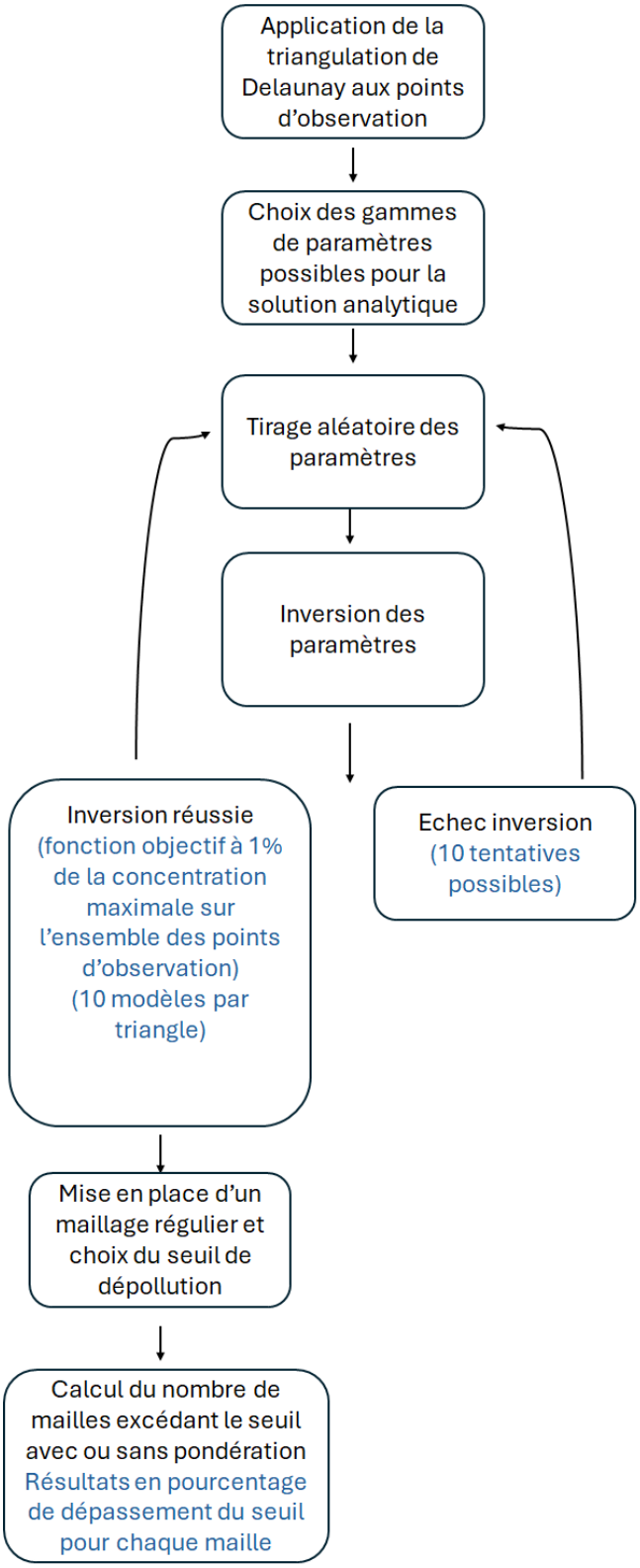


Figure 3 : Logigramme de la méthodologie

Le modèle analytique utilisé dans ce travail est la solution développée par Troldbord et al. (2009) pour l'équation de transport (2.3) , laquelle prend en compte l'advection, la dispersion et la diffusion dans l'eau, la diffusion dans les gaz du sol, la sorption et la dégradation :

$$\frac{\partial(R\theta_w+H\theta_a)C_w}{\partial t} = \nabla(\theta_w \mathrm{D_w} + H\theta_a \mathrm{D_a})\nabla C_w \; - Ieff\frac{\partial C_w}{\partial \mathrm{z}} - \theta_w \lambda C_w \tag{2.3}$$

où R est le coefficient de retard [-], θ_w est la teneur en eau volumique [-], H la constante de Henry du polluant [-], θ_a est la teneur volumique en air [-], C_w est la concentration du soluté dans l'eau [$M.L^{-3}$], *Ieff* est la vitesse d'infiltration [$L.T^{-1}$], λ est la constante de dégradation de premier ordre [T^{-1}], ρb est la masse volumique apparente du sol [$M.L^{-3}$], z est l'axe vertical [L]. (positif vers le haut)

La solution analytique proposée par Troldbord et al. (2009), qui considère une injection continue de concentration constante C_{w0} [$M.L^{-3}$] et de dimension X et Y [L], s'écrit :

$$C_w\,(x,y,z,t) = \frac{Cw0}{8}\int_{\tau=0}^{\tau=t} f'x(x,\tau)\, f'y\,(y,\tau)\, f'z(z,\tau)d\tau \tag{2.4}$$

Avec :

$$f'x(x,\tau) = \mathrm{erf}\left[\frac{(xobs-Xs)+\frac{X}{2}}{2\sqrt{D'_x\tau}}\right] - \mathrm{erf}\left[\frac{(xobs-Xs)-\frac{X}{2}}{2\sqrt{D'_x\tau}}\right] \tag{2.5}$$

$$f'y(y,\tau) = \mathrm{erf}\left[\frac{(yobs-YS)+\frac{Y}{2}}{2\sqrt{D'_y\tau}}\right] - \mathrm{erf}\left[\frac{(yobs-YS)-\frac{Y}{2}}{2\sqrt{D'_y\tau}}\right] \tag{2.6}$$

$$f'z(z,\tau) = \frac{z}{\sqrt{\pi D'_z}}\exp\left(\frac{v'z}{2D'_z}\right)\exp\left[-\tau\left(\frac{(v')^2}{4D'_z}+\lambda'\right)-\frac{z^2}{4D'_z\tau}\right]\tau^{-\frac{3}{2}} \tag{2.7}$$

$$\lambda' = \frac{\theta_w\lambda}{R'}$$

$$v = \frac{Ieff}{\theta_w} \text{ et } v' = \frac{Ieff}{R'} \tag{2.8}$$

(2.9)

$$D'_x = D'_y = \frac{\theta_w D^e{}_w + \theta_w \alpha_t v + H\theta_a D^e{}_a}{R'} \quad (2.10)$$

$$D'_z = \frac{\theta_w D^e{}_w + \theta_w \alpha_l v + H\theta_a D^e{}_a}{R'} \quad (2.11)$$

$$R' = R\theta_w + H\theta_a \quad (2.12)$$

où *xobs* et *yobs* sont les coordonnées horizontales [L], *Xs* et *Ys* sont les coordonnées du centre de la source [L], D^e_w est le coefficient de diffusion de la phase liquide dans les pores [$L^2.T^{-1}$], α_l est la dispersivité longitudinale [L], α_T est la dispersivité transverse [L], H la constante de Henry du polluant [-], θ_a est la teneur volumique en air [-], D^e_a est le coefficient de diffusion effectif dans la phase gazeuse du milieu poreux [$L^2.T^{-1}$]

Les hypothèses sur lesquelles repose cette solution analytique sont les suivantes : (i) teneur en eau volumique, masse volumique apparente, porosité et fraction de carbone organique spatialement uniformes et constantes dans le temps, (ii) advection unidirectionnelle selon la direction verticale, (iii) réactions de sorption réversibles, instantanées, et limitées aux interactions entre la phase aqueuse et la phase solide, (iv) réactions de dégradation limitées à la phase aqueuse et pouvant être décrites comme un processus du premier ordre.

2.2. LIMITES DE LA METHODE

La méthodologie est développée pour estimer les concentrations dans les sols dépassant un seuil de dépollution (seuil critique) en zone non saturée. Elle repose sur une solution analytique qui intègre des hypothèses simplificatrices (voir section 2.1). Il convient de préciser que la solution de Troldborg et al. (2009), bien qu'adaptée aux sites pollués, s'appuie sur le formalisme mathématique de Leij et al. (1991) qui suppose un écoulement uniforme et stationnaire. Cette hypothèse constitue une limite intrinsèque car elle ne reflète pas la dynamique réellement transitoire des flux en zone non saturée caractérisée par l'alternance des pluies et de l'évapotranspiration. Des comparaisons bibliographiques (Schotanus et al., 2013 ; Kuntz et Grathwohl, 2009) indiquent que cette simplification est acceptable (pas ou peu de différence) pour des bilans de masse à très long terme. Elle peut cependant induire des erreurs significatives sur les concentrations de pointe et les temps de transfert pour des polluants très mobiles ou lors

d'événements de recharge extrêmes. Les principales limites de l'IAM sont celles de la solution analytique à savoir qu'elle est adaptée uniquement à des polluants organiques, en zone non saturée. Les concentrations mesurées dans l'eau peuvent présenter des valeurs anormales, dépassant parfois la limite de solubilité du composé. Cette anomalie résulte de la nature triphasique de la solution analytique, qui rend impossible le calcul de la teneur dans la phase organique, qui est partiellement incluse dans les concentrations estimées pour la phase aqueuse. La méthodologie peut être transposée à toute solution analytique, y compris le transport dans une zone saturée (changement de solution analytique).

2.3. SYNTHESE DU CHAPITRE

Une nouvelle méthodologie pour l'estimation des volumes de terres polluées, fondée sur l'inversion de solutions analytiques, est introduite. L'implémentation et la vérification de cette approche seront détaillées dans les chapitres suivants, d'abord via une confrontation à des cas d'étude synthétiques (Chapitre 3), puis via une application à des cas réels (Chapitre 4).

CHAPITRE 3 : APPLICATION A DES CAS SYNTHETIQUES

Dans ce chapitre, la méthode IAM présentée au chapitre 2 est appliquée à différents scénarios synthétiques qui se distinguent par leurs caractéristiques d'hétérogénéité spatiale et leur dimensionnalité (2D homogène, 2D hétérogène et 3D hétérogène). Pour chaque cas, un scénario de pollution synthétique est simulé à l'aide d'un modèle numérique de transport 3D sur une durée de deux ans afin d'établir un état de référence. Ce temps d'observation a été retenu car il permet d'observer le panache en phase transitoire : la migration du contaminant est amorcée, mais le système n'a pas encore atteint un régime permanent, garantissant ainsi une variabilité spatiale des concentrations.

Les surfaces et les volumes de sols contaminés sont donc connus. La deuxième étape consiste en l'échantillonnage des valeurs de concentration en certains points des modèles numériques, puis à appliquer l'IAM. Différents schémas de répartition spatiale des points d'échantillonnage sont utilisés. Enfin, les volumes de pollution estimés sont comparés aux « volumes réels » des cas synthétiques ainsi qu'aux méthodes d'interpolation spatiale.

3.1. CONSTRUCTION DES CAS SYNTHETIQUES

Les cas d'études synthétiques ont été conçus à l'aide du logiciel Groundwater Vistas (Rumbaugh et Rumbaugh, 2024). Les simulations ont été réalisées avec le code de calcul MODFLOW USG (Panday, 2024) qui permet de résoudre les équations d'écoulement et de transport en zones saturée et non saturée par une méthode numérique de type volumes finis.

Pour cette modélisation spécifique, les hypothèses simplificatrices suivantes ont été adoptées :

i) Le polluant est considéré comme non réactif (i.e. aucune sorption et/ou dégradation),

ii) Le transport par diffusion moléculaire est considéré comme négligeable par rapport aux transports advectif et dispersif,

iii) La teneur en eau volumique (θ_w) est supposée uniforme verticalement et constante dans le temps, i.e., en régime permanent,

iv) l'infiltration dans le sol est également en régime permanent (I_{eff} constant).

Le domaine de modélisation pour chaque cas d'étude synthétique s'étend sur des dimensions de 56 m de longueur × 56 m de largeur × 6 m de profondeur. La discrétisation

spatiale repose sur un maillage horizontal régulier de 0,7 × 0,7 m et des couches verticales de 0,1 m d'épaisseur.

Afin d'établir un régime d'écoulement permanent à travers la zone non saturée, une recharge constante de 160 mm.an^{-1}, correspondant au paramètre I_{eff} de la solution analytique, est appliquée sur la première couche du modèle tandis qu'une charge hydraulique de -50 m est imposée sur la couche inférieure du modèle. Cette condition à la limite inférieure simule l'existence d'une zone saturée profonde et garantit un écoulement non contraint (drainage libre) à travers le domaine modélisé. Elle permet ainsi le maintien d'un profil de teneur en eau volumique (θ_w) quasi uniforme sur la verticale du modèle (Šimůnek et al., 1999b).

Les modèles de capillarité de Brook et Corey ainsi que de van Genuchten peuvent être utilisées dans le logiciel MODFLOW USG. Pour cette étude, le modèle de van Genuchten a été retenu. Ce modèle est l'un des plus couramment implémentés dans les outils de modélisation de la zone non saturée, tels que le logiciel HYDRUS (Šimůnek, et al., 2005).

Pour le cas synthétique n°1, représentant un milieu homogène, les paramètres suivants ont été appliqués à l'ensemble du maillage : porosité = 0,25, α = 0,2 m^{-1}, qui est une valeur typique pour les sols sableux-limoneux (Nemes et al., 2001), et β = 5, qui correspond à une valeur élevée (Nemes et al., 2001). Le choix d'une valeur élevée pour le paramètre β du modèle de van Genuchten permet de limiter l'influence de la zone saturée sur la dynamique de l'écoulement dans la partie non saturée du modèle, en induisant une transition plus marquée au niveau de la frange capillaire. Les paramètres utilisés pour les autres cas synthétiques (milieux hétérogènes) seront discutés dans les sections correspondantes.

3.2. APPLICATION A UN CAS SYNTHETIQUE HOMOGENE

3.2.1. Caractérisation de la pollution du cas synthétique

Une concentration constante C_{w0} = 100 000 µg.l^{-1} a été imposée dans les mailles situées au centre de la couche de surface du modèle, délimitant une zone source de 14 m × 14 m. Les dispersivités longitudinale et transversale ont été fixées respectivement à 25 m et 5 m. La Figure 4 présente les distributions spatiales des concentrations simulées après deux années de transport ; c'est à ce temps que les comparaisons sont effectuées.

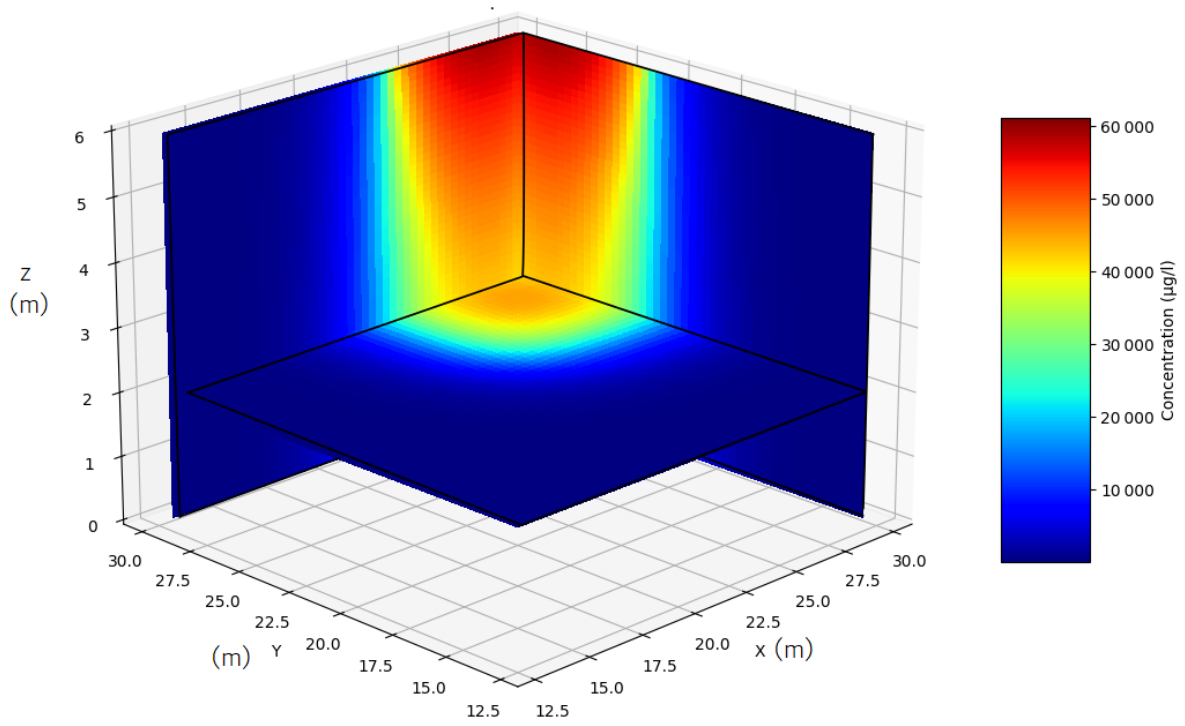


Figure 4 : Panache de pollution après deux ans pour le cas homogène

Quatre seuils de dépollution potentiels (rapports C_w/C_{w0}) ont été définis à 1 %, 5 %, 10 % et 20 % (correspondant respectivement à des concentrations de 1 000, 5 000, 10 000 et 20 000 µg.l^{-1}). Les comparaisons entre volumes pollués réels et estimés ont été conduites à trois profondeurs distinctes : 1 m, 3 m et 5 m.

Les surfaces polluées aux 3 profondeurs considérées et pour les différents seuils sont synthétisées dans le Tableau 2. Les surfaces polluées sont comprises entre 1028,5 m^2 pour un seuil à 1 000 µg.l^{-1} et 310,2 m^2 pour un seuil à 20 000 µg.l^{-1}.

Tableau 2 : Surfaces pour lesquelles les concentrations sont supérieures aux différents seuils de dépollution considérés pour le cas synthétique homogène

Profondeur	1 m	3 m	5 m	Total
Seuil de dépollution (µg.l^{-1})	Surface supérieure au seuil (m²)			
1 000	331,7	344,1	352,7	1028,5
5 000	195,5	205,8	206,6	607,9

Profondeur	1 m	3 m	5 m	Total
Seuil de dépollution ($\mu g.l^{-1}$)	Surface supérieure au seuil (m^2)			
10 000	145,5	151,6	154,3	451,4
20 000	111,4	105,1	93,7	310,2

3.2.2. Plan d'échantillonnage

Les points d'observation ont été répartis de manière identique sur les 3 couches situées à 1 m, 3 m et 5 m de profondeur.

Deux scénarios différents d'échantillonnage ont été construits : i) une distribution régulière des points d'observation avec 25 points d'observation par profondeur et 75 points d'observation au total et ii) une distribution aléatoire des points d'observation avec 7 points d'observation par profondeur et 21 points d'observation au total. La distribution des points est présentée sur la Figure 5. L'application des triangles de Delaunay conduit à 32 triangles pour la distribution régulière et 8 triangles pour la distribution aléatoire.

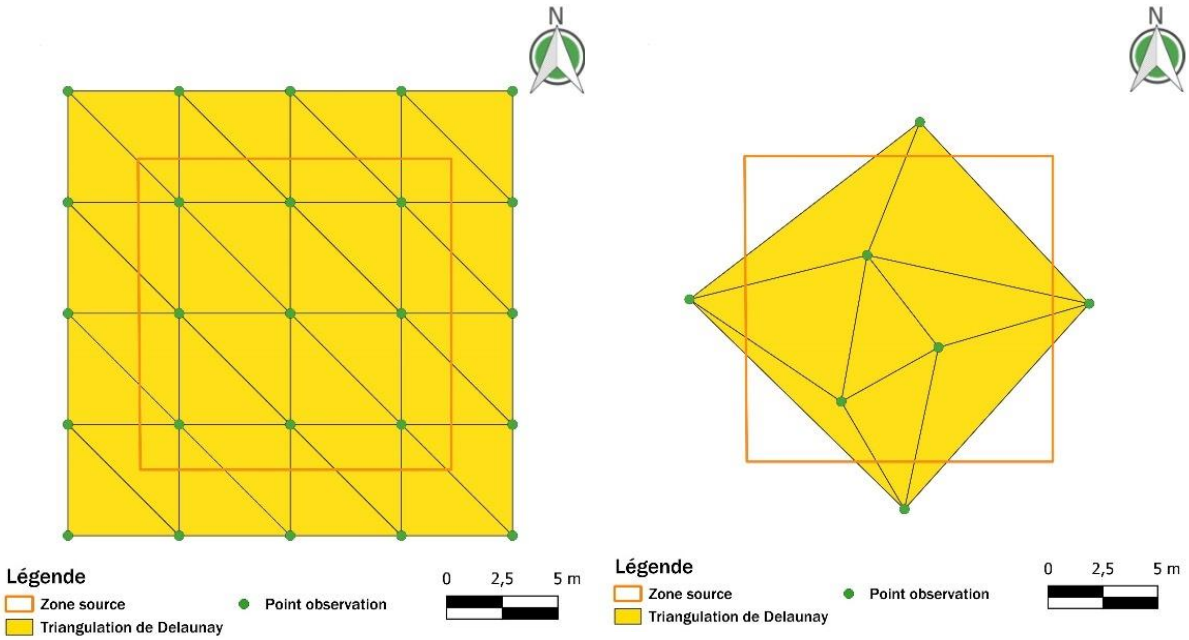


Figure 5 : Distributions spatiales des points d'observation pour le cas synthétique homogène

3.2.3. Méthode de comparaison des surfaces polluées

Pour ce premier cas, les méthodes d'interpolation choisies sont uniquement en 2D. Trois méthodes d'interpolation spatiale ont été sélectionnées et appliquées aux données

d'observation : la méthode RBF, la méthode MQS et la méthode KRG. Un critère essentiel pour la sélection est la capacité des méthodes à réaliser des extrapolations, c'est-à-dire à estimer des valeurs au-delà de la portée spatiale des observations disponibles. Par conséquent, les méthodes strictement limitées à l'interpolation comme la triangulation ou les techniques basées sur le plus proche voisin ont été exclues de cette étude comparative. Étant donné qu'un nombre limité de points d'observation ne permet pas de réaliser une analyse variographique robuste (Demougeot-Renard et al., 2012), un modèle de variogramme exponentiel, avec une portée fixée empiriquement à 20 % de l'étendue du domaine modélisé, a été sélectionné pour les interpolations par krigeage après évaluation comparative de différentes valeurs par rapport au cas synthétique. Pour la méthode RBF, une fonction de type multiquadratique a été employée, avec le paramètre R2 fixé à sa valeur par défaut de 2,2 dans le logiciel Surfer (Golden Software, 2024) . Les paramètres par défaut pour la méthode MQS utilisés pour cette méthode sont 13 voisins pour l'ajustement quadratique local et 15 voisins pour la pondération finale. Les interpolations sont donc utilisées de façon standard, les paramètres de chacune des méthodes pourraient être modifiés. Afin de comparer la solution analytique avec ces trois techniques d'interpolation, l'erreur moyenne sur le volume de sol pollué estimé a été calculée pour chaque configuration spatiale des points d'observation considérés.

Bien que l'approche de modélisation inverse analytique permette d'estimer les concentrations à n'importe quelle profondeur, l'évaluation comparative avec les méthodes d'interpolation sélectionnées a été restreinte aux profondeurs correspondant aux points d'observation (1 m, 3 m et 5 m) afin d'assurer la cohérence de l'analyse.

Un maillage régulier avec une résolution de 0,7 m × 0,7 m a été superposé au domaine d'étude pour représenter les résultats de la modélisation analytique inverse. Les concentrations dans chaque maille ont été calculées à partir de l'IAM ou des interpolations. Pour chaque méthode, si une valeur de concentration calculée dépassait le seuil fixé, la maille était considérée comme polluée. Concernant la méthode analytique inverse, et comme discuté dans le chapitre 2.1, une pondération (σ) a été appliquée afin d'attribuer un poids plus important aux mailles situées dans le triangle de Delaunay où le modèle a été inversé (équation (2.2)). La valeur de pondération retenue distribue l'influence comme suit : σ=1/3 pour les valeurs issues des triangles, et σ=2/3 pour l'ensemble des valeurs hors triangles. Suite au calcul de la probabilité de dépassement $C>C_{crit}$ pour chaque maille du modèle, cette valeur de probabilité doit être comparée à un seuil. Ce seuil de probabilité permet de décider si la maille est catégorisée comme « polluée » ou « non polluée ». En toute logique, la valeur optimale du seuil

de probabilité est celle qui permettrait d'inclure un maximum des points d'observation (idéalement l'ensemble des points) dépassant la concentration critique. Pour satisfaire au principe de précaution, cette valeur a été majorée afin d'étendre les limites de la zone identifiée comme polluée. Un exemple est fourni en Figure 6. Pour cet exemple, le seuil critique est imposé à 5000 µg.l^{-1}. La Figure 6a montre les résultats correspondant à un seuil de probabilité à 50% alors que la Figure 6b correspond à un seuil à 70%. La surface estimée dépassant le seuil critique pour la Figure 6a est supérieure à celle de la Figure 6b. Toutefois, les deux estimations respectent l'ensemble des concentrations des points d'observations. Dans les prochaines sections, le scénario correspondant à la Figure 6a a été sélectionné pour la suite de l'étude. À noter que dans la section 3.4, les deux seuils de probabilité (appelé bornes minimum et maximum) seront considérés, permettant d'obtenir une incertitude sur la surface polluée.

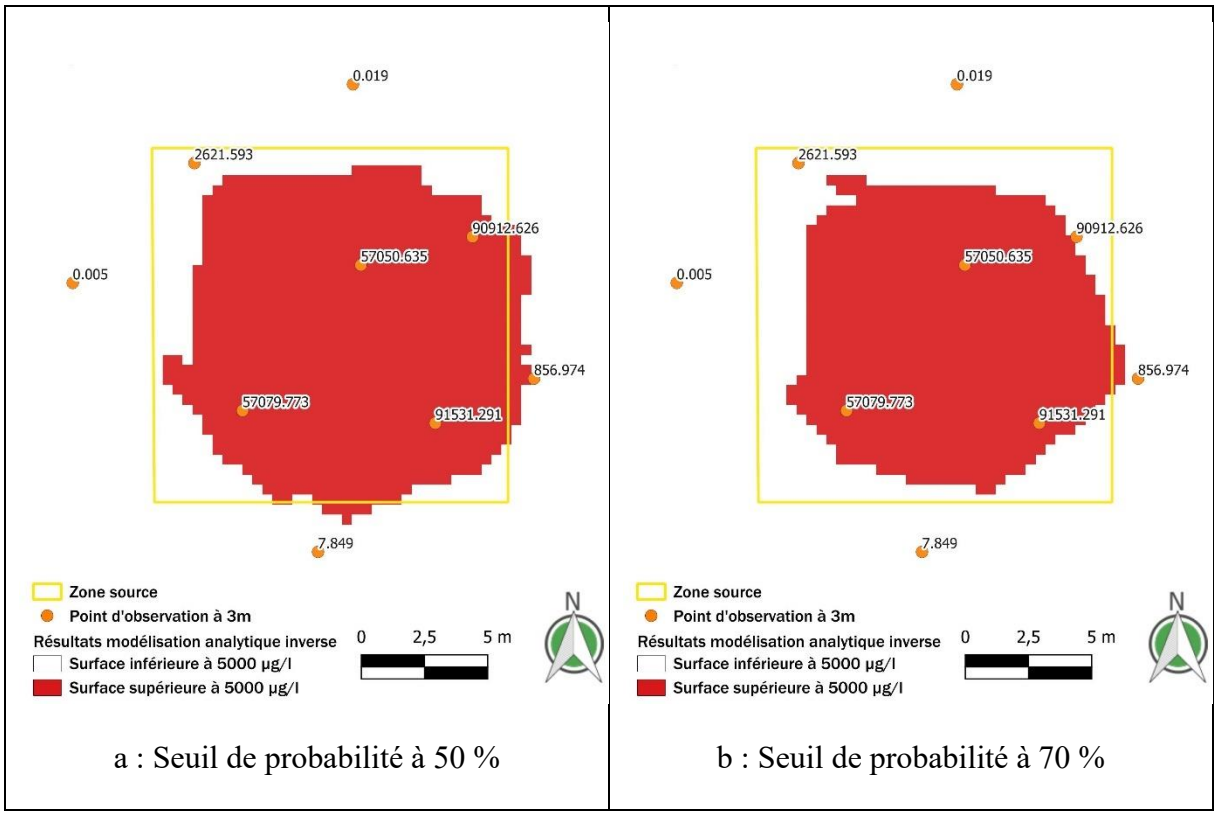


Figure 6 : Exemple de seuil de probabilité différent

La première étape pour réaliser les comparaisons entre les différentes méthodes a consisté en la comptabilisation maille par maille des erreurs commises aux trois profondeurs considérées, en distinguant :

- les erreurs de type « faux négatif », i.e. les mailles avec une concentration supérieure au seuil critique ($C_w >= C_{crit}$) dans le modèle synthétique mais identifiées comme non

polluées ($C_w < C_{crit}$) avec les méthodes d'interpolation ou de modélisation analytique inverse ;

- les erreurs de type « faux positif » correspondant au cas opposé, i.e. des mailles non polluées identifiées comme polluées.

Un pourcentage d'erreur moyenne pour chaque profondeur a ensuite été calculé en divisant la somme des surfaces faux positifs et faux négatifs par la surface « réelle » polluée du cas synthétique. Cette procédure a été répétée pour les 4 seuils définis. Finalement, la moyenne des erreurs associées aux 4 seuils a été retenue comme mesure de performance de chaque méthode. Ce terme d'erreur moyenne utilisé ci-après correspond au pourcentage d'erreur moyenne associant toutes les profondeurs et tous les seuils.

3.2.4. Comparaison des surfaces estimées par les différentes méthodes

Le modèle étant considéré comme non réactif, les paramètres suivants ont été inversés : la concentration C_{w0} de la source de contamination, ses dimensions X et Y ainsi que la localisation X_s et Y_s du centre, l'infiltration effective I_{eff}, la teneur en eau volumique θ_w, la dispersivité longitudinale α_L, la dispersivité transversale α_T et le temps t (temps écoulé depuis l'apparition de la source de polluant). Les gammes de paramètres pour l'inversion sont précisées dans le tableau 5 ci-après avec l'analyse de sensibilité des paramètres.

Les résultats obtenus par l'IAM correspondent à la probabilité de dépasser le seuil de concentration (C_{crit}) fixé. La délimitation du panache de pollution dépend donc, pour cette méthode, 1) du seuil de concentration considéré et 2) du seuil de probabilité de dépassement retenu. Le Tableau 3 présente les résultats obtenus par l'IAM ainsi que les comparaisons avec les trois autres méthodes d'interpolation considérées dans cette étude.

Tableau 3 : Pourcentage d'erreur moyenne associée à chaque distribution spatiale pour la méthode IAM et pour les 3 méthodes d'interpolation

Distribution des points d'observation	Erreur moyenne IAM	Erreur moyenne RBF	Erreur moyenne MQS	Erreur moyenne KRG
Régulière	14,2%	18,3%	23,5%	22,2%
Aléatoire	25,2%	30,7%	50,7%	85,6%

L'analyse détaillée des performances (seuil par seuil et profondeur par profondeur, non présentée dans le Tableau 3), révèle que l'IAM ne surpasse pas systématiquement les méthodes d'interpolation de référence pour chaque seuil de concentration et pour chaque profondeur analysée. De manière ponctuelle, les méthodes RBF ou MQS peuvent présenter des performances légèrement supérieures. Toutefois, malgré ces nuances observées dans les résultats détaillés, l'IAM développée s'avère globalement plus performante que les méthodes d'interpolation classiques et permet de sensiblement minimiser l'erreur d'estimation (section 3.3.1) des surfaces polluées. Il est particulièrement notable que cette erreur moyenne se révèle faible (14,2 %) pour la configuration d'échantillonnage régulier. L'inter-comparaison des différentes techniques d'interpolation entre elles indique que c'est la méthode RBF qui conduit aux meilleures estimations, pour les deux configurations d'échantillonnage testées.

3.3. CAS SYNTHETIQUE HETEROGENE AVEC DIVERSES REPARTITIONS SPATIALES DES POINTS D'OBSERVATION

3.3.1. Implémentation de l'hétérogénéité et caractérisation de la pollution simulée

Dans ce cas d'étude synthétique, l'hétérogénéité du milieu souterrain a été introduite en définissant des variations spatiales de la teneur en eau volumique résiduelle comprises entre 0,075 et 0,27, et de la porosité variant de 0,25 à 0,45. Deux motifs de zonation latérale (Fig. 7) ont été utilisés sur les différentes couches du modèle afin de définir un profil vertical stratifié (Fig. 8) :

- de 0 à 0,5 m de profondeur : couches homogènes ;
- de 0,5 à 2,5 m : zonation de type 1 ;
- de 2,5 à 3,5 m : couches homogènes ;
- de 3,5 à 6 m : zonation de type 2.

Les caractéristiques de la zone source sont similaires au cas synthétique homogène. Les dispersivités longitudinale et transversale sont fixées respectivement à 0,25 et 0,05 m. Ces valeurs ont été abaissées par rapport au cas homogène afin de ne pas homogénéiser le panache.

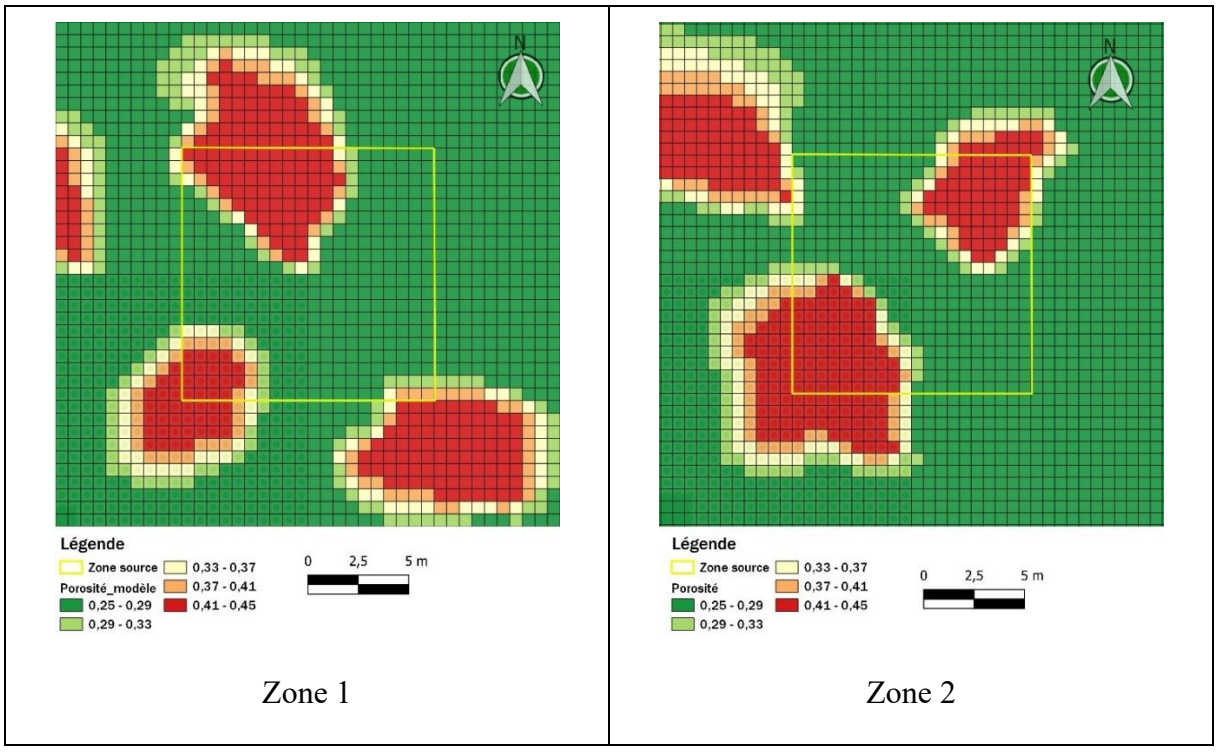


Figure 7 : Distribution spatiale de la porosité entre 0,5 m et 2,5 (zone 1) et entre 3,5 et 6 m (zone 2)

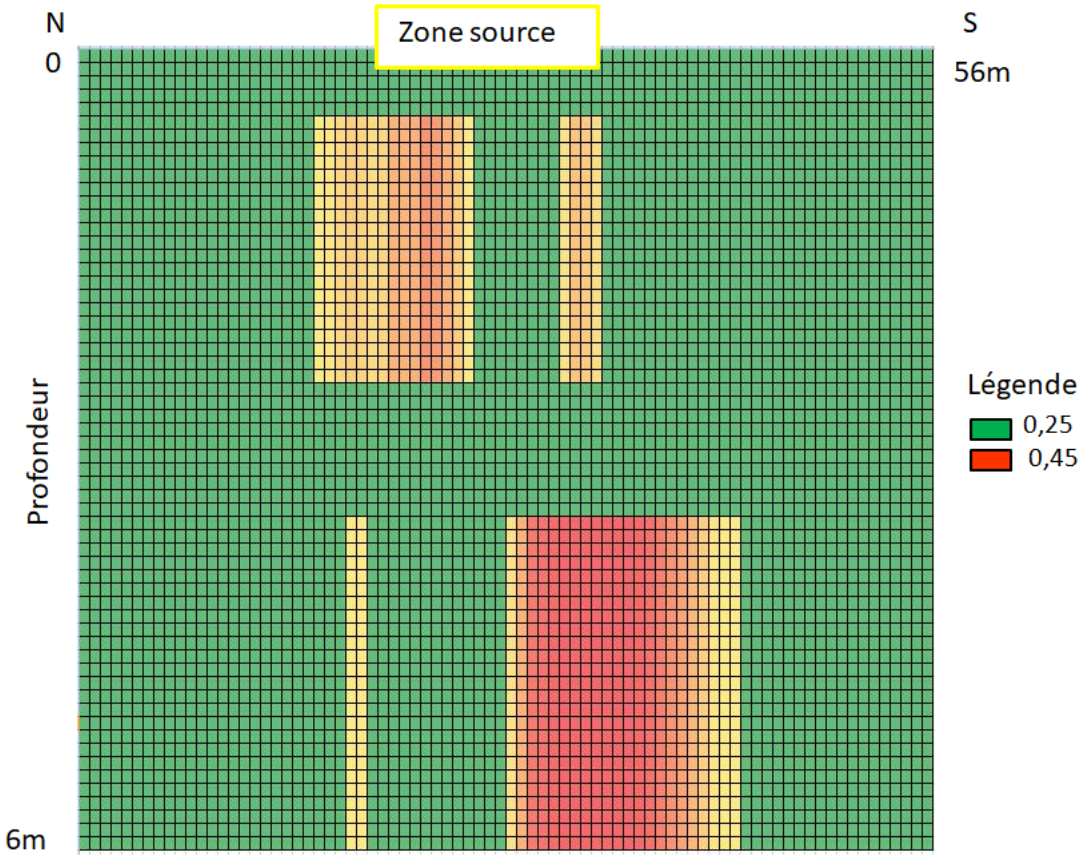


Figure 8 : Profil Nord-Sud de la porosité

La Figure 9 présente les distributions spatiales des concentrations simulées après deux années de transport.

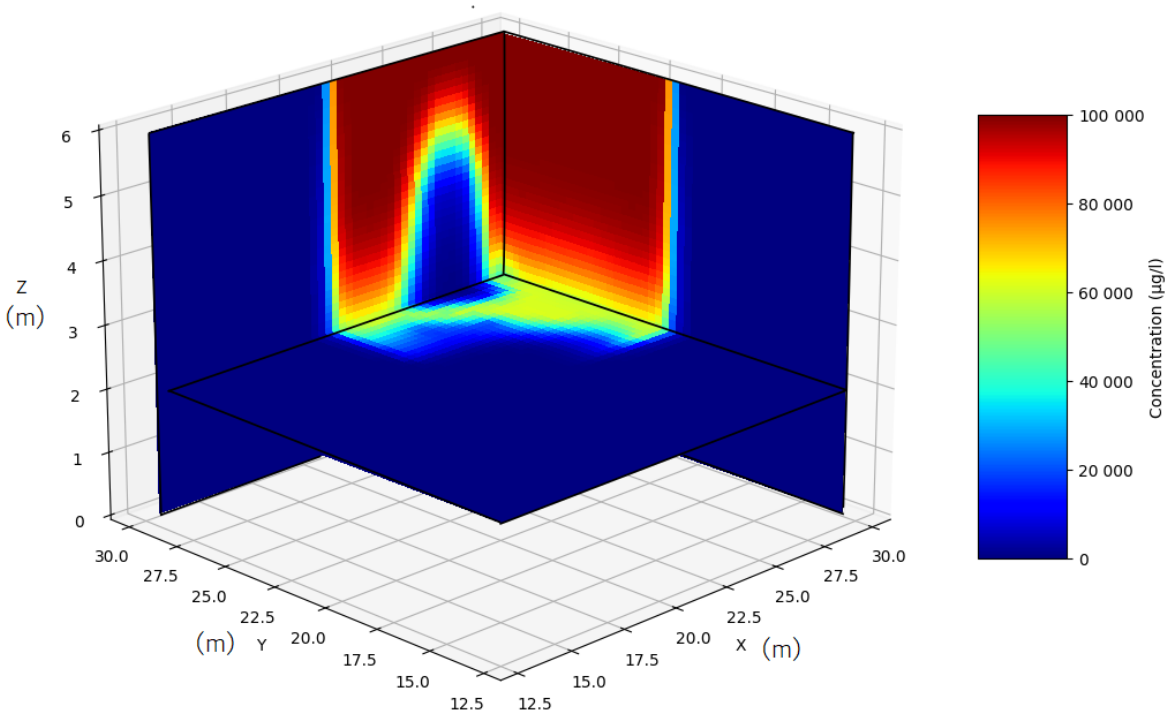


Figure 9 : Vue en plans du panache de pollution après deux ans pour le cas hétérogène

Les surfaces de dépassement des différents seuils de concentrations aux trois profondeurs ciblées sont synthétisées dans le Tableau 4. Les surfaces sont comprises entre 561,6 m^2 (seuil à 1000 $\mu g.l^{-1}$) et 327,9 m^2 (seuil à 50 000 $\mu g.l^{-1}$).

Tableau 4 : Surfaces de dépassement des différents seuils de pollution pour le cas synthétique hétérogène

Profondeur	1 m	3 m	5 m	1, 3 et 5 m
Seuil de dépollution ($\mu g.l^{-1}$)	Surface pour laquelle la concentration est supérieure au seuil (m^2)			
1000	236,3	222,8	102,6	561,6
5000	196,5	183,5	73,1	453,2
10000	196,5	152,4	61,6	410,5
20000	196,5	144,2	25,0	365,7
50000	196,5	131,4	0	327,9

3.3.2. Plan d'échantillonnage

Dix configurations spatiales des points d'observation ont été considérées, dont cinq impliquant une distribution régulière de 9 à 81 points par couche (soit 27 à 243 points d'observation au total, Fig. 10), et cinq avec une distribution irrégulière de 7 à 25 points par couche (soit 21 à 75 points d'observation au total, Fig. 11).

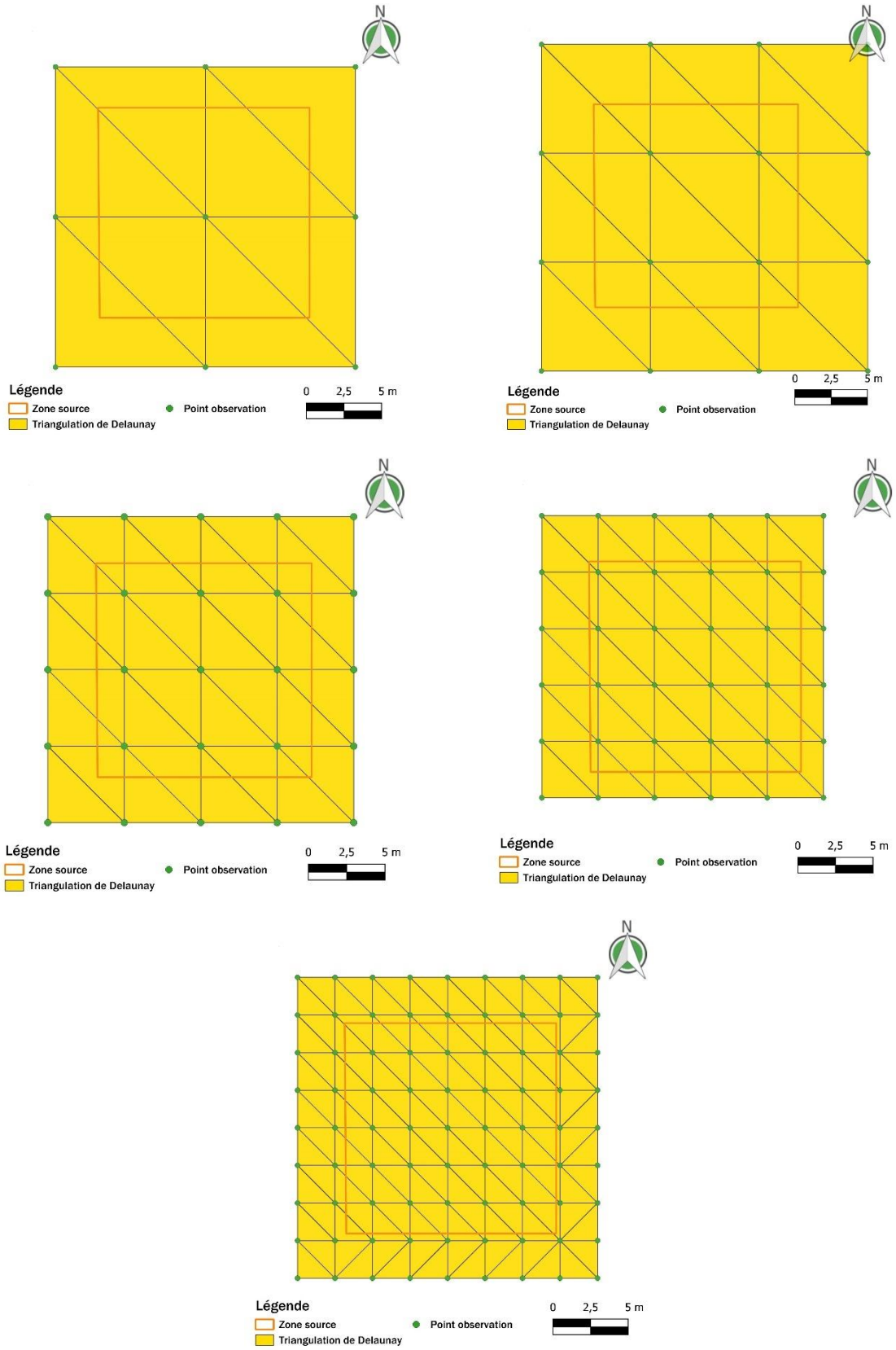


Figure 10 : Distributions spatiales régulières des points d'observation pour le cas synthétique hétérogène

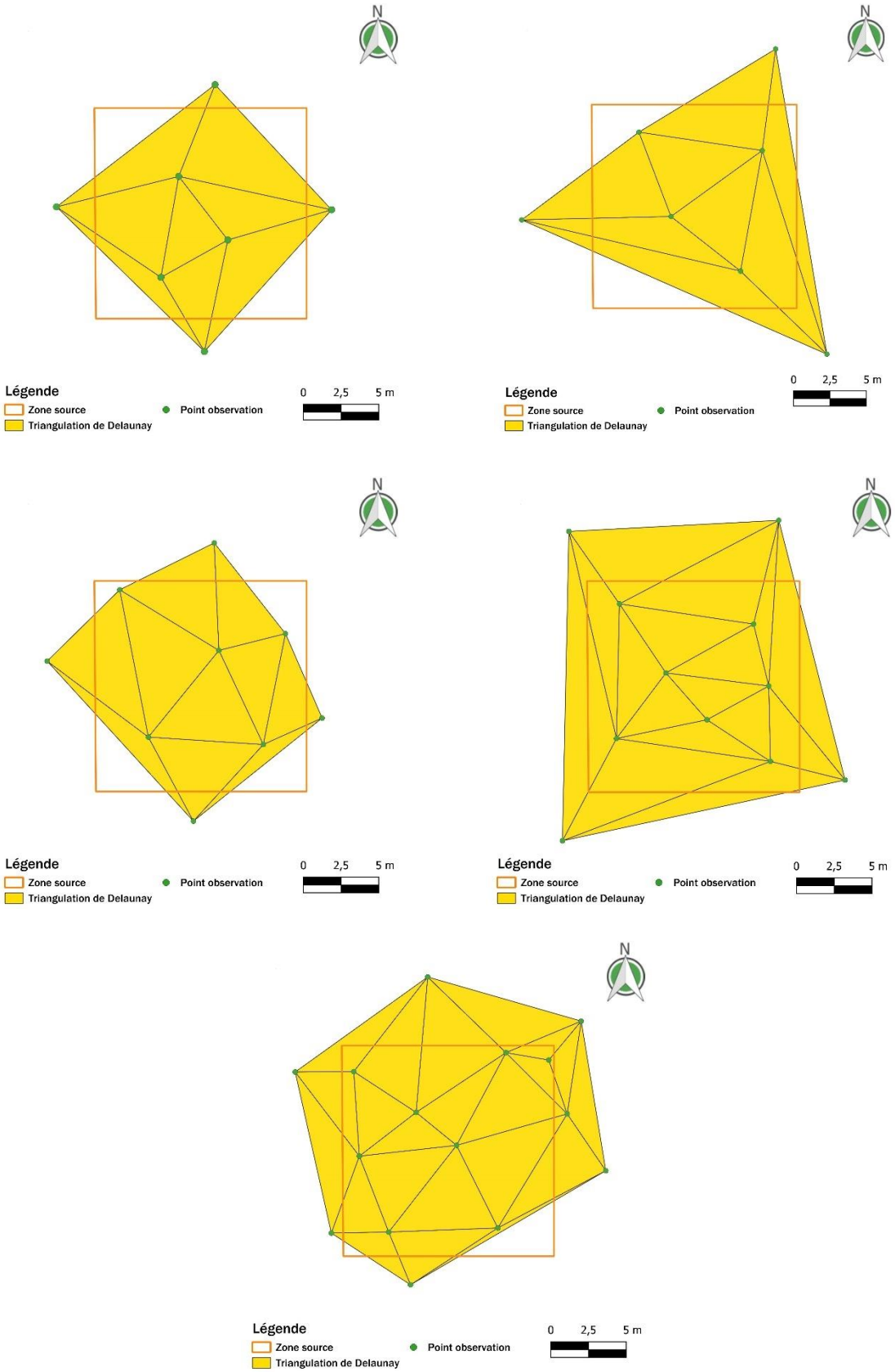


Figure 11 : Distributions spatiales aléatoires des points d'observation

3.3.3. Méthodologie de comparaison des surfaces estimés par les différentes méthodes

La procédure de comparaison est identique à celle suivie pour le cas synthétique homogène (section 3.2.3) à l'exception de l'ajout d'un 5ème seuil de concentration critique à 50 000 µg.l^{-1} (ratio C_w/C_{w0} = 50%). Ce seuil n'était pas pertinent pour le cas homogène car les concentrations étaient plus faibles compte tenu des valeurs de dispersivité plus élevées.

3.3.4. Analyse de sensibilité du modèle analytique

Une étude de sensibilité des paramètres du modèle analytique de Troldborg et al. (2009) a été réalisée selon la méthode de Sobol (1993). Cette analyse est abordée spécifiquement ici car, pour être pertinente, elle doit être menée en utilisant des gammes de paramètres représentatives des conditions réelles et des incertitudes associées aux propriétés du site, plutôt que des valeurs universelles.

Cette approche, adaptée aux modèles non linéaires (Homma et Saltelli, 1996), permet de décomposer la variance des résultats de simulation induite par les plages de variation des paramètres d'entrée. Les indices de Sobol quantifient la contribution de chaque paramètre, ainsi que ses interactions avec les autres paramètres, à cette variance totale.

Différents types d'indices peuvent être calculés. Dans le cas présent, deux indices ont été retenus :

- L'indice de premier ordre qui mesure la contribution directe de la variance d'un paramètre d'entrée à la variance totale de la sortie du modèle, sans inclure les effets d'interaction ;
- L'indice d'ordre total (Saltelli, 2002) qui évalue l'influence globale d'un paramètre d'entrée sur la sortie en englobant à la fois son effet direct et tous les effets liés à ses interactions avec les autres paramètres.

Les paramètres varient selon les gammes de valeurs indiquées dans le Tableau 5. La taille du nombre de tirages aléatoires effectués dans les gammes de valeurs du tableau a été fixée à 1 million. Les résultats de l'étude de sensibilité sont présentés dans la Figure 12.

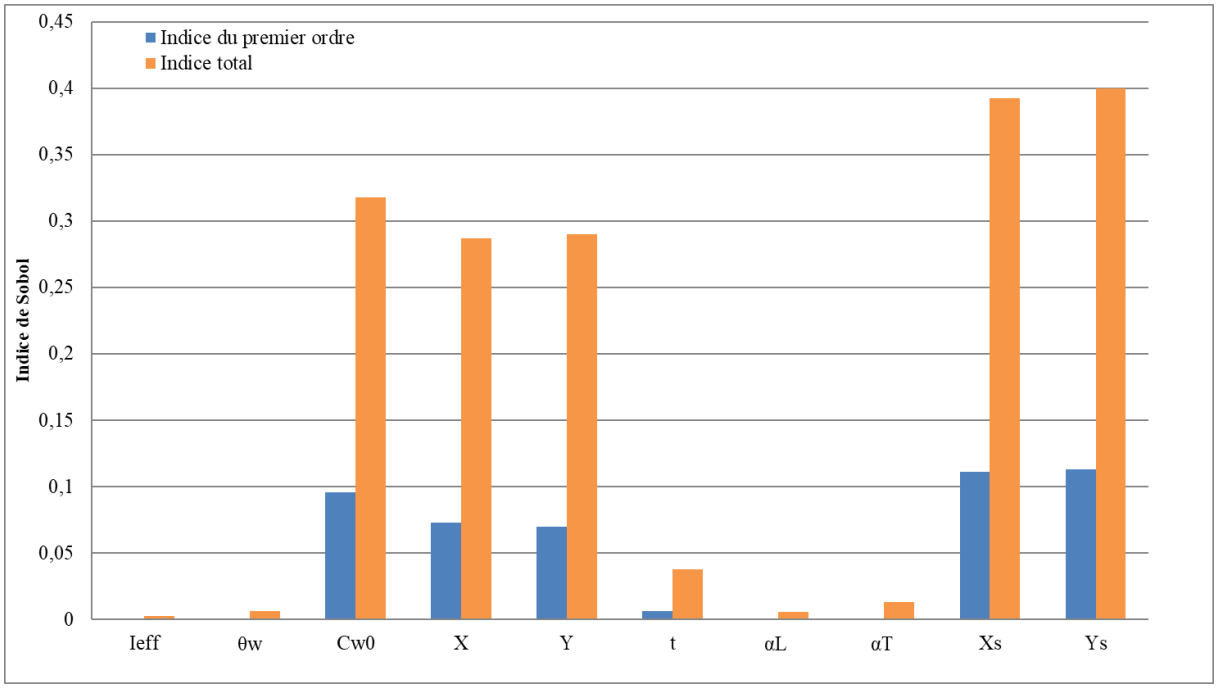


Figure 12 : Analyse de sensibilité - indice de Sobol

L'analyse révèle, conformément aux attentes, que les paramètres exerçant l'influence la plus significative sur les prévisions du modèle sont ceux définissant la source de contamination : la localisation du centre (X_s et Y_s), ses dimensions (X et Y) et la concentration imposée (C_{w0}). Ces paramètres présentent les indices de sensibilité de premier ordre et d'ordre total les plus élevés.

A l'inverse, les paramètres de moindre sensibilité sont :

- L'infiltration effective I_{eff} ;
- La teneur en eau volumique θ_w ;
- La dispersivité longitudinale α_L et la dispersivité transversale α_T ;
- Le temps t.

La faible sensibilité apparente de ces paramètres pourrait être en partie liée à l'étendue limitée des plages de variation considérées dans cette étude.

3.3.5. Calibration du modèle analytique

Les valeurs initiales des paramètres utilisées pour la calibration du modèle analytique ont été générées par tirage aléatoire au sein des plages de valeurs spécifiées dans le Tableau 5. Les plages de variation de ces paramètres ont été définies sur la base d'une revue exhaustive de la littérature scientifique et sont établies de manière à correspondre à des grandeurs physiques soit

mesurables (*Ieff* et θ_w), soit fixées empiriquement (C_{w0}, α_l et α_t). A noter que, pour un cas réel, les concentrations aux points d'observation pour chaque profondeur peuvent fournir une première indication de la taille maximale de la zone source (aucune zone source au-delà des points où la concentration est proche de zéro).

Tableau 5 : Plages de paramètres possibles pour la calibration

Paramètre	Unité	Valeur minimale	Valeur maximale
I_{eff}	$m.an^{-1}$	0,09	0,25
θ_w	-	0,05	0,5
C_{w0}	$\mu g.l^{-1}$	1 000	9 000 000
X	m	1	18
Y	m	1	18
t	an	0,1	50
α_l	m	0,01	10
α_t	m	0,001	1
X_s	m	0,1	20
Y_s	m	0,1	20

Le graphique de calibration du modèle d'estimation, ajusté sur l'une des distributions spatiales aléatoires et comprenant 33 points d'observation répartis sur les 3 profondeurs, est présenté sur la Figure 13. La distribution spatiale comprend 16 triangles de Delaunay, chacun inversé 10 fois pour chacune des 3 profondeurs. Au total, 8 triangles de Delaunay sur 480 présentaient des erreurs de calibration importantes (supérieures à 1 % de valeur maximale de concentration sur l'ensemble des points d'observation), ce qui correspond à 10 points d'observation sur 1 380 (certains points correspondant au même triangle de Delaunay).

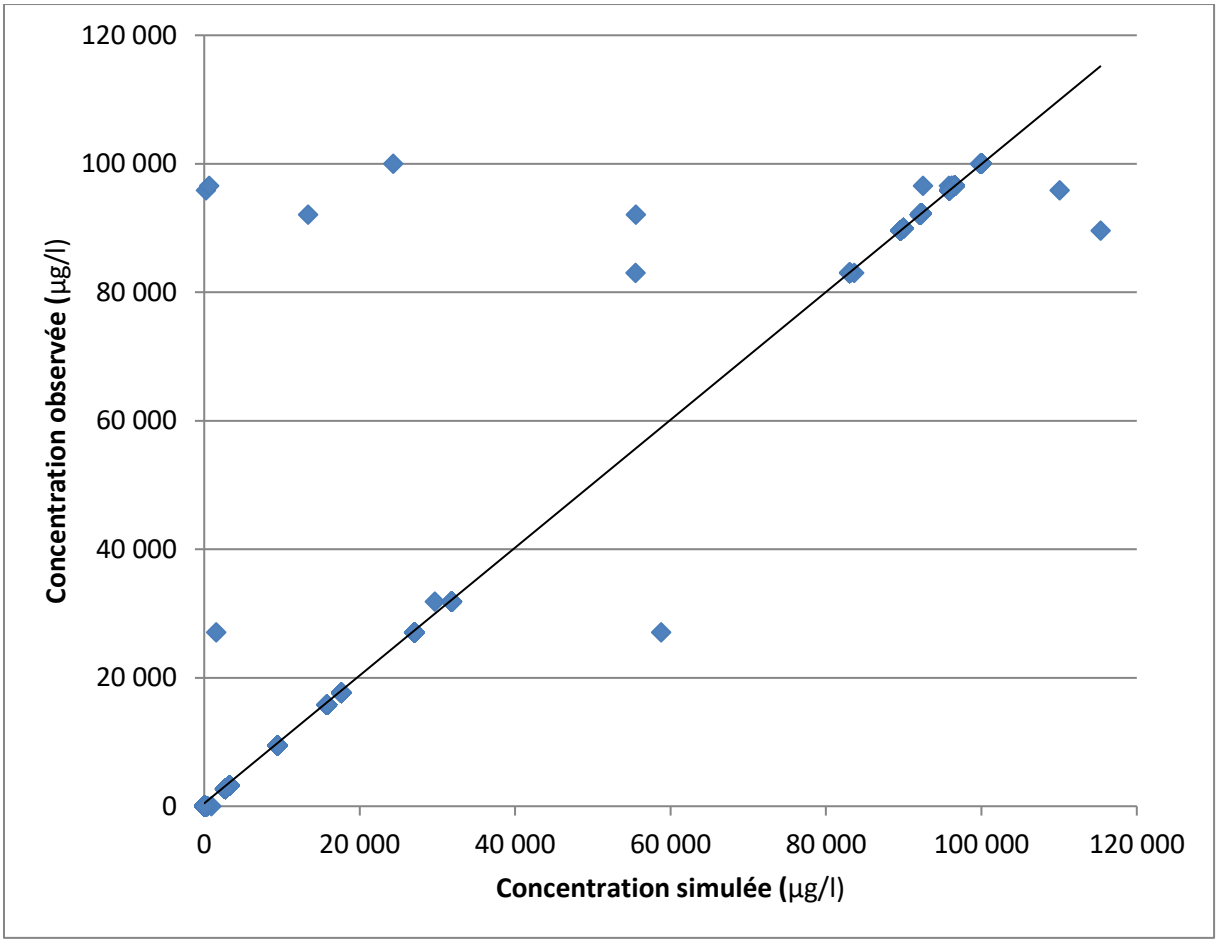


Figure 13 : Exemple de calibration

Les valeurs des paramètres inversés, synthétisées dans le Tableau 6, peuvent être comparées aux valeurs du cas synthétique (pseudo-réalité).

Tableau 6 : Statistique concernant les paramètres inversés

	I_{eff}	θ_w	C_{w0}	X	Y	t	α_L	αT	Xs	Ys
Moyenne	0,17	0,26	727352	11,4	11,1	25,5	4,4	0,48	1,5	-2,5
Percentile 25	0,13	0,14	100284	7,1	6,9	13,4	2,0	0,23	-1,1	-5,5
Médiane	0,17	0,24	135081	12,1	11,9	24,0	3,9	0,42	1,9	-3,1
Percentile 75	0,22	0,38	563694	16,3	16,1	38,1	6,7	0,71	4,3	-1,5

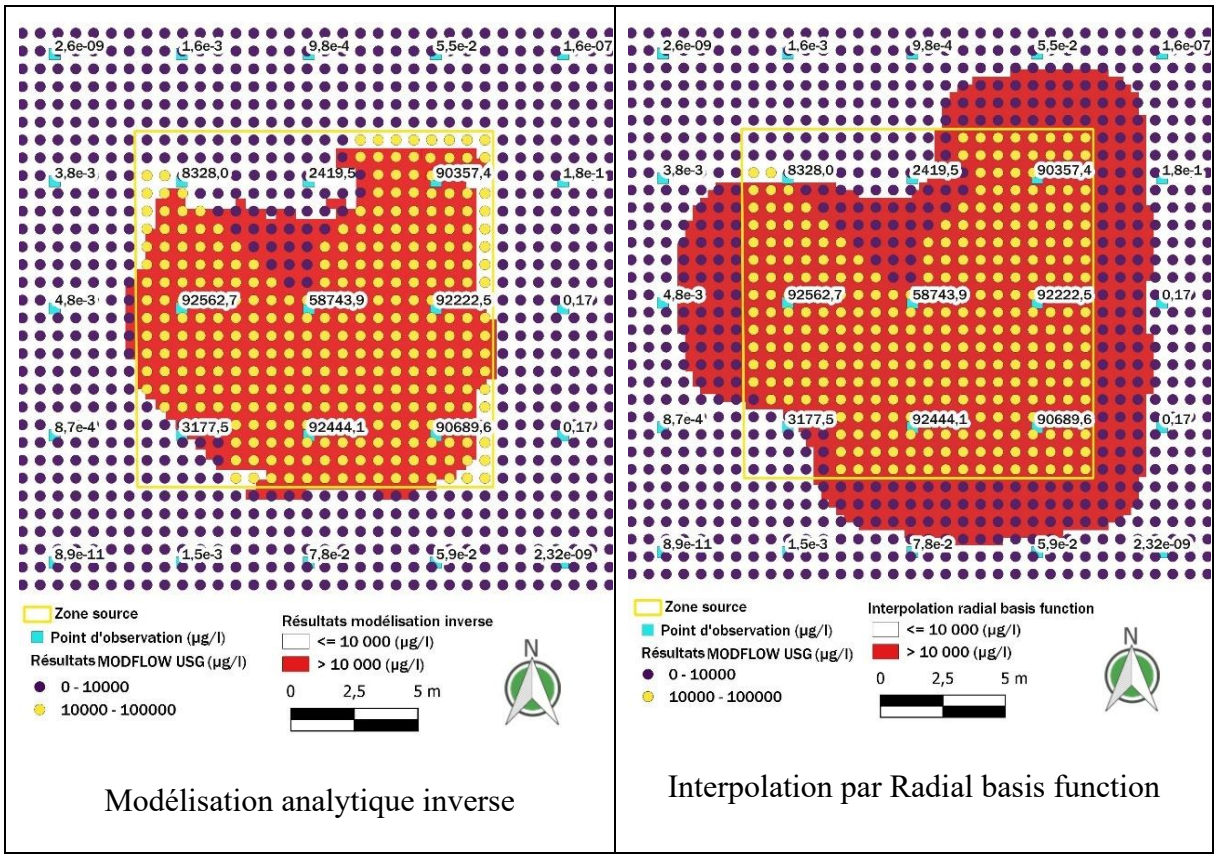
Les comparaisons sont effectuées avec la médiane, qui n'est pas influencée par les valeurs extrêmes contrairement à la moyenne. Les comparaisons révèlent que la valeur I_{eff} est proche de la réalité (0,17 m/an contre 0,16 m/an en réalité), tandis que la taille de la source est légèrement plus petite (12 x 12 m contre 14 x 14 m). La valeur C_{w0} est légèrement plus élevée (35 % de plus que la valeur réelle). Les valeurs de dispersivité sont beaucoup plus élevées dans les inversions que dans la réalité (3,85 et 0,42 contre 0,25 et 0,05). Le temps était également

nettement plus long (24ans contre 2 ans dans la pseudo-réalité). Comme les points d'observation sont inversés par plan dans ce cas, les données fournies peuvent ne pas être suffisantes pour déterminer ces valeurs. Les paramètres estimés peuvent donc être approximatifs. Les valeurs Xs et Ys concernant le centre de la source sont décalés de 1,9 m en X soit 15% et -3,1 m en Y soit 22%.

3.3.6. Comparaison des surfaces estimées par les différentes méthodes

La présentation des résultats de cette section est structurée en deux parties distinctes : (i) les distributions spatiales régulières des points d'observation et (ii) les distributions aléatoires des points d'observation.

La Figure 14 présente un exemple des résultats obtenus avec la méthode IAM et par interpolation à une profondeur de 3 m, en considérant un seuil critique de 10 000 µg.l^{-1}. Pour ce cas précis, c'est l'interpolation RBF qui conduit au pourcentage d'erreur le plus faible parmi les méthodes d'interpolation (78,2 %). La méthode MQS donne un pourcentage d'erreur de 81,0 % et KRG 81,4 %. La performance de la modélisation analytique inverse a été caractérisée par un indice d'erreur de 27,9%.



Modélisation analytique inverse

Interpolation par Radial basis function

Figure 14 : Cartes présentant la modélisation analytique inverse et une interpolation RBF à une profondeur de 3 m et un seuil de 10 000 µg.l^{-1}

3.3.6.1. Distributions régulières

Les interpolations spatiales, de même que la méthode IAM, se sont montrées particulièrement sensibles au nombre de points d'observation associés à de fortes valeurs de concentration relative. Cet indicateur a donc été retenu (nombre de points d'observation avec une valeur supérieure à 1% de la valeur maximale) comme élément de comparaison entre les différentes méthodes (Fig 15 et 17).

L'erreur moyenne (section 3.1.1) a été comparée pour les distributions spatiales régulières des points d'observation entre la méthode analytique inverse et les trois méthodes d'interpolation spatiales (Fig. 15).

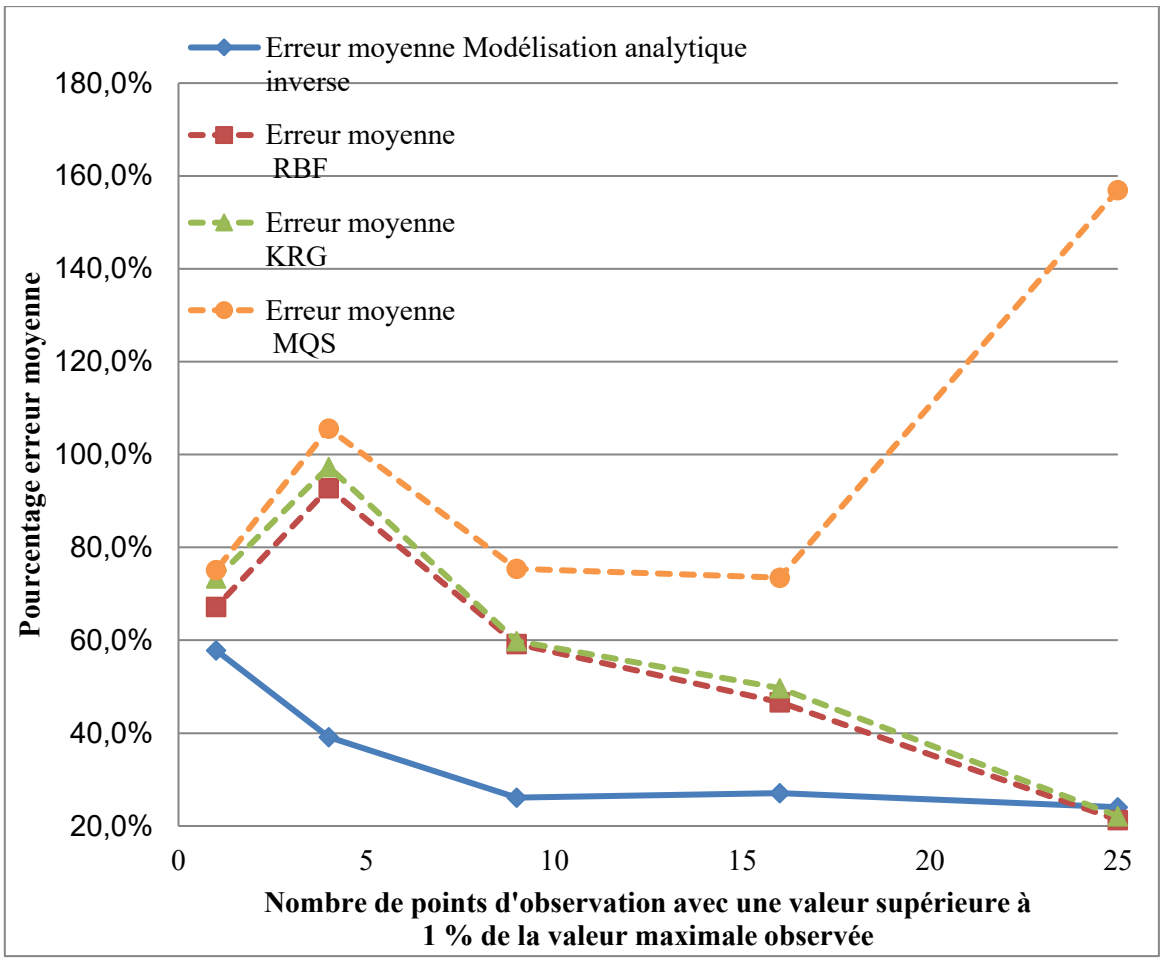


Figure 15 : Erreurs moyennes de surfaces estimées par la méthode IAM et par les méthodes d'interpolation à partir de distributions régulières des points d'observations

Comme le montre la Figure 15, la performance du modèle analytique inverse est généralement supérieure à celle des méthodes d'interpolation spatiale évaluées (RBF, MQS, KRG), sauf pour le scénario de très haute densité d'échantillonnage (25 points d'observation supérieure à 1% de la valeur maximale par couche, 81 points par couche et 243 points d'observation au total). Pour ce dernier cas, le nombre de points semble suffisant pour que les méthodes d'interpolation puissent caractériser finement l'hétérogénéité spatiale du milieu. Il est toutefois important de souligner que de telles densités d'échantillonnage sont rarement considérées dans les investigations de sites réels, principalement pour des raisons budgétaires. À l'inverse, la performance du modèle analytique inverse montre une stabilisation de l'erreur à partir d'un nombre relativement faible de points d'observation (8 points dans ce cas). Ceci suggère un avantage pratique notable, indiquant qu'une estimation robuste du volume pourrait être obtenue avec un nombre limité d'observations (< 10 points, par exemple).

Les pourcentages de surface contaminés faux positifs (Fig. 16a) et faux négatifs (Fig. 16b) sont basés sur les trois profondeurs (1, 3 et 5 m) et sur l'ensemble des seuils. Les tendances pour les différents seuils sont similaires.

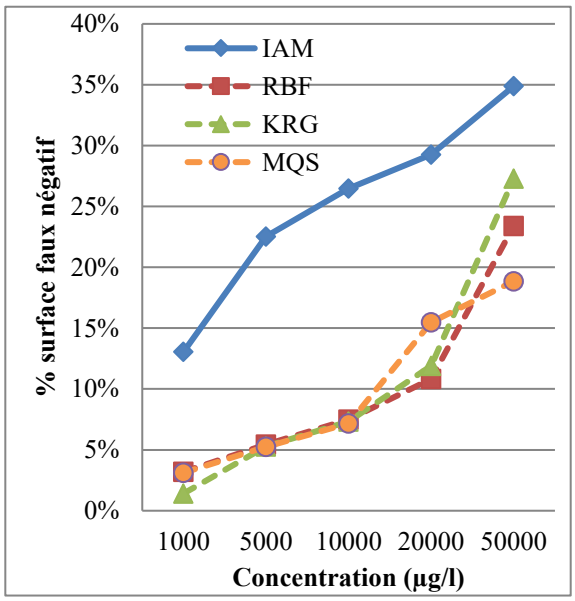


Fig.16.a

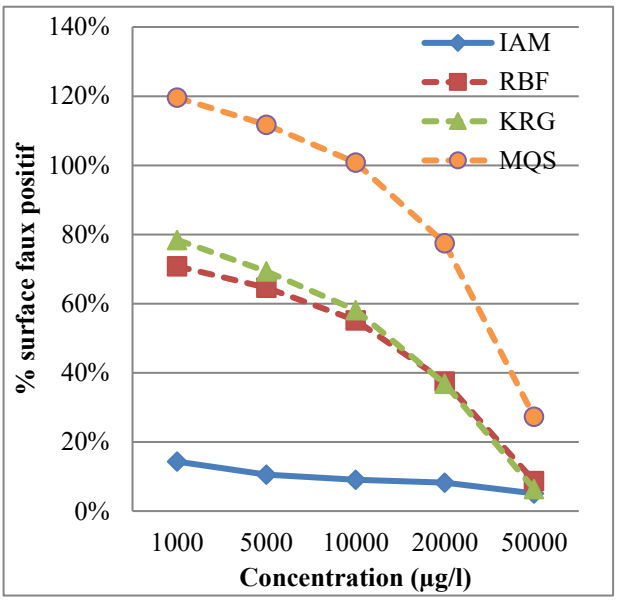


Fig.16.b

Figure 16 : Pourcentages de surfaces faussement négatives (a) et faussement positives (b) en fonction du seuil de dépollution pour les distributions régulières des points d'observation

La proportion de surfaces faussement négatifs (Fig. 16a) est plus importante avec l'approche IAM qu'avec les trois méthodes d'interpolation (5 à 15 points de pourcentage en plus), et ce pour les cinq seuils de pollution considérés. Avec l'augmentation du seuil, le pourcentage de valeurs faussement négatives pour toutes les courbes a augmenté.

En ce qui concerne le pourcentage de valeurs faussement positives (Fig. 16b), il existe une différence significative entre l'IAM et les méthodes d'interpolation (entre 1,3 % à 50 000 $\mu g.l^{-1}$ et 58 % à 1 000 $\mu g.l^{-1}$ pour les meilleurs résultats d'interpolation), en particulier pour les concentrations les plus faibles. Cette différence s'atténue à des niveaux de concentration plus élevés. Cette convergence est probablement due au fait que ces fortes concentrations se rapprochent des valeurs maximales mesurées aux points d'observation, ce qui réduit l'effet de lissage généralement induit par les méthodes d'interpolation. Les trois méthodes d'interpolation ont toutes abouti à une surestimation de la surface de pollution (de 1 point à 57 points de pourcentage).

3.3.6.2. Comparaison des surfaces estimés pour les distributions aléatoires

La figue 17 montre les erreurs moyennes d'estimation associées à la méthode IAM et aux trois techniques d'interpolation avec les distributions aléatoires des points d'observation.

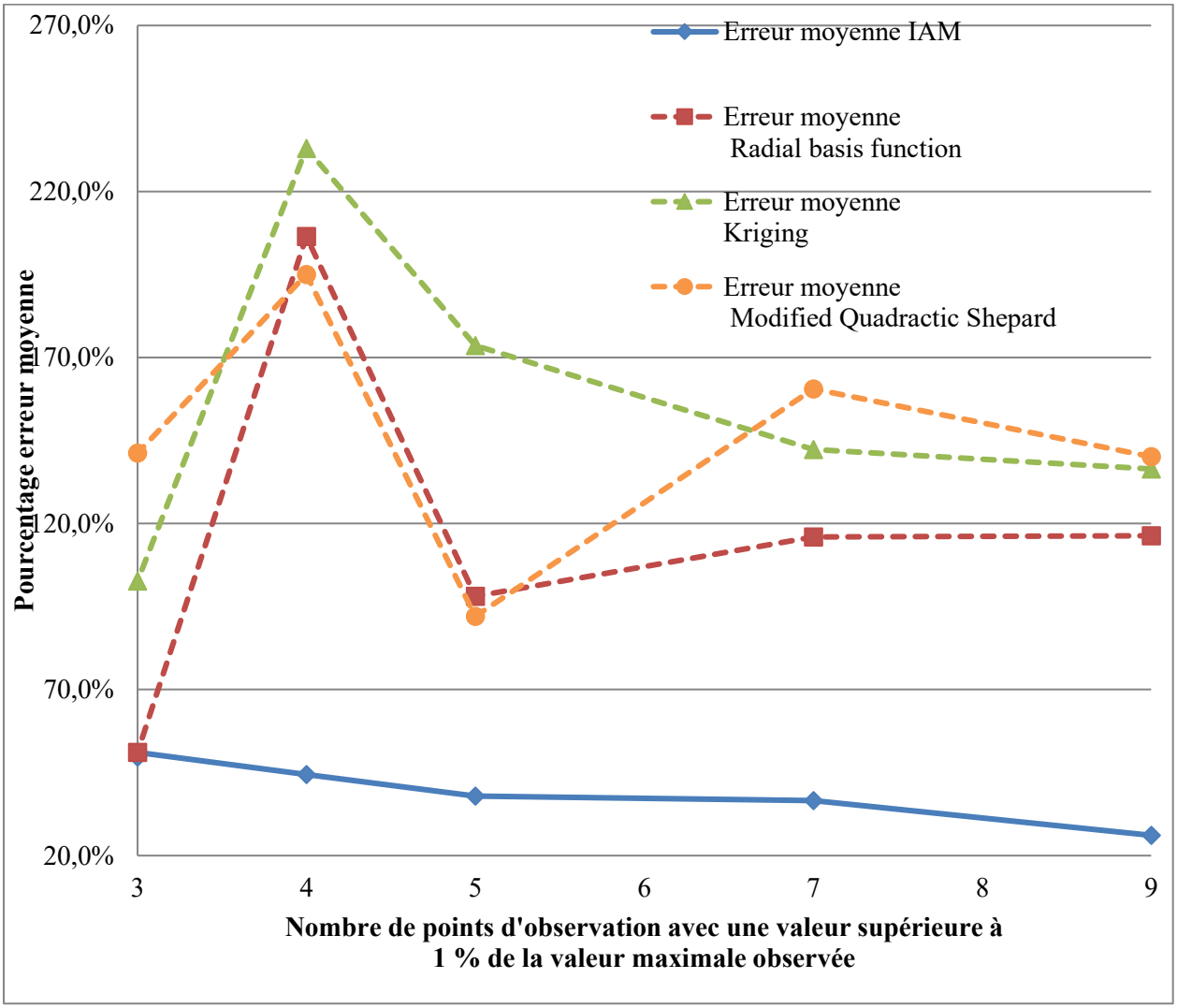


Figure 17 : Erreurs moyennes de surfaces estimées par la méthode IAM et par les méthodes d'interpolation à partir de distributions aléatoires des points d'observations

Là encore, les performances de la méthode IAM sont meilleures que celles des méthodes d'interpolation spatiale. L'erreur moyenne du modèle analytique inverse varie de 51,1 % pour le nombre minimum de points supérieur à 1% de la valeur maximale (3 points d'observation par couche) à 26,1 % pour la distribution spatiale comprenant 9 points d'observation supérieurs à 1% de la valeur maximale. Concernant les méthodes d'interpolation, la méthode RBF a enregistré la plus faible erreur moyenne (51,1 %), et ce uniquement avec le nombre minimal de points d'observation. Pour les autres méthodes, les erreurs obtenues sur l'ensemble des

distributions spatiales étaient beaucoup plus importantes, allant de 92 % à 195 % pour la meilleure approche d'interpolation.

La modélisation analytique inverse s'avère beaucoup plus robuste que les différentes méthodes d'interpolation. Le pourcentage d'erreur est beaucoup plus stable avec l'approche de modélisation inverse qu'avec les trois méthodes d'interpolation, dont les erreurs varient fortement en fonction du nombre et de l'emplacement des points situés en dehors et dans la zone source.

Les pourcentages de surfaces contaminés faux positifs (Fig. 18a) et faux négatifs (Fig. 18b) sont basés sur les trois profondeurs (1, 3 et 5 m) et sur l'ensemble des seuils.

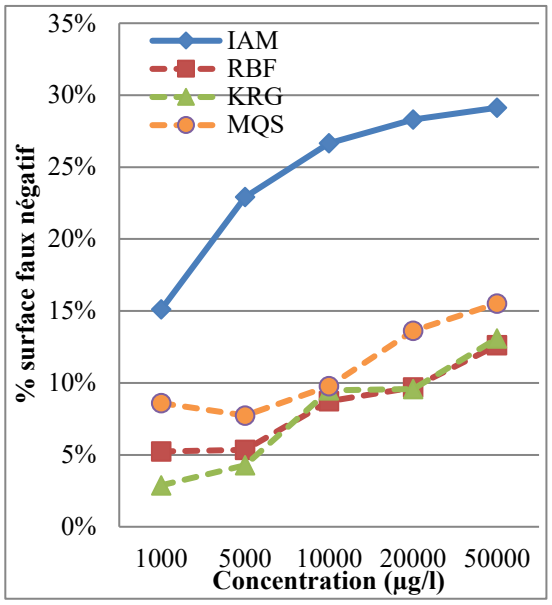


Fig.18a

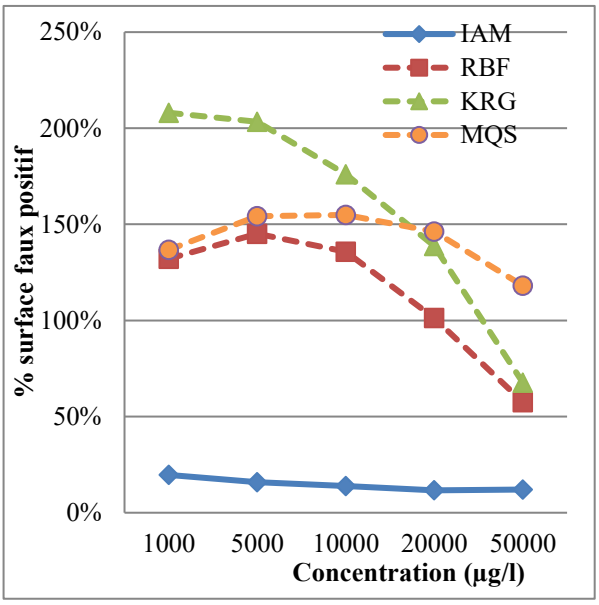


Fig.18b

Figure 18 : Pourcentages de valeurs faussement négatives (a) et faussement positives (b) en fonction du seuil pour les distributions aléatoires des points d'observation

Les proportions de surfaces faussement négatifs (Fig.18a et b) sont plus importantes avec le modèle analytique inverse qu'avec les 3 méthodes d'interpolation pour les 5 seuils de pollution de la pollution définis. Le pourcentage de surfaces faux négatifs augmente avec le seuil, quelle que soit la méthode d'estimation. Les proportions de surfaces faux positifs sont nettement plus élevées pour les trois méthodes d'interpolation que pour les autres méthodes (46 % à 129 % de surfaces en plus pour la meilleure méthode d'interpolation). Les méthodes d'interpolation ont notamment conduit à une surestimation des surfaces avec une distribution aléatoire des points, quel que soit le seuil de pollution considéré.

3.4. ÉVALUATION DE LA PERFORMANCE DES DIFFERENTES METHODES SUR DES CAS SYNTHETIQUES HETEROGENES AVEC ESTIMATION EN 3D

Afin de conforter les résultats obtenus dans la section précédente, 4 nouveaux cas synthétiques (nommés de cas n°1 à cas n°4) avec une répartition différente du panache de pollution ont été implémentés avec deux distributions différentes des points d'observation. Pour ces nouveaux cas, une estimation du volume total est réalisée, contrairement aux sections précédentes qui se basaient sur différentes surfaces.

L'objectif est de caractériser la variabilité potentielle de la distribution du panache de pollution plutôt que de représenter de manière exhaustive l'intégralité des configurations géologiques possibles.

3.4.1. Paramètres et caractérisation de la pollution simulée

Pour la construction de ces modèles synthétiques, seuls les paramètres I_{eff}=160 mm par an et la charge hydraulique = -50m imposée sur la dernière couche du modèle sont conservés. L'hétérogénéité spatiale du milieu a été implémentée en définissant des champs de variation pour : i) la porosité, ii) la teneur en eau volumique résiduelle, iii) les paramètres α et β de l'équation de van Genuchten. La génération de ces champs de paramètres avec une distribution spatiale variable a été réalisée à l'aide du module Python "GSTools" (Müller et al., 2022). Ce module est conçu pour la génération de champs aléatoires offrant la possibilité de définir et d'intégrer des structures d'hétérogénéité spatialement maîtrisées au sein d'un modèle numérique. GSTools offre la flexibilité de spécifier des propriétés telles que la variance, la portée (*range*), le type de modèle de variogramme et l'anisotropie. Ceci permet de construire des simulations qui respectent les structures de dépendance spatiale et les paramètres statistiques observés, afin de mieux représenter la variabilité spatiale des propriétés du milieu. Un exemple de configuration pour la construction du variogramme est fourni (cas n°2) dans le Tableau 7.

Tableau 7 : Exemples de valeurs pour construire le variogramme du cas n°2

Paramètres	Valeurs
Dimension	3
Variance	1
Portée du variogramme (x,y,z)	(7,7,8)
Effet pépite	0,1
Paramètre de régularité	2 = Gaussien

Les paramètres précisés dans le Tableau 7 montrent une portée des structures géologiques de 7×7×8 m avec un effet pépite très faible limitant le bruit de fond. Le paramètre de régularité fixé à 2 correspond à un modèle gaussien, le champ généré sera donc régulier.

La configuration de la source de pollution a été modifiée et sa forme a été ajustée pour être moins régulière et plus représentative d'une géométrie complexe (la cartographie de la zone source est présentée avec l'échantillonnage en Fig.27). La concentration d'injection du contaminant a également été augmentée, fixée à 400 000 $\mu g.l^{-1}$.

Les Figures 19 à 22 illustrent les champs de porosité générés. Il est à noter que la configuration spatiale des champs de la teneur en eau volumique résiduelle et des paramètres de van Genuchten reprend l'architecture spatiale et les mêmes caractéristiques géostatistiques que les champs de porosité présentés**.** Plusieurs configurations synthétiques ont été élaborées pour simuler différents types et degrés d'hétérogénéité du sous-sol. Sur les différentes figures présentées ci-après, l'échelle verticale (z) est différente de l'échelle x et y.

Le Cas n°1 représente une distribution de porosité caractérisée par une courte portée de corrélation spatiale, avec des valeurs variant entre 0,1 et 0,5. L'hétérogénéité est caractérisée par des structures de quelques décimètres d'extensions latérale et verticale. Bien que cette configuration soit peu probable dans des conditions géologiques naturelles, elle peut se rencontrer sur des sites ayant fait l'objet de remblaiements hétérogènes avec des matériaux divers. Ces contextes peuvent notamment être observés dans les grandes métropoles françaises qui ont fait l'objet de remblaiement massif notamment aux abords des cours d'eau pour prévenir les risques d'inondations.

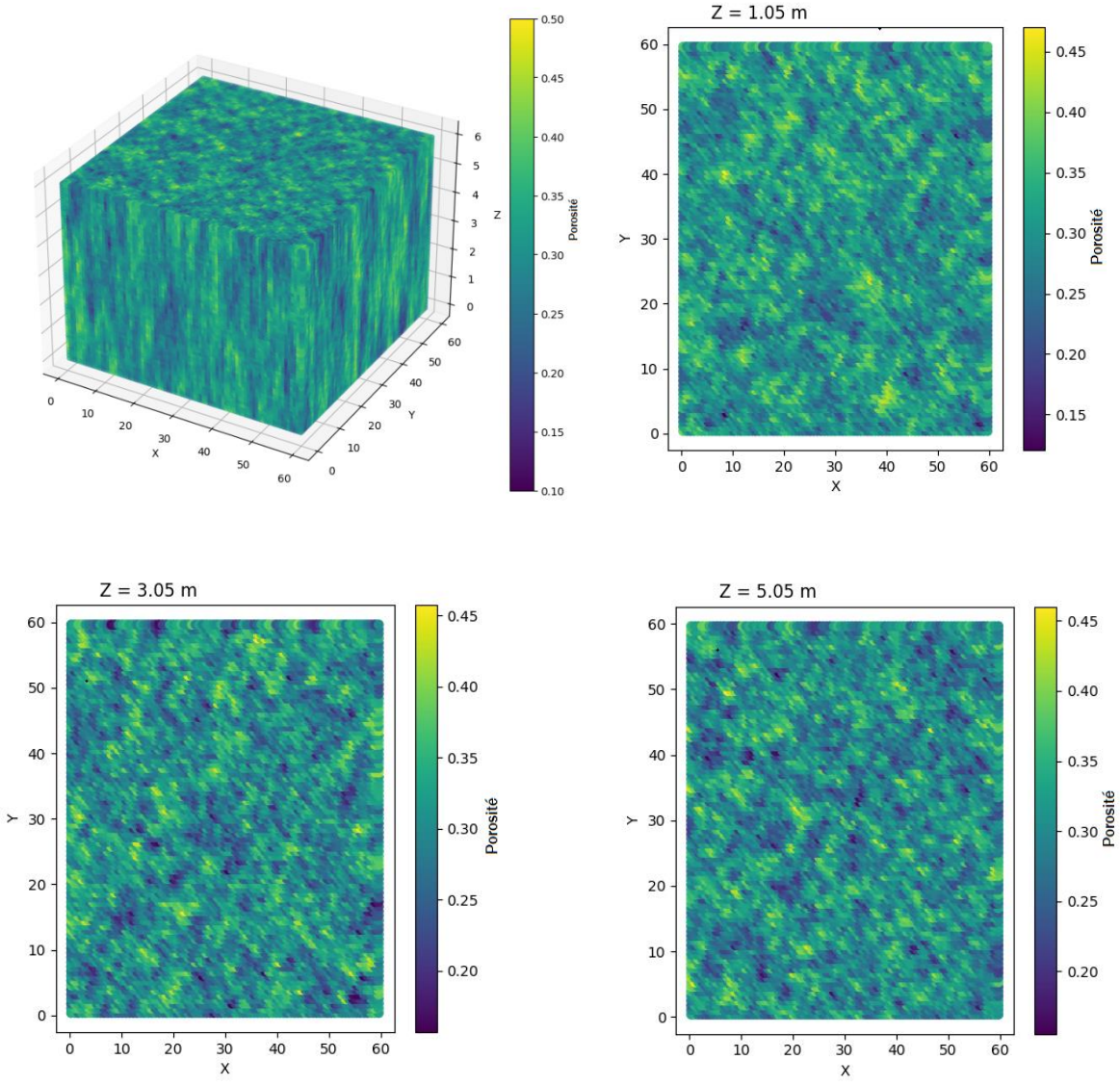


Figure 19 : Vue en 3D et en coupes des champs de porosité aux profondeurs 1 m, 3 m et 5 m pour le cas hétérogène n°1

Le Cas n°2 (Figure 20) correspond à une hétérogénéité où les matériaux présentent une corrélation verticale. Les structures géologiques représentées dans ce cas représentent une continuité spatiale d'environ une dizaine de mètres dans la direction horizontale et plus de 6 m dans la direction verticale. Les contrastes de propriétés entre les différentes entités d'hétérogénéité sont peu marqués. Ce type de structure pourrait s'apparenter, par exemple, à des profils d'altération de la roche mère, comme des arènes granitiques.

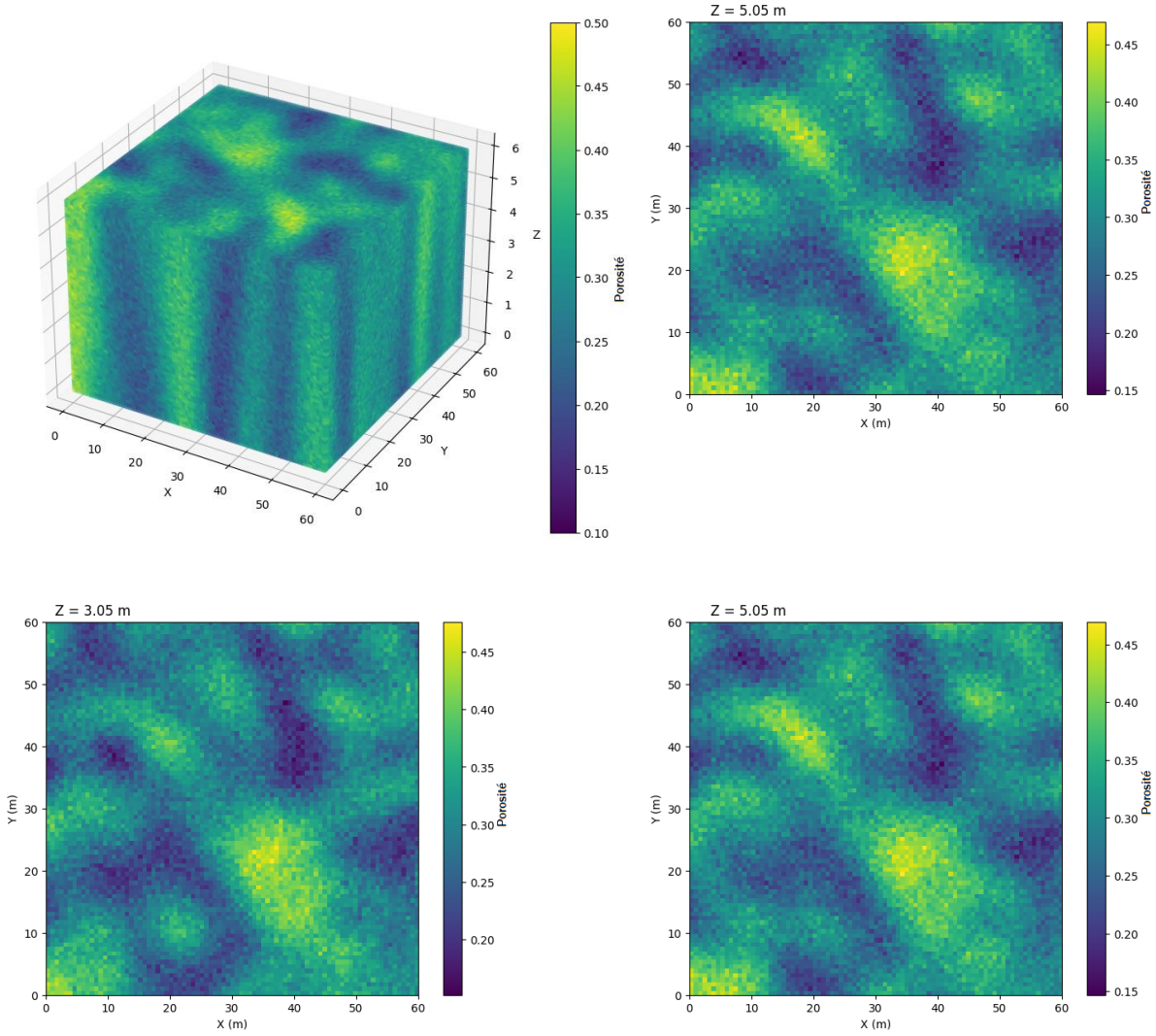


Figure 20 : Vue 3D et en coupes des champs de porosité aux profondeurs 1 m, 3 m et 5 m pour le cas hétérogène n°2

Le cas n°3 a été spécifiquement conçu pour simuler des conditions d'hétérogénéité prononcée et est possiblement moins réaliste vis-à-vis du milieu naturel. Dans ce scénario, des hétérogénéités aux contrastes marqués sont implémentées, s'étendant sur des dimensions de 2 à 4 m tant horizontalement que verticalement. Cette configuration se distingue par la présence de transitions abruptes entre les différentes unités hétérogènes qui la composent. De plus, une caractéristique notable est l'existence de contrastes significatifs dans les valeurs de la porosité (passage de 0,15-0,20 à 0,4-0,5) entre les unités adjacentes. Ces contrastes ont été obtenus à l'aide d'une transformation sigmoïde des valeurs permettant des transitions rapides entre les entités géologiques. Les fonctions sigmoïdes permettent au modèle de représenter avec précision des détails et des discontinuités (comme les fronts de pollution abrupts) que des transformations purement linéaires ou gaussiennes ne seraient pas capables de modéliser. Cela confère à l'approche la flexibilité nécessaire pour simuler des transitions non-graduelles dans le milieu synthétique.

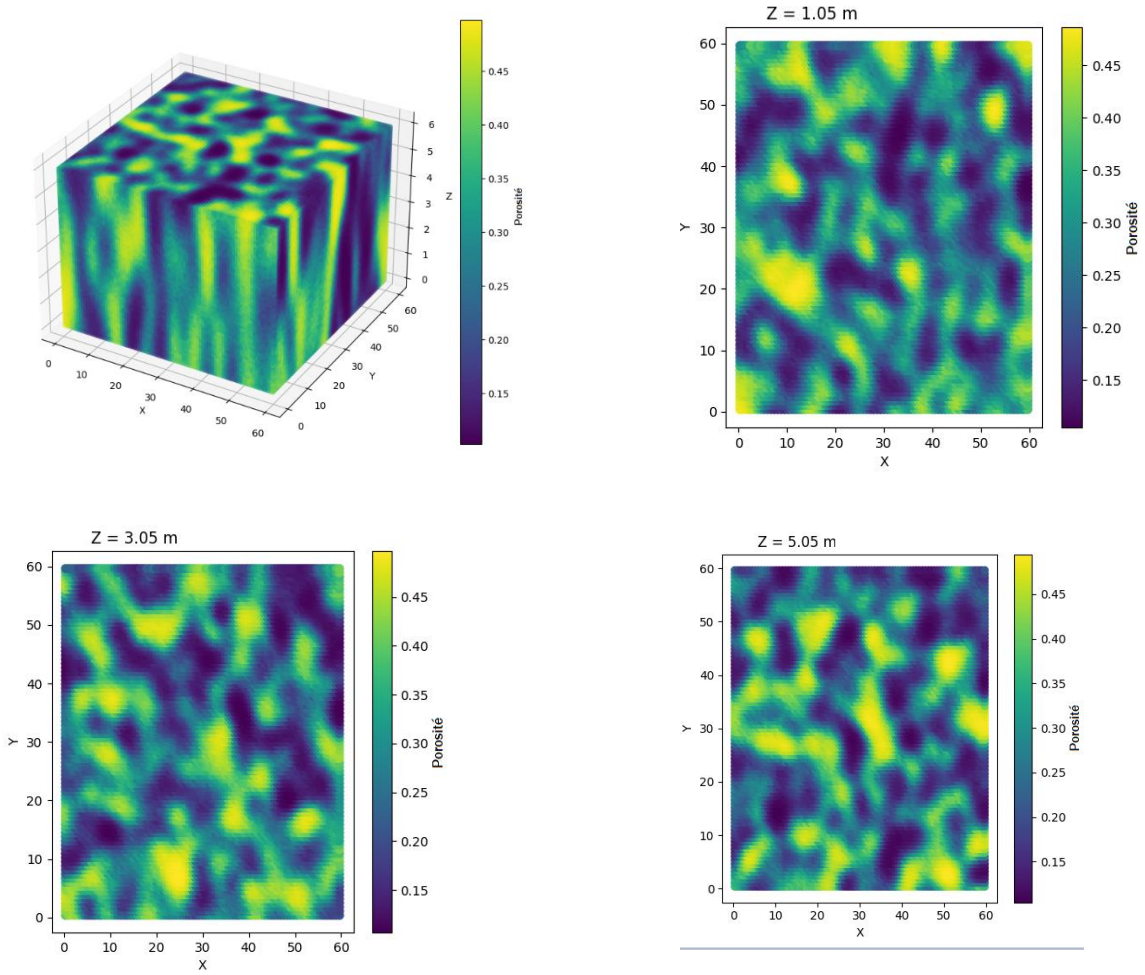


Figure 21 : Vue 3D et en coupes des champs de porosité aux profondeurs 1 m, 3 m et 5 m pour le cas hétérogène n°3

Le cas n°4 cas représente une situation géologiquement plus plausible, caractérisée par la superposition de deux couches distinctes. Ces couches présentent un léger pendage et des valeurs de porosité très différentes, simulant ainsi une interface entre deux formations aux propriétés géologiques (ex. conductivité hydraulique) contrastées.

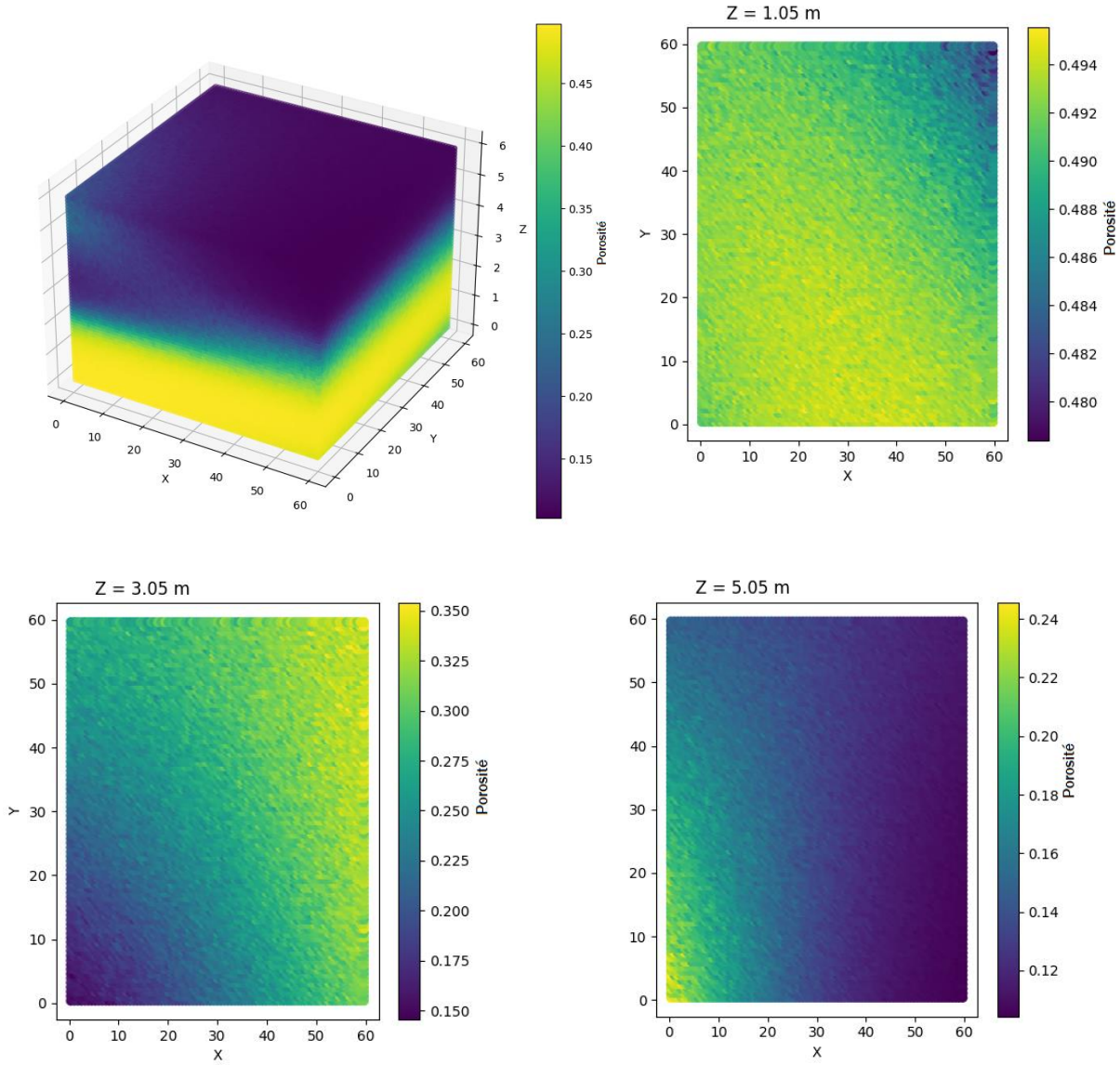


Figure 22 : Vue 3D et en coupes des champs de porosité aux profondeurs 1 m, 3 m et 5 m pour le cas hétérogène n°4

A l'image des cas synthétiques présentés dans les sections précédentes, les champs de paramètres ont été implantés dans le logiciel MODFLOW USG. Les panaches de pollution correspondants, simulés à partir de ces configurations hétérogènes après 2 ans sont présentés sur les Figures 23 à 26.

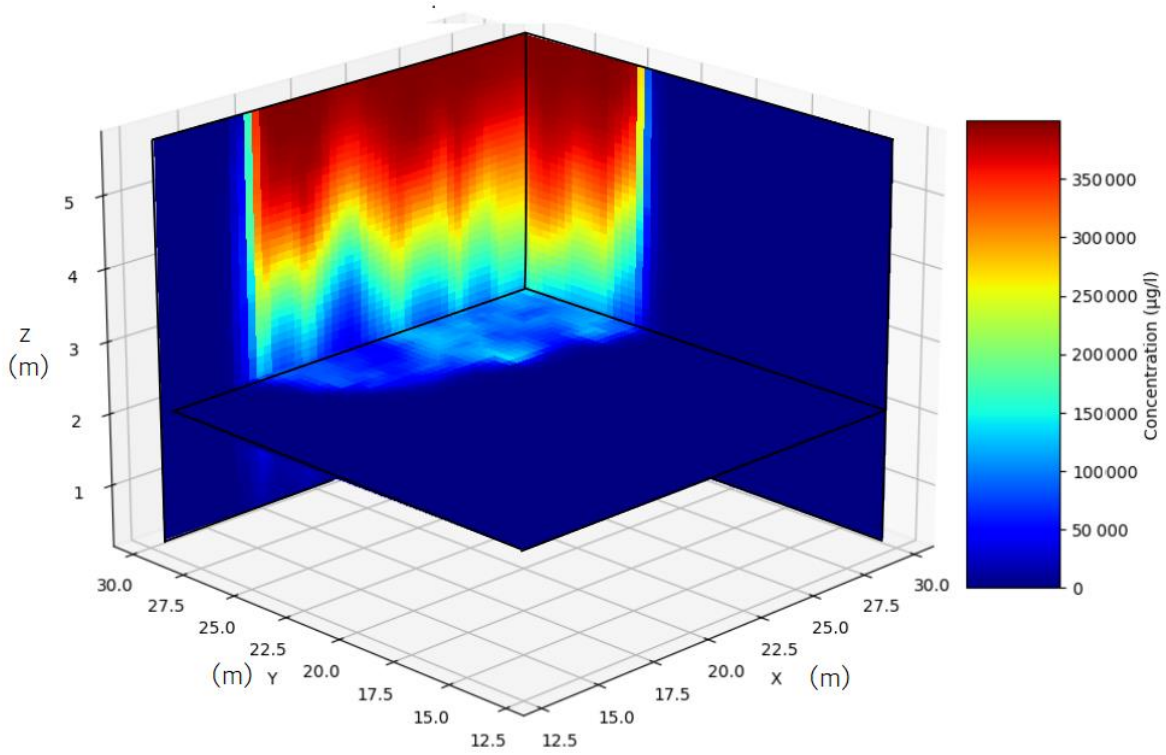


Figure 23 : Vue en plans du panache de pollution après deux ans pour le cas hétérogène n°1

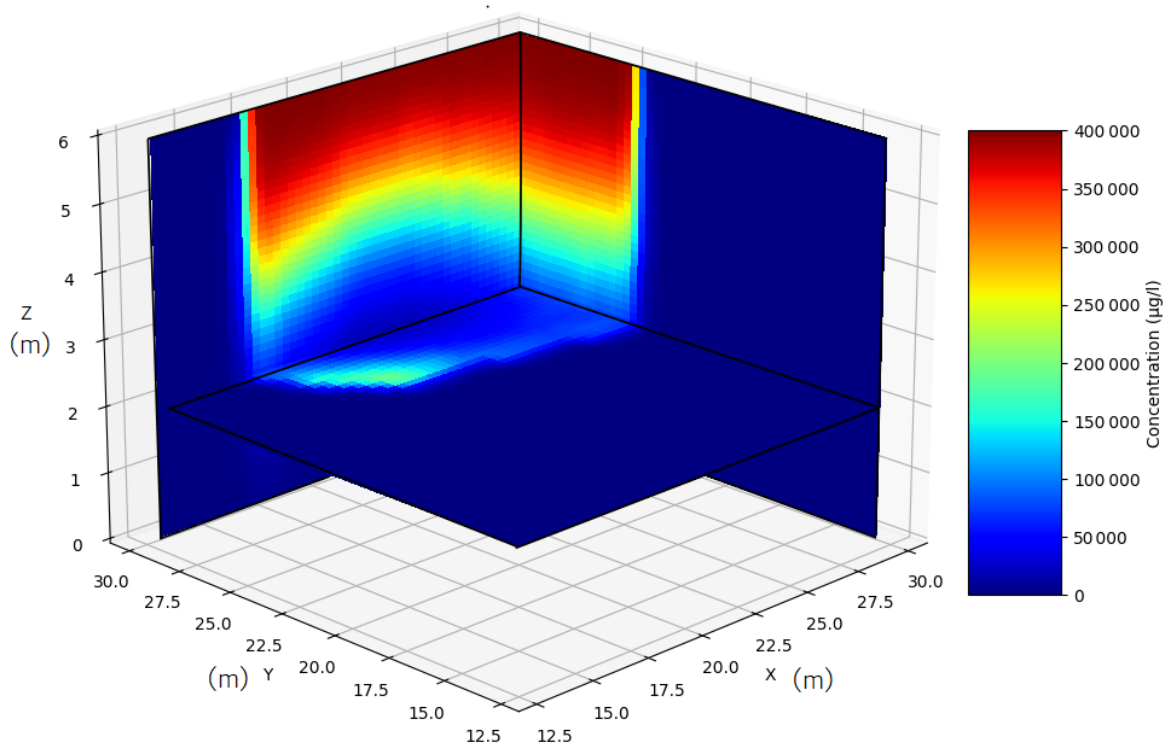


Figure 24 : Vue en plans du panache de pollution après deux ans pour le cas hétérogène n°2

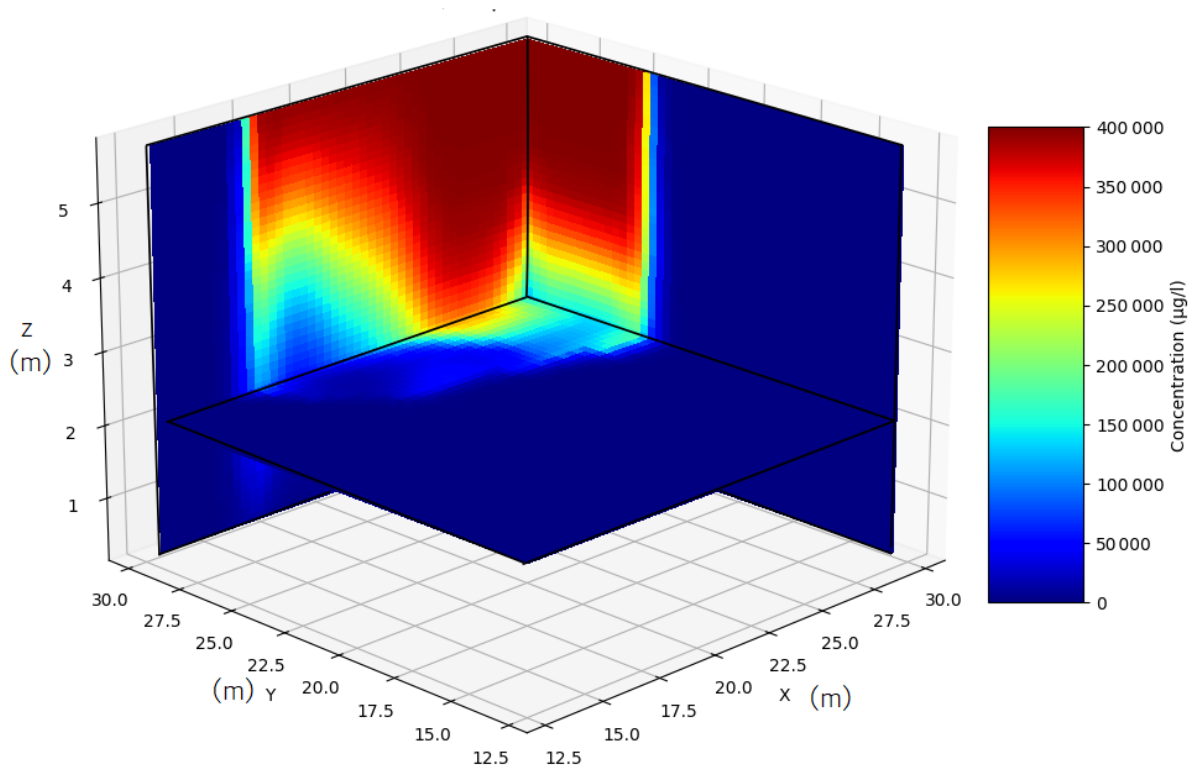


Figure 25 : Vue en plans du panache de pollution après deux ans pour le cas hétérogène n°3

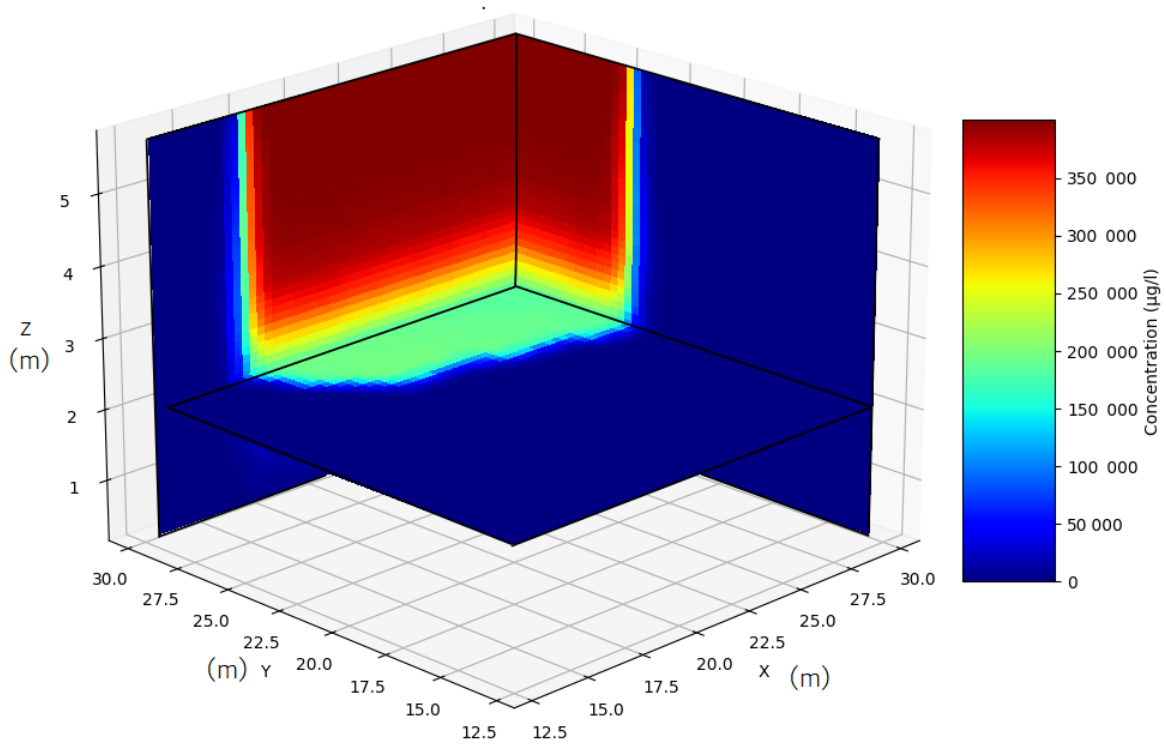


Figure 26 : Vue en plans du panache de pollution après deux ans pour le cas hétérogène n°4

Le Tableau 8 détaille les volumes de sol contaminés dont les concentrations simulées excèdent les seuils de pollution, pour chacune des configurations hétérogènes modélisées.

Tableau 8 : Volumes de terres en m^3 dont les concentrations sont supérieures aux différents seuils de dépollution considérés pour les 4 cas synthétiques hétérogènes

Seuil de concentration	Cas n°1	Cas n°2	Cas n°3	Cas n°4
5000	1562	1385	1442	1509
10000	1453	1276	1344	1441
25000	1245	1082	1162	1279
50000	1052	903	1003	1083
100000	887	755	857	1010
200000	700	581	687	899

Les volumes contaminés présentés dans le Tableau 8, révèlent des divergences notables entre les différents cas. Il est observé que l'ampleur des écarts volumétriques tend à s'accentuer avec l'augmentation des seuils de concentration des polluants considérés.

3.4.2. Comparaison des résultats

Six seuils de dépollution potentiels (C_{crit}) (rapports C_w/C_{w0}) ont été définis à 1,25 %, 2,5%, 6,25%, 12,5 %, 25 % et 50 % (correspondant respectivement à 5 000, 10 000, 25 000, 50 000, 100 000 et 200 000 µg.l^{-1}). L'évaluation a été conduite de 0 à 6 m de profondeur sur l'ensemble de la zone potentiellement polluée, ce qui permet de considérer le volume total de terres contaminées. Pour permettre la comparaison entre la solution analytique inverse et une méthode d'interpolation, une interpolation 3D utilisant la méthode des fonctions de base radiale (RBF) a été spécifiquement sélectionnée. Ce choix est justifié par les résultats obtenus dans le chapitre 3.3, où la méthode RBF a démontré les meilleures performances parmi les techniques d'interpolation évaluées.

3.4.3. Plan d'échantillonnage et stratégie de la modélisation analytique inverse

Deux configurations distinctes, avec des densités d'échantillonnages différentes ont été appliquées à chacun des quatre cas d'étude synthétiques. Ces configurations comprennent respectivement 10 et 20 points d'observation par profondeur (1, 3 et 5 m) soit entre 30 et 60 points d'observation au total. La localisation spatiale (x, y) des points d'observation a été déterminée par un processus de placement aléatoire mis en œuvre à l'aide d'un logiciel de Système d'Information Géographique (SIG), avec comme unique contrainte un espacement minimum de 4 m entre les points (Fig. 27).

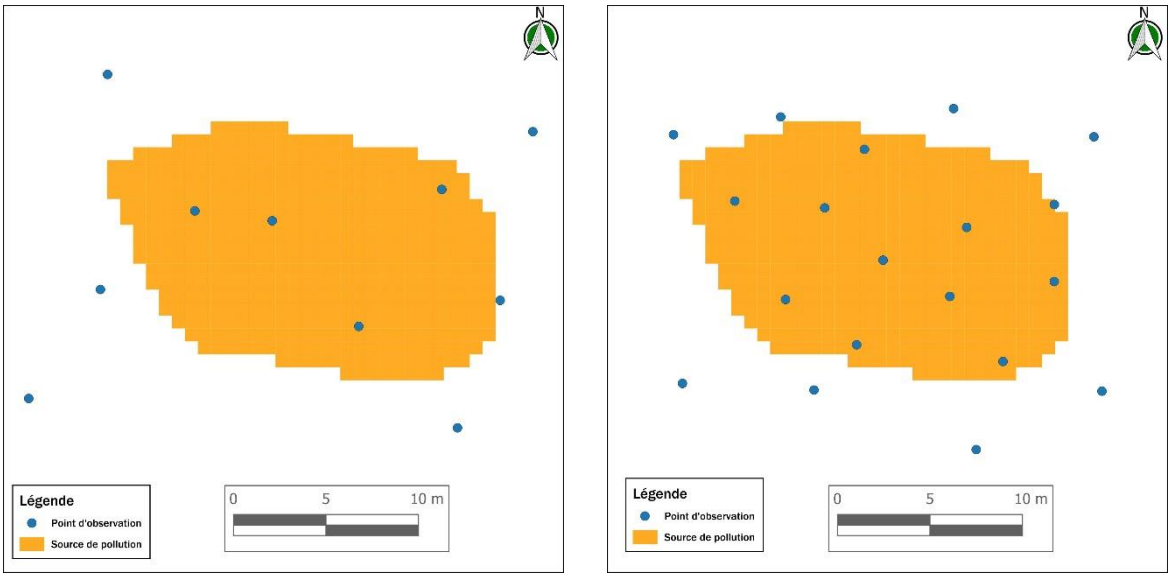


Figure 27 : Carte des configurations d'échantillonnage appliquées aux 4 cas synthétiques

L'échantillonnage à 10 points d'observation a généré 13 triangles de Delaunay, tandis que celui à 20 points d'observation en a produit 30. Le nombre total de modèles inversés est donc de 130 pour l'échantillonnage à 10 points et 300 pour l'échantillonnage à 20 points.

Afin d'obtenir une estimation volumétrique complète sur l'ensemble du profil vertical, incluant les profondeurs non échantillonnées, il est nécessaire d'abord d'appliquer une discrétisation verticale du milieu par intervalles de 50 cm. Ensuite, une méthode d'assignation est appliquée pour caractériser chaque couche. Pour les cotes de profondeur non échantillonnées, les valeurs de concentrations sont estimées en utilisant les données de la profondeur échantillonnée la plus proche. Dans les cas où une zone se situe à mi-distance entre les zones d'échantillonnage, une combinaison des deux zones d'échantillonnage les plus proches est utilisée (Fig. 28).

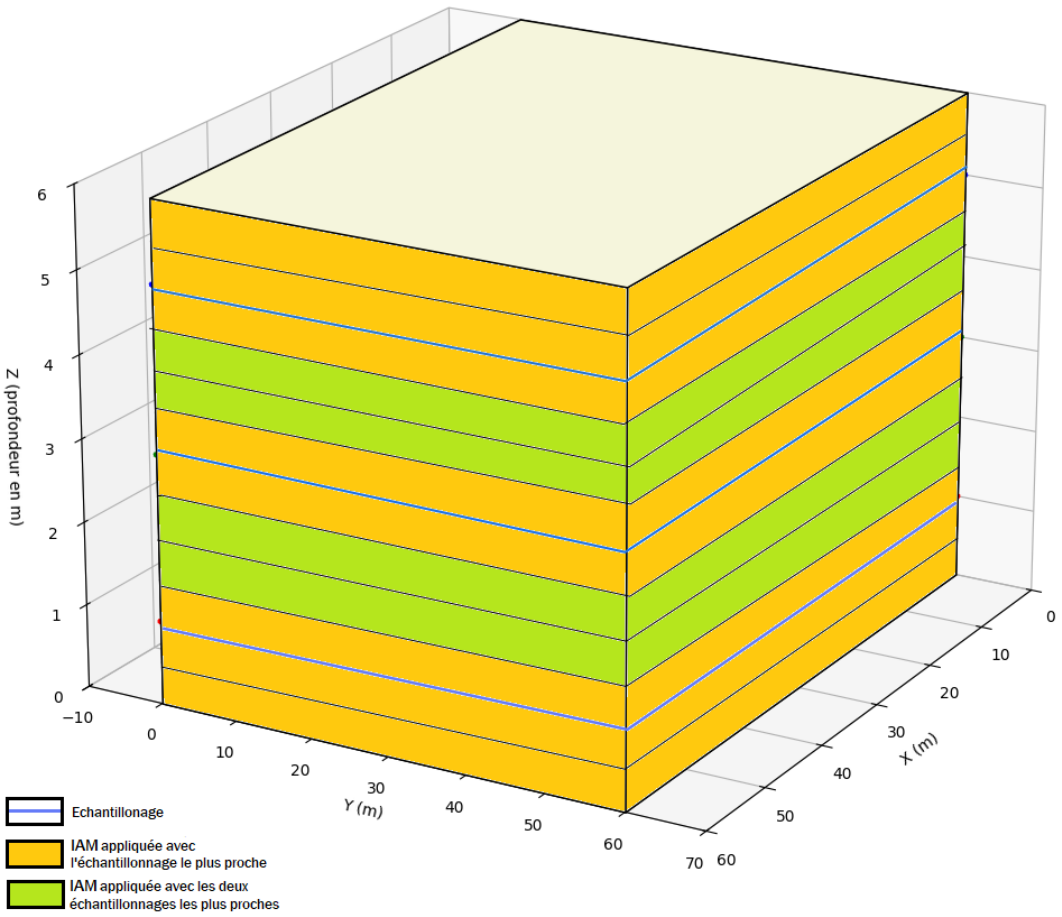


Figure 28 : Données utilisées pour estimer le volume de terres polluées

3.4.4. Comparaison des volumes estimés par modélisation analytique inverse avec une interpolation spatiale

Le Tableau 9 synthétise les erreurs moyennes associées à la méthode IAM et à l'interpolation. Contrairement aux sections antérieures traitant des cas homogène et hétérogène, l'approche adoptée ici n'emploie pas un seuil de probabilité unique pour la délimitation des volumes polluées mais un intervalle de probabilité permettant d'expliciter l'incertitude inhérente aux estimations. Une borne moyenne est également considérée correspondant à la moyenne entre les bornes minimum et maximum. Le Tableau 9 récapitule l'ensemble des erreurs pour les bornes minimum, moyenne et maximum du seuil de probabilité ainsi que l'interpolation.

Tableau 9 : Récapitulatif des Erreurs Moyennes d'Estimation pour la Modélisation Analytique Inverse (borne moyenne) et l'interpolation (4 Cas, 2 Échantillonnages)

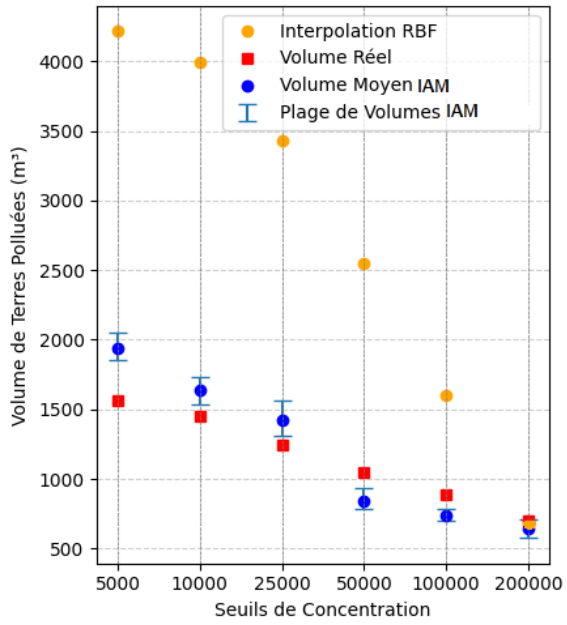
	Cas n°1		Cas n°2		Cas n°3		Cas n°4	
Nombre de points d'échantillonnage	**10**	**20**	**10**	**20**	**10**	**20**	**10**	**20**
% erreur moyenne – seuil de probabilité à la borne moyenne - IAM	37,0	27,0	49,0	42,7	43,8	25,5	30,5	17,0
% erreur moyenne – seuil de probabilité à la borne minimum - IAM	36,2	25,6	51	42,7	42,5	25,6	28,5	15,6
% erreur moyenne – seuil de probabilité à la borne maximum - IAM	40,3	26,5	54	42,9	48,1	27.5	32	19
% erreur moyenne interpolation	135,6	69,1	161,6	79,9	175,5	62,7	92,7	59,1

L'erreur moyenne d'estimation pour la méthode IAM varie de 15,6 % (observée pour le cas synthétique n°4 avec la configuration d'échantillonnage la plus dense et la borne minimum) à

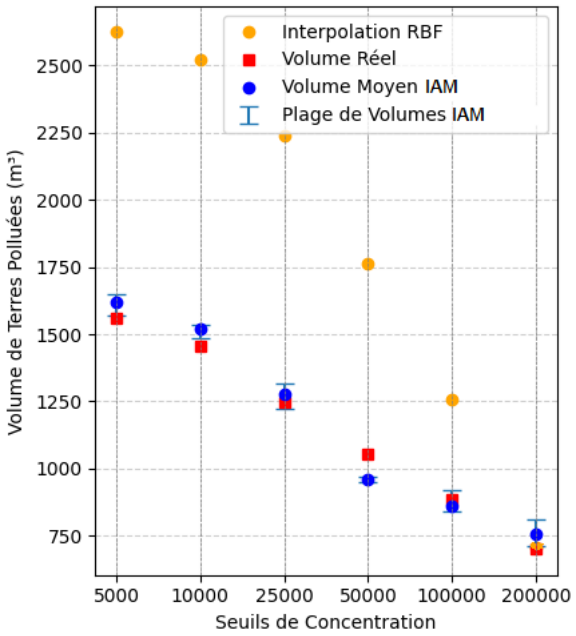
54 % (pour le cas n°2 avec la configuration d'échantillonnage la plus faible et la borne maximum). L'erreur diminue logiquement lorsque la densité d'échantillonnage augmente (correspondant à l''échantillonnage n°2), et ce, indépendamment du cas synthétique considéré.

La comparaison avec l'interpolation révèle une performance globalement meilleure de l'IAM. Le pourcentage d'erreur moyen est plus faible pour l'ensemble des cas pour la modélisation analytique inverse, avec des écarts de performance souvent significatifs (diminution de l'erreur moyenne d'estimation du volume pollué de 47 à 75%). Sur un total de 48 comparaisons (résultant de 4 cas synthétiques, 2 densités d'échantillonnage et 6 seuils de concentration), la modélisation analytique inverse a fourni des résultats plus précis dans 45 de ces comparaisons. Dans une comparaison, la performance était similaire à celle de l'interpolation, et dans deux cas, la méthode d'interpolation a montré une erreur moyenne plus faible. Ces résultats sur les 48 comparaisons sont similaires en considérant la borne minimum, la borne moyenne ou la borne maximum de l'approche IAM. Il est toutefois observé que la borne minimum conduit occasionnellement à des pourcentages d'erreur d'estimation légèrement plus faibles que ceux obtenus avec la valeur moyenne. Cette tendance peut s'expliquer par une surestimation des volumes plus faibles avec cette borne.

Les Figures 29 à 32 illustrent les estimations des volumes de sols contaminés pour chacun des cas synthétiques étudiés. Afin de mieux visualiser les résultats, l'échelle en ordonnée est adaptée aux valeurs maximales observées, elle diffère donc entre les graphiques. Les comparaisons se concentrent exclusivement sur les erreurs de volume total par rapport aux volumes « réels » (c'est-à-dire l'erreur chiffrée, sans prendre en compte l'erreur de localisation spatiale) par seuil pour chaque cas.

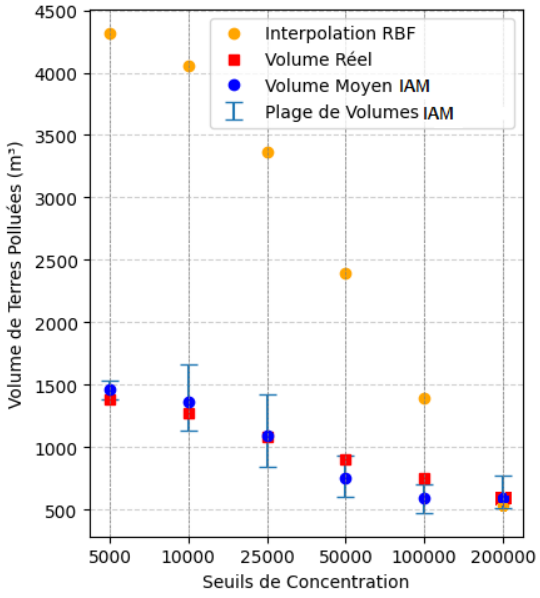


Cas 1 échantillonnage 1

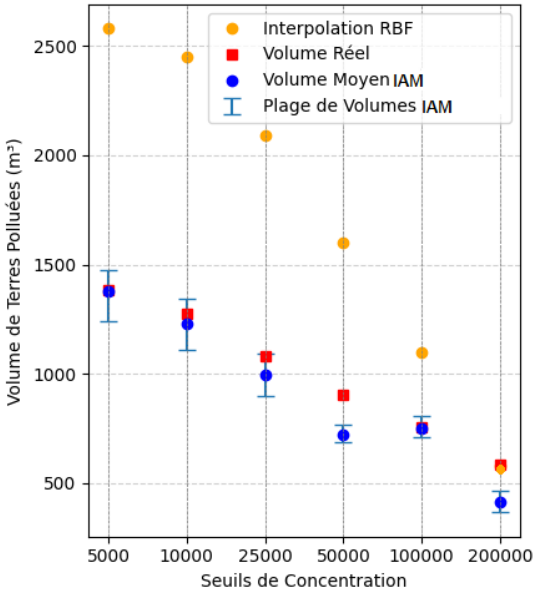


Cas 1 échantillonnage 2

Figure 29 : Graphiques de comparaison des volumes pour le cas synthétique n°1 et l'ensemble des seuils

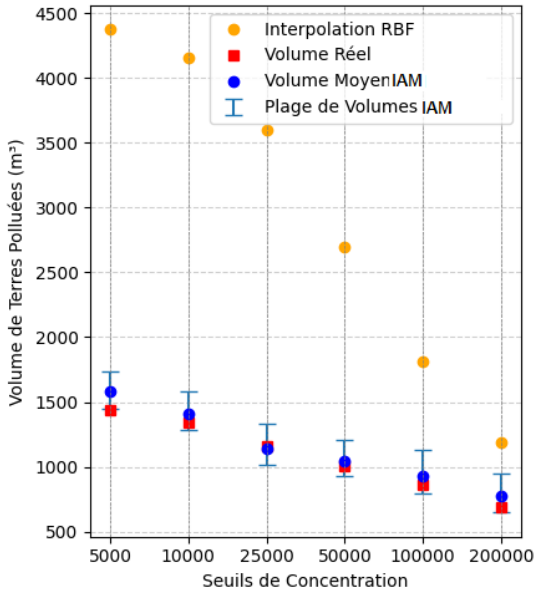


Cas 2 échantillonnage 1

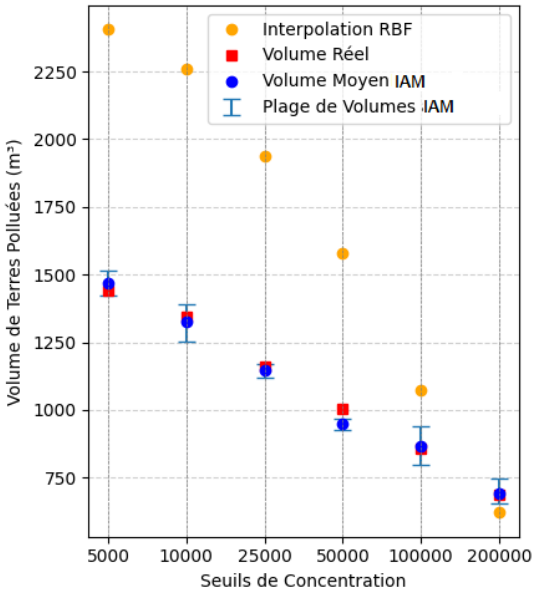


Cas 2 échantillonnage 2

Figure 30 : Graphiques de comparaison des volumes pour le cas synthétique n°2 et l'ensemble des seuils

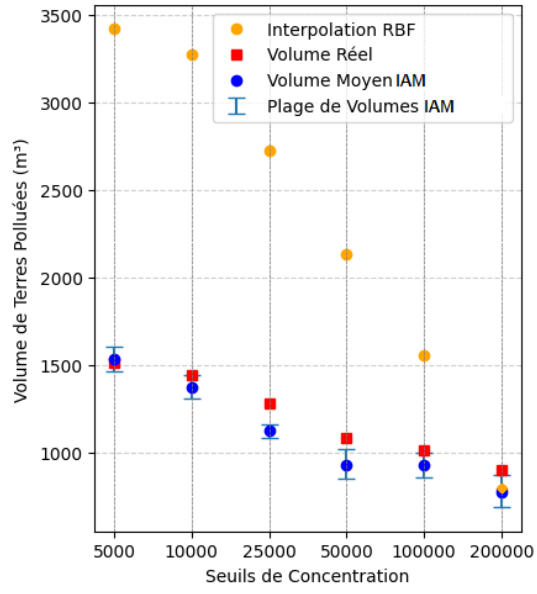


Cas 3 échantillonnage 1

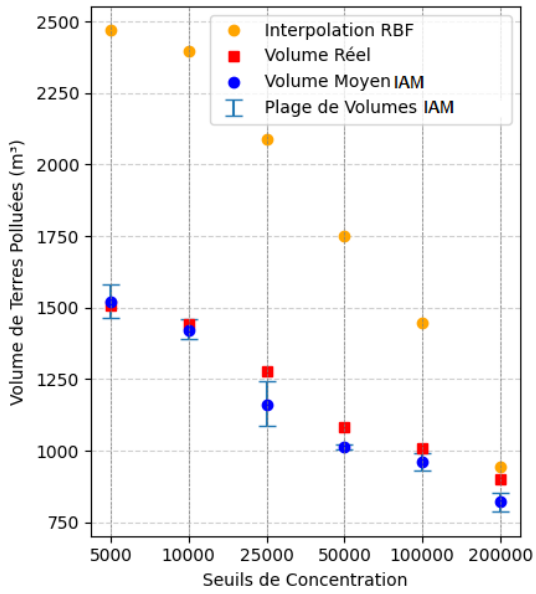


Cas 3 échantillonnage 2

Figure 31 : Graphiques de comparaison des volumes pour le cas synthétique n°3 et l'ensemble des seuils



Cas 4 échantillonnage 1



Cas 4 échantillonnage 2

Figure 32 : Graphiques de comparaison des volumes pour le cas synthétique n°4 et l'ensemble des seuils

Les résultats montrent une meilleure adéquation des volumes estimés pour la modélisation analytique inverse. L'erreur de volume pour la borne moyenne est comprise entre 0,4 % et 29 % pour la solution analytique inverse. L'interpolation se rapproche du volume réel uniquement

pour le seuil de 200 000 µg.l^{-1}. Cette observation concorde avec les résultats obtenus dans les précédentes sections : les interpolations telles que la méthode RBF ont tendance à mieux estimer les fortes valeurs et surestiment les faibles valeurs. Le volume « réel » est compris entre les bornes minimum et maximum dans 27 configurations sur les 48 testées. L'analyse de l'erreur volumétrique en termes de faux positifs et de faux négatifs (Fig.33) révèle que la méthodologie IAM développée présente une répartition relativement équilibrée entre ces deux types d'erreurs. En revanche, la comparaison de ces volumes IAM et des volumes de l'interpolation, pour l'ensemble des configurations testées, met en évidence une tendance marquée des méthodes d'interpolation à surestimer de manière importante le volume réel de sol contaminé. Cette surestimation est particulièrement marquée par les volumes de faux positifs qui excèdent fréquemment 100 % pour l'échantillonnage n°1 et 50 % pour la configuration d'échantillonnage n°2. Afin de pouvoir permettre une meilleure lisibilité du graphique, l'échelle des abscisses est différente de l'échelle des ordonnées.

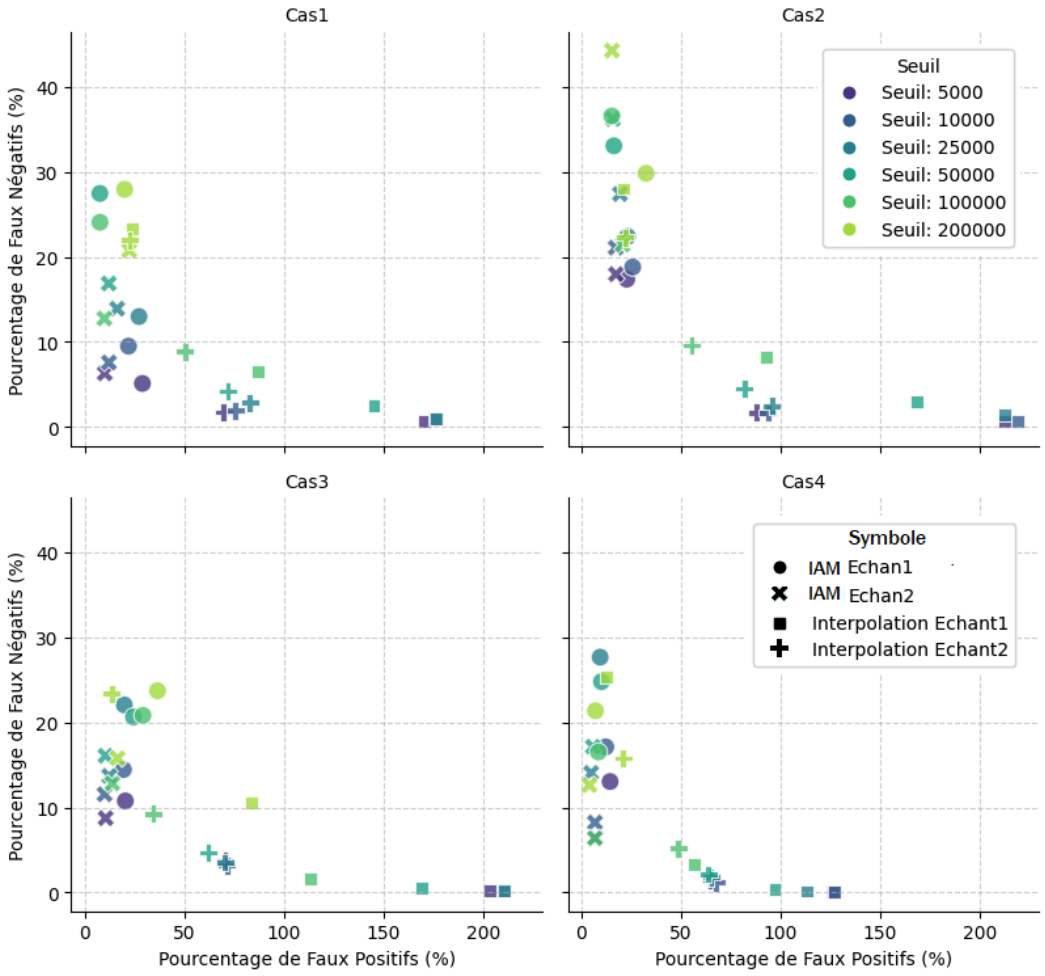
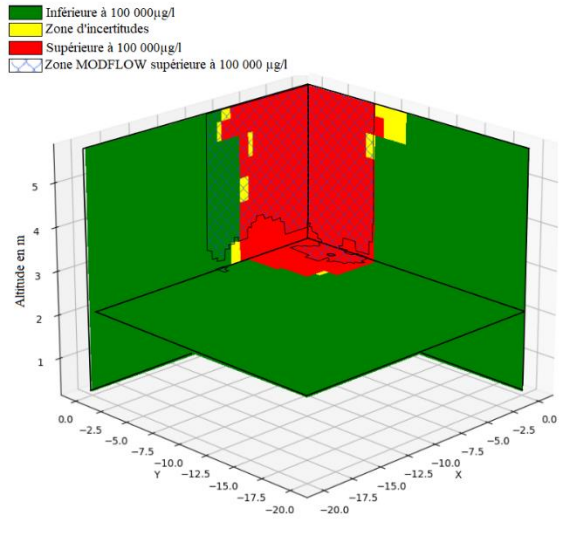
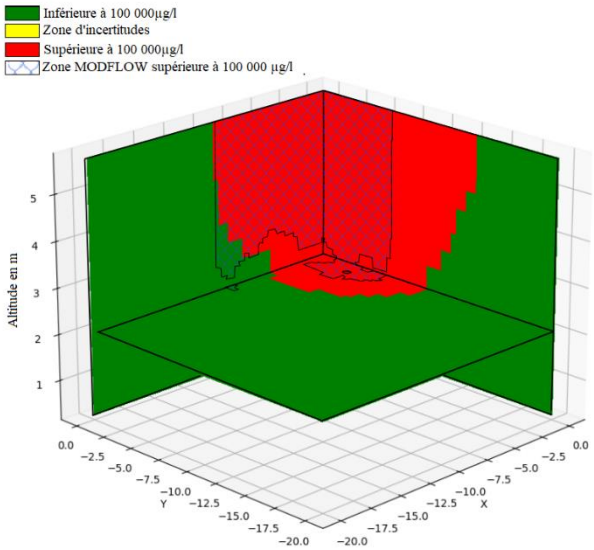


Figure 33 : Comparaison entre l'interpolation et la modélisation analytique des volumes faux positifs et faux négatifs

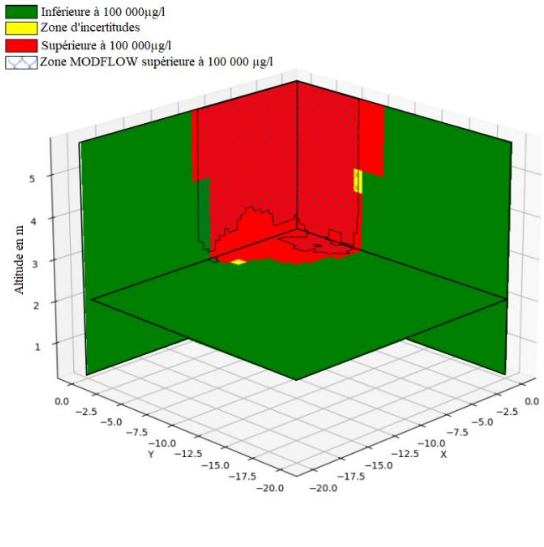
Les Figures 34 et 35 permettent de comparer les résultats obtenus par l'IAM et l'interpolation RBF. Cette comparaison est présentée sous forme de trois coupes dans l'ensemble des directions pour les cas synthétiques n°1 et n°3, en considérant les deux configurations d'échantillonnage précédemment définies et un seuil de concentration de 100 000 µg.l^{-1}.



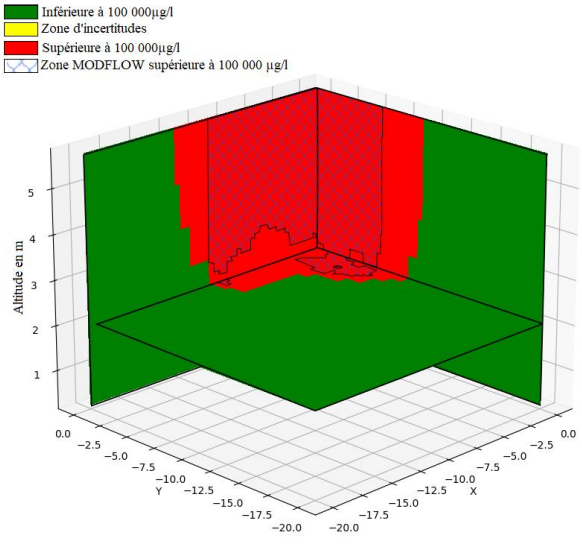
Cas1 - Echantillonnage 1 modèle analytique inverse



Cas1 - Echantillonnage 1 Interpolation

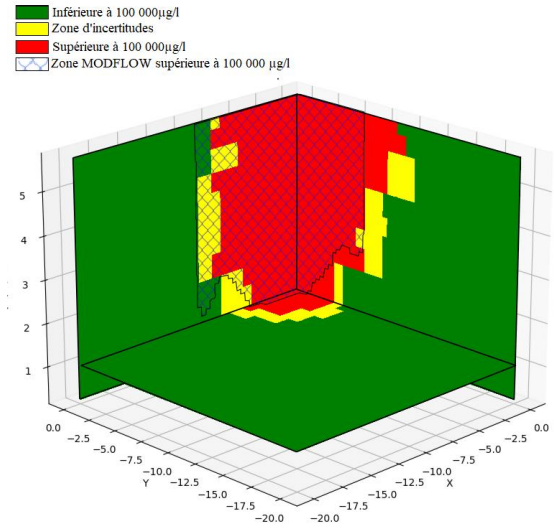


Cas1 - Echantillonnage 2 modèle analytique inverse

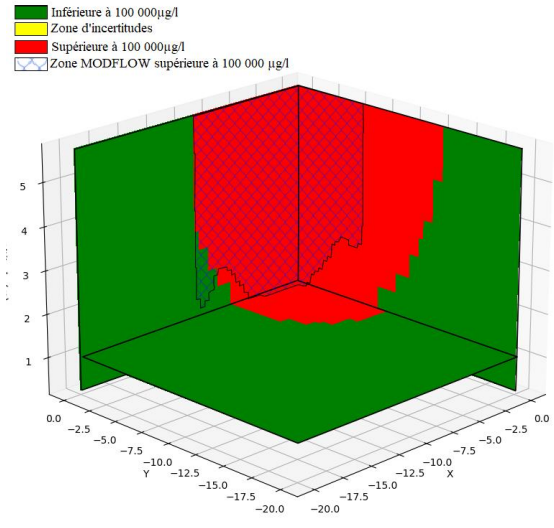


Cas1- Echantillonnage 2 Interpolation

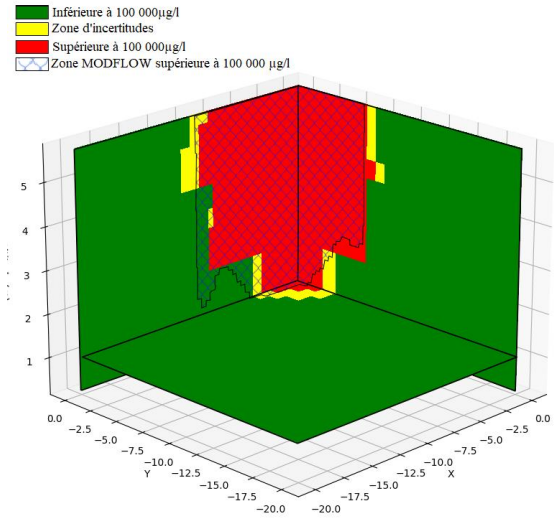
Figure 34 : Comparaison entre l'interpolation et la modélisation analytique inverse avec un seuil de 100 000 µg.l^{-1} pour le cas n°1



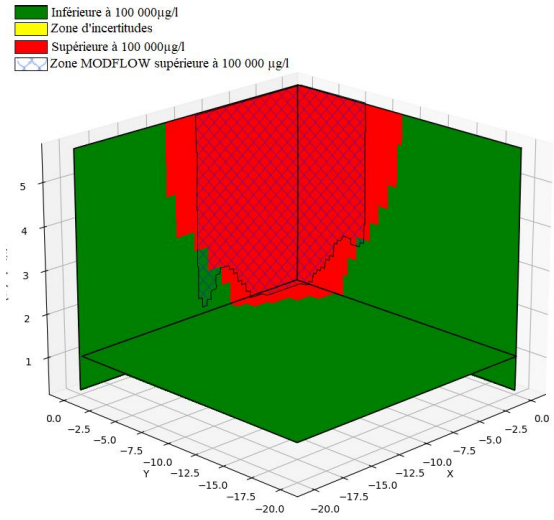
Cas3 - Echantillonnage 1 modèle analytique inverse



Cas3 - Echantillonnage 1 Interpolation



Cas3 - Echantillonnage 2 modèle analytique inverse



Cas3 - Echantillonnage 2 Interpolation

Figure 35 : Comparaison entre l'interpolation et la modélisation analytique inverse avec un seuil de 100 000 µg.l^{-1} pour le cas n°3

Les Figures 34 et 35 illustrent la surestimation des volumes pour l'interpolation pour l'ensemble des cas avec les deux échantillonnages. Les volumes d'incertitudes sont logiquement plus importants pour l'échantillonnage avec le nombre de points le plus faible.

3.5. SYNTHESE DU CHAPITRE

Ce chapitre applique la modélisation analytique inverse (IAM) à divers cas synthétiques, homogènes et hétérogènes, pour estimer les surfaces ou volumes de sols contaminés. La méthode de modélisation analytique inverse a ensuite été appliquée à des triplets de points d'observation (triangulation de Delaunay) et comparée à des interpolations spatiales classiques (RBF, MQS, Krigeage).

Les résultats montrent que la modélisation analytique inverse réduit significativement l'erreur moyenne d'estimation (15-25% en milieu homogène, 40-75% en hétérogène) par rapport aux méthodes d'interpolation. Elle se révèle particulièrement robuste avec un nombre limité de points d'observation (moins de 10 points). Bien que la performance des interpolations puisse ponctuellement surpasser l'IAM pour certaines configurations très denses, la modélisation analytique inverse offre globalement une meilleure performance et un équilibre plus fiable entre les faux positifs et les faux négatifs, les interpolations tendant à surestimer le volume réel.

La méthodologie appliquée peut donc être testée sur des cas réels.

CHAPITRE 4 : APPLICATION A DES CAS REELS

Compte tenu des performances encourageantes de la méthode analytique inverse sur les cas synthétiques, son application et sa vérification sur des cas d'étude réels sont présentées dans ce chapitre. L'évaluation est conduite de manière progressive :

i) Construction et description d'une base de données de 32 sites avec des sols pollués ;

ii) Application à un cas réel spécifique et détaillé ;

iii) Mise en œuvre sur un ensemble plus large et diversifié de 32 sites distincts ;

iv) Enfin, la démarche est également testée sur un scénario de panache de pollution affectant les eaux souterraines.

L'objectif de l'application de l'IAM à des cas réels est de tester sa performance en utilisant des données directement issues d'études de terrain indépendantes. Les résultats de ces différents diagnostics, qui servent de base pour l'évaluation de la méthode, ne seront pas commentés dans les sections suivantes.

Les investigations de site sont souvent contraintes par des facteurs tels que le budget, le planning du projet, ainsi que les limitations physiques et d'accessibilité du terrain.

4.1. CONSTRUCTION ET DESCRIPTION DE LA BASE DE DONNEES

Une base de données a été spécialement conçue à partir de divers sites, en respectant les critères de sélection suivants :

- Travaux de dépollution par excavation et suivi : les sites devaient avoir fait l'objet d'un terrassement avec la réalisation de prélèvements en bords de fouille ;
- Données de localisation précises : les fouilles devaient obligatoirement faire l'objet d'un relevé de géomètre pour assurer une cohérence des données spatiales ;
- Source de pollution : Seuls les sites présentant une ou plusieurs sources de pollution avérées ont été retenus. Les sites où une dépollution a été réalisée sans source distincte (par exemple, en raison d'une mauvaise qualité des remblais correspondant à une pollution diffuse) ont été exclus ;
- Prélèvement individuel : les diagnostics pris en compte devaient être basés sur des prélèvements individuels par sondage. Les diagnostics utilisant des échantillons composites provenant de multiples sondages ont été écartés ;
- Une seule fouille a été considérée par site : dans le cas où plusieurs zones distinctes sur site ont été terrassées, seule une fouille a été conservée afin de ne pas tronquer les statistiques. Cependant, dans des cas spécifiques, la conservation de deux zones

d'excavation par site a été justifiée par la présence de polluants distincts, permettant une analyse différenciée ;

- Exclusion des données ne concernant pas la source de pollution : les autres sondages/prélèvements réalisés sur site ne concernant pas la source de pollution ont été exclus de la base de données ;
- La surface des fouilles considérée correspond à la surface finale terrassée, les éventuelles reprises après un premier terrassement sont donc intégrées ;
- Non intégration des fonds de fouille : pour les besoins de cette étude comparative avec des cas réels, il a été délibérément choisi de ne pas intégrer les données relatives aux prélèvements en fonds de fouille, malgré leur disponibilité à l'issue des travaux de dépollution pour certains sites. Cette décision est motivée par plusieurs facteurs. Premièrement, pour près de 30% des cas étudiés, le fond de fouille se situait en zone saturée, ce qui introduit des dynamiques hydrogéologiques complexes distinctes de celles des bords de fouille. Deuxièmement, pour 15% des sites, l'interruption de l'excavation en profondeur était dictée par un changement lithologique, indiquant une hétérogénéité géologique significative.

Au total, 32 sites ont été inclus dans la base de données. Les données proviennent majoritairement de diagnostics réalisés par la société Apave (23), les autres données ont été fournies par d'autres bureaux d'étude (9). Tous les sites sont localisés en France métropolitaine. La Figure (Fig 36) présente les sites par département.

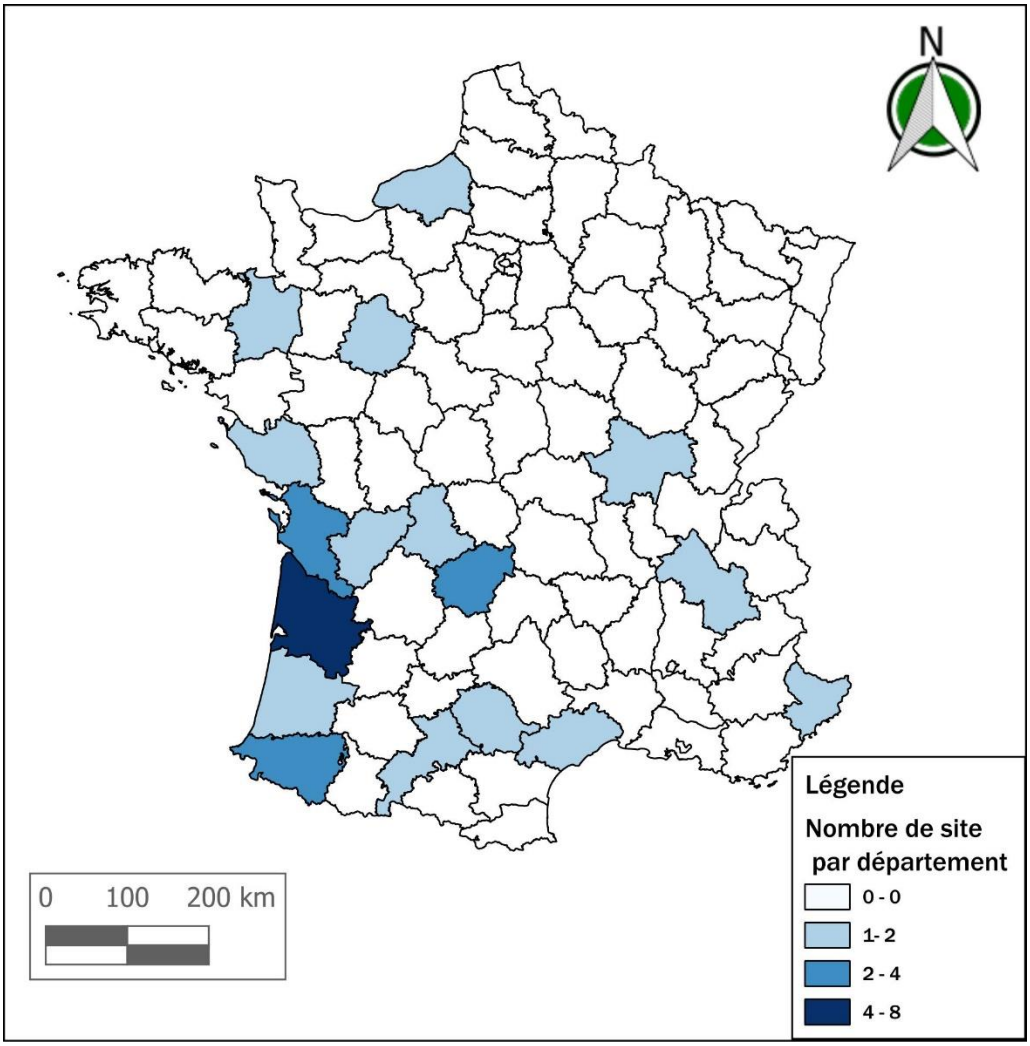


Figure 36 : La cartographie des sites inclus dans la base de données par département

La base de données a été mise en ligne (https://doi.org/10.5281/zenodo.16100900). Le Tableau 10 présente la base de données avec : le numéro du site, le nom des sondages, les coordonnées X et Y en relatif afin de conserver l'anonymat du site, la profondeur moyenne des prélèvements, les concentrations en $mg.kg^{-1}$ MS, le composé chimique concerné et la lithologie principale du site.

Tableau 10 : Présentation de la base de données (uniquement le site 1)

Site	Nom	X (m)	Y (m)	Profondeur moyenne (m)	Concentration ($mg.kg^{-1}$ MS)	Composé	Seuil de dépollution ($mg.kg^{-1}$ MS)	Lithologie majoritaire
1	1	2,626	1,986	1	0	HCT C10-C40	500	Remblais sableux
1	2	3,184	0,004	4	83	HCT C10-C40	500	Remblais sableux
1	3	-1,259	-0,885	2,25	51	HCT C10-C40	500	Remblais sableux
1	5	0,918	0,481	2,6	2180	HCT C10-C40	500	Remblais sableux
1	6	-0,119	2,173	1,75	65	HCT C10-C40	500	Remblais sableux
1	7	1,016	-0,976	3	1240	HCT C10-C40	500	Remblais sableux
1	9	0,317	-1,518	1	44	HCT C10-C40	500	Remblais sableux
1	8	2,212	-1,259	2,5	15	HCT C10-C40	500	Remblais sableux
1	A1	-0,008	-0,655	2,75	3130	HCT C10-C40	500	Remblais sableux

Les polluants majoritaires dans la base de données correspondent aux hydrocarbures (29 sites sur 32) : différentes pollutions aux hydrocarbures sont présentes telles que le fioul (6 sites), le gazole (12 sites) et les huiles (11 sites). Les trois autres composés correspondent à des pollutions aux COHV, en PFOS et en pentachlorophénols. Une faible proportion de polluants volatils est mentionnée dans le jeu de données analysé. Il convient de souligner que ces types de contaminants sont fréquemment traités par des méthodes in situ, ce qui explique leur exclusion de la présente base de données. Le traitement « in situ » fait référence à des techniques de dépollution qui sont mises en œuvre directement dans le sol ou la nappe souterraine, sans avoir besoin d'excaver les matériaux. Cela inclut, par exemple, l'injection de produits chimiques pour dégrader les polluants, ou la bioremédiation qui utilise des micro-organismes pour décomposer les substances polluantes. Par ailleurs, bien que les BTEX soient régulièrement associés aux pollutions hydrocarbonées, les hydrocarbures C10-C40 ont été retenus comme traceur principal dans les cas étudiés.

Le nombre moyen de sondages réalisés par site s'élève à 9,8, avec un écart type de 4,6 et une médiane de 9, ce qui suggère une distribution relativement symétrique. Le nombre de sondages par site a varié de 4 à 23. Concernant les prélèvements associés, leur nombre moyen est d'environ 14, avec un écart type de 8,5 et une plage s'étendant de 4 à 42 prélèvements. L'écart type plus élevé pour les prélèvements (8,5) par rapport à celui des sondages (4,6) indique une

plus grande dispersion dans le nombre de prélèvements effectués par rapport au nombre de sondages.

Les lithologies majoritaires (lithologie la plus représentative de la zone dépolluée) correspondent à des sables (37,5%), des remblais hétérogènes (25%), des argiles (18,7%), des limons (9,3%) et autres lithologies (alluvions, micaschiste altéré, marne altérée…) (9,3%).

Les méthodes préalablement utilisées pour dimensionner les différentes zones contaminées correspondent à :

- Tracé à main levé (47%) ;
- Interpolation : polygones de Voronoï (22%) ;
- Interpolation : polygones de Voronoï modifiée (13 %) ;
- Zone comprise dans des travaux prévus (la zone d'excavation est déjà définie par un plan de terrassement pour un projet de construction) (6%) ;
- Autres méthodes comme la réalisation de maillage supplémentaire (prélèvements supplémentaires avant les travaux de dépollution) (3%), interpolations plus proche voisin (3%) ou méthodes non communiquées (6%).

L'évaluation des pratiques en matière de délimitation des zones de pollution révèle une prédominance de l'approche manuelle (tracé à main levée), qui constitue la méthode la plus fréquemment employée. En seconde position, la méthode des polygones de Voronoï est couramment appliquée pour définir l'étendue spatiale des contaminations. Il convient de noter qu'une variante de cette approche, impliquant une modification des polygones de Voronoï pour optimiser la géométrie en vue de faciliter l'excavation, représente la troisième méthodologie la plus utilisée. Dans les plans de gestion examinés, d'autres techniques d'interpolation sont parfois présentées. Ces alternatives n'ont généralement pas été retenues pour le choix de la zone à excaver, leur pertinence ayant été jugée insuffisante. Cette situation souligne un besoin non satisfait en outils de dimensionnement robustes pour les sources de pollution de petite et moyenne taille notamment.

La moyenne des seuils de dépollution est de 1375 $mg.kg^{-1}$ MS, avec un écart type aux alentours de 1345 $mg.kg^{-1}$ MS. Le seuil maximum est de 5000 $mg.kg^{-1}$ MS et le minimum est de 0,06 $mg.kg^{-1}$ MS. Ce seuil très faible s'explique par la nature du polluant identifié (PFOS).

Les principales statistiques concernant les travaux de dépollution sont présentées dans le tableau suivant (Tableau 11).

Tableau 11 : Principales statistiques concernant les travaux de dépollution

Statistiques	Surface de fouille (m^2)	Profondeur (m)	Volume (m^3)	Densité de m^2/ sondage/fouille	Nombre de prélèvements en bords de fouille	Nombre de dépassements seuil en bords de fouille
Minimum	10	0,6	6	5,2	2	0
Maximum	2000	5	8000	250	15	6
Moyenne	275	2,8	787	49,4	6,1	1,3
Médiane	137	2	291	35	4,5	1
Ecart type	390	1,3	1480	50,0	3,1	1,7
Percentile 80	325	3,9	987	66,9	8,8	2,8
Percentile 90	574	4	1690	87,1	10	3,9

L'écart type pour la dimension des fouilles (surface, profondeur et volume) est assez éloigné de la moyenne, ce qui indique une variabilité importante avec des opérations de dépollution allant de faibles ampleurs à des travaux d'envergure importante. Ces statistiques indiquent que la base de données est représentative de cas de différentes tailles. Le nombre total de reprises (terrassement supplémentaire par rapport au programme initial) est de 17. Quatorze sites ont fait l'objet d'une reprise de travaux à minima. Elles pouvaient concerner un ou plusieurs bords de fouille. La moyenne en m^2/ sondage /fouille (sondage réalisé lors du diagnostic au droit de la futur zone terrassée) est d'un sondage pour 49 m^2, la densité minimum est d'un sondage pour 250 m^2 et la densité maximum d'un sondage pour 5 m^2. Le nombre total de bords de fouille pour l'ensemble des sites est de 194. À la fin des opérations de dépollution, 41 des bords de fouille n'ont pas atteint l'objectif de réhabilitation préalablement établi. La moyenne des dépassements est de 1,3 bords de fouille par fouille, avec une plage variant de 0 à 6 prélèvements excédant l'objectif. Des classes de densités d'échantillonnage ont été définies

(Tableau 12). Les résultats ne suggèrent aucune diminution significative du pourcentage de non-conformité lorsque la densité spatiale d'échantillonnage augmente.

Tableau 12 : Pourcentage de conformité entre les bords de fouille réalisés et les seuils de dépollution pour la base de données

Gamme de densité de sondage/m²/fouille	Nombre de site	% de non-conformité entre les bords de fouille réalisés et les seuils de dépollution
5-15	6	14%
15-30	7	31%
30-50	8	24%
50-70	6	9%
70 et +	5	17%

4.2. APPLICATION DE L'IAM A UN CAS REEL SPECIFIQUE ET DETAILLE

Le site étudié est localisé dans la ville de Saint Vincent de Tyrosse dans le département des Landes (40). La localisation du site est présentée sur la Figure 37. Le terme zone d'étude correspond aux limites parcellaires du site.

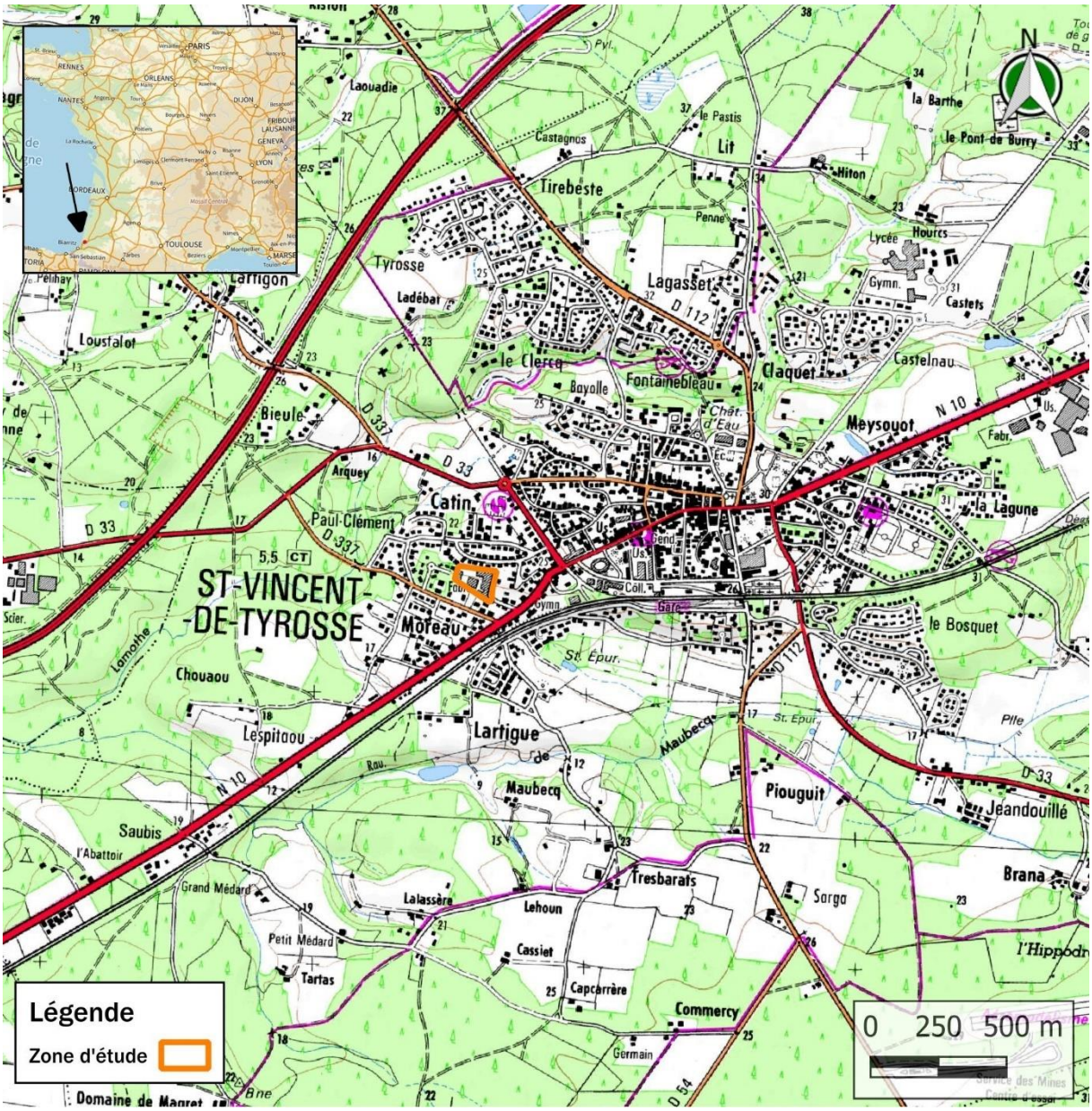


Figure 37 : Localisation du site où l'IAM a été appliquée

4.2.1. Contexte géologique

La géologie de Saint Vincent de Tyrosse et ses alentours correspond majoritairement à des dépôts éoliens du type « sables des Landes », datant principalement du Pléistocène supérieur et plus spécifiquement de la période glaciaire du Würm III. Ces dépôts sableux résultent du remaniement éolien de matériaux plus anciens, notamment des alluvions fluviatiles issues des bassins de l'Adour et de la Garonne, ainsi que de sédiments côtiers. Ces sédiments sont décrits comme des sables grossiers, avec une médiane granulométrique supérieure à 0,2 mm avec un excellent classement granulométrique. Ils sont majoritairement de nature quartzeuse. Ces formations sableuses recouvrent des terrains plus anciens, tels que les "Sables fauves" du Pliocène, des alluvions Quaternaires plus anciennes, ou des sables de formations marines du Tertiaire (Miocène, Oligocène). Des limons argileux datant du Pléistocène ainsi que des colluvions provenant de sables éoliens plus ou moins argileux ou limoneux sont aussi présents. Les limons argileux présentent d'après la carte géologique de Saint Vincent de Tyrosse (Fig.38) : 22,5 % de sables, 54 % de limons et 23,5 % d'argiles

La zone d'étude est localisée dans la zone des limons argileux, en contact au Nord avec les sables des Landes.



Légende

Zone d'étude	
Formation du Sable des Landes	
Limons argileux	
Colluvions. Remaniements de sables éoliens plus ou moins argileux ou limoneux.	
Remaniements de sables éoliens plus ou moins argileux sur graviers, sables et vases fossilifères	

Figure 38 : Extrait de la carte géologique de Saint Vincent de Tyrosse (BRGM, 1987)

4.2.2. Contexte hydrogéologique

Le contexte hydrogéologique est caractérisé par la forte conductivité hydraulique (de l'ordre de 1×10^{-4} $m.s^{-1}$) des formations géologiques sableuses, qui abritent un aquifère libre dont l'épaisseur varie généralement de 10 à 20 m et peut atteindre 40 m par endroits. La profondeur de la surface piézométrique est généralement comprise entre 0,5 et 5 m sous la surface du sol, à l'exception des zones dunaires où elle peut être plus profonde. La porosité des sédiments constituant l'aquifère est estimée entre 10 et 12 %. Cette valeur est considérée comme relativement uniforme quelle que soit la profondeur. L'alimentation de cette nappe est assurée essentiellement par l'infiltration des eaux de surface. L'écoulement souterrain s'effectue principalement de l'Est vers l'Ouest (Fig.39).

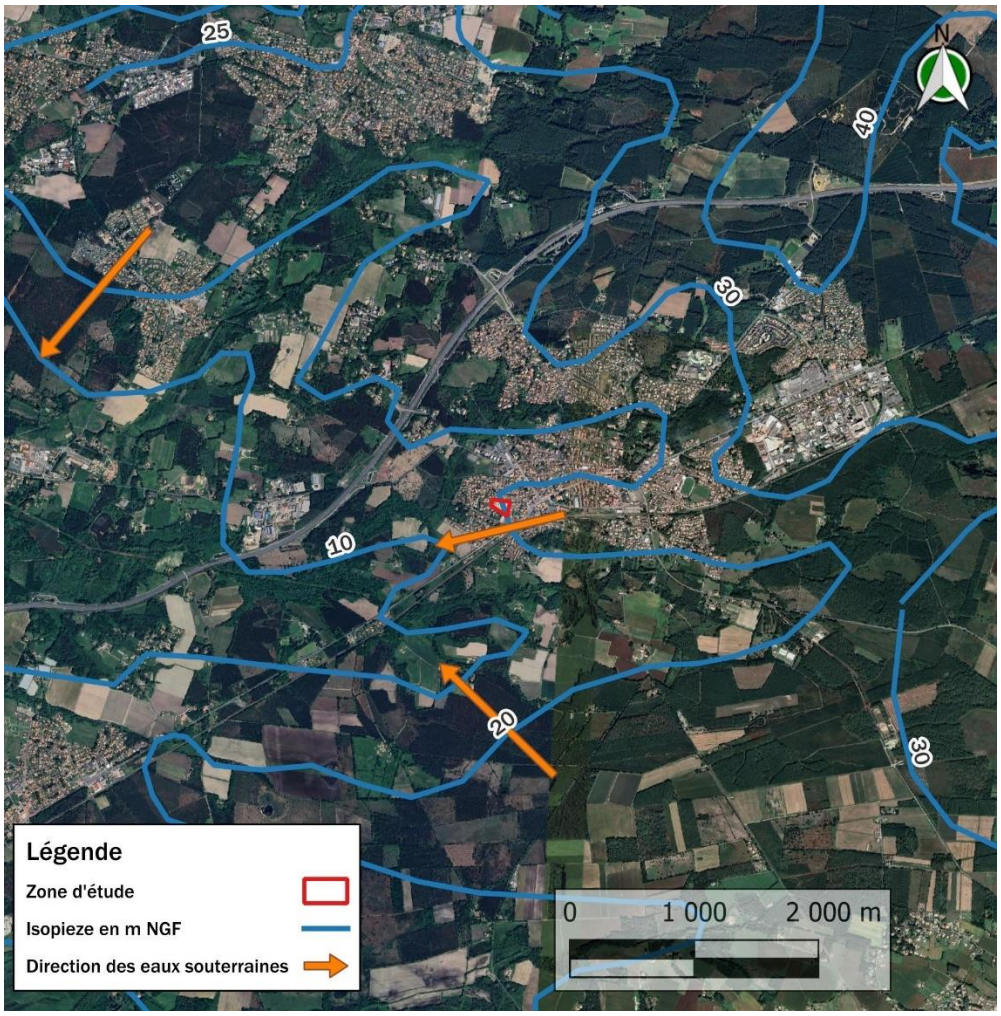


Figure 39 : Carte piézométrique (SIGES, 2025)

D'après le log géo-hydrogéologique local (Fig. 40, base LISA), l'épaisseur de l'aquifère au droit du site d'étude est estimée à 15 m.

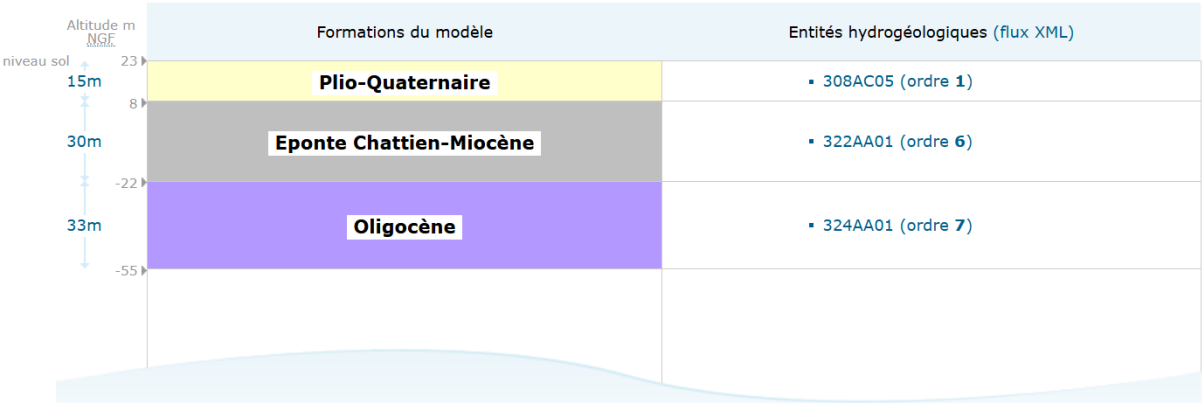


Figure 40 : Log géo-hydrogéologique (base LISA, SIGES 2025)

4.2.3. Historique du site et investigations réalisées

Le site n'est plus en activité au moment de la réalisation du diagnostic. La majorité des bâtiments et des installations industrielles ont déjà été démantelées ou détruites. Le diagnostic initial de pollution a été réalisé en deux étapes :

i) Une étude historique et documentaire. Celle-ci a mis en évidence la présence d'activités potentiellement polluantes. La présence d'une cuve de fioul de 7000 litres a notamment été avérée. La cuve avait déjà été retirée lors de la réalisation des investigations. Des doutes subsistent quant à sa nature (cuve aérienne ou semi-enterrée), ainsi que sur son emplacement exact, i.e. seule sa localisation approximative est connue. Deux sondages ont été réalisés à proximité de la localisation indiquée. D'autres sources de pollution ont également été identifiées mais ne seront pas détaillées dans ce manuscrit car elles n'ont pas fait l'objet d'une dépollution ;

ii) La réalisation des investigations. A la suite de l'étude documentaire, deux sondages de 3 m de profondeur ont été réalisés. Les lithologies observées correspondent essentiellement à des sables bruns ou beiges. Des niveaux de limons sableux et d'argiles sableuses ont également été mis en évidence. Les eaux souterraines ont été rencontrées à partir de 2 m de profondeur. Des indices organoleptiques, notamment des odeurs caractéristiques d'hydrocarbures, ont été constatés au niveau des deux

sondages. Parallèlement, des mesures réalisées à l'aide d'un détecteur à Photoionisation (PID) ont révélé des teneurs significatives de 97 et 202 ppm de composés organiques volatils (COV) au sein des horizons identifiés comme pollués. Des échantillons (1 par sondage) ont été prélevés entre 1 et 1,75 m de profondeur pour le sondage S12 et entre 1 et 2 m de profondeur pour le sondage S13. Ces échantillons ont fait l'objet d'analyses chimiques ciblant spécifiquement les hydrocarbures totaux (gamme C10-C40). Les analyses chimiques quantitatives ont révélé des concentrations en hydrocarbures totaux de 1790 $mg.kg^{-1}$ de matière sèche (MS) pour le sondage S12 et de 4860 $mg.kg^{-1}$ MS pour le sondage S13 (Fig. 41).

Au vu de ces résultats initiaux, des investigations complémentaires ont été décidées. Sept sondages additionnels ont ainsi été réalisés afin de préciser la délimitation du panache de contamination autour de la zone précédemment échantillonnée. La présence de gravats et de zones surcreusées sur le site a imposé des contraintes significatives quant au positionnement des points de sondage. Des investigations sur les eaux souterraines et sur les gaz du sol ont parallèlement été réalisées. La foration des nouveaux ouvrages s'est arrêtée à l'atteinte des horizons saturés en eau, soit à environ 2 m de profondeur. L'ensemble des sondages a fait l'objet de prélèvements entre 1 et 2 m de profondeur. La cartographie des teneurs en polluants mesurées dans les échantillons prélevés est présentée sur la Figure 41.



Figure 41 : Cartographie des teneurs en polluants pour l'horizon 1-2 m

Les investigations sur les eaux souterraines montrent un impact au droit de la zone. Une contamination des gaz du sol est également relevée (un piézair a été réalisé entre les deux sondages du diagnostic initial) avec des valeurs qui excèdent les valeurs de référence pour les hydrocarbures aromatiques C8-C12 et aliphatiques C6-C12. Le piézomètre PzB réalisé à proximité du sondage avec la teneur la plus élevée (7770 $mg.kg^{-1}$ MS) présente une teneur en hydrocarbures C10-C40 dissous de 4200 $\mu g.l^{-1}$.

4.2.4. Plan de gestion

Les résultats acquis dans le diagnostic de pollution initial et le diagnostic approfondi ont permis l'établissement d'un plan de gestion. La démarche de définition de la pollution concentrée est résumée dans la Figure 42 :

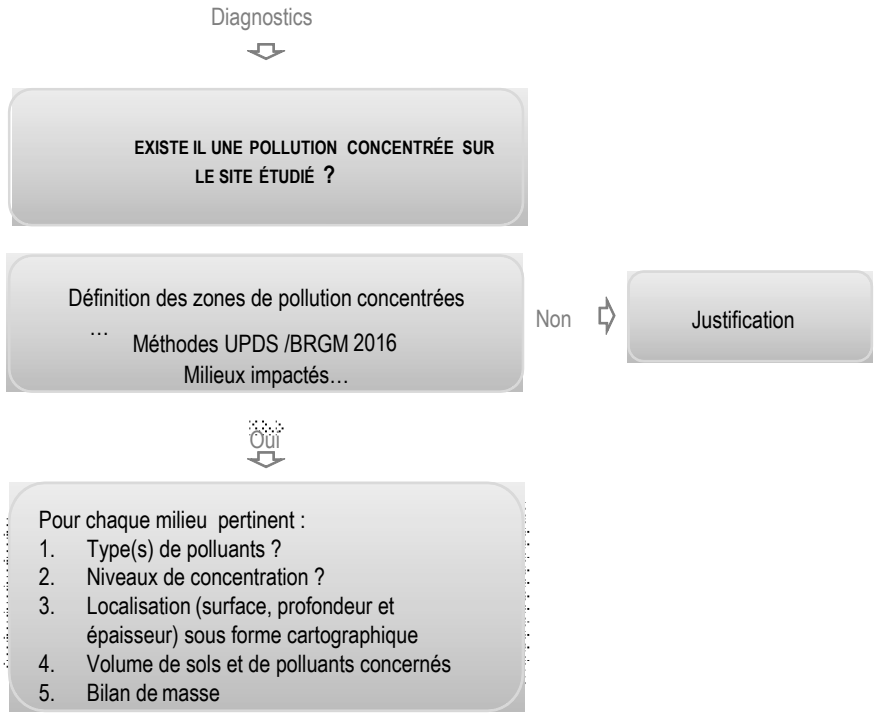


Figure 42 : Résumé de la démarche de définition de la pollution concentrée

La « boîte à outils » proposée par l'Union Pour la Dépollution des Sols (UPDS, 2016) est constituée des six méthodes d'interprétation des données de terrain et de laboratoire les plus couramment utilisées par la profession. Ces méthodes permettent de caractériser et d'identifier une pollution concentrée dans la zone non saturée ou dans la zone saturée. Les outils méthodologiques sont les suivants :

- Méthode n°1 - interprétation des constats de terrain ;
- Méthode n°2 - analyse statistique ;
- Méthode n°3 - détermination de la présence d'une phase organique ;
- Méthode n°4 - cartographie fondée sur des méthodes déterministes ;
- Méthode n°5 - bilan massique fondé sur des méthodes déterministes (voir aussi Loi Pareto Guide BRGM, Lion et al., 2016) ;
- Méthode n°6 - approche intégrée géostatistique.

Le Tableau 13 rappelle la relation entre les méthodes proposées et leurs objectifs. Les différentes méthodes ne seront pas détaillées dans le manuscrit.

Tableau 13 : Méthodes de définition d'une pollution concentrée (extrait guide UPDS)

Objectifs	Indice qualitatif d'une PC*	Définition d'un seuil de coupure	Visualisation 2D ou 3D de la PC*	Estimation du volume de sols PC*	Estimation du volume de polluant
Méthode n°1 : interprétation des constats de terrain	✓Sols				
Méthode n°2 : analyse statistique		✓Sols			
Méthode n°3 : détermination de la présence d'une phase organique					
Méthode n°4 : cartographie fondée sur des méthodes déterministes (outils logiciels…)		✓Sols	✓Sols	✓Sol	
Méthode n°5 : bilan massique fondé sur des méthodes déterministes		✓Sols		✓Sols	✓Sols
Méthode n°6 : approche intégrée géostatistique (3D/ probabilité…)					

*PC : Pollution Concentrée / Méthode n°5 : cf. Loi Pareto sol Guide BRGM 2016)

	Capacité de la méthode
✓Sols	Méthode retenue dans le plan de gestion pour cette étude

Les principales statistiques concernant les données acquises à proximité de la cuve de fioul sont résumées dans le Tableau 14. Le nombre de valeurs pour réaliser ces statistiques est faible (11 valeurs). La moyenne (1865 mg/kg MS) est très au-dessus de la médiane (108 $mg.kg^{-1}$ MS) ce qui indique une asymétrie importante des données.

Tableau 14 : Principales statistiques pour les données acquises autour de la cuve de fioul en hydrocarbures C10-C40

Paramètre	Valeur (mg/kg MS)
Concentration maximum	7770
Concentration moyenne	2419
Médiane	1003
Écart type	3057
Percentile 25	67
Percentile 80	5304
Percentile 85	6303
Percentile 90	6703
Percentile 95	7460

Le seuil de concentration de 500 $mg.kg^{-1}$ MS en hydrocarbures C10−C40 a été défini à l'issue d'une démarche multi-méthodes pour caractériser la pollution concentrée. Cette démarche a notamment inclus le bilan massique, l'interprétation des constats de terrain et des analyses statistiques (détails et cartographie non présentés dans ce document).

Deux représentations cartographiques distinctes ont été produites en appliquant des méthodes d'interpolation déterministes. Ces deux interpolations bidimensionnelles (2D) ont été réalisées à partir des données de concentration mesurées dans l'intervalle de profondeur 1–2 m. L'application de techniques géostatistiques n'a pas été retenue en raison du faible nombre d'observations, insuffisant pour une analyse robuste par ces méthodes.

La première méthode d'interpolation mise en œuvre est celle des plus proches voisins. L'application de cette technique révèle une incertitude notable en limites nord-est et sud-est du site, en raison de l'absence de points de sondage dans ce secteur. Cette incertitude est inhérente à la méthode des plus proches voisins, qui n'est pas en mesure d'extrapoler des valeurs au-delà du domaine spatial couvert par les points d'observation. Le volume de sol dont les concentrations sont supérieures à 500 $mg.kg^{-1}$ MS établi entre 1 et 2 m est de 645 m^3. Il est considéré que l'horizon 0-1 m est pollué sur la même superficie que l'horizon 1-2 m. Le volume total considéré est donc de 1290 m^3. La cartographie présentant la surface dont la concentration est supérieure au seuil considéré est présentée en Figure 43.

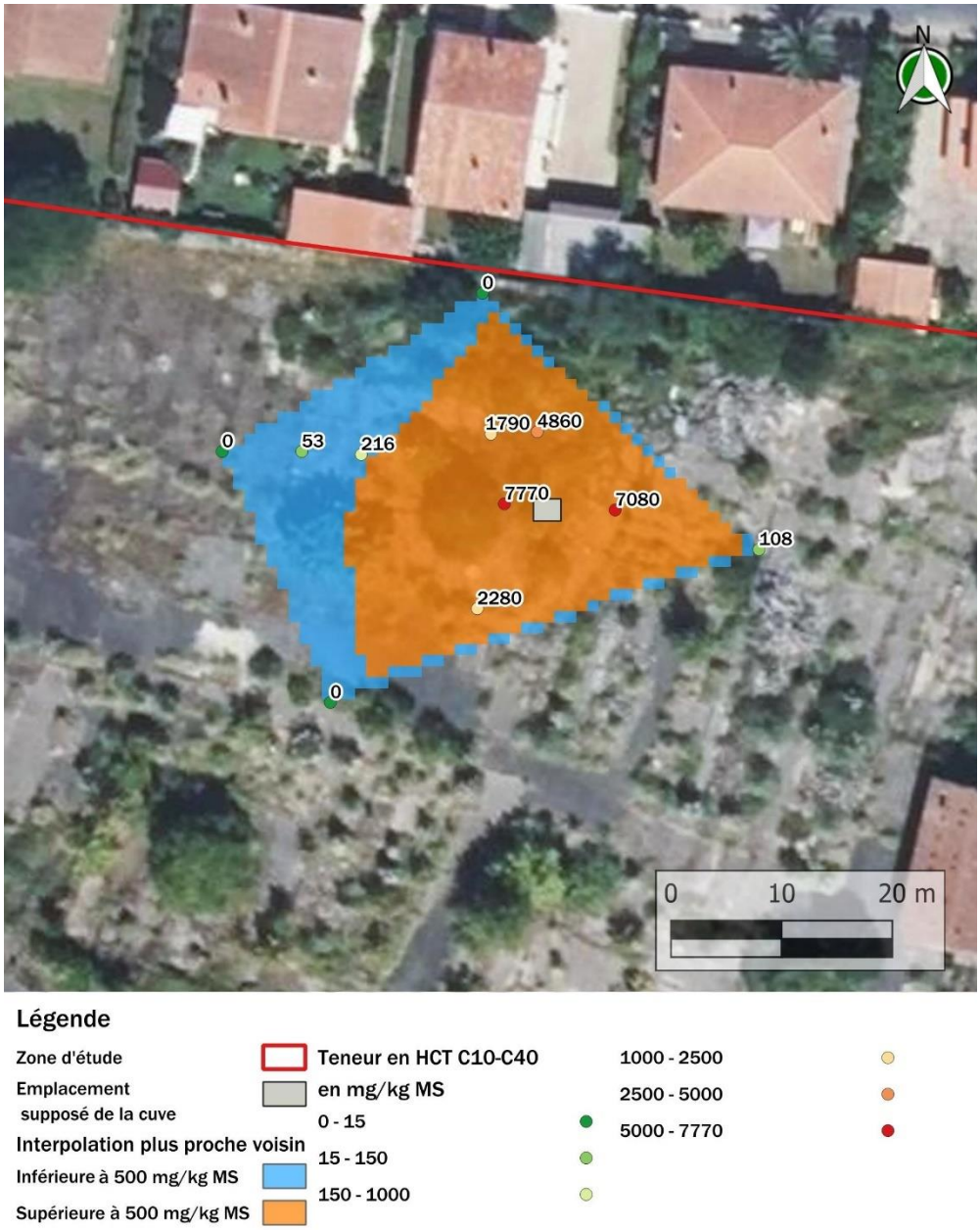


Figure 43 : Interpolation avec la méthode plus proche voisin pour l'horizon 1-2 m

La deuxième méthode utilisée est celle des polygones de Voronoï. Comme pour la méthode des plus proches voisins, l'absence de sondage dans les zones nord-est et sud-est provoque des incertitudes importantes. Il est possible d'extrapoler au-delà des points d'observation en réalisant une zone tampon, cependant cela ne permet pas de réduire les incertitudes. Le volume dont la concentration est supérieure à 500 $mg.kg^{-1}$ MS (Fig. 44) entre 1 et 2 m est de 993 m^3. Il est considéré que l'horizon 0-1 m est pollué sur la même superficie que l'horizon 1-2 m. Le volume total considéré est donc de 1986 m^3.

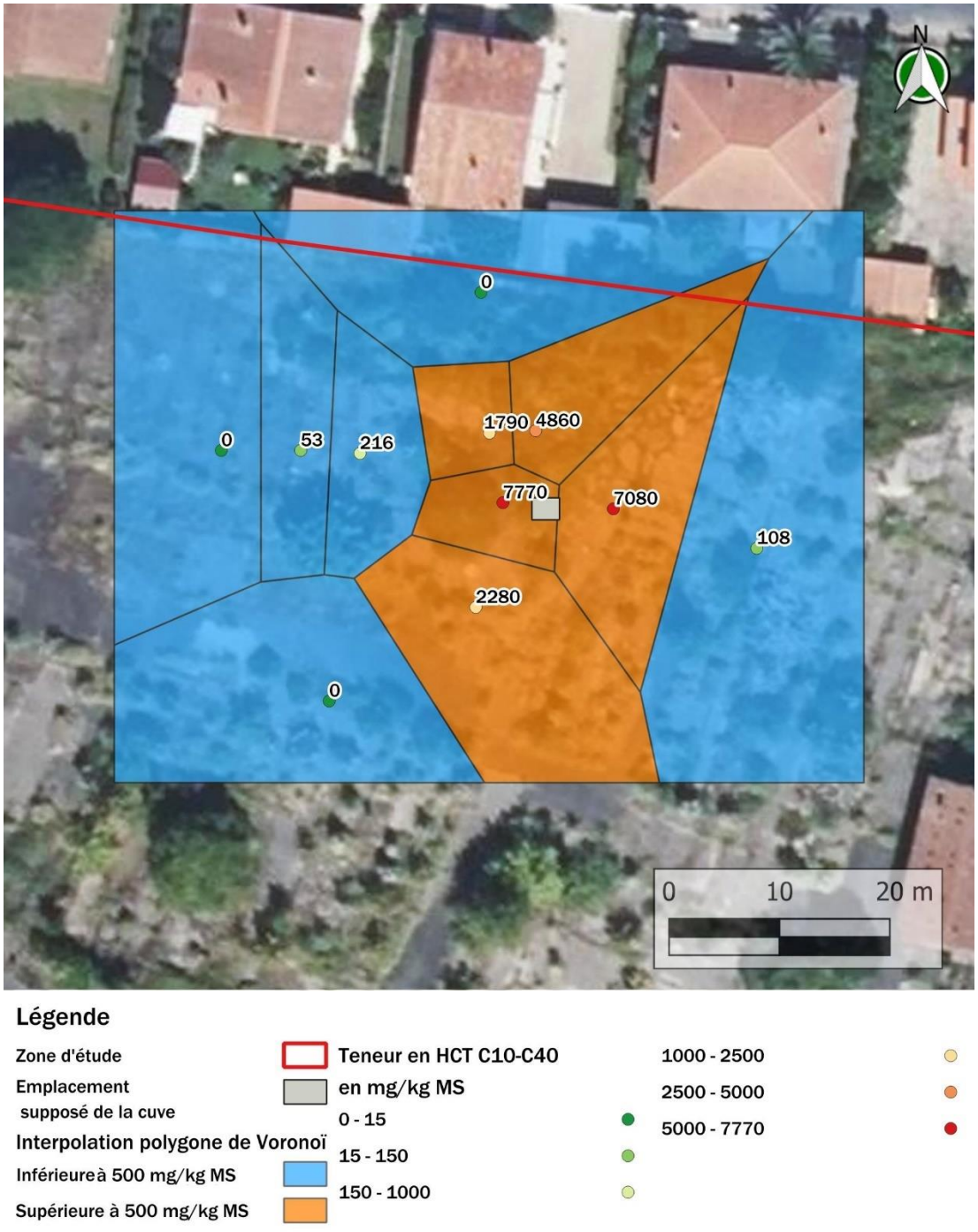


Figure 44 : Interpolation avec la méthode polygone de Voronoï pour l'horizon 1-2 m

La méthode des plus proches voisins a été appliquée pour délimiter la zone à dépolluer dans le plan de gestion. L'étude a néanmoins mis en évidence une incertitude significative dans les secteurs nord-est et sud-est, nécessitant une attention particulière. Cette méthode évalue de manière moins importante les volumes de terre à excaver, justifiant son choix. Concernant la technique de traitement, 3 scénarios de dépollution ont été proposés : i) excavation, tri sur site et traitement hors site en filière autorisée (ISDI, ISDD, ISDND ou biocentre) de la pollution concentrée, ii) excavation et traitement sur site en biopiles ventilées de la pollution concentrée.

Ce procédé consiste à regrouper les sols pollués excavés en tas et à les ventiler pour stimuler la dégradation biologique des polluants, et iii) ventilation de la zone non saturée (venting) au droit de la zone de pollution concentrée.

En considération les limitations (délai, absence d'électricité, odeurs potentielles) et les spécificités du site (polluant, géologie, hydrogéologie), la stratégie de remédiation sélectionnée est l'excavation des sols impactés, tri sur site et traitement hors site en filière autorisée.

Pour délimiter plus précisément les zones à traiter et optimiser les travaux de dépollution, une caractérisation par maillage a été réalisée en amont du démarrage des opérations. Ces travaux ont été menés par la société de dépollution mandatée pour le projet. Au total, six mailles (Fig. 45) ont fait l'objet d'une caractérisation impliquant l'utilisation d'une pelle mécanique pour l'excavation. Pour chacune de ces six mailles, la stratégie d'échantillonnage a consisté en le prélèvement de deux échantillons composites (un entre 0 et 1 m puis un entre 1 et 2 m de profondeur), obtenus par le mélange de deux prélèvements distincts effectués au sein de la maille. Les résultats de cette caractérisation sont présentés sur la Figure 45. Les résultats entre 0 et 1 m puis 1 et 2 m sont similaires concernant le dépassement du seuil de dépollution. Seules les mailles centrales ont des concentrations supérieures au seuil de dépollution de 500 $mg.kg^{-1}$ MS. Cette zone sera donc celle prise en compte pour les travaux de dépollution. Le volume total de terres supérieur à 500 $mg.kg^{-1}$ MS entre 0 et 2 m de profondeur est estimé à 730 m^3.



Figure 45 : Cartographie du maillage et des investigations complémentaires de sol réalisées pour la caractérisation affinée de la zone source concentrée par la société de dépollution

4.2.5. Application de la méthode analytique inverse

L'intégration de la modélisation analytique inverse pourrait être envisagée dès la phase d'élaboration du plan de gestion d'un site (ce n'est pas le cas pour ce site car la méthode IAM a été développée après). Cette approche méthodologique servirait de méthode de comparaison ou de validation aux six méthodes d'estimation des volumes existantes (telles que les cartographies

d'interpolation). L'IAM contribuerait ainsi à affiner la détermination du seuil de pollution concentrée avant même le début des travaux.

Les sections antérieures présentaient des approches simplifiées. Pour cette application, tous les processus physiques, chimiques et biologiques de la solution analytique (définis par les formules 2.6 à 2.14) sont intégralement pris en compte dans la modélisation. Les triangles dont la somme des concentrations associées aux trois sommets était inférieure à 500 $mg.kg^{-1}$ MS n'ont pas été inversés. Cette exclusion se justifie par le fait que les modélisations issues de ces zones à faibles concentrations ne contribueraient que de manière négligeable à la délimitation du panache de contamination principal. Le maillage utilisé est ainsi constitué de 12 triangles (Fig.46). Les inversions ont été réalisées sur une base bidimensionnelle (plan par plan), en utilisant les données de concentration des horizons situés entre 0 et 1 m, et entre 1 et 2 m de profondeur. Dans les cas où des données étaient absentes pour un horizon spécifique (principalement pour l'intervalle 0-1 m), la valeur de l'horizon le plus proche échantillonné a été utilisée par substitution. Les profondeurs assignées aux échantillons prélevés sur un intervalle vertical correspondent au point médian de cet intervalle (par exemple, 0,5 m pour un échantillon prélevé entre 0 et 1 m).



Figure 46 : Cartographie représentant les triangles de Delaunay pour le cas réel

De manière similaire à l'approche adoptée pour les cas synthétiques, une analyse de sensibilité a été conduite sur l'ensemble des paramètres de la solution analytique. Les résultats de cette analyse sont présentés et illustrés sur la Figure 47.

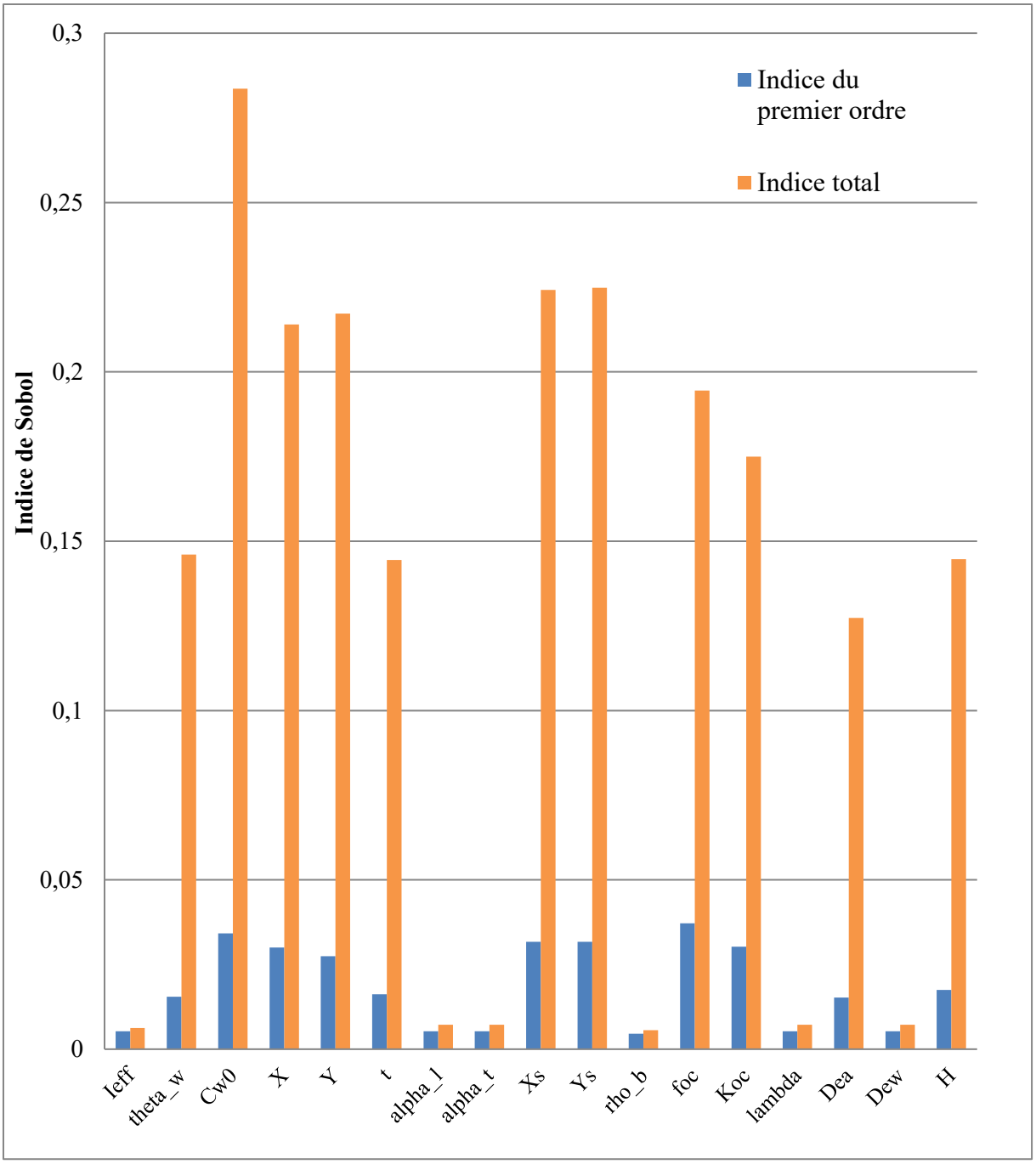


Figure 47 : Indice de Sobol avec l'ensemble des paramètres pour la solution analytique de Troldbord

L'analyse révèle, comme pour la solution analytique sans la phase réactive, que les paramètres exerçant l'influence la plus significative sur les prédictions du modèle sont ceux définissant la source de contamination : la concentration initiale (C_{w0}), sa localisation (X_s et Y_s) et ses dimensions (X et Y). D'autres paramètres exercent une influence moindre comme la teneur en carbone total (*foc*), le coefficient de partage carbone organique/eau (*Koc*), le temps (t), la teneur en eau volumique (θ_w), la constante de Henry (H) et le coefficient de diffusion

dans la phase gazeuse (D^e_a). Ces paramètres présentent les indices de sensibilité d'ordre total les plus élevés.

L'analyse de sensibilité a identifié plusieurs paramètres exerçant une influence mineure sur la sortie du modèle. Ceux-ci incluent :

- L'infiltration effective (I_{eff}) ;
- Les coefficients de dispersion longitudinale (α_L) et transversale (α_T) ;
- Le coefficient de diffusion dans la phase liquide (D^e_w) ;
- La constante de dégradation de première ordre (λ).

Pour calibrer le modèle analytique inverse, les paramètres ont été tirés au hasard à partir des plages de valeurs minimales et maximales prédéfinies (Tab.15). Ces plages ont été choisies soit à partir de gammes de valeurs documentées dans la littérature (pour I_{eff} et θ_w), soit déterminées empiriquement (pour C_{w0}, t, α_L et α_T). Concernant les paramètres liés à la source (X, Y, X_s et Y_s), les concentrations aux points de mesure peuvent fournir une première indication de la taille maximale de la zone (aucune zone source au-delà des points où la concentration est proche de zéro). Les hydrocarbures C10-C40 sont constitués de diverses fractions aliphatiques et aromatiques, chacune possédant des caractéristiques physico-chimiques distinctes. Afin de permettre une inversion des paramètres d'écrivant ces hydrocarbures, nous avons sélectionné une gamme de valeurs représentatives des fractions majoritaires (mesurées en laboratoire) par l'analyse TPH (Hydrocarbures Totaux Pétroliers) de cette famille de composés.

Tableau 15 : Plages de paramètres possibles pour l'étalonnage

Paramètre	Unité	Valeur minimale	Valeur maximale	Source
I_{eff}	$m.an^{-1}$	0,09	0,25	Météo France, 2025
θ_w	/	0,05	0,5	Valeur empirique
C_{w0}	$\mu g.l^{-1}$	1,000	90,000,000	Valeur empirique
X	M	1	10	Longueur possible à partir des données disponibles

Paramètre	Unité	Valeur minimale	Valeur maximale	Source
Y	M	1	10	Largeur possible à partir des données disponibles
t	Année	1	70	Temps maximum depuis le début de l'activité
α_l	M	0,01	10	Valeur empirique
α_t	M	0,001	1	Valeur empirique
X_s	M	-7	7	Valeur en fonction des observations déjà disponibles
Y_s	M	-7	7	Valeur en fonction des observations déjà disponibles
ρb	$kg.m^{-3}$	1500	2500	Gamme de masse volumique classiquement mesurée dans les sols Das et Sobhan, 2006.
foc	/	0,001	0,03	US EPA, 1996
Koc	$m^3.kg^{-1}$	100	5000	Gustafson et al, 1997
λ	$Année^{-1}$	0,03	3,65	Gustafson et al, 1997
D^e_a	$m^2.an^{-1}$	31	1575	Gustafson et al, 1997
D^e_w	$m^2.an^{-1}$	0,003	0,29	Gustafson et al, 1997
H	/	0,000082	160	Gustafson et al, 1997

L'ensemble des triangles ont été inversés et ont atteint l'objectif fixé à 1% de la valeur maximale.

Comme pour le cas synthétique, une synthèse statistique des paramètres estimés a été effectuée. Les résultats sont présentés dans le Tableau (Tab.16). Les valeurs obtenues sont difficilement comparables car hormis la fraction organique, aucun de ces paramètres n'a été mesuré. Les valeurs pour la fraction organique sont cohérentes avec celle mesurée en laboratoire (0,0017).

Tableau 16 : Statistique concernant les paramètres inversés pour le cas réel

	ieff	C_{w0}	X	Y	t	α_l	α_t	X_s	Y_s	ρ_b	foc	Koc	θ_w	λ	d_{ea}	d_{ew}	H
Unité	m.an-1	µg.l^{-1}	m	m	an	m	m	m	m	kg.m^{-3}	/	m^3.kg^{-1}	/	an^{-1}	m^2.an^{-1}	m^2.an^{-1}	/
Moyenne	0,17	3234659	5,3	5,2	42,6	5,0	0,49	0,09	0,31	1994,8	0,010	1752	0,22	1,80	890,4	0,14	79,5
Percentile 25	0,13	106541	3,0	3,3	30,3	2,7	0,21	-2,67	-3,75	1764,3	0,005	648	0,12	0,80	624,4	0,07	29,9
Médiane	0,17	305545	5,1	5,0	41,4	5,3	0,52	0,89	0,45	2020,0	0,009	1599	0,21	1,70	893,8	0,14	71,3
Percentile 75	0,21	2477929	7,5	7,0	59,8	7,6	0,77	1,98	3,99	2190,9	0,016	2689	0,31	2,73	1221,7	0,22	130,6

Les estimations basse et haute du volume de sol présentant des concentrations supérieures à 500 mg.kg^{-1} MS entre 0 et 2 m de profondeur sont respectivement de 939 m³ et 1218 m³. La valeur médiane de ce volume est estimée à 1078 m³.

Les Figures 48 et 49 fournissent une représentation graphique 2D ainsi qu'une visualisation en 3 plans de la pollution estimée.

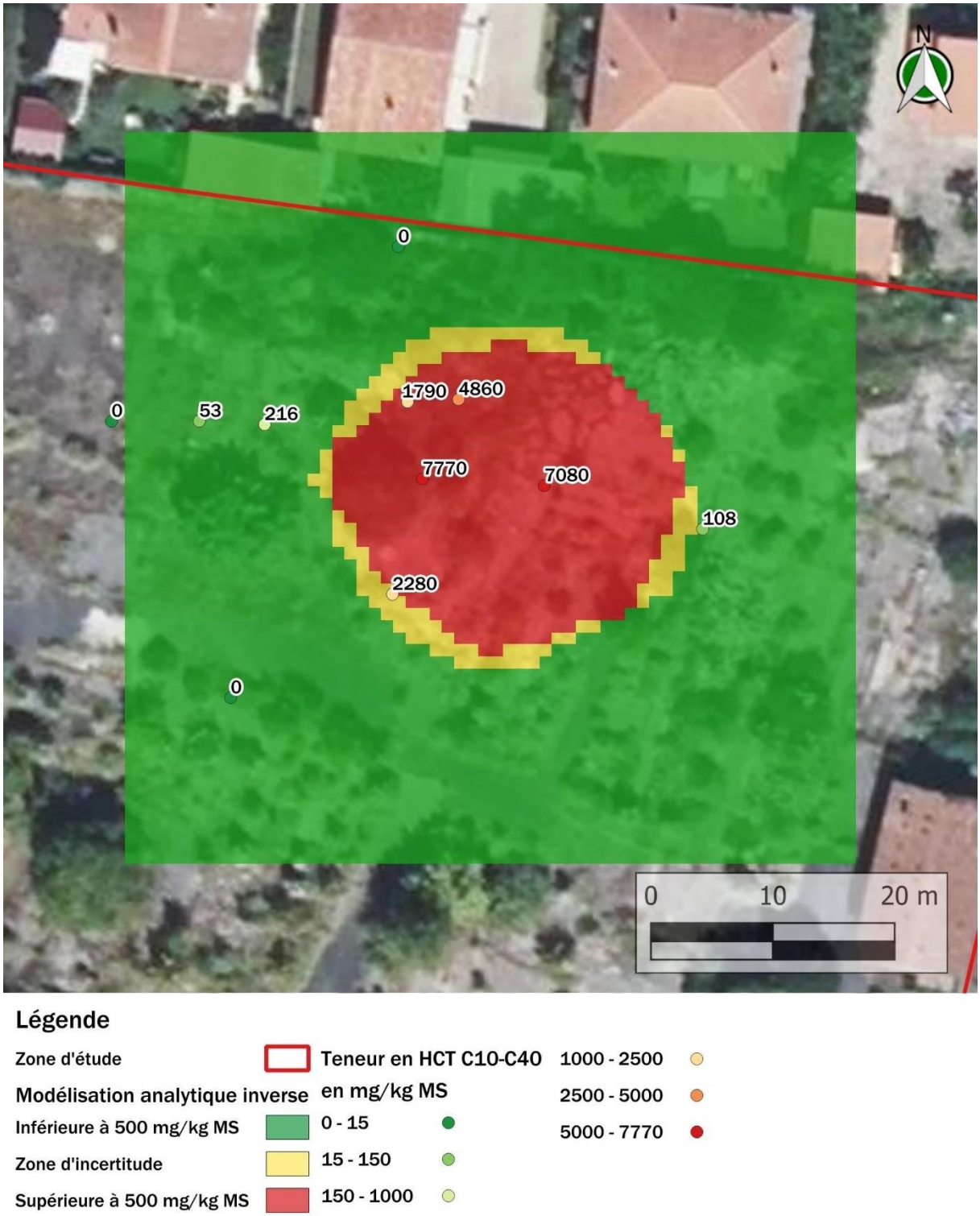


Figure 48 : Vue 2D à 1,5 m de profondeur de l'estimation des concentrations en hydrocarbures C10-C40 avec un seuil de pollution à 500 mg.kg^{-1} MS avec l'IAM

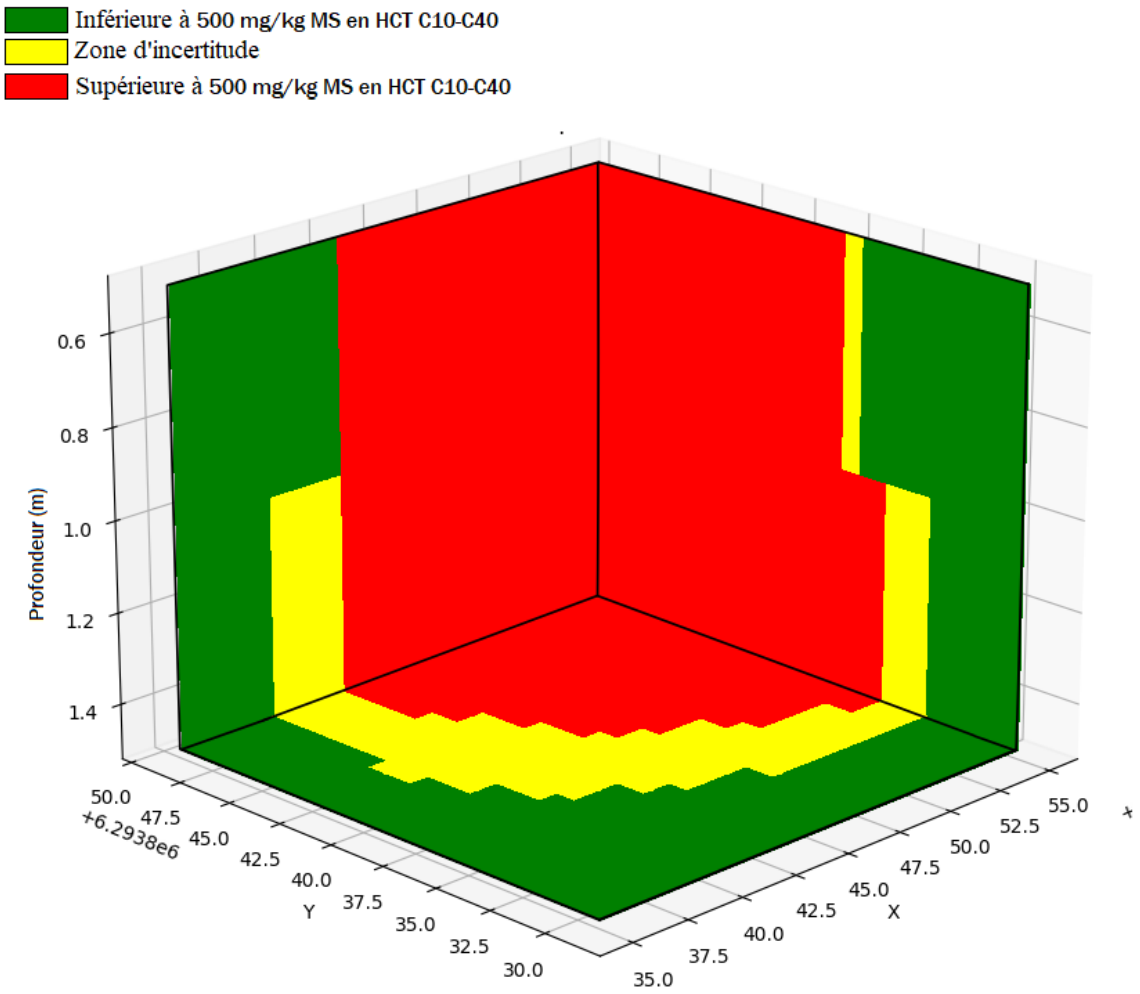


Figure 49 : Vue en 3D de l'estimation des concentrations en hydrocarbures C10-C40 avec un seuil de pollution à 500 mg.kg^{-1} MS avec la modélisation analytique inverse

4.2.6. Comparaison des résultats

À la suite d'une première phase de travaux de dépollution par excavation, une campagne de prélèvements de contrôle a été menée. Les campagnes de prélèvements de contrôle sont menées en fonction de plusieurs contraintes opérationnelles, notamment l'accessibilité du site, le matériel de prélèvement disponible et la nature du polluant ciblé. Ces prélèvements sont souvent réalisés à l'aide d'une pelle mécanique et visent spécifiquement à collecter des échantillons sur les parois de la fouille afin de vérifier la conformité des concentrations dans les sols aux seuils de dépollution définis avant les travaux. Deux stratégies peuvent être adoptées :

- Prélèvements ponctuels à localisation précise : cette méthode est privilégiée lorsque les sols présentent une hétérogénéité marquée ou lorsqu'une zone spécifique a été la cible privilégiée de l'opération de dépollution ;
- Prélèvements et constitution d'échantillons composites : cette méthode permet d'obtenir une caractérisation plus représentative d'une zone étendue. Plusieurs échantillons sont collectés sur une même paroi ou un même horizon, puis homogénéisés pour former un échantillon composite.

Le nombre et la localisation des échantillons sont déterminés en fonction de critères tels que la présence d'indices de contamination visuels sur le terrain, l'hétérogénéité ou l'homogénéité du sol, l'épaisseur de l'horizon potentiellement contaminé, ainsi que les exigences spécifiées dans le cahier des charges ou formulées par l'administration compétente.

Dans le cas présent, les prélèvements ont été effectués à l'aide d'une pelle mécanique et la stratégie de constitution d'échantillons composites a été mise en œuvre. Cette campagne a consisté en la collecte de 6 échantillons composites sur les parois de la fouille et de 2 échantillons en fond de fouille. Les analyses chimiques de ces échantillons ont révélé des dépassements des concentrations dans les sols, parfois importants, des objectifs de dépollution fixés pour les parois nord, ouest et est de l'excavation. En revanche, les deux prélèvements effectués sur la paroi sud se sont avérés conformes aux objectifs.

Compte tenu de ces résultats, une campagne de terrassement complémentaire a été programmée, ciblant spécifiquement les terres au nord et à l'est de la fouille initiale. La reprise de la paroi ouest n'a pas été effectuée en raison de contraintes techniques. Le volume de terres excavées lors de cette phase complémentaire est estimé à environ 200 m³ supplémentaires. Des échantillons composites complémentaires ont ensuite été prélevés sur les nouvelles parois de la fouille (deux prélèvements au nord et deux à l'est). Les résultats d'analyse de ces quatre échantillons ont de nouveau indiqué des dépassements du seuil de dépollution.

Une troisième campagne de terrassement a donc été entreprise, durant laquelle 100 m³ supplémentaires de terres ont été excavés. Les analyses réalisées sur les échantillons prélevés après cette nouvelle phase ont encore montré des dépassements des objectifs au niveau des limites ouest et nord de l'excavation. Face à des contraintes techniques, il a été décidé d'arrêter les travaux d'excavation. Au total, un volume d'environ 1040 m³ de terres a été terrassé au cours de l'ensemble des opérations.

La Figure 50 illustre l'emprise finale de la fouille, les zones de reprises successives, ainsi que les résultats des analyses effectuées sur les échantillons prélevés en parois.



Figure 50 : Cartographie de la zone terrassée et des résultats en bords de fouille pour le cas réel.

Quatre méthodes d'estimation des volumes ont été comparées au volume final de l'excavation qui, malgré les incertitudes discutées ci-dessus, sera considéré comme représentatif du volume de sol dont la concentration excède 500 mg.kg^{-1} MS : i) deux approches

issues du plan de gestion, ii) une interpolation RBF 3D multiquadratique (cf. chapitres 3.4) et iii) la modélisation analytique inverse. Certaines méthodes d'estimation étant limitées à l'intervalle de profondeur de 1 à 2 m (méthodes des plus proches voisins et polygones de Voronoï sont en 2D), la comparaison sera spécifiquement focalisée sur cet intervalle (en considérant la surface à 1,5 m de profondeur). Le pourcentage d'erreur est calculé en divisant la somme des surfaces faux positifs et faux négatifs à la surface réellement dépolluée.

L'ensemble des surfaces faussement positives et faussement négatives par rapport à la fouille terrassée est répertorié dans le Tableau 17. Ce tableau révèle un pourcentage d'erreur plus faible pour la modélisation analytique inverse (34%) que les autres méthodes. La méthode se classant en deuxième position est celle du plus proche voisin (48,1 %). Cependant, cette interpolation est limitée uniquement entre les points de prélèvements, l'erreur est donc minorée tout comme pour la méthode des polygones de Voronoï. A titre de comparaison, le maillage réalisé avant la dépollution à partir des données déjà disponibles a montré une erreur de 34%, ce qui est similaire à la modélisation analytique inverse avec des données supplémentaires.

Tableau 17 : Plages de paramètres possibles pour l'étalonnage

Méthode	Surface totale (m²)	Surface faux positif (m²)	Surface faux négatif (m²)	% erreur	Observations
Plus proche voisin	640	212	106	49,6%	La méthode n'a pas permis d'estimation sur la zone nord-est.
Polygones de Voronoï	993	487	30	99,8%	L'interpolation est limitée aux bordures de site dans le plan de gestion.
Radial basis function	1776	1256	0	241%	/
Modélisation analytique inverse	600	121	62	34%	/

La Figure 51 illustre la cartographie des différentes surfaces dont les concentrations sont estimées supérieures à 500mg.kg^{-1} MS pour l'ensemble des méthodes testées.

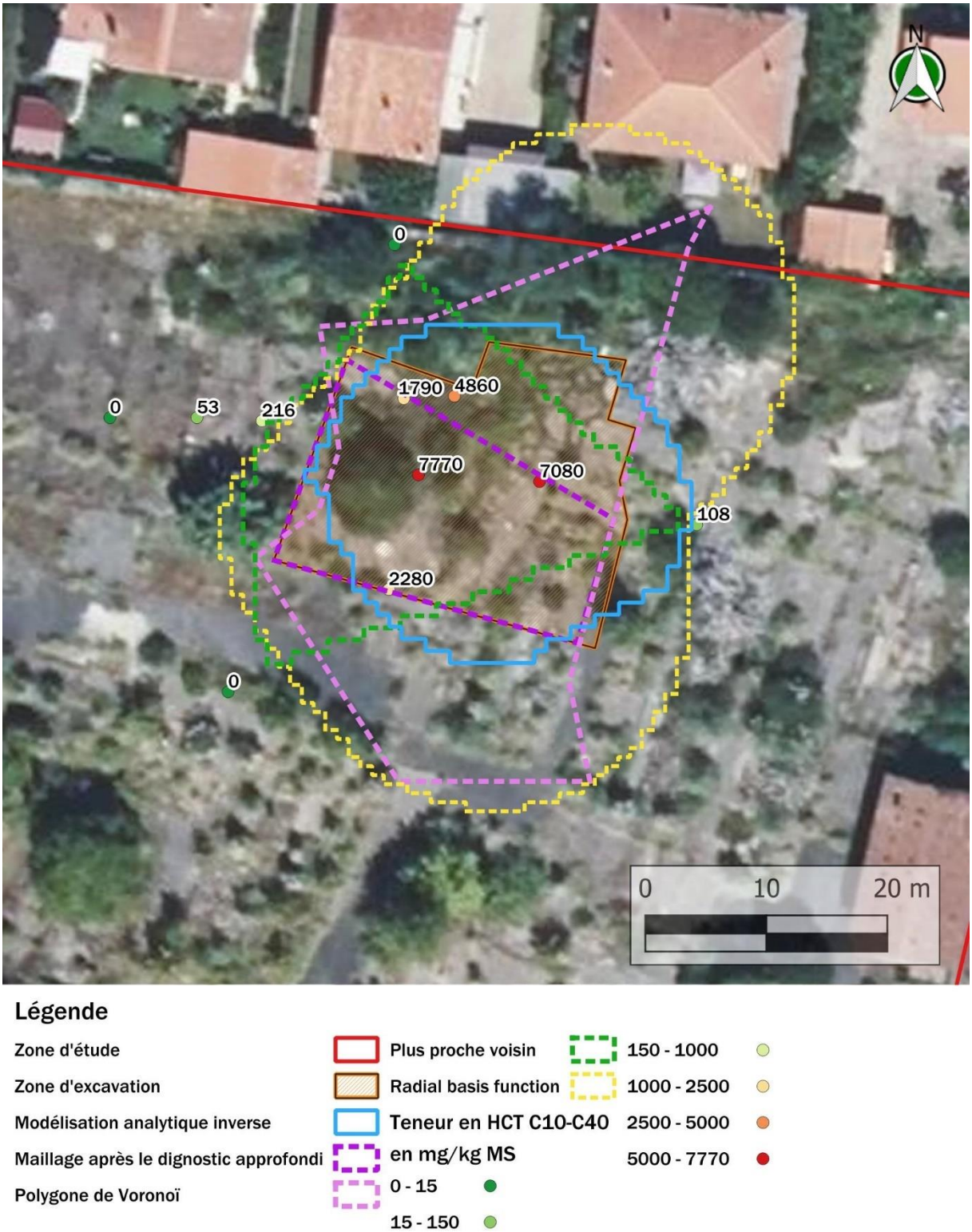


Figure 51 : Cartographie des surfaces polluées obtenues par différentes méthodes d'estimation réalisée à 1,5 m de profondeur

4.3. APPLICATION DE L'IAM AUX AUTRES CAS DE LA BASE DE DONNEES

La modélisation analytique inverse est dans cette section appliquée à l'ensemble des cas de la base de données présentée en section 4.1.

4.3.1. Comparaison des bords de fouille estimés pollués

La modélisation analytique inverse a été appliquée de manière à pouvoir être comparée aux prélèvements en bords de fouille. Les profondeurs d'application de la méthode analytique inverse ont été définies selon les critères suivants :

1. Profondeur spécifique : la priorité a été donnée aux profondeurs explicitement mentionnées dans les documents de référence fournis (avec des prélèvements composites ou ponctuels), tels que les rapports de fin de travaux ou les comptes rendus d'intervention ;
2. Profondeur de l'horizon pollué : en l'absence d'une profondeur spécifique, le milieu de l'horizon identifié comme contaminé a été ciblé ;
3. Profondeur médiane de la paroi : lorsqu'aucune information de profondeur n'était disponible, la profondeur médiane de la paroi de fouille a été retenue pour l'application de la modélisation analytique inverse.

Les résultats obtenus à partir de la méthode IAM correspondent à une zone définie comme polluée, délimitée par la borne maximum (3.2.3), et à une zone d'incertitude située entre les bornes minimum et maximum. Les statistiques présentées intègrent à la fois la zone d'incertitude et la zone polluée définie. Le bord de fouille est considéré comme pollué si l'estimation IAM indique qu'au moins 50 % de ce bord dépasse le seuil de pollution.

L'analyse statistique du nombre de triangles de Delaunay par site révèle une moyenne de 10 triangles, avec un écart-type de 6,4 et une médiane de 8,5. La gamme observée s'étend de 2 à 34 triangles, ce qui témoigne d'une variabilité significative dans la complexité spatiale des maillages générés pour les différents sites.

Des gammes de paramètres spécifiques à chaque cas ont été utilisées pour l'inversion, en tenant compte du polluant considéré, de la configuration du site et de la distribution spatiale des sondages d'investigation.

La modélisation analytique inverse présente un taux de cohérence de 80,4 % pour l'ensemble des prélèvements en bords de fouille (évaluation pour chaque bord de fouille). Ce

taux monte à 82,4 % lorsque la cohérence est évaluée par site (où chaque site, et non chaque prélèvement, est comptabilisé de manière égale). L'analyse de la conformité aux seuils de dépollution révèle que la méthode IAM a sous-estimé la contamination dans 6 des 41 bords de fouilles non conformes, ce qui représente un taux d'erreur de 14,6%. Inversement, la méthode a surestimé la contamination à 32 reprises parmi les 153 prélèvements réalisés en bords de fouille (20,9 % d'erreur). Les résultats, évalués site par site, montrent que le taux de cohérence varie entre 50 % (3 sites, soit 9,3 % des sites) et 100 % (12 sites, soit 37,5 %).

Les zones d'incertitude comprennent 20 bords de fouille estimés (soit 10,3 % du total des bords de fouille). Une analyse de ces cas révèle que 11 des 20 bords compris dans cette fourchette ne sont pas conformes aux résultats sur le terrain (c'est-à-dire qu'il s'agit de prédictions incorrectes). Ces 11 erreurs représentent 28,9 % du total des erreurs observées (11 sur un total de 38 erreurs), ce qui souligne fortement le potentiel d'amélioration réalisable grâce à une gestion optimisée ou à une réévaluation de la zone d'incertitude.

Les statistiques présentées dans le Tableau 18 résument la performance de la méthode en fonction des différentes lithologies rencontrées. Le nombre de sites par catégorie lithologique est limité, ce qui restreint la robustesse statistique de ces observations. Néanmoins, les résultats indiquent que les lithologies argileuses sont associées aux correspondances les plus faibles, avec un taux de 74,3%. Les limons affichent les taux de correspondance les plus élevés, atteignant 90%. À l'exception des lithologies argileuses, toutes les autres catégories lithologiques présentent un pourcentage de conformité supérieur à 80%. Ces tendances suggèrent une influence potentielle de la nature géologique sur la performance de la modélisation, avec des milieux plus fins (argiles) apparaissant plus complexes à caractériser avec la modélisation analytique inverse. Des investigations supplémentaires sur un échantillon plus important de sites seraient nécessaires pour confirmer ces observations et établir des corrélations.

Tableau 18 : Pourcentage de conformité pour la modélisation analytique inverse par lithologie

	Argile	Limon	Sables	Remblais	Autres
Nombre de sites	6	3	12	8	3
% Conformité Modélisation analytique inverse	74,3%	90,0%	82,0%	84,6%	80,6%

Le Tableau 19 montre qu'il n'existe pas de corrélation significative entre la densité de sondage et l'efficacité de la méthode. Contre-intuitivement, les sites caractérisés par un maillage d'échantillonnage plus resserré n'ont pas démontré de meilleures performances ; en effet, ils ont même présenté des résultats de conformité parfois inférieurs comparativement aux sites où le maillage était moins dense. Cette non-correspondance est à mettre en relation avec le nombre de triangles de Delaunay qui n’a donc pas d’influence sur les résultats de la méthode.

Tableau 19 : Pourcentage de conformité pour la modélisation analytique inverse réalisés et les seuils de dépollution pour la base de données

Gamme de m²/sondage/fouille	**Nombre de sites**	% Conformité Modélisation analytique inverse
5-15	6	78%
15-30	7	86%
30-50	8	74%
50-70	6	87%
70 et +	5	89%

Quelques exemples des résultats d'application sont présentés dans les Figures suivantes (Figure 52 à 54). L'ensemble des cartographies est présenté en annexe 1.

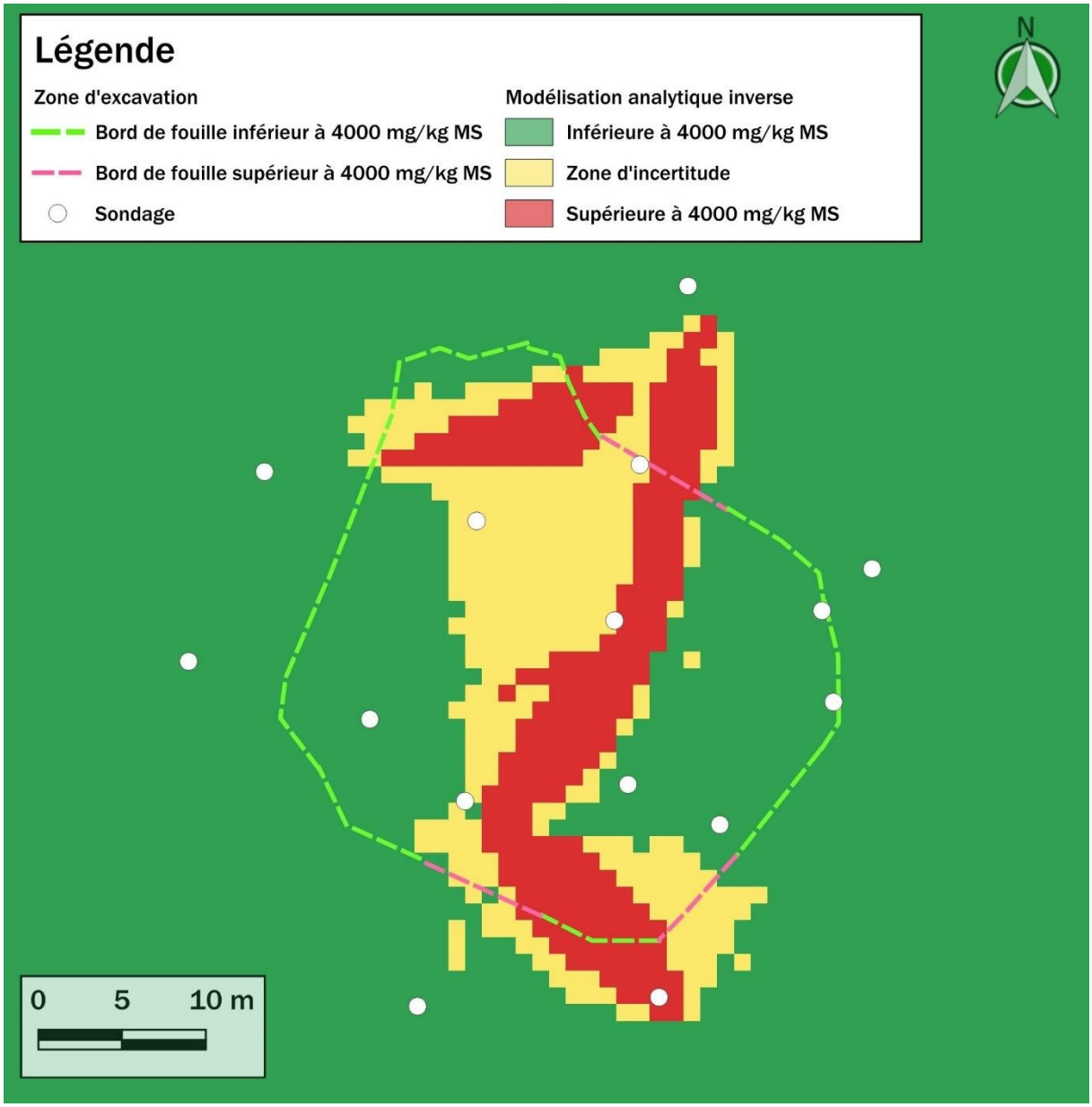


Figure 52 : Exemple n°1/3 de résultats pour la modélisation analytique inverse sur des cas réels

La Figure 52 illustre un exemple où l'opération de dépollution visait plusieurs sources de contamination au sein d'une même excavation. Les analyses complémentaires des terres excavées ont révélé qu'une proportion significative des lots (terres excavées regroupées par volumes de 50 m^3 ayant fait l'objet d'analyses) avait une concentration en deçà du seuil de dépollution, indiquant ainsi une surestimation initiale de la zone à excaver. Cependant, malgré cette surestimation volumétrique, des concentrations supérieures au seuil de dépollution ont été identifiées sur trois des bords de fouille (un au nord et deux au sud). Ce constat souligne l'importance d'une délimitation précise pour éviter à la fois un sur-traitement coûteux et la persistance de poches de contamination non identifiées par les méthodes conventionnelles.

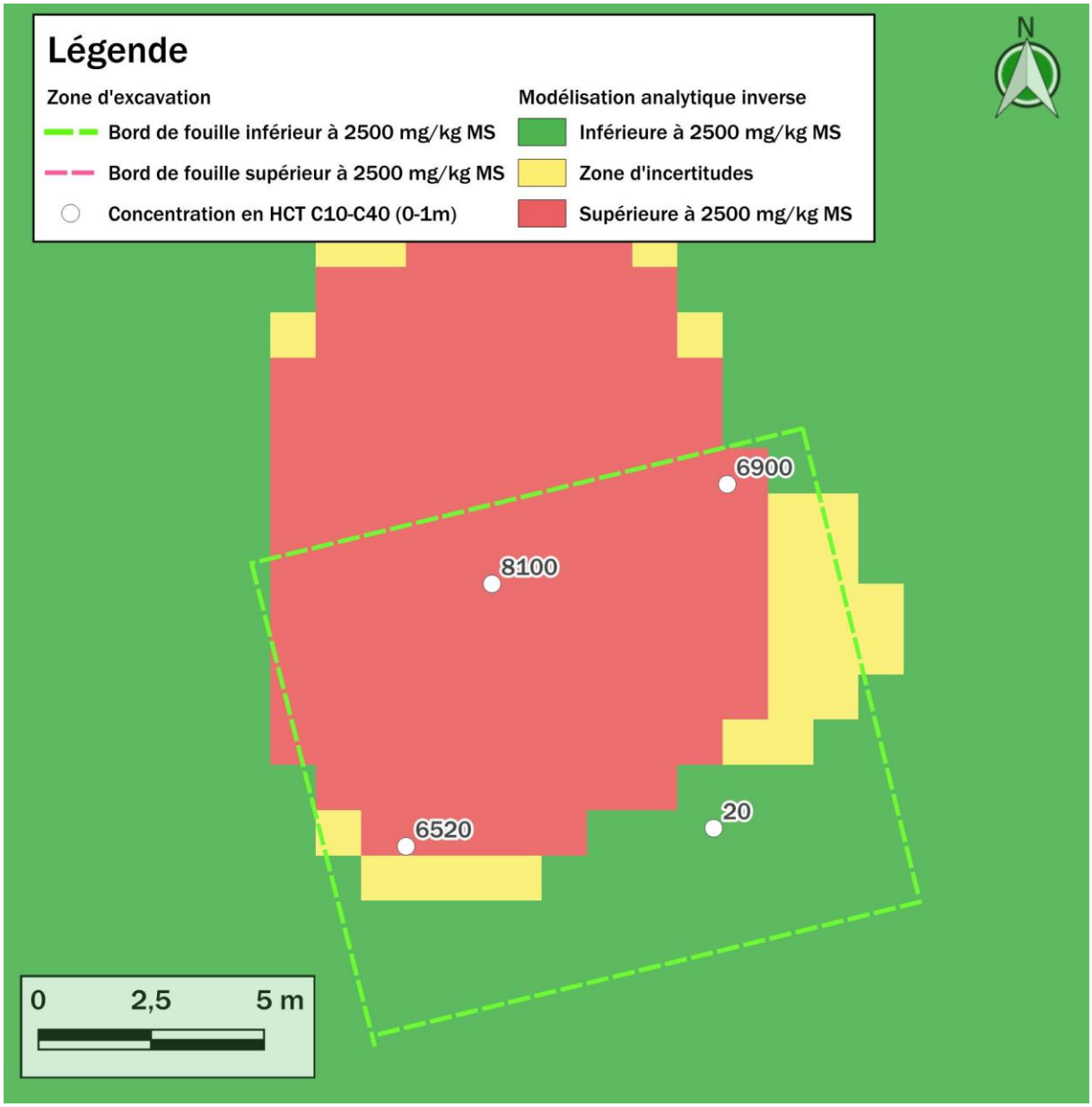


Figure 53 : Exemple n°(2/3) de résultats pour la modélisation analytique inverse sur des cas réels

La Figure 53 présente un exemple où la modélisation analytique inverse a conduit à une surestimation des concentrations de polluants sur deux bords de fouille. Au centre de la zone indiquée en rouge sur la Figure 53, une concentration de 8100 $mg.kg^{-1}$ MS en hydrocarbures C10-C40 a été analysée. Ce point de sondage est situé à seulement 1,5 mètres de la paroi de fouille, et le seuil de dépollution applicable est de 2500 $mg.kg^{-1}$ MS.

Il est très probable que les méthodes d'interpolation classiques auraient également surestimé les concentrations dans un tel scénario, compte tenu de la forte valeur du point

d'échantillonnage proche de la limite d'excavation et du manque d'informations détaillées sur le gradient de concentration.

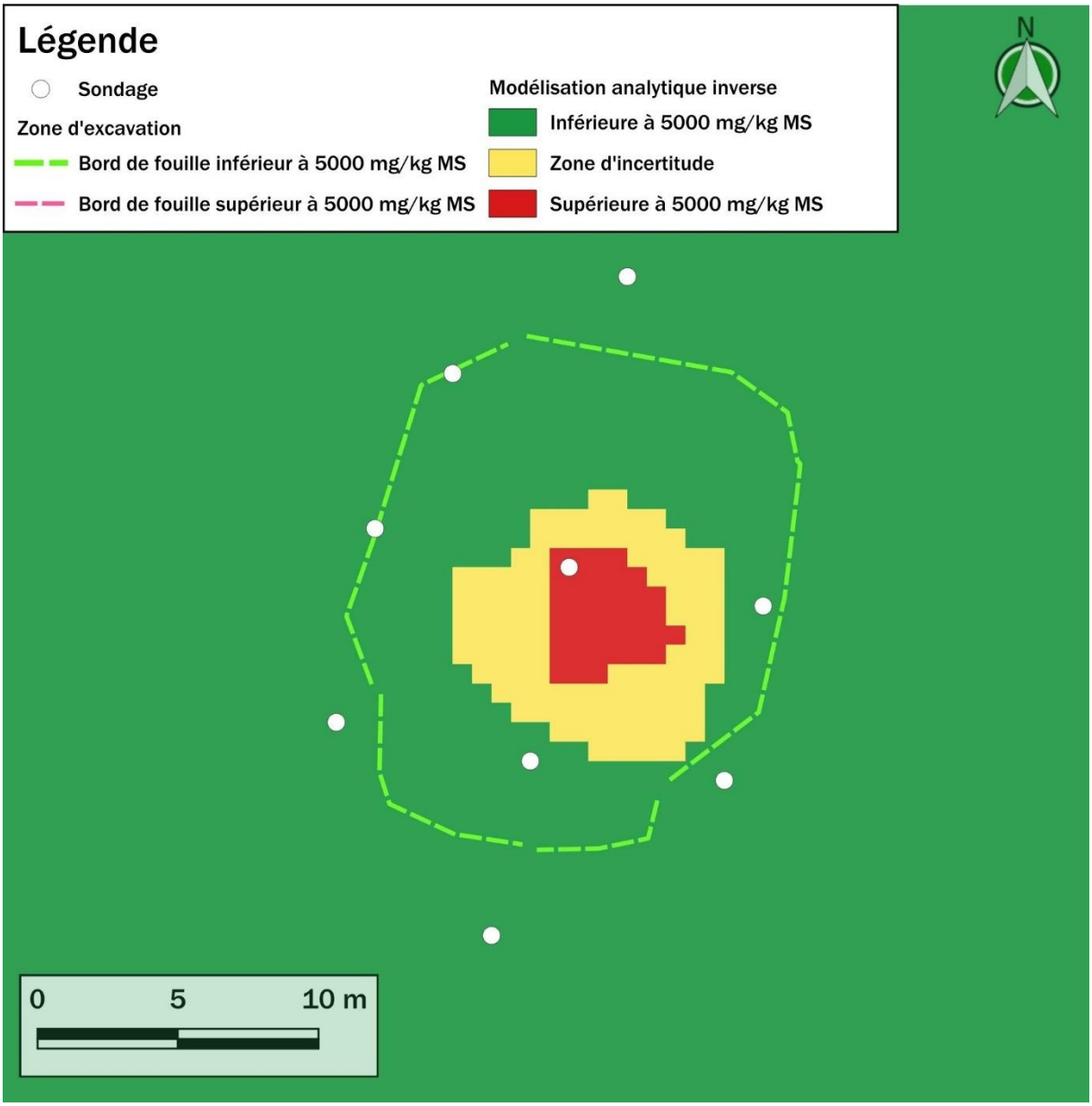


Figure 54 : Exemple n°(3/3) de résultats pour la modélisation analytique inverse (3/3) sur des cas réels

L'analyse des données post-dépollution d'un autre exemple issu de la base de données, illustrée en Figure 54, révèle une potentielle surestimation du volume de sols à traiter. Les travaux ont couvert une surface de 225 m², tandis que la modélisation analytique inverse prévoyait une zone contaminée de seulement 75 m². Les prélèvements de contrôle effectués après intervention en bords de fouille attestent de concentrations résiduelles significativement inférieures au seuil de dépollution défini confirmant l'efficacité de la dépollution mais interrogeant l'optimisation des volumes excavés.

4.4. APPLICATION DE LA MODELISATION ANALYTIQUE INVERSE A UN PANACHE DE POLLUTION DES EAUX SOUTERRAINES

L'objectif de cette section est de démontrer la polyvalence et la robustesse de la méthode IAM en l'appliquant à la zone saturée pour la caractérisation d'un panache de pollution dissout. Bien que la localisation de la source et la direction d'écoulement soient connues pour le cas présenté ci-après, l'enjeu majeur réside dans la capacité de l'approche à reconstruire fidèlement l'étendue spatiale et la morphologie du panache en deux dimensions à partir d'un nombre restreint de points d'observation.

4.4.1. Description du jeu de données

Cette étude s'appuie sur un jeu de données réel mais confidentiel, raison pour laquelle une description exhaustive du site ne pourra pas être fournie. Une contamination par des composés organiques halogénés volatils (COHV) a été identifiée sur ce site il y a plusieurs années. La zone source a été localisée. Le polluant principal est le tétrachloroéthylène, dont le panache de contamination, y compris ses sous-produits de dégradation, est présent dans les eaux souterraines. La nappe est libre, la géologie qui compose l'aquifère est majoritairement composée de sables et de graviers. L'épaisseur moyenne de l'aquifère est d'environ 12 m et le niveau piézométrique se situe à environ 3 m de profondeur. Afin de pouvoir appréhender l'ensemble des phénomènes de dégradation, l'analyse porte sur la somme totale des concentrations en COHV. La Figure 55 illustre les concentrations mesurées en COHV dans les eaux souterraines, ainsi que la direction d'écoulement des eaux souterraines. Il est à noter que, en raison de contraintes liées aux activités du site, peu d'ouvrages sont disponibles à proximité immédiate de la zone source. Au total, 17 piézomètres sont présents, situés pour la plupart en limite de site. Les concentrations en COHV sont comprises entre 0 (les valeurs mesurées inférieures à la limite de quantification sont considérées à 0) et 607 $\mu g.l^{-1}$. Le seuil réglementaire retenu pour la somme des COHV est de 50 $\mu g.l^{-1}$.

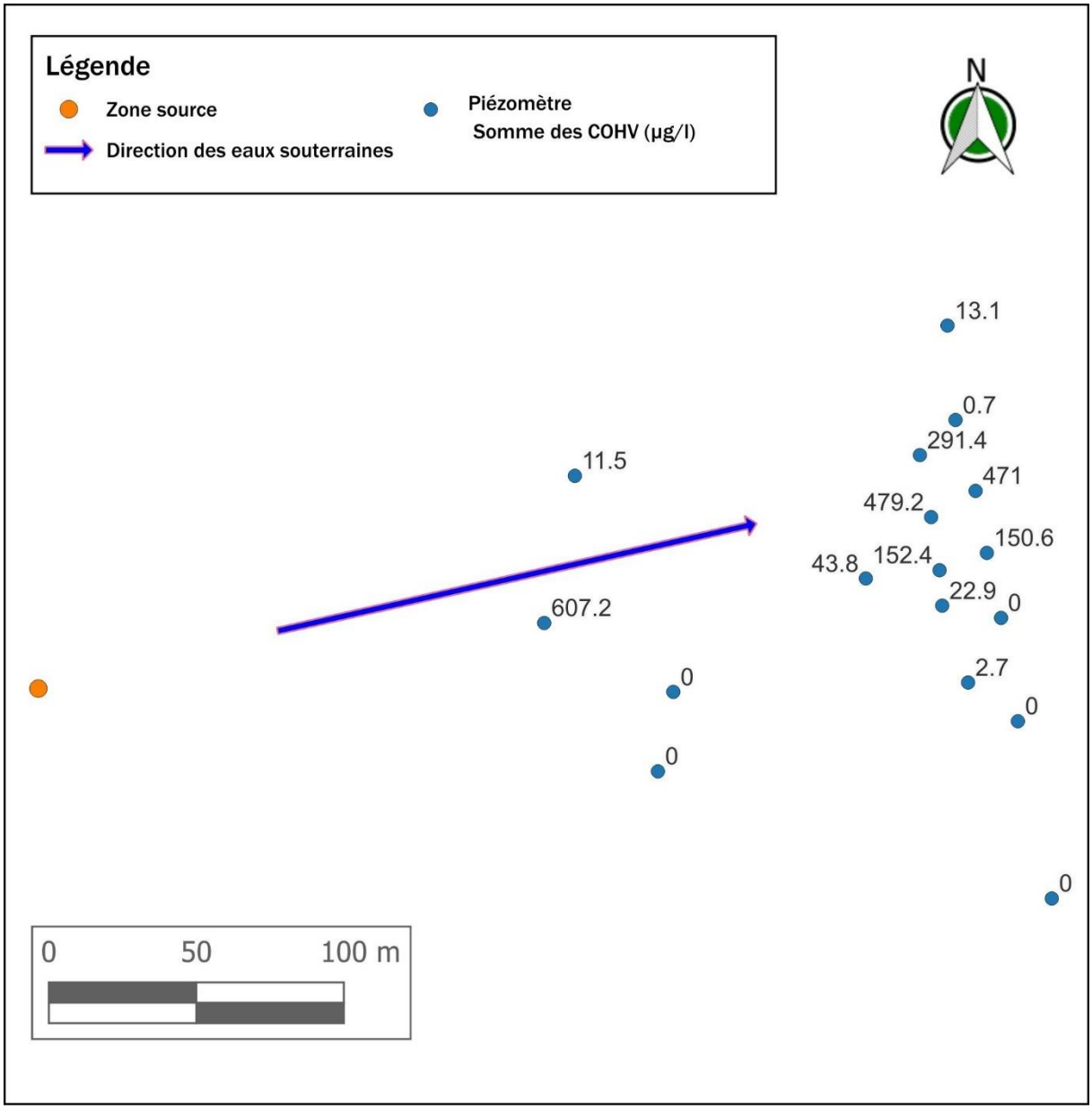


Figure 55 : Cartographie des concentrations en COHV et direction d'écoulement des eaux souterraines

4.4.2. Sélection d'une solution analytique adaptée

L'approche utilisée est similaire à celle décrite dans les sections précédentes. Cependant, la solution analytique doit être adaptée à la problématique des eaux souterraines. Compte-tenu des effets densitaires pouvant influencer la migration des polluants de types hydrocarbures, BTEX ou COHV, les prélèvements concernant ces composés sont réalisés à une même profondeur dans la colonne d'eau (surface pour la majorité des hydrocarbures et BTEX et en profondeur pour la majorité des COHV). Il a donc été jugé pertinent d'opter pour une solution

analytique en 2D plutôt qu'en 3D. La solution analytique sélectionnée est celle de Wilson et Miller (1978) (équations 4.1 à 4.6). Cette solution analytique implique un milieu homogène, avec un écoulement en régime permanent, uniforme selon la direction x. La source de pollution est considérée comme ponctuelle et continue dans le temps, ce qui semble cohérent avec la réalité du cas d'étude.

$$C_w = \frac{\frac{QC_{w0}}{b}\exp\left(\frac{x}{B}\right)}{4\pi\,\theta\left(D_x D_y\right)^{\frac{1}{2}}}\; W\left(u,\frac{r}{B}\right) \qquad 4.1$$

Avec :

$$B = \frac{2*Dx}{V} \qquad 4.2$$

$$\kappa = 1 + \left(\frac{2*B*\lambda}{V}\right) \qquad 4.3$$

$$\mathrm{u} = \frac{\mathrm{r}^2}{4\,\kappa\,\mathrm{D_x t}} \qquad 4.4$$

$$r = \left[\left(\mathrm{x}^2 + \left(\frac{\mathrm{D_x}}{\mathrm{D_y}}\right)\mathrm{y}^2\right)\kappa\right]^{\frac{1}{2}} \qquad 4.5$$

$$\mathrm{W}\left(\mathrm{u},\frac{\mathrm{r}}{\mathrm{B}}\right) = \int_u^{\infty}\frac{1}{\psi}\exp\left(-\left(\psi + \frac{\mathrm{r}^2}{4\mathrm{B}^2\psi}\right)\right)\mathrm{d}\psi \qquad 4.6$$

Où : Q est de débit d'injection du polluant [$L.T^{-3}$], b est l'épaisseur de l'aquifère [L], θ est la porosité effective [-], ψ est la variable d'intégration [-] et V est la vitesse d'écoulement dans la direction x [$L.T^{-1}$]

4.4.3. Application de la modélisation analytique inverse

Pour évaluer la performance de la modélisation analytique inverse, le jeu de données a été séparé en deux sous-ensembles distincts. Neuf points ont été alloués à la calibration du modèle et 8 huit points ont été réservés pour la vérification a posteriori. La séparation de ces points a été effectuée aléatoirement à l'aide de la fonction « point aléatoire dans une sélection » du logiciel QGIS. La triangulation de Delaunay a ensuite été appliquée aux données, générant 12 triangles (Figure 56). Parmi ceux-ci, 3 triangles présentaient des concentrations totales en COHV proches de la limite de quantification du laboratoire. Par cohérence avec la méthode employée pour la zone non saturée, ces trois triangles à très faibles concentrations n'ont pas été inclus dans les processus d'inversion. Ainsi, les données composant les 9 triangles restants ont été utilisées pour l'inversion. Chaque triangle a fait l'objet de 10 inversions.

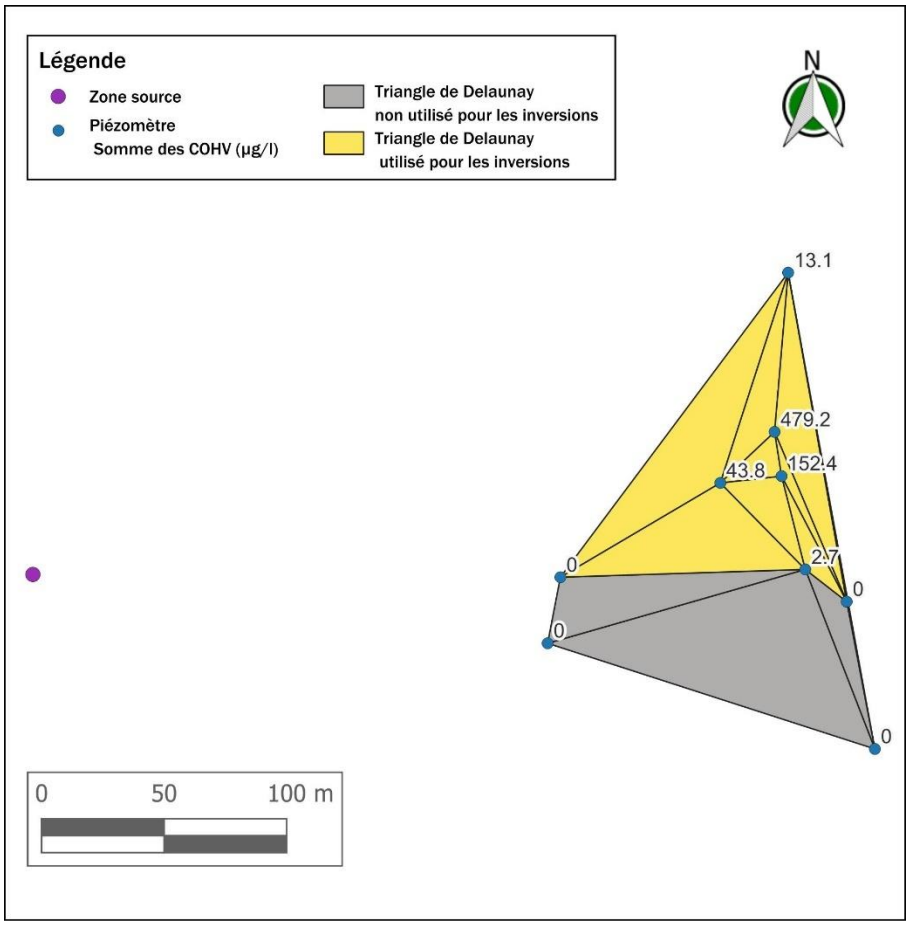


Figure 56 : Triangulation de Delaunay appliquée au panache de contamination des eaux souterraines

La combinaison des modèles a été appliquée avec la même pondération que celle décrite dans les sections précédentes (1/3).

Pour calibrer le modèle analytique, les paramètres ont été tirés au hasard à partir des plages de valeurs minimales et maximales prédéfinies (Tableau 20). Ces plages ont été choisies soit à partir de valeurs courantes de la littérature (pour I_{eff} et θ_w), soit déterminées empiriquement (pour $C_{w0,}$ t, α_L et α_T).

Tableau 20 : Plages de paramètres pour l'étalonnage pour le modèle analytique en zone saturée

Paramètre	Unité	Valeur basse	Valeur haute	Source
Q	$m.j^{-1}$	0,2	1	Valeur empirique
C_{w0}	$\mu g.l^{-1}$	0,01	90 000 000	Valeur empirique
θ	-	0,1	0,5	Valeur empirique

Paramètre	Unité	Valeur basse	Valeur haute	Source
t	jour	3650	25 550	Temps maximum depuis le début de l'activité et temps minimum depuis la découverte de la pollution
u	m/jour	0,01	100	Valeur empirique
α_l	m	0,01	10	Valeur empirique
α_t	m	0,001	1	Valeur empirique
λ	Anné$^{-1}$	0,00000365	0,0365	Mariot, 2005

Une analyse des paramètres estimés a été effectuée. Les résultats sont présentés dans le Tableau 21. Les valeurs obtenues sont difficilement comparables car aucun de ces paramètres n'a été mesuré. La médiane de la porosité est de 0,38, celle du temps est de 44 ans. La vitesse d'écoulement semble élevée avec une valeur de 88 m.j^{-1}.

Tableau 21 : Statistiques concernant les paramètres inversés pour le cas réel en zone saturée

	C_{w0}	Q	θ	t	u	α_L	α_T	λ
Moyenne	3,9E+06	4,6E-01	3,6E-01	1,6E+04	7,7E+01	6,4E+00	5,1E-01	4,8E-06
Percentile 25	4,1E+05	2,9E-01	2,7E-01	1,0E+04	5,5E+01	1,1E+00	3,7E-01	2,0E-06
Médiane	9,5E+05	4,3E-01	3,8E-01	1,6E+04	8,8E+01	4,8E+00	5,1E-01	4,8E-06
Percentile 75	2,2E+06	5,9E-01	4,7E-01	2,2E+04	9,9E+01	8,6E+00	6,6E-01	7,3E-06

4.4.4. Résultats

Les résultats de la modélisation analytique inverse (Fig. 57) montrent une cohérence significative avec les données de vérification. Ainsi, 7 piézomètres sur 8 (87,5 % de correspondance) ont révélé des concentrations en accord avec les simulations. Un taux de cohérence identique a été obtenu en réitérant la démarche avec deux autres seuils de pollution (10 µg.l^{-1} et 500 µg.l^{-1}). Pour ces différents scénarios, la valeur de concentration mal reproduite par le modèle correspond au point à 0,7 µg.l^{-1} situé à l'Est de la zone d'étude, et qui est entouré de valeurs de concentrations significativement plus élevées.

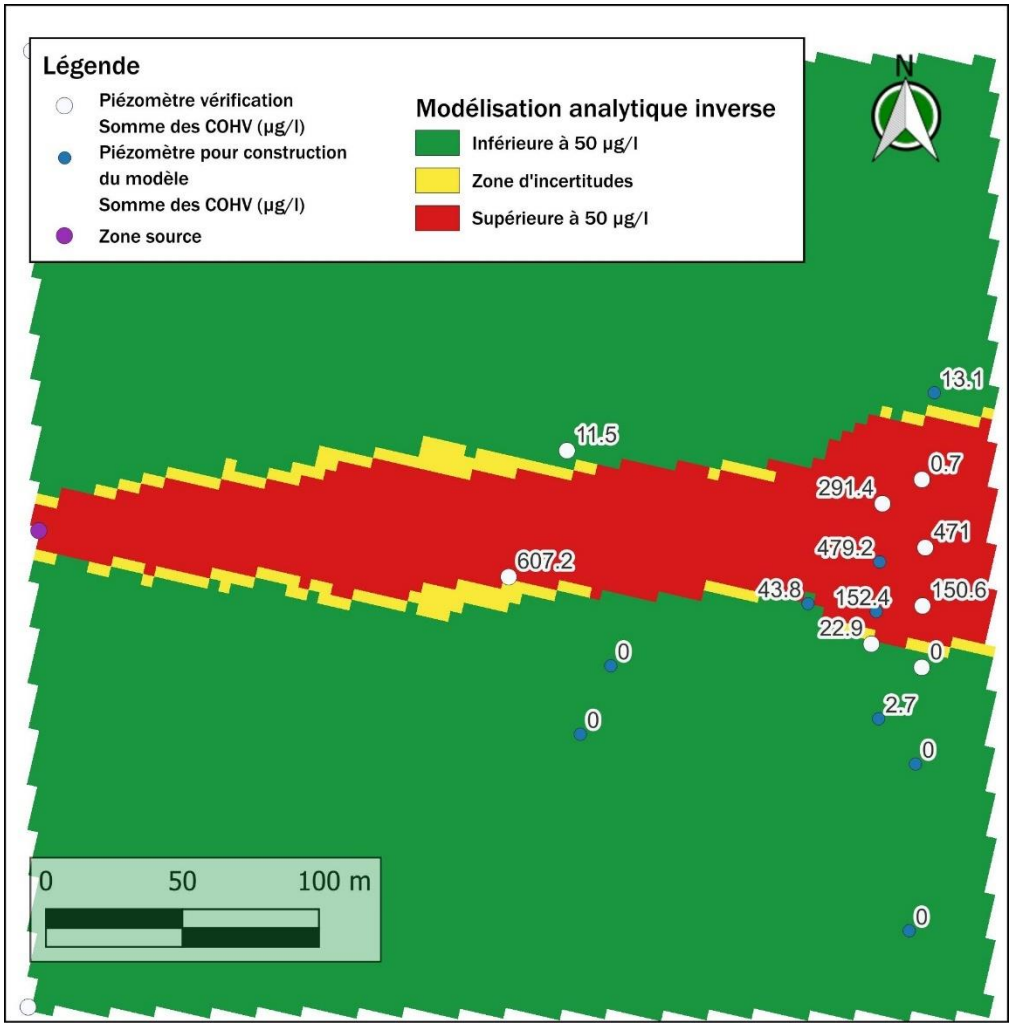


Figure 57 : Cartographie des concentrations pour l'application de la modélisation analytique inverse à un panache de COHV (seuil à 50 µg.l^{-1})

CHAPITRE 5 : COMPARAISON DE L'IAM AVEC LA MODELISATION NUMERIQUE

Dans ce chapitre, l'IAM est comparée à une approche d'inversion mettant en œuvre un modèle numérique spatialement distribué (maillé). L'objectif est d'évaluer la capacité de cette approche à reconstituer un panache de pollution à partir de points d'observations. Le scénario de pollution considéré correspond au cas synthétique utilisé dans la section 3.3. Suite à l'échec de cette première approche, les différentes stratégies de modélisation testées sont présentées. L'objectif principal de ce chapitre est de définir une méthodologie robuste pour initialiser un modèle d'inversion afin de résoudre une problématique de pollution dans un contexte réel. La méthode ayant été appliquée sur un cas synthétique, des biais positifs sont présents tels qu'un maillage identique ou des valeurs de paramètres similaires.

5.1. CONSTRUCTION ET INITIALISATION D'UN MODELE MAILLE POUR L'INVERSION

Un modèle maillé de 80 lignes, 80 colonnes et 60 couches est créé dans Groundwater Vistas (ref). Les dimensions de chaque maille sont de 0,7 m×0,7 m×0,1 m. La côte altimétrique va de 0 à 6 m. Comme pour les cas synthétiques présentés au chapitre 3, une charge hydraulique de -50m est imposée sur la couche inférieure du modèle. Le code de calcul MODFLOW USG est utilisé pour simuler l'écoulement et le transport à partir des paramètres suivants :

- Porosité ;
- Paramètre α de van Genuchten ;
- Paramètre β de van Genuchten ;
- Teneur en eau volumique résiduelle θ_r ;
- Recharge ;
- Concentration de la zone source ;
- Dispersivité longitudinale α_L ;
- Dispersivité transversale α_T.

Le modèle maillé est initialisé en attribuant des valeurs de paramètres homogènes sur l'ensemble du maillage. Le logiciel PEST est utilisé pour inverser les paramètres en cherchant à minimiser autant que possible l'écart entre les concentrations simulées et observées.

5.2. DISTRIBUTION SPATIALE DES POINTS D'OBSERVATION

Le cas synthétique considéré ici comporte 27 points d'observation répartis sur 3 profondeurs distinctes (1, 3 et 5 m, Fig.58). La densité spatiale des points est dans la moyenne des différentes configurations étudiées dans la section 3.3.

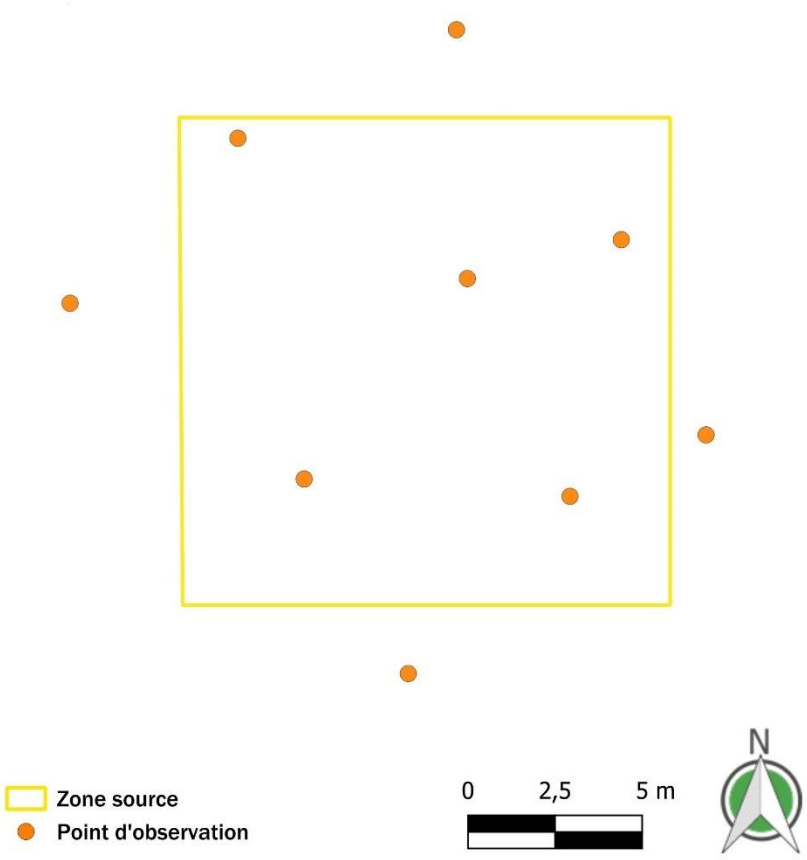


Figure 58 : Distribution spatiale des points d'observation pour l'inversion du modèle maillé

5.3. STRATEGIES D'INVERSION

Avec un modèle maillé, les deux approches classiques d'inversion sont i) l'inversion par zones (pouvant correspondre à des mailles individuelles ou bien à des ensembles de mailles) et ii) la méthode des points pilotes (Lavenue et de Marsilly, 2001, Doherty, 2003, Fienen et al., 2009). La méthode des points pilotes permet d'inverser les paramètres à des localisations sélectionnées afin d'assurer une représentation adéquate de la variabilité spatiale avec un nombre minimal de paramètres. Ces points agissent comme des « points de contrôle » dont la valeur est distribuée sur

l'ensemble du maillage par une interpolation spatiale. Cette approche limite ainsi le nombre de paramètres à inverser.

Il est à noter que, parmi les paramètres listés au chapitre 5.1, seule la porosité peut être inversée sous la forme de points pilotes dans le logiciel Groundwater Vistas. Les autres paramètres ne peuvent être inversés que sous forme de zones.

Deux stratégies d'inversion distinctes ont été testées pour cette étude : la méthode PEST Iterative Ensemble Smoother (PEST-IES), qui constitue la dernière avancée du logiciel PEST (Doherty, 2023), et la méthode PEST « standard ». Ces stratégies sont présentées dans les sections suivantes.

5.3.1. Inversions avec Pest IES

Les algorithmes d'optimisation traditionnels tels que ceux implémentés dans PEST requièrent à minima une exécution du modèle pour chaque paramètre à inverser, ce qui peut être très coûteux en temps de calcul, surtout lorsque le nombre de paramètres est important. L'algorithme IES (*Iterative Ensemble Smoother*) implémenté dans PEST++ offre une approche plus efficace en calculant des sensibilités approximatives à partir d'un ensemble de jeux de paramètres initiaux (ou échantillons de paramètres), cette méthode permet d'obtenir une incertitude. Une seule exécution du modèle est nécessaire par réalisation. Cela rend la méthode PEST-IES en théorie particulièrement avantageuse pour les problèmes comportant de nombreux paramètres ajustables. Le nombre de jeux de paramètres échantillonnés (ou réalisations d'ensemble) constitue un élément clé de la méthode IES, car il devrait idéalement refléter la dimensionnalité de l'espace des solutions, une notion cependant souvent difficile à déterminer en pratique en raison du nombre limité de données d'observation (White, 2018, Hayley et al., 2019, Farnum, 2025).

Dans le présent travail, la modélisation inverse a été appliquée à l'ensemble des paramètres du modèle numérique spatialement distribué. Cette approche est extrême, aucune grille d'interpolation (point pilote) n'est utilisée. Les hypothèses suivantes ont été retenues :

- Une zone uniforme a été définie pour la recharge, négligeant des hétérogénéités sur les précipitations à cette échelle ;
- Une affectation des paramètres maille par maille a été utilisée pour la porosité, les paramètres de van Genuchten et de rétention d'eau (α, β, θ_r), les dispersivités longitudinale et transversale et les valeurs de concentration de la zone source (sur la première couche uniquement).

Le nombre total de paramètres à inverser s'élevait à 2 694 401. Pour réaliser cette inversion, les paramètres par défaut de PEST et PEST IES ont été appliqués, et le nombre de simulations a été fixé

à 200. Ce test s'est soldé par un échec, i.e. les 200 simulations PEST-IES n'ont pas permis d'identifier des jeux de valeurs permettant de diminuer significativement la valeur de la fonction objectif.

Le problème a été simplifié avec différentes approches testées successivement en regroupant les mailles par 9 (zones de 3×3) réduisant ainsi le nombre de paramètres à environ de 300 000, puis en les regroupant par couches. D'autres simplifications ont été envisagées telles qu'une réinitialisation du modèle avec les paramètres de transport optimaux (concentration de la source, Infiltration efficace, dispersivités longitudinales et verticales) et une inversion limitée aux paramètres hydrauliques (porosité, θ_r, α, et β). Malgré ces simplifications, l'algorithme n'a pas convergé. Des modifications des options par défaut de PEST ou PEST-IES ont également été testées, e.g. ajout d'une régularisation supplémentaire via le module ies_reg_factor, modification du nombre de simulations, augmentation des facteurs de recherche lambda … Ces différents tests n'ont pas donné de meilleurs résultats et l'outil PEST-IES a finalement été abandonné.

5.3.2. Inversions avec PEST

L'estimation des paramètres avec PEST est directement influencée par le nombre de paramètres à optimiser. Il est donc essentiel de limiter ce nombre pour que les temps de calculs de la procédure d'inversion demeurent raisonnables et efficaces. Le temps de calcul nécessaire à une exécution unique du modèle numérique est de 14 secondes sur un ordinateur équipé d'un processeur Intel i9-12900H et de 32 Go de RAM. Pour maîtriser le temps de calcul et rendre l'inversion praticable, l'utilisation de 540 points pilotes au total a été adoptée pour l'inversion des paramètres (soit 9 points par couche pour la porosité). Le temps de calcul total de l'inversion peut être estimé en multipliant le temps d'une exécution directe du modèle par le nombre de simulations requis par PEST (qui dépend, notamment, de l'ensemble des points pilotes et des autres paramètres). Pour accélérer l'ensemble du processus, le logiciel PEST est couplé à l'outil BeoPEST qui permet de paralléliser les calculs. Cette approche a permis d'obtenir une durée de calcul raisonnable (plusieurs heures) sur un ordinateur standard. La localisation des points pilotes est présentée sur la Figure 59.

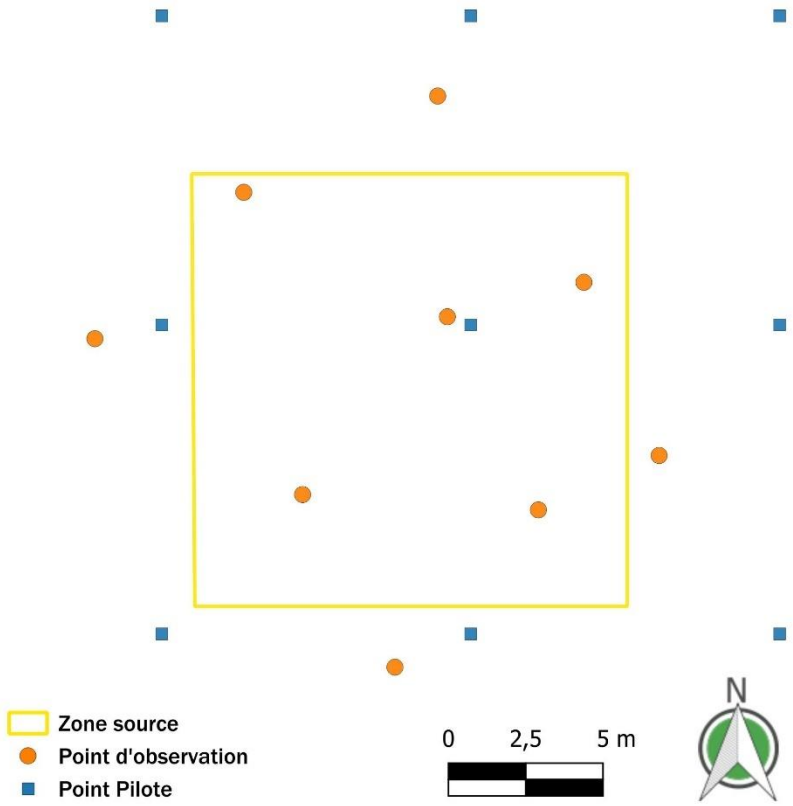


Figure 59 : Localisation des points pilotes de porosité

Compte tenu de l'impossibilité d'appliquer la méthode des points pilotes aux paramètres tels que α, β, θ_r, α_L et α_T, il a été décidé d'inverser ces paramètres individuellement par couche verticale. Parallèlement, la zone source a été structurée pour l'inversion en étant divisée en plusieurs lots de neuf mailles (3x3), regroupés sur la première couche verticale du modèle.

L'inversion initiale des paramètres a abouti, malgré la variation des valeurs initiales, à un échec de PEST à identifier des valeurs permettant de diminuer significativement la fonction objectif.

Pour pallier cette problématique, il a été décidé de scinder le processus d'inversion en deux étapes distinctes, séparant les paramètres hydrauliques des paramètres liés à la source de pollution. De plus, les paramètres de transport non-hydrauliques, comme les dispersivités (α_L et α_T), ont été fixés aux valeurs du modèle synthétique 'réalité'.

Seule la porosité a été retenue pour l'inversion par PEST. Les autres paramètres hydrauliques (α, β, et θ_r) ont été fixés (ce ne sont donc plus des valeurs de départ à inverser) à des valeurs jugées représentatives : β=2, α=0,1 m, et θ_r=0,05. L'inversion ne portant que sur la porosité, celle-ci a été traitée comme un paramètre effectif global (incluant l'effet de l'incertitude sur les autres paramètres

hydrauliques fixés). Pour permettre à l'algorithme d'explorer un large espace de solution, les bornes d'inversion pour la porosité ont été élargies au-delà des limites physiques (de 1×10^{-9} à 1×10^{9}).

Des contraintes de régularisation ont été ajoutées dans le but d'homogénéiser spatialement la solution (réduire les variations extrêmes entre les points). Ces contraintes ne visent pas à réduire l'erreur de manière absolue, mais à trouver un compromis entre la fidélité aux données mesurées et la stabilité mathématique de la solution. Elles sont essentielles pour obtenir un modèle qui soit stable et interprétable, surtout lorsque le nombre d'observations est limité.

Les meilleurs résultats (Fig. 60) ont été obtenus à l'aide de l'outil « SVD-Assist » qui combine les points pilotes en « super-paramètres » en s'appuyant sur une décomposition de la matrice jacobienne en valeurs singulières. Cette méthode permet à la fois d'accélérer le processus d'inversion en réduisant le nombre de paramètres à inverser (passant ici de 540 points pilotes à 22 super-paramètres) et de minimiser les effets de corrélations entre les paramètres, préjudiciables au processus d'inversion. La Figure 61 illustre le panache de concentration simulé à partir des paramètres inversés.

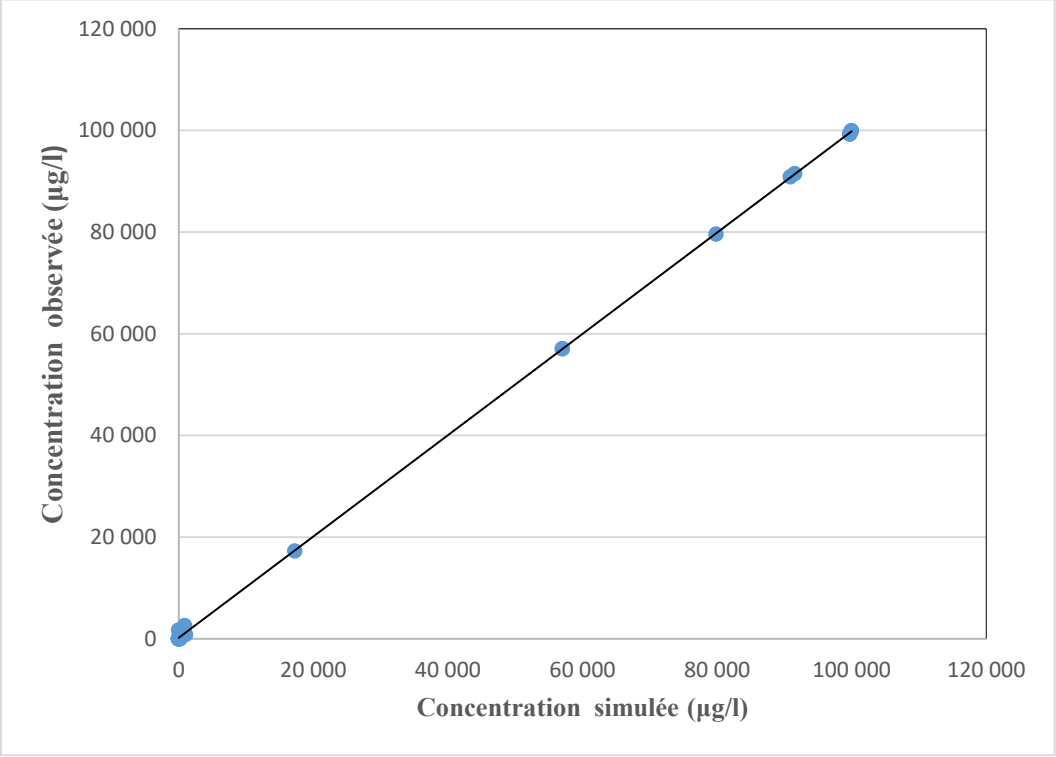
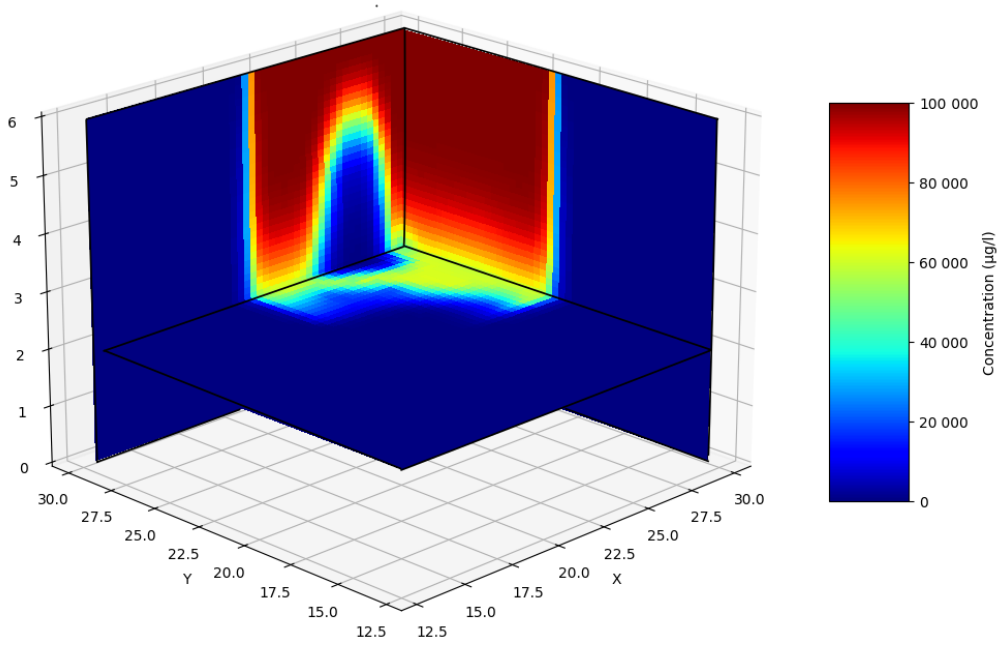
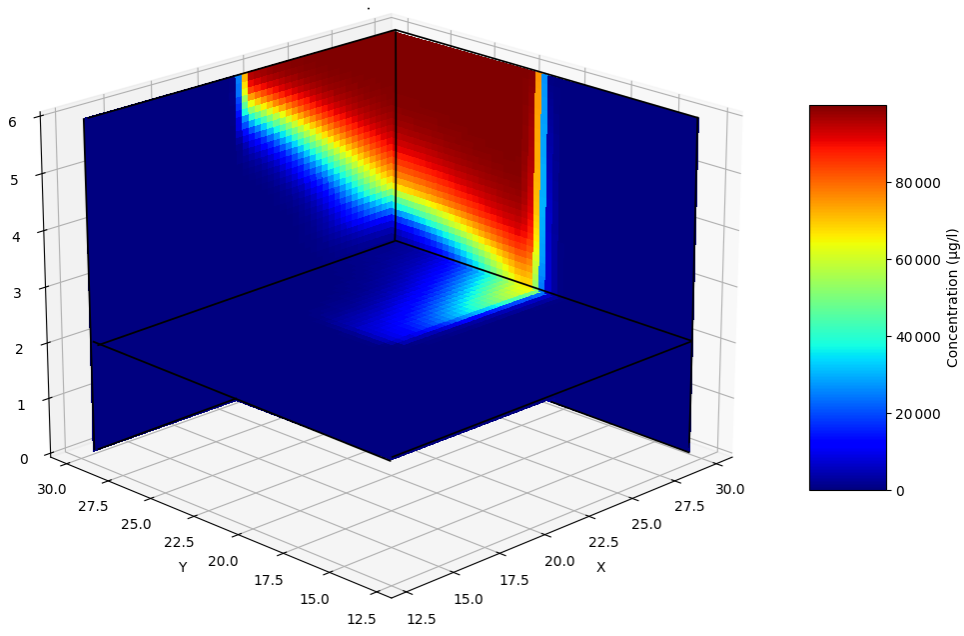


Figure 60 : Graphique de calibration des concentrations correspondant à l'inversion des porosités du modèle maillé avec la méthode SVD-Assist



a : Cas synthétique « réalité »



b : Modèle inversé (porosité)

Figure 61 : Coupes 3D des concentrations simulées à partir des porosités inversées

Bien que l'ensemble des paramètres n'ait pas été inversé, un calcul des volumes pollués a tout de même été réalisé. Cette étape, bien qu'elle ne résolve pas la problématique de l'inversion complète, a permis d'évaluer la cohérence du modèle simplifié et de fournir une base de réflexion sur sa pertinence pour estimer les volumes de terre à excaver.

Les Figure 62 et Tableau 22 comparent les volumes entre le modèle où l'inversion des paramètres hydrauliques a été réalisée (uniquement par la porosité) et le modèle correspondant à la « réalité » (cas synthétique).

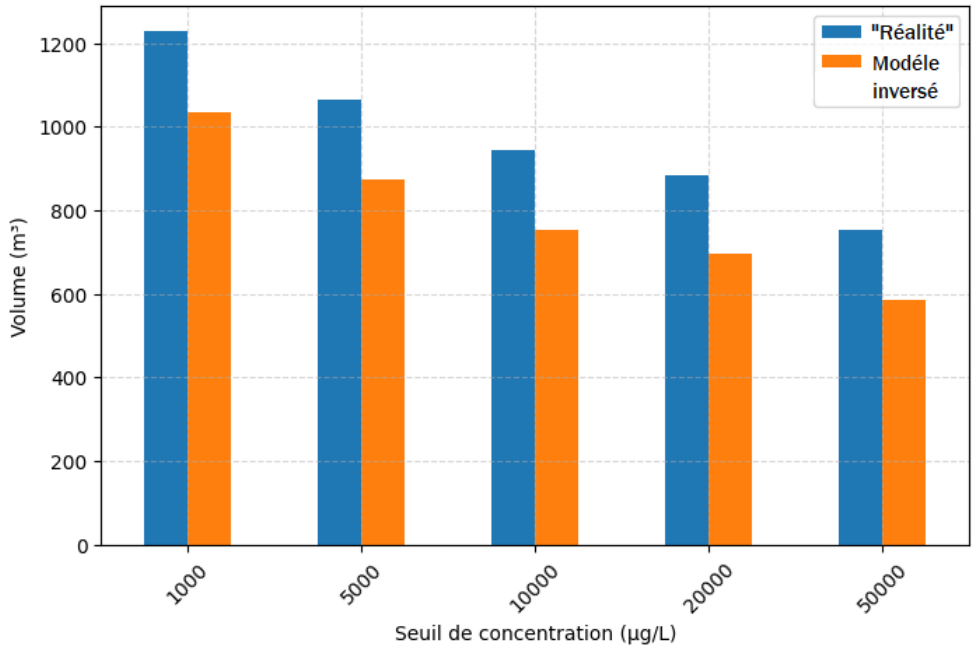


Figure 62 : Comparaison des volumes au-dessus des seuils critiques pour le modèle inversé et la « pseudo-réalité » (cas synthétique)

Le modèle inversé a tendance à sous-estimer les volumes de terres contaminées, un phénomène observé pour chacun des seuils critiques considérés. Cette erreur de volume devient particulièrement significative à des profondeurs supérieures à 3 m (Tab.22). Le modèle présente une meilleure performance entre 0 et 1 mètre de profondeur. Ceci s'explique par la configuration homogène des paramètres hydrauliques dans le modèle cas synthétique « réalité » sur ce même horizon, ce qui permet au modèle inversé de reproduire fidèlement cette zone.

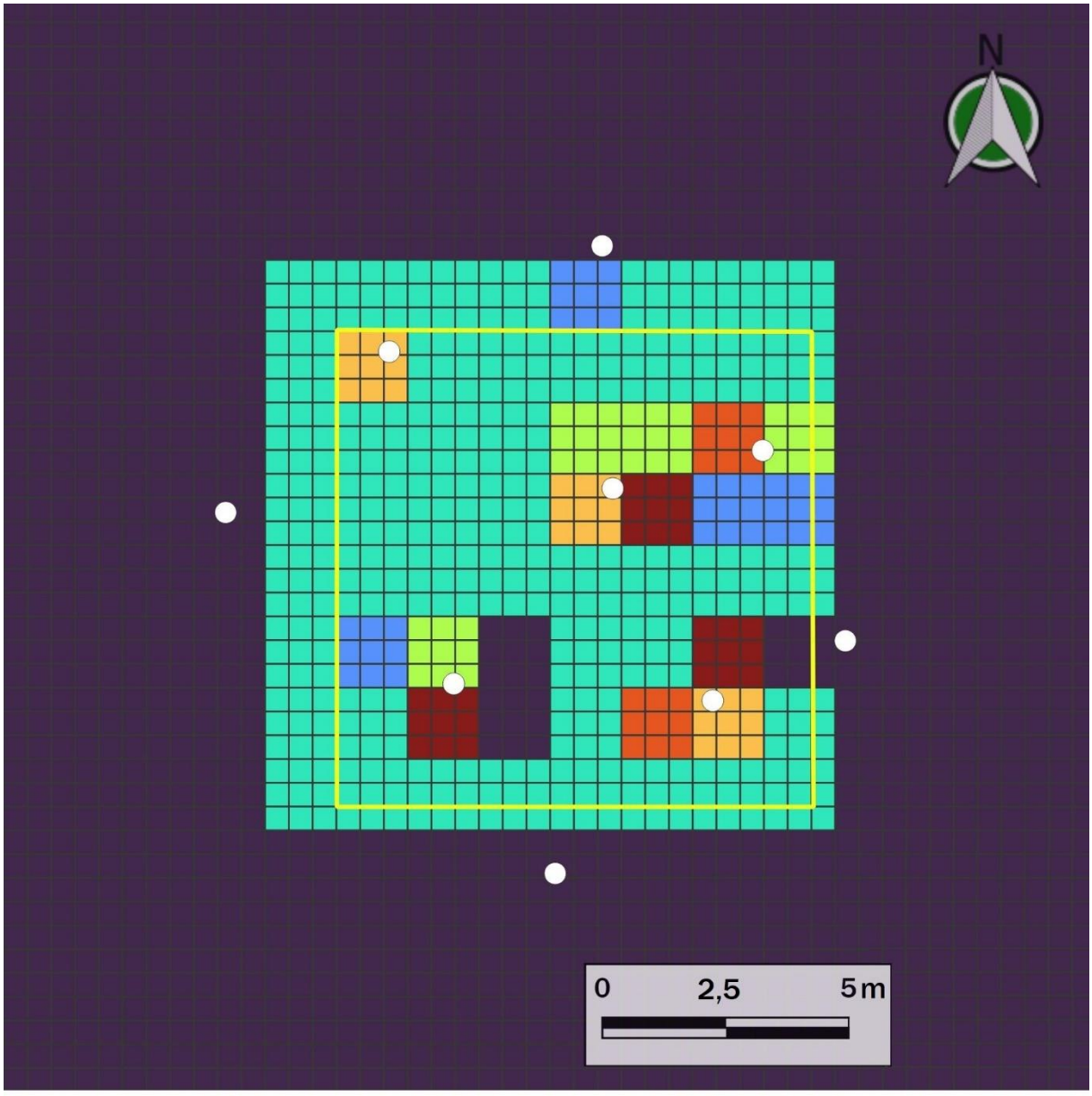
Une comparaison a été réalisée avec l'approche IAM. Pour les surfaces situées à 3 et 5 m de profondeur, le modèle inversé conduit à un pourcentage d'erreur moyen de 35 % et 89 % respectivement. En comparaison, le modèle analytique inverse conduit à des erreurs équivalentes ou inférieures, soit 34 % et 80 % pour les mêmes profondeurs. Le pourcentage d'erreur est bien plus faible dans l'horizon 0-1 m (2,5% contre 27% pour IAM). Les résultats obtenus par modélisation

numérique inverse sont à minoriser par rapport à ceux de la modélisation analytique inverse, car un sous-ensemble des paramètres a été contraint aux valeurs exactes du cas synthétique (pseudo-réalité).

Tableau 22 : Erreur moyenne en pourcentage pour l'inversion du modèle numérique

	% moyen faux négatif	% moyen faux positif	% Erreur moyenne
1 m	0	2,5%	2,5%
3 m	18%	17%	35,0%
5 m	64,0%	24,0%	89,0%

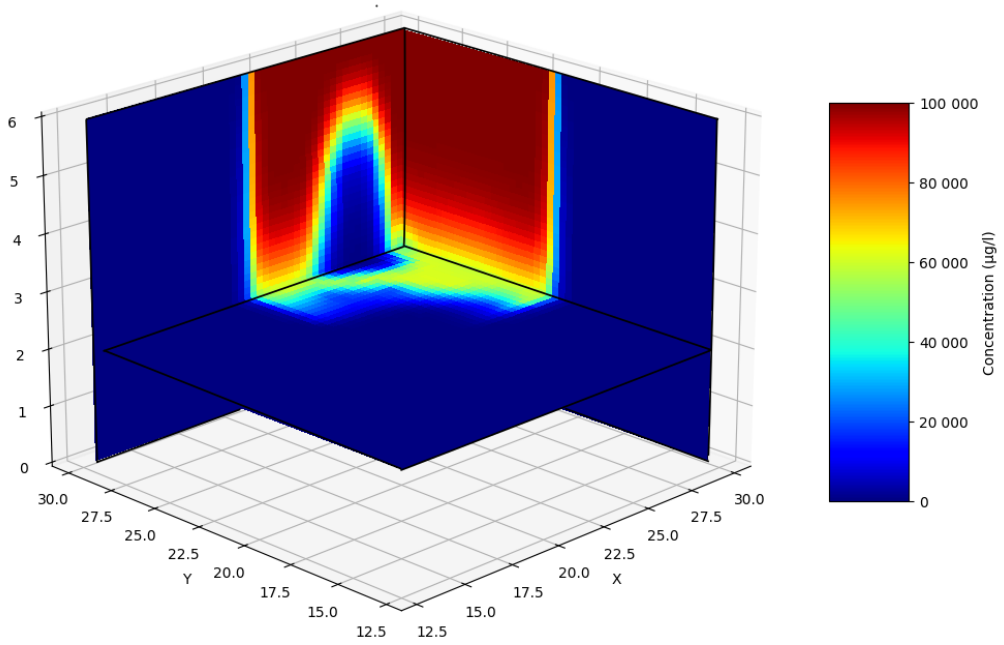
Concernant les paramètres de la zone source, seules les concentrations associées aux flux de recharge ont été inversées avec le modèle numérique dans une première approche. Le processus d'inversion a convergé mais la solution obtenue n'est pas conforme au cas synthétique. Les mailles identifiées comme source de pollution sont discontinues et placées uniquement à la verticale des points d'observation (Fig.63). L'ajout de contraintes de régularisation pénalisant la fonction objectif en cas de contrastes de concentrations sources entre mailles contigües n'a pas permis d'obtenir une solution spatialement plus homogène. Ainsi, bien que le processus d'inversion ait convergé, les paramètres identifiés sont erronés. D'un autre côté, si la valeur initiale du paramètre à inverser, telle que la concentration de la source, est trop éloignée des concentrations mesurées aux points d'observation, l'algorithme d'inversion ne converge pas. Par exemple, pour le cas présenté en Figure 63, l'inversion a échoué lorsque la valeur de départ était fixée en dessous de 50 000 $\mu g.l^{-1}$. Des tests ont également été effectués en utilisant des inversions par zone, par maille, et en démarrant avec des valeurs initiales aléatoires. Aucun calcul de volume n'a été réalisé avec ce cas compte tenu des résultats obtenus jugés non satisfaisants. La Fig.64 illustre la distribution spatiale des concentrations simulées à partir des zones sources inversées.



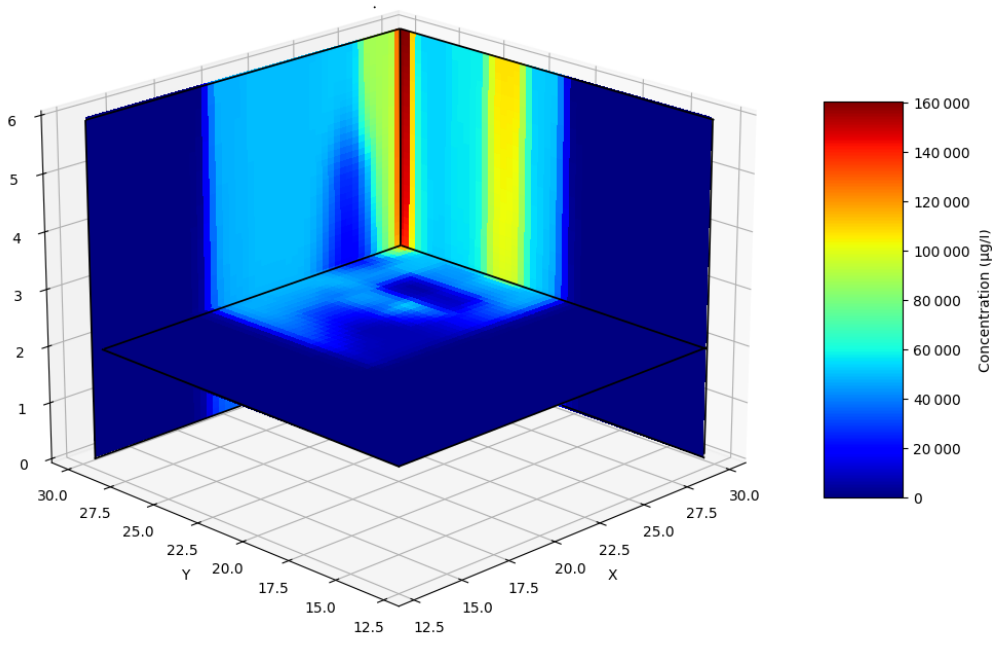
Légende

Concentration (µg/l)
0 - 20000
20000 - 40000
40000 - 60000
60000 - 80000
80000 - 100000
100000 - 120000
120000 - 161000
Point d'observation
Zone source "réelle"

Figure 63 : Distribution spatiale des concentrations sources identifiées par modélisation inverse



a : Cas synthétique « réalité »



b : Modèle inversé (zone source)

Figure 64 : Coupes 3D des concentrations simulées à partir des zones sources inversées

5.4. SYNTHESE DU CHAPITRE

Les résultats obtenus montrent que, dans le contexte de cette étude, l'inversion de modèles maillés n'a pas permis de caractériser efficacement le panache de pollution en zone non saturée. Ces observations suggèrent que cette approche pourrait ne pas être la plus appropriée pour des cas présentant une complexité similaire. L'objectif initial d'inverser l'ensemble des paramètres n'a pas été atteint, en raison des difficultés liées à la représentation spatiale de la zone source et à la non-unicité de la solution. Des simplifications ont été apportées en n'inversant que la porosité comme paramètre effectif global représentant les paramètres hydrauliques. Une approche similaire a été faite avec la concentration en zone source.

L'inversion simplifiée (porosité seule) permet une estimation satisfaisante des volumes à faible profondeur (erreur de 3 % à 1 m), grâce à une paramétrisation initiale proche du modèle de référence. Cependant, les erreurs s'accroissent fortement en profondeur (35 % à 3 m, 89 % à 5 m), dépassant ou égalant les performances de la modélisation analytique inverse. Ces performances, limitées par l'inversion d'un seul paramètre, soulignent la complexité du problème, notamment la non-unicité de la distribution de la pollution.

DISCUSSION GENERALE, CONCLUSION ET PERSPECTIVES

RAPPEL DES OBJECTIFS ET METHODOLOGIE

L'objectif principal de cette thèse était de développer une nouvelle méthode pour la caractérisation d'un panache de pollution en zone non saturée en se basant sur la modélisation analytique inverse (IAM). La méthodologie développée a discrétisé le domaine de pollution en utilisant une triangulation de Delaunay. Ce découpage spatial permet d'appliquer des solutions analytiques de manière séquentielle sur chaque triangle, ce qui permet de modéliser les variations de propriétés et ainsi de s'affranchir de la simplification d'un milieu homogène. Cette méthode a été évaluée à la fois sur des cas synthétiques et sur des cas réels afin de valider sa robustesse et sa pertinence pratique.

DISCUSSION GENERALE

Application à des cas synthétiques

Les études menées sur les cas synthétiques, en appliquant la méthode de modélisation analytique inverse avec une distribution standard des points d'observation, ont révélé une diminution significative de l'erreur moyenne (de 40 à 70%) par rapport aux interpolations spatiales. La méthodologie employée s'est avérée robuste, en particulier face à la rareté des données. Alors que les méthodes d'interpolation ont montré des erreurs moyennes très variables et nettement plus élevées (par exemple, 59 % pour une distribution régulière et 116 % pour une distribution aléatoire), conduisant souvent à une surestimation substantielle des volumes pollués, l'erreur moyenne de la méthode IAM s'est stabilisée avec un nombre limité de points d'observation (moins de 6 à 10 points présentant des valeurs supérieures à 1 % de la concentration maximale). Cela suggère qu'une estimation robuste du volume de terres polluées peut être obtenue avec un ensemble de données relativement limité. Cela constitue un avantage pratique significatif dans les diagnostics de pollution où un échantillonnage exhaustif peut conduire à des coûts prohibitifs. Ces résultats quantitatifs confirment que l'IAM pourrait être une alternative compétitive et souvent avec une performance supérieure aux techniques d'interpolation traditionnelles pour l'estimation volumétrique de terres polluées.

Dans le domaine de la remédiation, pour laquelle la protection de l'environnement et de la santé publique est primordiale, la gestion de l'incertitude est un enjeu crucial. Un équilibre doit être trouvé entre les faux négatifs (contamination non détectée) et les faux positifs (surestimation des zones polluées). Une contamination non détectée représente un risque plus important, car elle peut entraîner une remédiation incomplète et des risques environnementaux persistants. La comparaison entre les méthodes d'interpolation et l'IAM a mis en évidence des approches fondamentalement différentes face à cet enjeu. Les méthodes d'interpolation se distinguent par un taux de faux négatifs bas. Cependant, cet avantage se fait au prix d'une surestimation conséquente des volumes pollués, avec un taux de faux positifs qui peut dépasser 100 %.

L'approche IAM, à l'inverse, offre un équilibre plus efficient. Bien que la modélisation analytique inverse puisse présenter un taux de faux négatifs supérieur par rapport aux techniques d'interpolation, elle permet une diminution systématique et significative des faux positifs. Cette réduction se traduit directement par une meilleure maîtrise des coûts d'excavation et de traitement, et une plus grande efficacité globale des projets. Le taux plus élevé de faux négatifs observé dans certains scénarios pour la méthode analytique inverse est un point critique pour son application pratique. Pour atténuer ce risque et garantir une estimation plus fiable, il est recommandé d'ajuster le seuil de probabilité utilisé pour la délimitation du panache. Cela pourrait impliquer de fixer un seuil plus bas, alignant ainsi l'IAM sur le principe de précaution tout en conservant son avantage en termes d'efficacité.

Application à des cas réels

L'application de l'IAM à 32 cas réels a consolidé les conclusions établies à partir des cas synthétiques. Les résultats ont montré une cohérence moyenne de plus de 80 % avec les prélèvements de contrôle effectués après excavation. L’équilibre entre les bords de fouille estimés en faux négatifs ou faux positifs est plutôt respecté avec 14,6 % des bords de fouille sous-estimés (6 sur 41) et 20,9 % surestimés (32 sur 153). De plus, l'analyse montre que 28,9 % de l'erreur totale (11 erreurs sur 38) sont concentrées dans les zones d'incertitude. Cela souligne que, si la gestion et la réévaluation de ces zones pouvaient être optimisées, la performance globale de l'IAM dépasserait encore les résultats présentés. Le nombre d’observations a une influence limitée sur les estimations : les sites présentant une densité de sondages plus élevée n’ont pas donné de meilleurs résultats (le pourcentage de non-conformité n'a pas diminué de manière significative à mesure que la densité d'échantillonnage augmentait). Il en est de même avec la lithologie, même si les sols argileux ont produit des taux de conformité légèrement inférieures (74 % contre plus de 80% pour les autres lithologies). L’étude du

cas spécifique localisé à Saint Vincent de Tyrosse a confirmé ces conclusions. Les méthodes d'interpolation utilisées dans le plan de gestion pour délimiter la zone polluée n'ont pas donné satisfaction (incertitudes élevées). Une caractérisation supplémentaire des terres a donc été réalisée avant les travaux de dépollution. L'estimation du volume de terres polluées par la modélisation analytique inverse avec un nombre de points limité (11 points d'observation) a donné des résultats satisfaisants avec une erreur de volume de l'ordre de 34% en comparaison avec la zone excavée. A noter que cette zone est probablement sous-excavée par rapport au panache de pollution réel compte tenu des dépassements de concentration en polluants toujours présents en bords de fouille après excavation.

Modélisation numérique inverse

La méthode IAM a été confrontée à une approche de modélisation numérique inverse. L'objectif était de déterminer si, en se basant sur les mêmes points d'observation, l'inversion d'un modèle maillé pouvait permettre de reconstruire une distribution spatiale plus précise du panache de pollution. Le cas d'étude hétérogène de la section 3.3 a été utilisé comme base de comparaison. Les différents tests réalisés n'ont pas permis de reconstruire un panache de pollution cohérent, que ce soit avec l'algorithme PEST « standard » et avec la variante PEST-IES. Différents problèmes ont été rencontrés avec notamment une problématique de régularisation pour la zone source de pollution. Bien que des contraintes de régularisation (homogénéisation de la solution) aient été ajoutées, le processus d'inversion conduit à une solution spatialement discontinue en identifiant une multiplicité de sources indépendantes, localisées à la verticale des points d'observation.

Les problématiques de temps de calcul et de non-unicité de la solution ont rendu le processus complexe et chronophage. L'IAM, malgré ses hypothèses simplificatrices, a démontré une efficacité remarquable pour estimer le volume de sols pollués, offrant une solution viable et plus facile à mettre en œuvre que les modèles numériques maillés.

LIMITES ET PERSPECTIVES

Les principales limites de la méthode IAM sont liées aux hypothèses simplificatrices associées à la solution analytique utilisée. Le modèle suppose l'existence d'une source unique dont les paramètres (tels que le flux d'infiltration (I_{eff}), la teneur en eau volumique (θ_w), ou le taux de dégradation (λ)) sont considérés comme uniformes dans l'espace et constants dans le temps (état stationnaire), alors qu'ils sont susceptibles de présenter une forte variabilité en réalité. Les concentrations prédites dans

l'eau (la concentration en solution du composé dissous dans l'eau) par la solution analytique peuvent parfois présenter des valeurs anormales, allant jusqu'à dépasser la limite de solubilité du composé. Cette anomalie résulte de l'absence de considération de la teneur de polluant dans la phase organique (calcul uniquement sur les phases sol, eau et air), celle-ci étant partiellement incluse dans les concentrations estimées pour la phase aqueuse. La solution analytique considère une source d'injection unique et continue dans le temps. Pour gérer des cas plus complexes, tels que la présence de sources de pollution multiples et distinctes ou des injections non continues, la solution analytique pourrait être modifiée en utilisant les principes de superposition spatiale et de convolution temporelle.

Compte tenu des résultats satisfaisants obtenus à partir de cas synthétiques et des cas réels, la modélisation analytique inverse pourrait être utilisée dans les plans de gestion de sites et sols pollués afin de déterminer l'étendue des pollutions présentes. Voici quelques recommandations pratiques :

- La planification stratégique des campagnes d'échantillonnage doit chercher à optimiser la répartition spatiale et la densité des points d'observation. Cette approche permet, notamment, d'obtenir des estimations fiables avec un nombre réduit de points placés stratégiquement. Pour cela, l'IAM pourrait être mise en œuvre une première fois (s'il y a à minima 3 valeurs) après un diagnostic initial afin de déterminer les zones avec une incertitude importante. Les sondages prévus pour le diagnostic approfondi pourraient être localisés aux emplacements définis par la méthode ;
- Ajustement du seuil : pour la délimitation du panache, un seuil de probabilité conservateur doit être mis en œuvre afin de minimiser le risque de faux négatifs, conformément aux objectifs de gestion des risques. L'intervalle entre deux seuils de probabilité peut aider à cibler les zones d'incertitude (un seuil de probabilité minimum et un seuil moyen ou maximum) ;
- Éviter la surcalibration des données : les limites de quantification du laboratoire sont souvent de l'ordre du $mg.kg^{-1}$ MS pour la majorité des composés. Lors de la calibration des paramètres du modèle à partir des données d'observation d'un cas réel via PEST, la surcalibration (atteindre une précision de l'ordre du $\mu g.kg^{-1}$ alors que la limite de quantification du laboratoire est en $mg.kg^{-1}$) peut générer du bruit de fond dans les résultats. Pour remédier à ce problème, il est possible de spécifier une valeur seuil pour la fonction objectif et de stopper le processus d'itération si cette valeur seuil est atteinte ; celle-ci pourrait correspondre à la limite de quantification en laboratoire au carré à multiplier par le nombre d'observations inversées.

Différentes perspectives sont envisagées à l'issue de ce travail :

- Outre l'estimation des volumes de sols pollués, l'approche de Modélisation Analytique Inverse (IAM) a démontré un potentiel significatif pour des applications plus larges. La méthode a été testée sur un cas de panache de contamination des eaux souterraines afin d'explorer son applicabilité à la délimitation de panaches en zone saturée. Les résultats de cette étude de cas sont très encourageants. Avec 7 piézomètres sur 8 (soit 87,5 %) respectant les seuils critiques, la méthode a permis de caractériser le panache de manière satisfaisante. Ce succès initial suggère que l'IAM pourrait être une solution innovante pour des problématiques allant au-delà de l'estimation des volumes de terres polluées ;
- La méthodologie pourrait être étendue pour prendre en compte d'autres types de polluants (métaux lourds avec des cinétiques de sorption/désorption plus complexes) ou des processus densitaires (présence de NAPL) ;
- L'IAM pourrait également permettre d'optimiser la localisation de nouveaux sondages après un diagnostic initial : à placer dans les zones avec les incertitudes les plus fortes ;
- Le développement d'un logiciel convivial faciliterait la diffusion de la méthode IAM au sein de la communauté professionnelle

REFERENCES

Alcolea, A, Carrera, J, Medina, A. (2008). Regularized pilot points method for reproducing the effect of small scale variability: Application to simulations of contaminant transport. Journal of Hydrology, Volume 355, Issues 1–4, Pages 76-90, ISSN 0022-1694. https://doi.org/10.1016/j.jhydrol.2008.03.004.

Amairic, L., Moreau, P. (2021). Analyse des sols en contexte Sites et sols pollués – sources d'incertitudes prise en compte par les laboratoires – Rapport final. BRGM/RP/-71307-FR, 36p.

Anderson, M. P., Woessner, W. W., et Hunt, R. J. (2015). Applied groundwater modeling: simulation of flow and advective transport. Academic press.

ASTM. (2000). Standard Guide for Risk-Based Corrective Action; ASTM: West Conshocken, PA, USA.

Beven, K. J. (1996). A discussion of distributed hydrological modelling. In Distributed hydrological modelling (pp. 255-278). Dordrecht: Springer Netherlands.

Beyer, C., Bauer, S., et Kolditz, O. (2006). Uncertainty assessment of contaminant plume length estimates in heterogeneous aquifers. Journal of contaminant hydrology, 87(1-2), 73-95.

Boudreault, J. P., Dubé, J. S., et Marcotte, D. (2016). Quantification and minimization of uncertainty by geostatistical simulations during the characterization of contaminated sites: 3-D approach to a multi-element contamination. Geoderma, 264, 214-226.

BRGM. (1987). *Carte géologique de la France au 1/50 000 n° 976, Saint-Vincent-de-Tyrosse*. (1ère éd., Échelle 1:50 000). Orléans, France : Bureau de recherches géologiques et minières.

Brooks, R. H., et Corey, A. T. (1966). Properties of porous media affecting fluid flow. *Journal of the irrigation and drainage division*, *92*(2), 61-88.

Brunner, P., et Simmons, C. T. (2012). HydroGeoSphere: A fully integrated, physically based hydrological model. Ground Water, 50(2), 170– 176.

Burnett, R. D., et Frind, E. O. (1987). Simulation of contaminant transport in three dimensions: 2. Dimensionality effects. Water Resources Research, 23(4), 695-705.

Burrough, P. A., McDonnell, R. A., et Lloyd, C. D. (2015). Principles of geographical information systems. Oxford University Press.

Carre, F., Caudeville, J., Bonnard, R, Bert, V., Boucard, P. (2015). Soil contamination and human health: A major challenge for global soil security. Global Soil Security Symposium, College Station, United States. pp.275-295, ff10.1007/978-3-319-43394-3_25ff. ffineris01864711f

Cavalcante Fraga, L-E. (2019). Caractérisation des sols pollués via des méthodes géophysiques : couplage entre le diagnostic conventionnel et les méthodes géophysiques pour estimer la distribution spatiale des polluants à l'aide du formalisme géostatistique. Interfaces continentales, environnement. Thèse de doctorat. Sorbonne Université.

Chiles, J. P., Delfiner, P. (2012). Geostatistics: modeling spatial uncertainty (Vol. 713). John Wiley et Sons.

Chukwunonso O. Aniagor, Marcel I. Ejimofor, Stephen N. Oba and Matthew C. (2022). Application of artificial intelligence in the mapping and measurement of soil pollution, Chapter 13, Current Trends and Advances in computer-Aided Intelligent Environmental Data Engineering

Ciampi, P., Esposito, C., Bartsch, E., Alesi, E. J., et Papini, M. P. (2021). 3D dynamic model empowering the knowledge of the decontamination mechanisms and controlling the complex remediation strategy of a contaminated industrial site. Science of the Total Environment, 793, 148649.

Colombano, S. (2010). *Le plan de gestion*. Ed. Techniques Ingénieur.

Cushman, J. H., et Tartakovsky, D. M. (Eds.). (2016). The handbook of groundwater engineering. CRC Press.

Dakins, M. E., Toll, J. E., et Small, M. J. (1994). Risk-based environmental remediation: Decision framework and role of uncertainty. Environmental Toxicology and Chemistry: An International Journal, 13(12), 1907-1915.

Das, B. M., et Sobhan, K. (2006). Principles of geotechnical engineering.

Davis, L. A.et Neuman, S. P. (1983). Documentation and user's guide: UNSAT2—Variably saturated flow model. Final Rep. WWL/TM-1791-1. Water, Waste et Land.

Demougeot-Renard, H, de Fouquet, C et Fritsch, M. (2004). Optimizing sampling for acceptable accuracy levels on remediation volume and cost estimations - An iterative approach, based on geostatistical methods, illustrated on a former smelting works polluted with lead. geoENV IV — Geostatistics for Environmental Applications. Quantitative Geology and Geostatistics, vol 13. Springer, Dordrecht. https://doi.org/10.1007/1-4020-2115-1_24

Demougeot-Renard, H., 2004. De la Reconnaissance à la Réhabilitation des Sols Industriels Pollués: Estimations Géostatistiques Pour une Optimisation Multicritère, Thèse de doctorat. Sciences Naturelles, Institut Polytechnique de Lorraine, Lorraine, France.

Demougeot-Renard, H., Garcia, M., Jeannée, N., 2012. What about the practical implementation of geostatistics for contaminated sites and soils, in: 9th International Geostatistics Congress, Oslo, Norway.

Deng, J., Yu, J., Wang, X., Yu, D., Ma, H., Wu, Y., ... et Pu, S. (2024). Spatial distribution and migration characteristics of heavy metals at an abandoned industrial site in the Southwest of China. *Journal of Hazardous Materials*, *480*, 136447.

Diallo, T M.O., Collignan, B., Allard, F. (2015). 2D Semi-empirical models for predicting the entry of soil gas pollutants into buildings, Building and Environment, Volume 85, Pages 1-16, ISSN 0360-1323,

Doherty, J. (2003). Ground water model calibration using pilot points and regularization. *Groundwater*, *41*(2), 170-177.

Doherty, J.E, and Hunt, R.J., (2010).a Approaches to highly parameterized inversion—A guide to using PEST for groundwater-model calibration: U.S. Geological Survey Scientific Investigations Report 2010–5169, 59 p.

Doherty, J.E, (2018).b. PEST, Model-independent parameter estimation—User manual (7th ed., with slight additions): Brisbane, Australia, Watermark Numerical Computing.

Doherty, J. E., Fienen, M. N., et Hunt, R. J. (2011). Approaches to highly parameterized inversion: Pilot-point theory, guidelines, and research directions. *Scientific investigations report*, (2010-5168).

Doherty, J. E. (2023). Model-independent parameter estimation user manual part II: PEST utility support software. *Watermark, Brisbane, Australia*, 268.

Domenico, P. A. and Robbins, G. A.: (1985). A new method of contaminant plume analysis, *Ground Water* **23**, 476-485.

D'or, D, Demougeot-Renard, H, Garcia, M. (2009). An Integrated Geostatistical Approach for Contaminated Site and Soil Characterisation, , Math.Geoci.41(3),307-322

El-Aassar, A. H., Hagagg, K., Hussien, R., Oterkus, S., et Oterkus, E. (2023). Integration of groundwater vulnerability with contaminants transport modeling in unsaturated zone, case study El-Sharqia, Egypt. Environmental Monitoring and Assessment, 195(6), 722.

Environnemental Protection Agency (EPA), (1996). Soil Screening Guidance: User's Guide, EPA540/R-96/018.

Ersoy, A. (2021). Critical review of the environmental investigation on soil heavy metal contamination. Applied Ecology et Environmental Research, 19(5).

Espinet, A. J., et Shoemaker, C. A. (2013). Comparison of optimization algorithms for parameter estimation of multi-phase flow models with application to geological carbon sequestration. Advances in water resources, 54, 133-148.

Essouayed ; E. (2019). Développement d'une stratégie de localisation d'une source de contaminants en nappe : mesures innovantes et modélisation inverse. Sciences de la Terre. Université Michel de Montaigne - Bordeaux III, Français. ffNNT : 2019BOR30006ff. fftel-02127045ff

Farnum, T. J., Leaf, A. T., et Fienen, M. N. (2025). PEST++ IES How Many Iterations and Realizations, Finding the Point of Diminishing Returns. *Groundwater*.

Favéreaux, S., Belbeze, S., Balon, P. (2019). Essai collaboratif sur l'échantillonnage de sols pollués – Exploitation des résultats et recommandations. Rapport final. BRGM/RP- 69237-FR, 113 p.

FEFLOW, (2015). User Guide, 220pfile:///C:/Users/s824694/Downloads/FEFLOW%20User%20Guide%20(1).pdf

Fienen, M. N., Muffels, C. T., et Hunt, R. J. (2009). On constraining pilot point calibration with regularization in PEST. *Groundwater*, *47*(6), 835-844.

Fitts, C. R. (2013). Groundwater Science. Academic Press.

Froger, C., Saby, N. P., Jolivet, C. C., Boulonne, L., Caria, G., Freulon, X., ... et Bispo, A. (2021). Spatial variations, origins, and risk assessments of polycyclic aromatic hydrocarbons in French soils. Soil, 7(1), 161-178.

Gautam, K, Sharma, P, Dwivedi, S, Singh, A, Gaur, V-K, Varjani, S, Srivastava, J-K, Pandey,A, Chang, J.S, Ngo, H.H. (2023). A review on control and abatement of soil pollution by heavy metals: Emphasis on artificial intelligence in recovery of contaminated soil, Environmental Research, Volume 225, 115592, ISSN 0013-9351. doi.org/10.1016/j.envres.2023.115592.

GeoSiPol, INERIS, ADEME. (2005). Géostatistique appliquée aux sites et sols pollués - Manuel méthodologique et exemples d'applications, 139 p.

Goovaerts, P. (1997). Geostatistics for natural resources evaluation. Oxford university press.

Goovaerts, P. (2000). Estimation or simulation of soil properties? An optimization problem with conflicting criteria. *Geoderma*, *97*(3-4), 165-186.

Gustafson, J. B., Tell, J. G., et Orem, D. (1997). *Selection of representative TPH fractions based on fate and transport considerations* (Vol. 3). Amherst, MA, USA: Amherst Scientific Publishers.

Guridi, I., Chassagne, R., Pryet, A., et Atteia, O. (2024). Uncertainty Quantification of Contaminated Soil Volume with Deep Neural Networks and Predictive Models. *Environmental Modeling et Assessment*, *29*(3), 621-640.

Guridi, I. (2024). Contaminated Sites Remediation under Uncertainty: Adaptive Strategies for Project Optimization [Thèse de doctorat – Université Michel de Montaigne Bordeaux 3]

Guyonnet, D., Dubois, D., de Fouquet, C., Ghestem, J.-P., Leprond, H., Atteia, O., Côme, J.-M., (2019). Prise en compte des notions d'incertitude dans la gestion des sites et sols pollués. Etat des lieux et perspectives. Rapport ADEME. Collection Expertises. Octobre 2019. 61 pages

Hayley, K., Valenza, A., White, E., Hutchison, B., et Schumacher, J. (2019). Application of the iterative ensemble smoother method and cloud computing: A groundwater modeling case study. *Water*, *11*(8), 1649.

He, Y., Ou, G. Z., Zhang, Z., Shen, Z. T., Wei, H., Ding, X. H., ... et Ye, W. M. (2024). On-site monitoring and numerical simulation on groundwater flow and pollution plume evolution in a hexavalent-chromium contaminated site. Journal of Hazardous Materials, 479, 135662.

Healy, R.W., (2010). Estimating Groundwater Recharge. Cambridge University Press, Cambridge, MA.

Homma, T., Saltelli, A., (1996). Importance measures in global sensitivity analysis of nonlinear models. Reliab. Eng. Syst. Saf. 52, 1–17. https://doi.org/10.1016/0951-8320(96)00002-6.

Huan, H., Hu, L., Yang, Y., Jia, Y., Lian, X., Ma, X., ... et Xi, B. (2020). Groundwater nitrate pollution risk assessment of the groundwater source field based on the integrated numerical simulations in the unsaturated zone and saturated aquifer. Environment international, 137, 105532.

Huang, W., Peng, P. A., Yu, Z., et Fu, J. (2003). Effects of organic matter heterogeneity on sorption and desorption of organic contaminants by soils and sediments. Applied geochemistry, 18(7), 955-972.

Hunt, R.J., Steuer, J.J., Mansor, M.T.C., Bullen, T.D. (2001). Delineating a recharge area for a spring using numerical modeling, Monte Carlo techniques, and geochemical investigation. Groundwater 39 (5), 702e712. doi.org/10.1111/j.1745-6584.2001.tb02360.x.

HydroGeoLogic, Inc. (1996) MODFLOW-SURFACT, A Comprehensive MODFLOW-Based Flow and Transport Simulator, Version 2.1 HydroGeoLogic, Inc,

IAEA. (2004). Soil sampling for environmental contaminants. IAEA Tec-Doc-1415.

INERIS. (2020). Prise en compte des incertitudes dans le cadre de dossiers conduits en contexte de sites et sols pollués, Ineris-20-177730-1814933-1.0

Isaaks, E.H. and Srivastava, R.M. (1989). Applied Geostatistics. Oxford University Press, New York, 561 pp.

Journel, A. G., Kyriakidis, P. C., et Mao, S. (2000). Correcting the smoothing effect of estimators: a spectral postprocessor. *Mathematical geology*, *32*, 787-813.

Jia, X., Cao, Y., O'Connor, D., Zhu, J., Tsang, D. C., Zou, B., et Hou, D. (2021). Mapping soil pollution by using drone image recognition and machine learning at an arsenic-contaminated agricultural field. *Environmental Pollution*, *270*, 116281.

Jury, W. A., Spencer, W. F., et Farmer, W. (1983). Behavior assessment model for trace organics in soil: I. Model description (Vol. 12, No. 4, pp. 558-564). American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America.

Kalman, D. (1996). A singularly valuable decomposition—The SVD of a matrix: College Mathematics Journal, v. 27, no. 1, p. 2–23

Karapanagioti, H. K., Gaganis, P., et Burganos, V. N. (2003). Modeling attenuation of volatile organic mixtures in the unsaturated zone: codes and usage. *Environmental modelling et software*, *18*(4), 329-337.

Kuntz, D., Grathwohl, P. (2009). Comparison of steady-state and transient flow conditions on reactive transport of contaminants in the vadose soil zone. *Journal of hydrology*, *369*(3-4), 225-233.

Lahvis, M. A., et Baehr, A. L. (1997). Documentation of R-UNSAT, a computer model for the simulation of reactive, multispecies transport in the unsaturated zone (No. 97-630). US Geological Survey.

Lambot, S, Javaux, M, Hupet, F, and Vanclooster, M, (2022). A global multilevel coordinate search procedure for estimating the unsaturated soil hydraulic properties, Water Resour. Res., 38(11), 1224, doi:10.1029/2001WR001224.

Lavenue, M., et De Marsily, G. (2001). Three-dimensional interference test interpretation in a fractured aquifer using the pilot point inverse method. *Water Resources Research*, *37*(11), 2659-2675.

Le Corfec, Y. (2011). *Sites et sols pollués : gestion des passifs environnementaux*. Dunod.

Le Coz, M, Pannecoucke, L, Freulon, X, Cazala, C, de Fouquet, C. (2020). Combining geostatistics and physically-based flow-and-transport simulations to characterize contaminated soils. EGU General Assembly 2020, EGU 2020, VIENNE, Austria. ff10.5194/egusphere-egu2020- 5486ff. ffhal-02929850ff

Leij, F. J., Skaggs, T. H., et Van Genuchten, M. T. (1991). Analytical solutions for solute transport in three-dimensional semi-infinite porous media. Water resources research, 27(10), 2719-2733.

Li, J., et Heap, A. D. (2008). A review of spatial interpolation methods for environmental scientists.

Li, J., et Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. Environmental Modelling et Software, 53, 173-189.

Li, S., Xi, W., Li, C., et Bi, T. (2020, March). Study on pollutant model construction and three-dimensional spatial interpolation in soil environmental survey. In IOP Conference Series: Earth and Environmental Science (Vol. 467, No. 1, p. 012154). IOP Publishing.

Li, Z., Tao, H., Zhao, D., et Li, H. (2022). Three-dimensional empirical Bayesian kriging for soil PAHs interpolation considering the vertical soil lithology. Catena, 212, 106098.

Li, Y., Sun, Q., Fu, Y., et Wei, J. (2025). Solving the Richards infiltration equation by coupling physics-informed neural networks with Hydrus-1D. *Scientific Reports*, *15*(1), 18649.

Lion, F., Colombano, S., Aubert, N., Boissard, G**.** (2015) – Définir une stratégie de dépollution : approche basée sur la masse de polluant et la capacité de relargage d'une pollution. Rapport final. BRGM/RP-64350-FR, 215 p., 91 Fig., 13 tabl., 11 ann.

Liu, X., Ma, E., Zhang, Y. K., et Liang, X. (2021). An analytical model of vapor intrusion with fluctuated water table. Journal of Hydrology, 596, 126085.

Lu, Z., et Zhang, D. (2004). Analytical solutions to steady state unsaturated flow in layered, randomly heterogeneous soils via Kirchhoff transformation. Advances in water resources, 27(8), 775-784.

Mariot, M. (2005). (INRS), Solvants chlorés, Données d'entrée des modèles analytiques ou numériques de transferts dans les sols et les eaux souterraines : Synthèse bibliographique relative aux paramètres Kd (sorption) et t1/2 (biodégradation), rapport d'étude (ref. INERIS-DRC-08-94669-08132A), Programme Transpol,

Martínez-Estudillo, A., Martínez-Estudillo, F., Hervás-Martínez, C., et García-Pedrajas, N. (2006). Evolutionary product unit based neural networks for regression. Neural Networks, 19(4), 477-486.

Ministère de la transition écologique et solidaire, 2017. *Méthodologie nationale de gestion des sites et sols pollués*.

Moore, S, McLaughlin, D. (1980). Mapping contaminated soil plumes by kriging. In: EPA national conference on management of uncontrolled hazardous wastes (Environmental Protection Agency Report). Resource Management Associates, Lafayette, California.

Moore, C., Doherty, J., (2005). Role of the calibration process in reducing model predictive error. Water Resources Research 41 (5), W05020. doi.org/10.1029/2004WR003501.

Müller, S., Schüler, L., Zech, A., et Heße, F. (2022). GSTools v1. 3: a toolbox for geostatistical modelling in Python. *Geoscientific Model Development*, *15*(7), 3161-3182.

Nemes, A. D., Schaap, M. G., Leij, F. J., et Wösten, J. H. M. (2001). Description of the unsaturated soil hydraulic database UNSODA version 2.0. *Journal of hydrology*, *251*(3-4), 151-162.

Oliver, M. A., et Webster, R. (2014). A tutorial guide to geostatistics: Computing and modelling variograms and kriging. Catena, 113, 56-69.

Oliver, M.A., Gregory, P.J., (2015). Soil, food security and human health: a review. Eur. J. Soil Sci. 66, 257–276. https://doi.org/10.1111/ejss.12216.

Panday, S., 2024. USG-transport version 2.3.0: transport and other enhancements to MODFLOW-USG, GSI environmental. http://www.gsi-net.com/en/software/free-software/USGTransport.html.

Pannecoucke, L. (2020). Combinaison de la géostatistique et de simulations à base physique – application à la caractérisation de panaches de contaminants. Thèse de doctorat, Université de Paris Sciences Lettres, Mines ParisTech

Paladino, O., Moranda, A., Massabò, M., et Robbins, G. A. (2018). Analytical Solutions of Three-Dimensional Contaminant Transport Models with Exponential Source Decay. Groundwater, 56(1), 96-108.

Parker, J. C., et Van Genuchten, M. T. (1984). Determining transport parameters from laboratory and field tracer experiments.

Parkhurst, D.L., et Appelo, C.A.J. (2013). Description of input and examples for PHREEQC version 3—A computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations: U.S. Geological Survey Techniques and Methods, book 6, chap. A43, 497 p.

Pellet, M., et Laville-Timsit, L. (1993). Echantillonnage de sols pour caractérisation d'une pollution : guide méthodologique. Ministère de l'Environnement, rapport BRGM, Service géologique national.

Pérez-Lucas, G., Vela, N., El Aatik, A., Navarro, S., 2019. Environmental risk of groundwater pollution by pesticide leaching through the soil profile, in: Larramendy, M.L., Soloneski, S. (Eds.), Pesticides-Use and Misuse and Their Impact in the Environment. IntechOpen, Croatia, Balkans, pp. 45–72.

Piotrowski, A. P., et Napiorkowski, J. J. (2011). Optimizing neural networks for river flow forecasting–Evolutionary Computation methods versus the Levenberg–Marquardt approach. Journal of hydrology, 407(1-4), 12-27.

Potrykus, D., Gumuła-Kawęcka, A., Jaworska-Szulc, B., Pruszkowska-Caceres, M., et Szymkiewicz, A. (2018). Assessing groundwater vulnerability to pollution in the Puck region (denudation moraine upland) using vertical seepage method.

Pu, H. F., Min, M., Deng, S. Y., Wen, X. J., et Xu, J. (2024). Transient analytical solution for coupled water–gas transport in unsaturated soil cover of landfill. Journal of Hydrology, 131515.

Rajabi Naraki, M. (2024). Development of a numerical solution for transport simulation in unsaturated porous media (Doctoral dissertation, Politecnico di Torino).

Ramsey, C. A., et Hewitt. A. D. (2005). A methodology for assessing sample representativeness. Environmental Forensics, 6(1), 71-75.

RECORD. (2013). Retour d'expérience critique sur l'utilisation de méthodes géostatistiques pour la caractérisation des sites et sols pollués. 135 pages, n°11-0514/1A. https://recordnet.org/storage/etudes/11-0514-1A/rapport/Rapport_record11-0514_1A.pdf

Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. physics, 1(5), 318-333.

Rumbaugh, J.O., Rumbaugh, D., (2024). Guide to Using Groundwater Vista, Version 9. Environmental Simulations, Inc., Reinholds, PA.

Saltelli, A., 2002. Making best use of model evaluations to compute sensitivity indices. Comput. Phys. Commun. 145, 280–297. https://doi.org/10.1016/S0010-4655(02)00280-1.

Schaap, M. G, Leij, F.J, et van Genuchten, M.T. (2001). ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. Journal of Hydrology, 251(3-4): 163–176, 2001.doi: 10.1016/S0022-1694(01)00466-8.

Schaap, M. G., Nemes, A., et Van Genuchten, M. T. (2004). Comparison of models for indirect estimation of water retention and available water in surface soils. Vadose Zone Journal, 3(4), 1455-1463.

Shu, X., Wu, Y., Zhang, X., et Yu, F. (2023). Experiments and models for contaminant transport in unsaturated and saturated porous media–A review. Chemical Engineering Research and Design, 192, 606-621.

Šimůnek, J., Van Genuchten, M. T., et Sejna, M. (1998). The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media. International Ground Water Modeling Center, Colorado School of Mines, Golden, CO 202 p

Simunek, J., et van Genuchten, M. T. (1999). Using the HYDRUS-1D and HYDRUS-2D codes for estimating unsaturated soil hydraulic and solute transport parameters. *Characterization and measurement of the hydraulic properties of unsaturated porous media*, *1*, 523-1.

Šimůnek, J., Van Genuchten, M. T., Sejna, M., Toride, N., et Leij, F. J. (1999a). The STANMOD computer software for evaluating solute transport in porous media using analytical solutions of convection-dispersion equation. Versions 1.0 and 2.0. International Ground Water Modeling Center.

Šimůnek, J., Šejna, M., Van Genuchten, M.T. (1999b). The HYDRUS-2D Software Package for Simulating Two-Dimensional Movement of Water, Heat, and Multiple Solutes in Variable

Saturated Media. Version 2.0. IGWMCTPS-53. Int. Ground Water Modeling Center, Colorado School of Mines, Golden, CO.

Šimůnek, J., Van Genuchten, M. T., et Sejna, M. (2005). The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media. University of California-Riverside Research Reports, 3, 1-240.

Šimůnek, J. (2006). Models of water flow and solute transport in the unsaturated zone. Encyclopedia of hydrological sciences.

Siregar, R., Tanjung, M., Roselina, T., et Arbi, Y. (2024). Analysis Distribution of Iron (Fe) and Hexavalent Chromium (Cr6+) Contaminants in Shallow Groundwater Due To Leachate Around Landfill Terjun, Medan City. In E3S Web of Conferences (Vol. 519, p. 03022). EDP Sciences.

Soboĺ, I., (1993). Sensitivity estimates for nonlinear mathematical models. Math. Model. Comput. Exp. 1, 407–414.

Sousa, M. R., Jones, J. P., Frind, E. O., et Rudolph, D. L. (2013). A simple method to assess unsaturated zone time lag in the travel time from ground surface to receptor. Journal of contaminant hydrology, 144(1), 138-151.

Spence, L. R., Walden, T., et Sunbury, U. K. (2001). RISC4. Risk integrated software for clean-ups. Bp-Amoco Oil, Sunbury, UK (464 pp.).

Stoppiello, M. G., Lofrano, G., Carotenuto, M., Viccione, G., Guarnaccia, C., et Cascini, L. (2020). A comparative assessment of analytical fate and transport models of organic contaminants in unsaturated soils. Sustainability, 12(7), 2949.

Srivastava, R., et Yeh, T. C. J. (1991). Analytical solutions for one-dimensional, transient infiltration toward the water table in homogeneous and layered soils. Water Resources Research, 27(5), 753-762.

Schotanus, D., van der Ploeg, M. J., & van der Zee, S. E. A. T. M. (2013). Spatial distribution of solute leaching with snowmelt and irrigation: measurements and simulations. Hydrology and Earth System Sciences, 17(4), 1547-1560.

Strack, O. D. (2003). Theory and applications of the analytic element method. Reviews of Geophysics, 41(2).

Szymkiewicz, A., Savard, J., et Jaworska-Szulc, B. (2019). Numerical analysis of recharge rates and contaminant travel time in layered unsaturated soils. Water, 11(3), 545.

Tao, H., Liao, X., Zhao, D., Gong, X., et Cassidy, D. P. (2019). Delineation of soil contaminant plumes at a co-contaminated site using BP neural networks and geostatistics. Geoderma, 354, 113878.

Tao, H., Liao, X., Cao, H., Zhao, D., et Hou, Y. (2022). Three-dimensional delineation of soil pollutants at contaminated sites: Progress and prospects. *Journal of Geographical Sciences*, *32*(8), 1615-1634.

Tikhonov, A. N, Goncharsky, A. V., Stepanov, V. V. Yagola, A. G, (1995), Numerical methods for the solution of ill-posed problems, Springer, Berlin, 1st edn., . https://doi.org/10.1007/978-94-015-8480-7

Tonkin, Matthew et Doherty, John. (2009). Calibration-constrained Monte Carlo analysis of highly-parameterized models using subspace methods. Water Resources Research. 45. 10.1029/2007WR006678.

Troldborg, M.; Binning, P.J.; Nielsen, S.; Kjeldsen, P.; Christensen, A.G. (2009). Unsaturated zone leaching models for assessing risk to groundwater of contaminated sites. J. Contam. Hydrol. 2009, 105, 28–37. doi.org/10.1016/j.jconhyd.2008.11.002

Tyson, H. N., et Weber, E. M. (1964). Use of electronic computer in the simulation of dynamic behaviour of groundwater basin. Proceedings of ASCE Water Resources Conference, Milwaukee, Wisconsin, USA.

Ukalska-Jaruga, A., Bejger, R., Smreczak, B., & Podlasiński, M. (2023). Sorption of organic contaminants by stable organic matter fraction in soil. Molecules, 28(1), 429.

Ünlü, K., Kemblowski, M. W., Parker, J. C., Stevens, D., Chong, P. K., et Kamil, I. (1992). A screening model for effects of land-disposed wastes on groundwater quality. Journal of contaminant hydrology, 11(1-2), 27-49.

UPDS. (2016). Pollution concentrée : Définition, outils de caractérisation et intégration dans la méthodologie nationale de gestion des sites et sols pollués (version mise à jour). Union professionnelle des entreprises de dépollution de sites.

United States. Environmental Protection Agency. Office of Emergency, & Remedial Response. (1996). *Soil screening guidance: Technical background document* (Vol. 9355). United States Environmental Protection Agency, Office of Solid Waste and Emergency Response.

Van Genuchten, M. T. (1980). closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Science Society of America Journal, 44(5): 892–898, doi: 10.2136/sssaj1980.03615995004400050002x.

Van Genuchten, M. T. (1981). Analytical solutions for chemical transport with simultaneous adsorption, zero-order production and first-order decay. Journal of hydrology, 49(3-4), 213-233.

Varado, N., Braud, I., Ross, P. J., et Haverkamp, R. (2006). Assessment of an efficient numerical solution of the 1D Richards' equation on bare soil. *Journal of Hydrology*, *323*(1-4), 244-257.

Viotti, P., Papini, M. P., Stracqualursi, N., et Gamba, C. (2005). Contaminant transport in an unsaturated soil: laboratory tests and numerical simulation model as procedure for parameters evaluation. Ecological modelling, 182(2), 131-148.

Vogel, T. (1987). SWMII: Numerical model of two-dimensional flow in a variably saturated porous medium. Res. Rep. 87. Dep. of Hydraul. and Catchment Hydrol., Agricultural Univ.

Voss, C. I., et Provost, A. M. (2010). SUTRA – A model for saturated-unsaturated, variable-density ground-water flow with solute or energy transport. U.S. Geological Survey Water-Resources Investigations Report 02-4231 (p. 291). U.S. Geological Survey.

Vrugt, J.A, Bouten, W et Weerts A.H. (2001). Information Content of Data for Identifying Soil Hydraulic Parameters from Outflow Experiments. Soil Science Society of America journal 65.1. 19–27.

Wang, S., Wang, J., Luo, Y., Shen, H., et Li, Y. (2023). Evaluation of heavy metal pollution with uneven spatial sampling distribution based on Voronoi area density. *Environmental Science and Pollution Research*, *30*(17), 50431-50443.

White, J. T. (2018). A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions. *Environmental Modelling et Software*, *109*, 191-201.

Wilson, J. L., et Miller, P. J. (1978). Two-dimensional plume in uniform ground-water flow. *Journal of the Hydraulics Division*, *104*(4), 503-514.

Winston, R. B. (2024). *Revision of ModelMuse to support the use of PEST software with MODFLOW and SUTRA models* (No. 6-A64). US Geological Survey.

Xie, Y., Chen,T., Lei, M., Yang, J., Guo, Q., Song, Bo., Zhou, X. (2011).Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: Accuracy and uncertainty analysis, Chemosphere, Volume 82, Issue 3, Pages 468-476, ISSN 0045-6535,

Yang, J., Tao, Y., Ren, W., et Cao, P. (2020). Numerical simulation of groundwater contaminant transport in unsaturated flow. Water Supply, 20(8), 3730-3738.

Yin, Y., et Sykes, J. F. (2006). Identification of Transport Parameters and Pollution Sources for a Physically Based Groundwater Contaminant Transport Model: A Comparison of Algorithms. In AGU Fall Meeting Abstracts (Vol. 2006, pp. H43B-0501).

Zhang, C. (2019). Soil and Groundwater Remediation: Fundamentals, Practices, and Sustainability. John Wiley et Sons. 1119393175, 9781119393177

Zhang, J., Peng, J., Chen, X., Shi, X., Feng, Z., Meng, Y., ... et Liu, Y. (2024). Comparative Study on Different Interpolation Methods and Source Analysis of Soil Toxic Element Pollution in Cangxi County, Guangyuan City, China. Sustainability, 16(9), 3545.

Zhang, Z., Wang, W., Yeh, T. C. J., Chen, L., Wang, Z., Duan, L., ... et Gong, C. (2016). Finite analytic method based on mixed-form Richards' equation for simulating water flow in vadose zone. Journal of Hydrology, 537, 146-156.

Zhang, Z., Wang, W., Gong, C., Yeh, T. C. J., Wang, Z., Wang, Y. L., et Chen, L. (2018). Finite analytic method for modeling variably saturated flows. Science of the Total Environment, 621, 1151-1162.

Zhou P, Zhao Y, Zhao Z, Chai T. (2015). Source mapping and determining of soil contamination by heavy metals using statistical analysis, artificial neural network, and adaptive genetic algorithm. Journal of Environmental Chemical Engineering.3(4):2569–2579

Site Géorisques (a), Pollution des sols -CASIAS https://www.georisques.gouv.fr/articles-risques/pollutions-sols-sis-anciens-sites-industriels/basias Consulté en août 2025.

Site Géorisques (b), Pollution des sols -CASIAS https://www.georisques.gouv.fr/donnees/bases-de-donnees/sites-et-sols-pollues-ou-potentiellement-pollues Consulté en août 2025.

Site de Santépubliquefrance, Pollution des sols https://www.santepubliquefrance.fr/determinants-de-sante/pollution-et-sante/sols Consulté en juin 2025

Site Infoterre, Méthodologie nationale des sites et sols pollués - https://ssp-infoterre.brgm.fr/fr/methodologie/methodologie-nationale-gestion-ssp Consulté en mai 2025

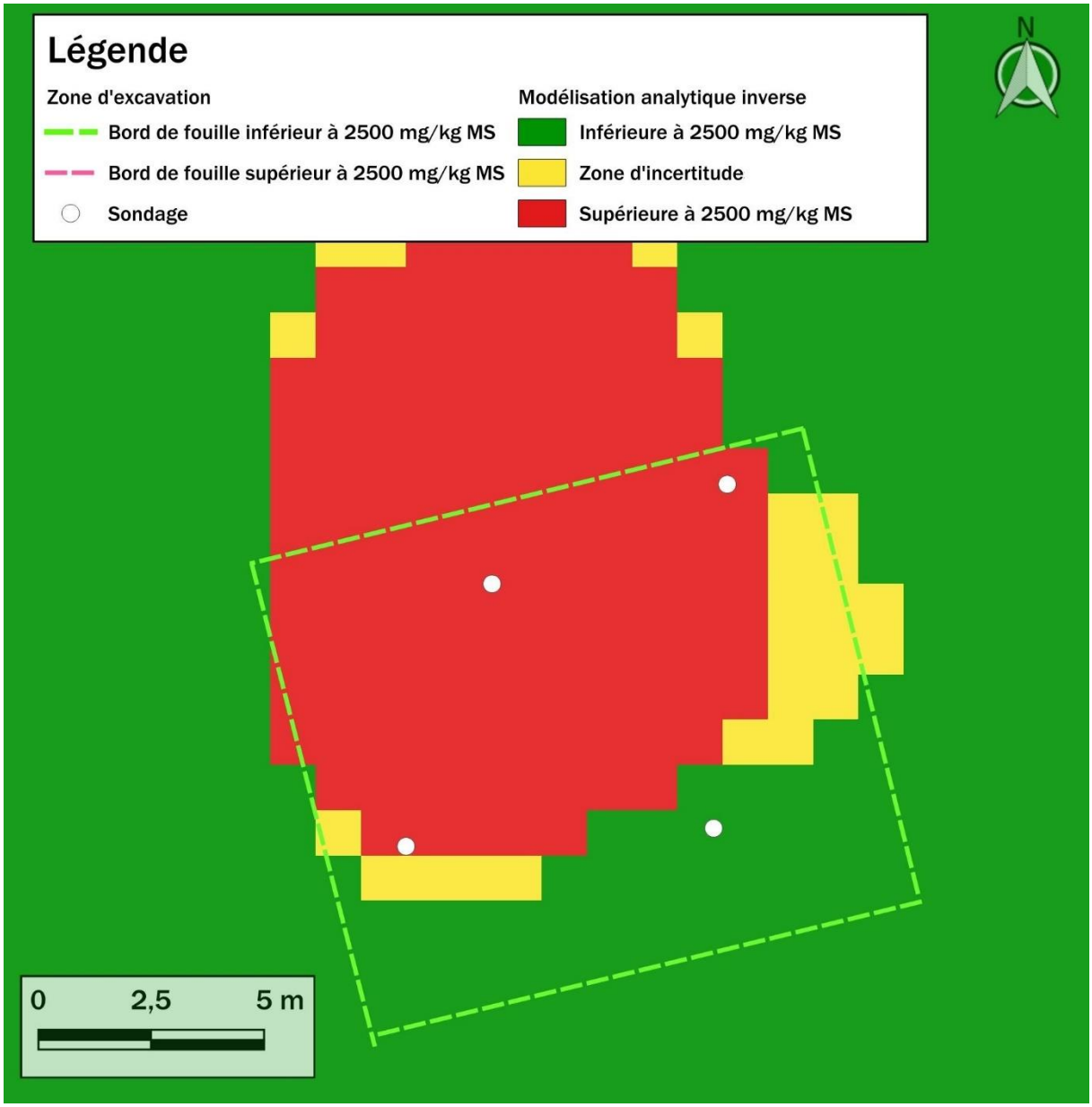
Site MétéoFrance, https://meteofrance.com/ Consulté en Février 2025

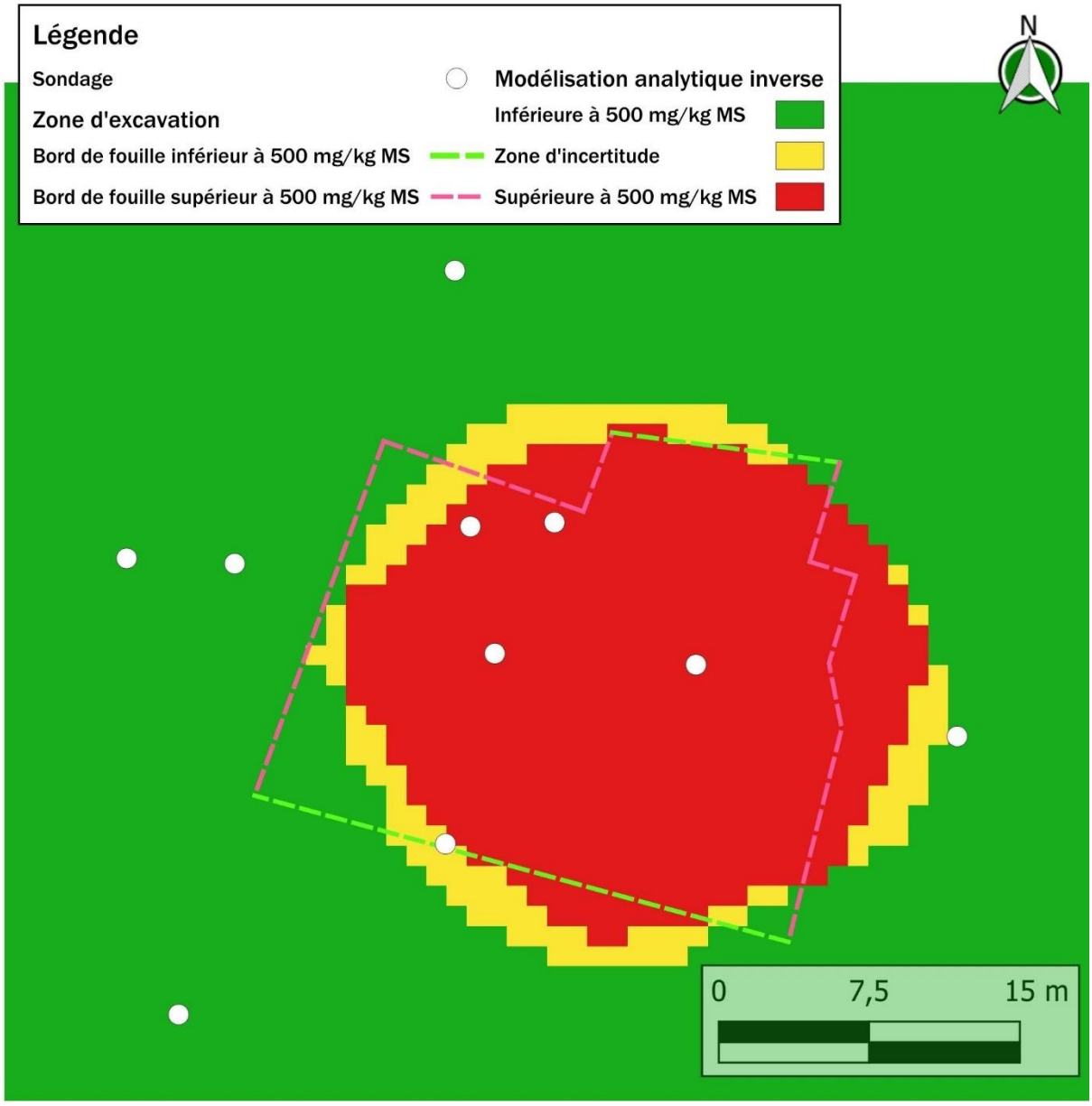
Site SIGES Aquitaine, https://sigesaqi.brgm.fr/ Consulté en Juillet 2025

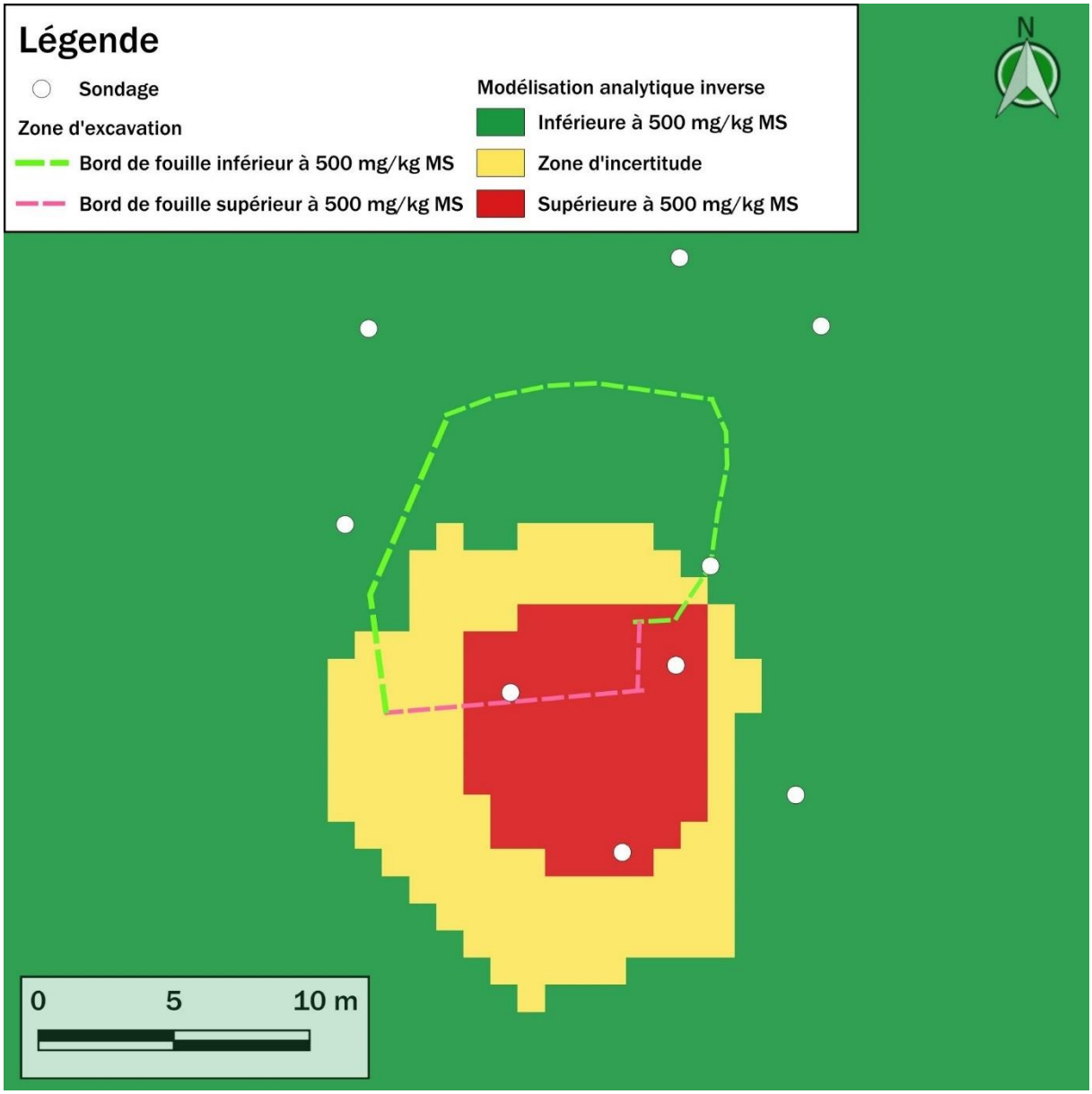
LISTE DES ANNEXES

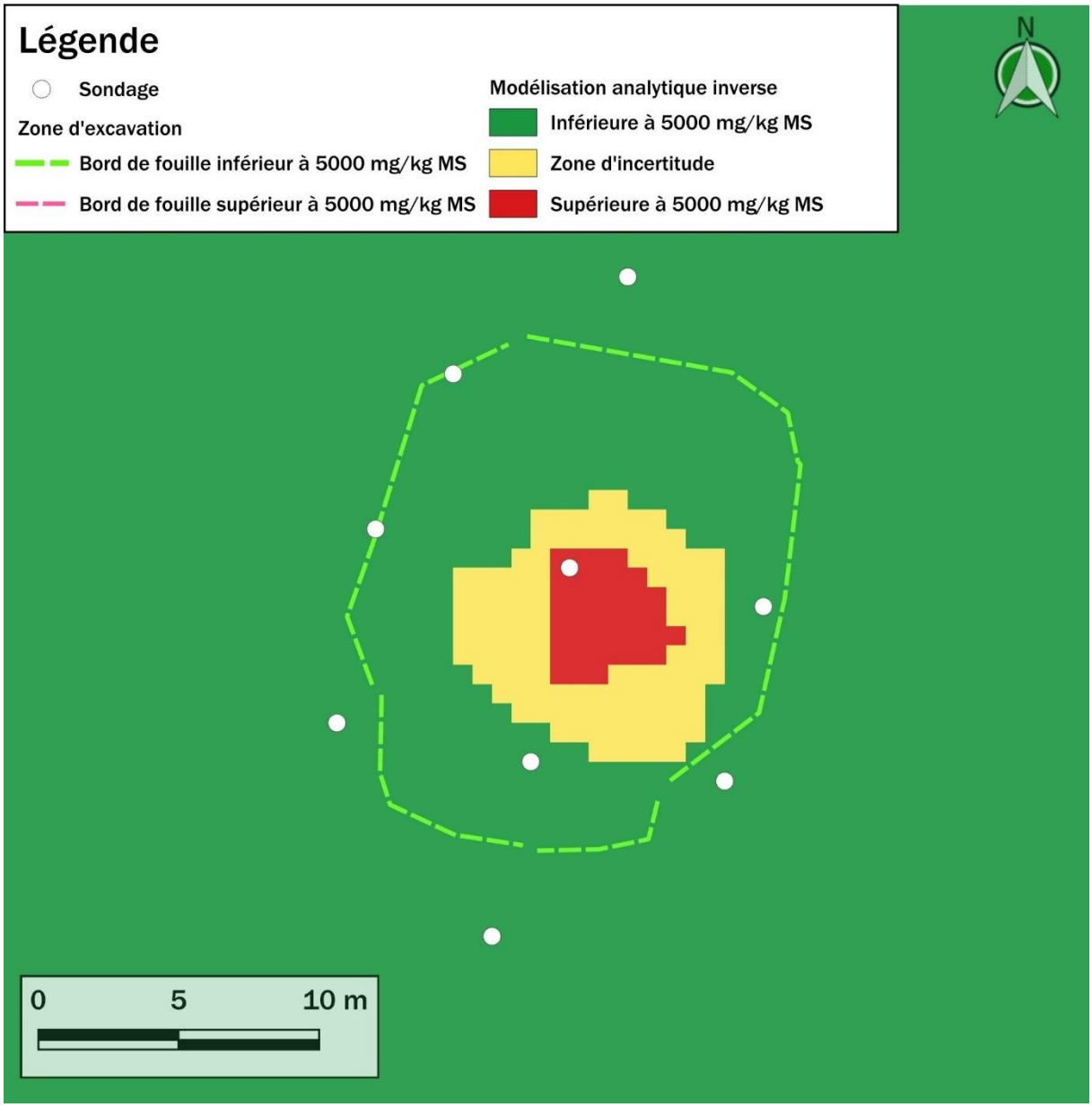
Annexe 1 : Cartographies des résultats de la modélisation analytique inverse pour les 32 sites de la base de données

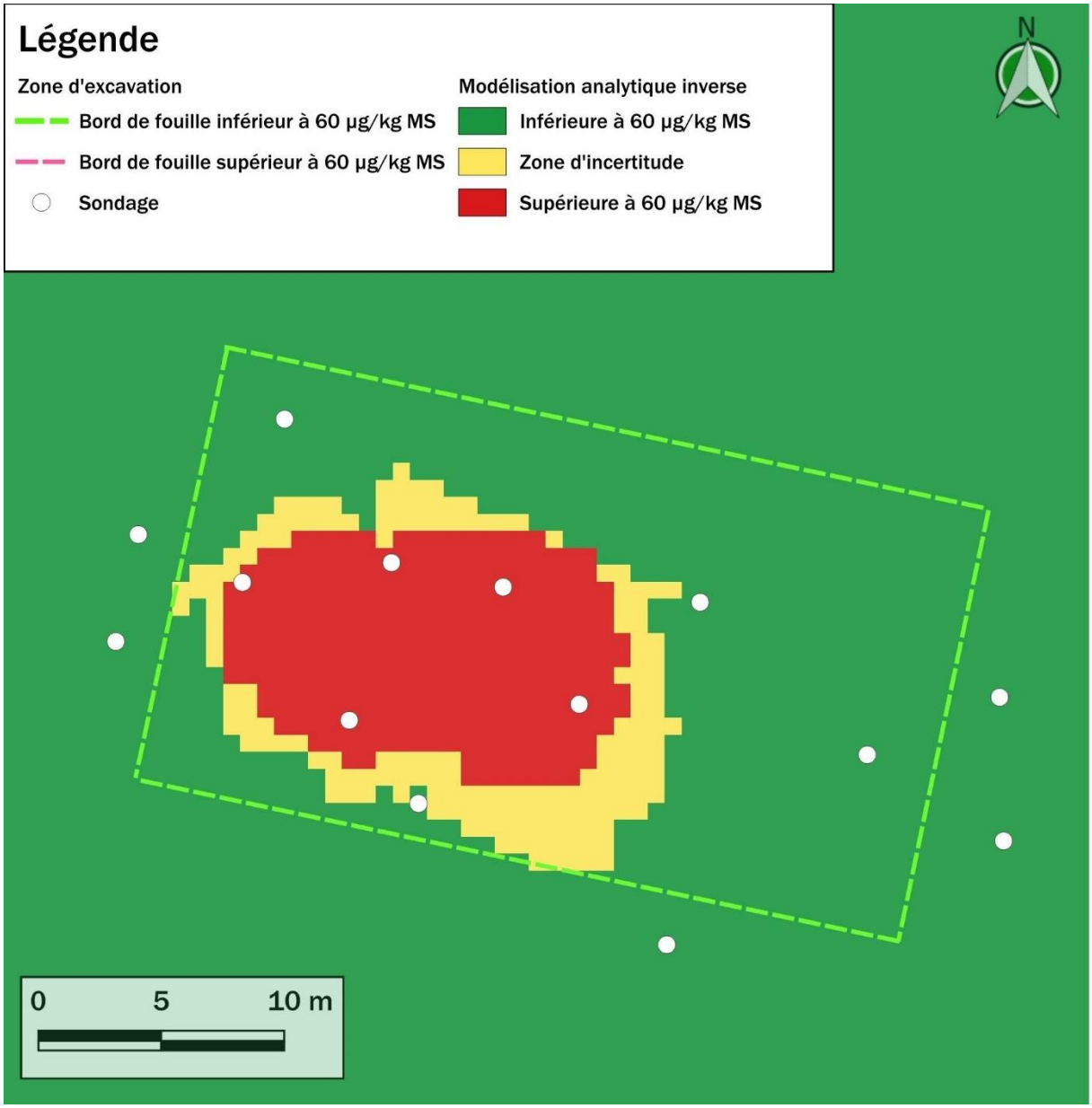
ANNEXE 1 Cartographies des résultats de la modélisation analytique inverse pour les 32 sites de la base de données

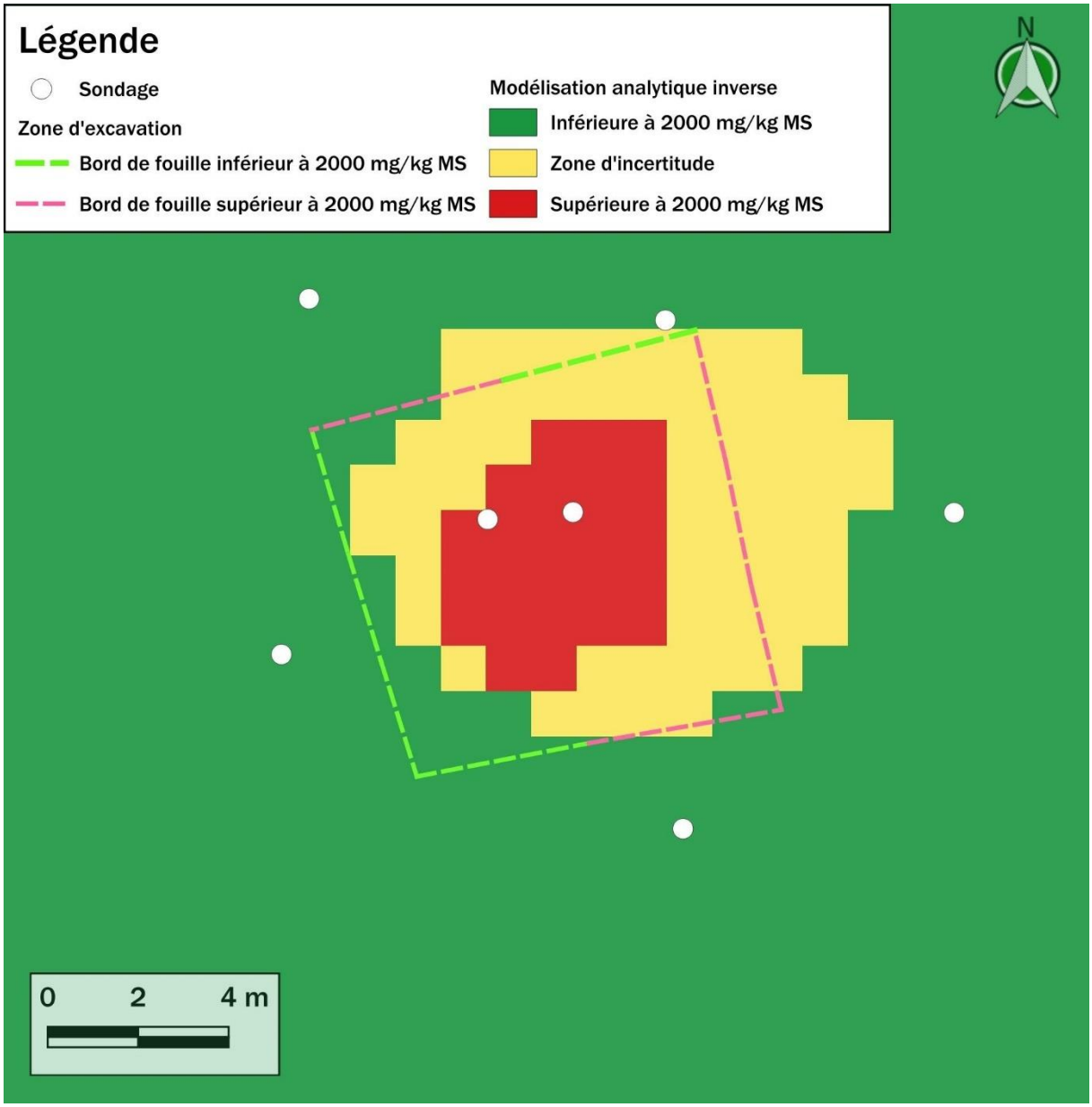

Légende
Zone d'excavation
Bord de fouille inférieur à 2500 mg/kg MS
Bord de fouille supérieur à 2500 mg/kg MS
Sondage
Modélisation analytique inverse
Inférieure à 2500 mg/kg MS
Zone d'incertitude
Supérieure à 2500 mg/kg MS
N
0
2,5
5 m

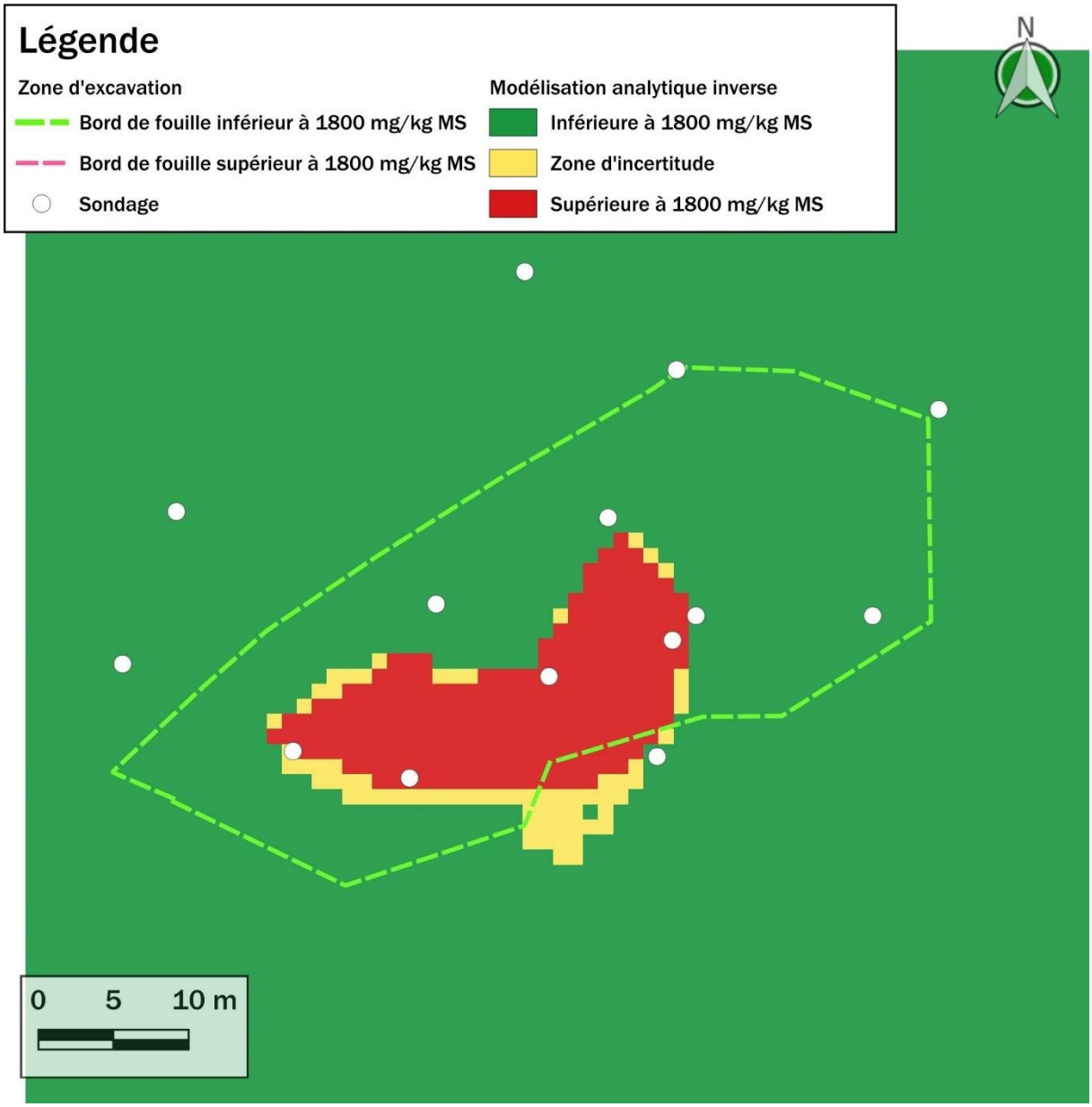

Légende
Sondage
Zone d'excavation
Bord de fouille inférieur à 500 mg/kg MS
Bord de fouille supérieur à 500 mg/kg MS
Modélisation analytique inverse
Inférieure à 500 mg/kg MS
Zone d'incertitude
Supérieure à 500 mg/kg MS
N
0
7,5
15 m

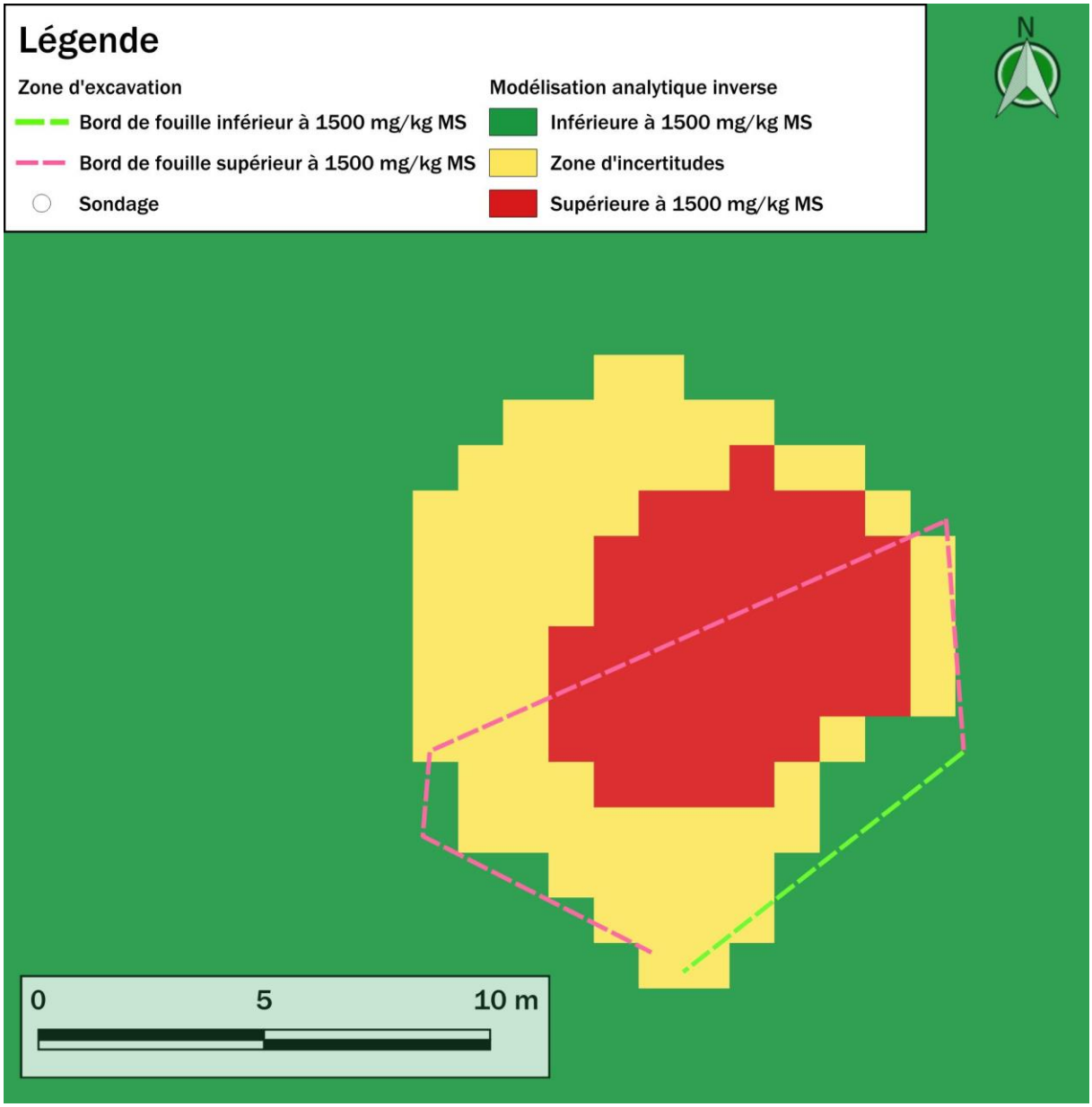

Légende
Sondage
Zone d'excavation
Bord de fouille inférieur à 500 mg/kg MS
Bord de fouille supérieur à 500 mg/kg MS
Modélisation analytique inverse
Inférieure à 500 mg/kg MS
Zone d'incertitude
Supérieure à 500 mg/kg MS
N
0
5
10 m

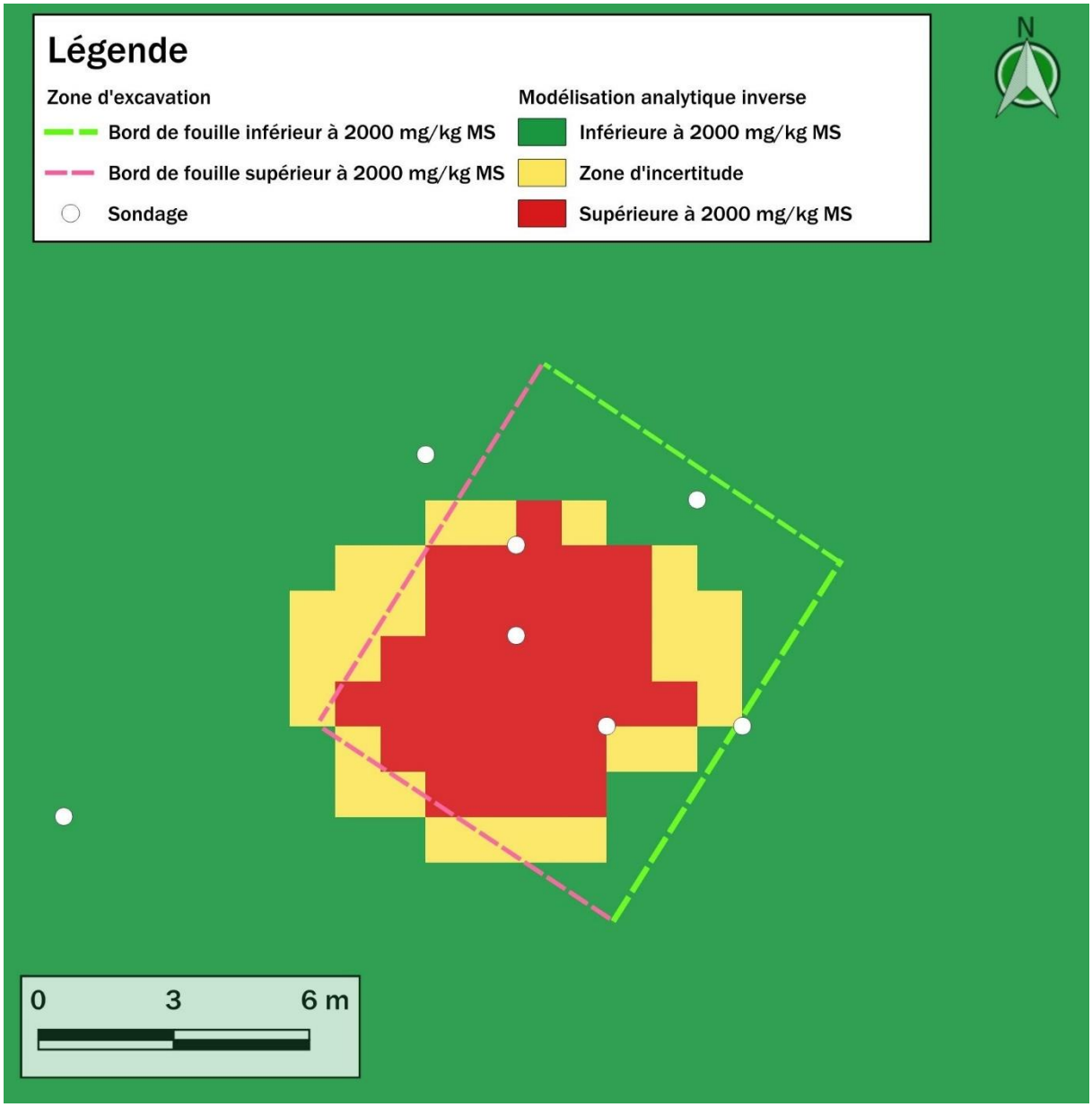

Légende
Sondage
Zone d'excavation
Bord de fouille inférieur à 5000 mg/kg MS
Bord de fouille supérieur à 5000 mg/kg MS
Modélisation analytique inverse
Inférieure à 5000 mg/kg MS
Zone d'incertitude
Supérieure à 5000 mg/kg MS
N
0
5
10 m

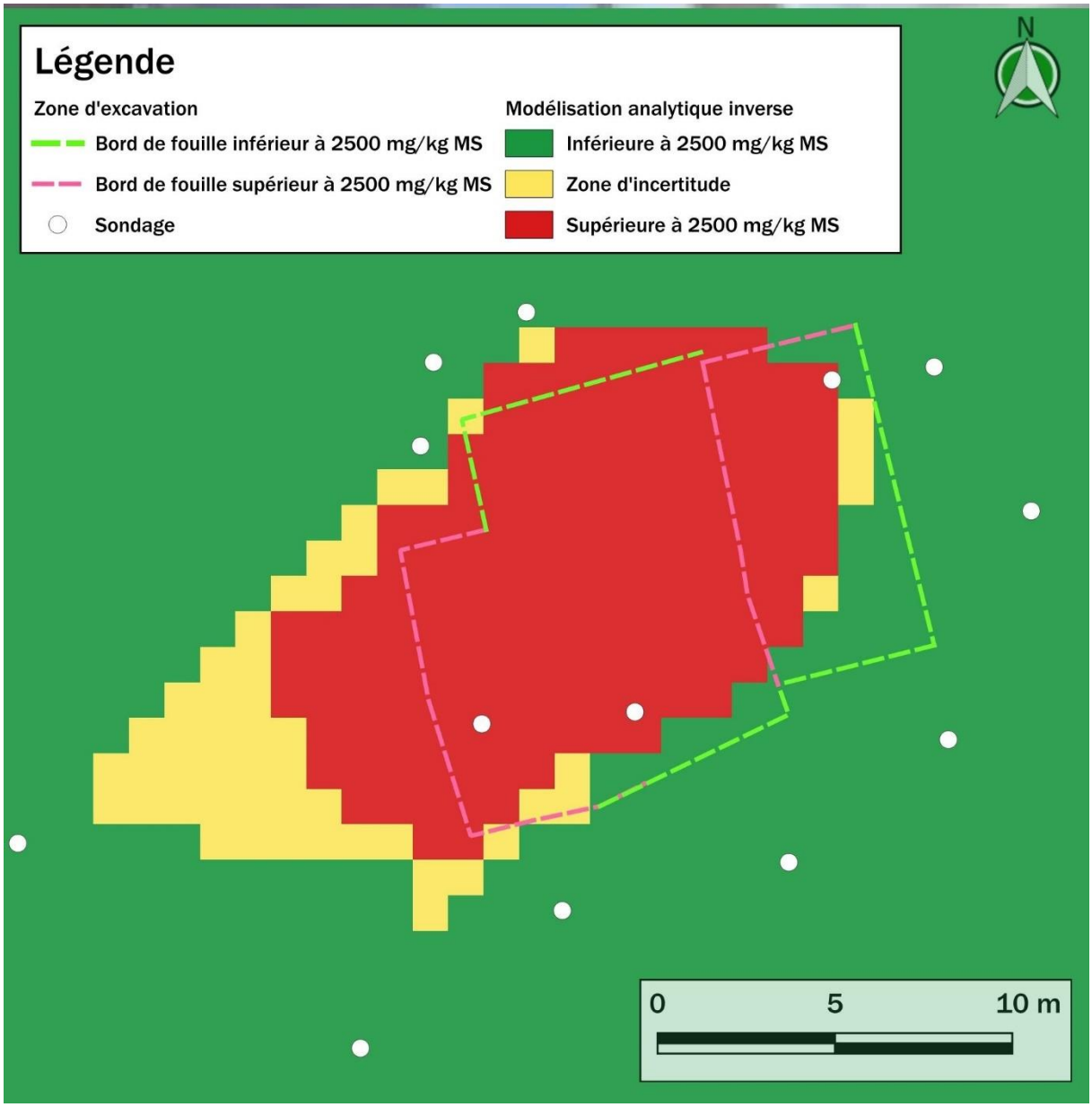

Légende
Zone d'excavation
Bord de fouille inférieur à 60 µg/kg MS
Bord de fouille supérieur à 60 µg/kg MS
Sondage
Modélisation analytique inverse
Inférieure à 60 µg/kg MS
Zone d'incertitude
Supérieure à 60 µg/kg MS
N
0
5
10 m

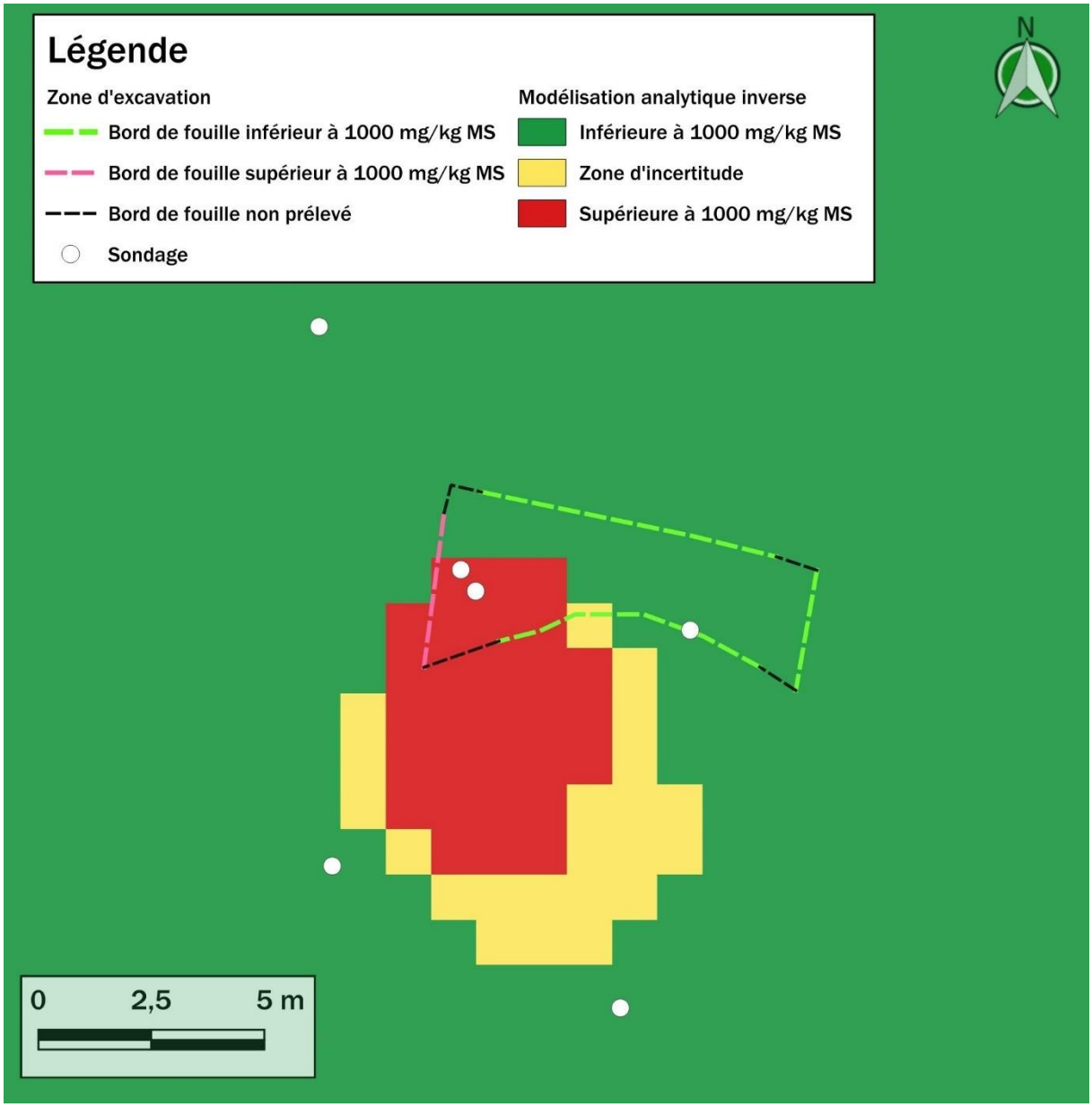

Légende
Sondage
Zone d'excavation
Bord de fouille inférieur à 2000 mg/kg MS
Bord de fouille supérieur à 2000 mg/kg MS
Modélisation analytique inverse
Inférieure à 2000 mg/kg MS
Zone d'incertitude
Supérieure à 2000 mg/kg MS
N
0
2
4 m

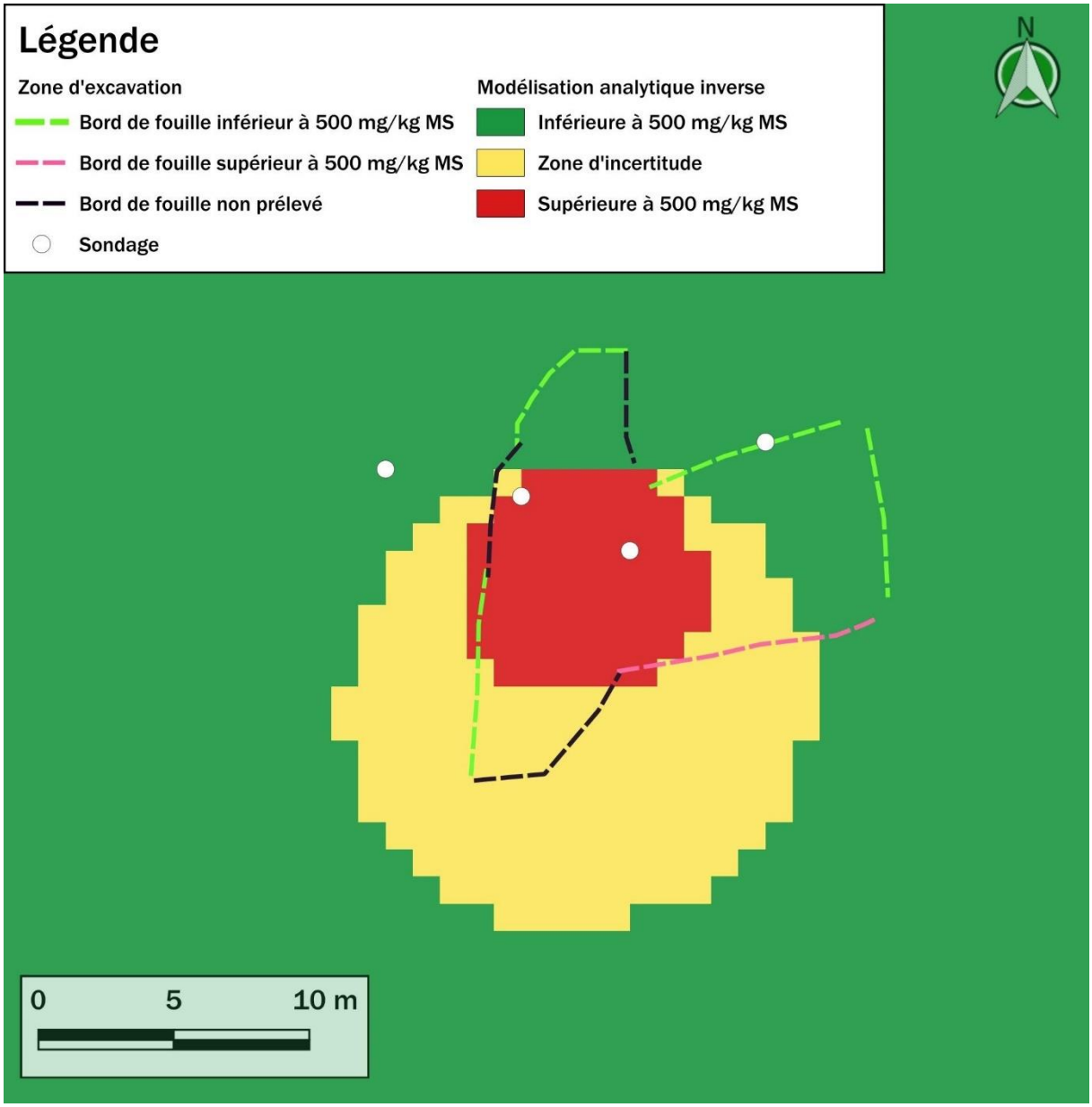

Légende
Zone d'excavation
Bord de fouille inférieur à 1800 mg/kg MS
Bord de fouille supérieur à 1800 mg/kg MS
Sondage
Modélisation analytique inverse
Inférieure à 1800 mg/kg MS
Zone d'incertitude
Supérieure à 1800 mg/kg MS
N
0
5
10 m

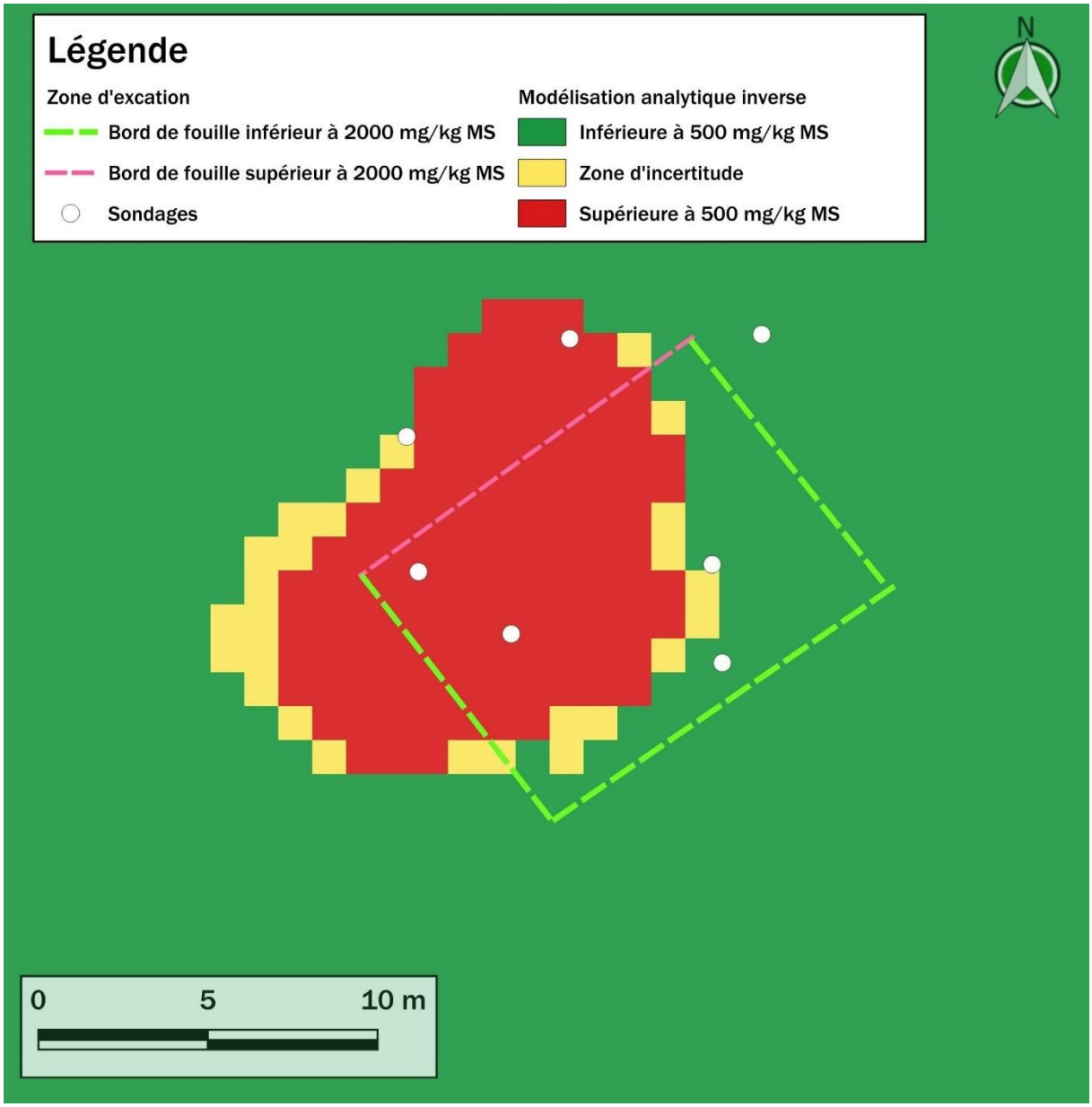

Légende
Zone d'excavation
Bord de fouille inférieur à 1500 mg/kg MS
Bord de fouille supérieur à 1500 mg/kg MS
Sondage
Modélisation analytique inverse
Inférieure à 1500 mg/kg MS
Zone d'incertitudes
Supérieure à 1500 mg/kg MS
N
0
5
10 m

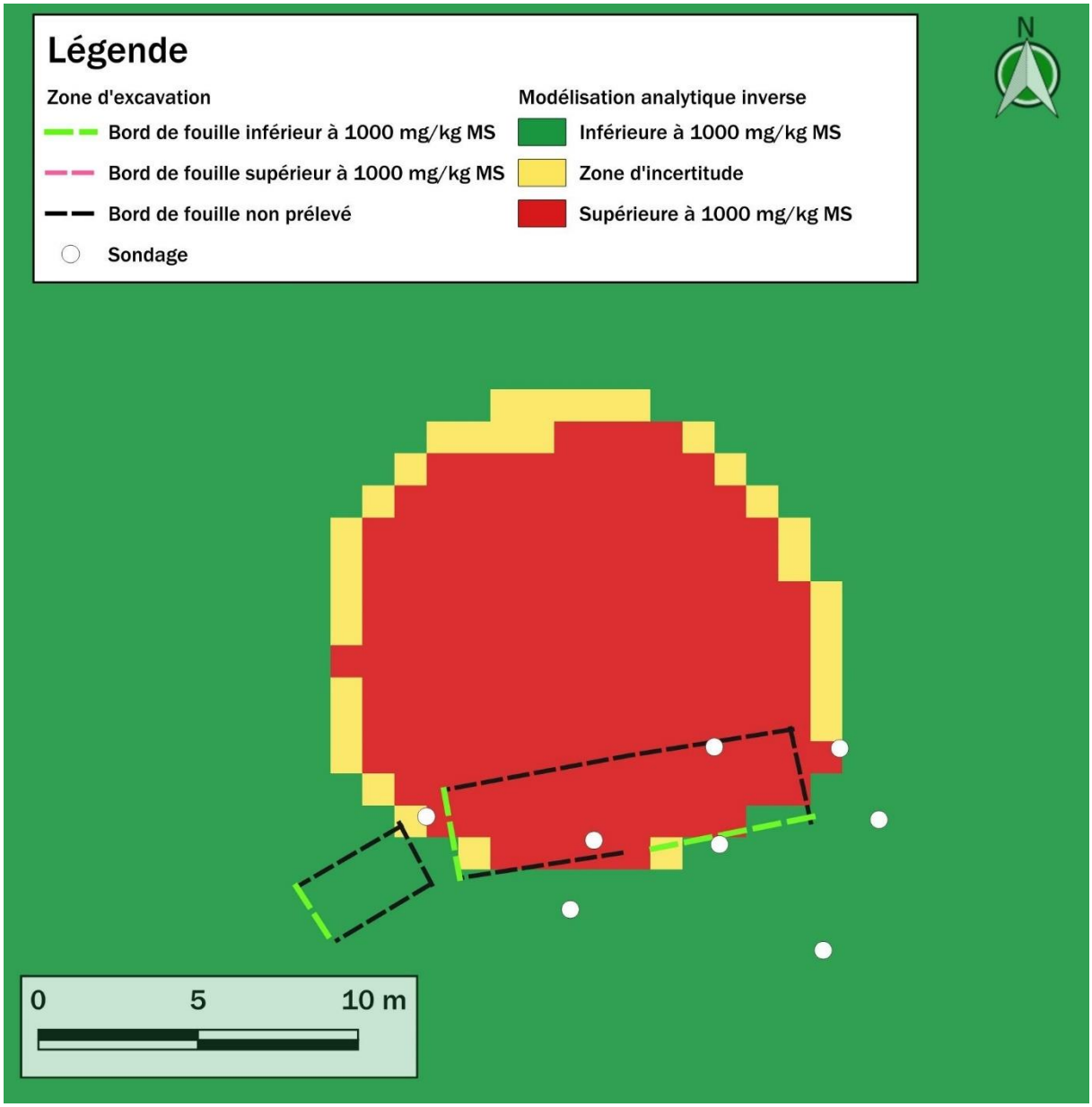

Légende
Zone d'excavation
Bord de fouille inférieur à 2000 mg/kg MS
Bord de fouille supérieur à 2000 mg/kg MS
Sondage
Modélisation analytique inverse
Inférieure à 2000 mg/kg MS
Zone d'incertitude
Supérieure à 2000 mg/kg MS
N
0
3
6 m

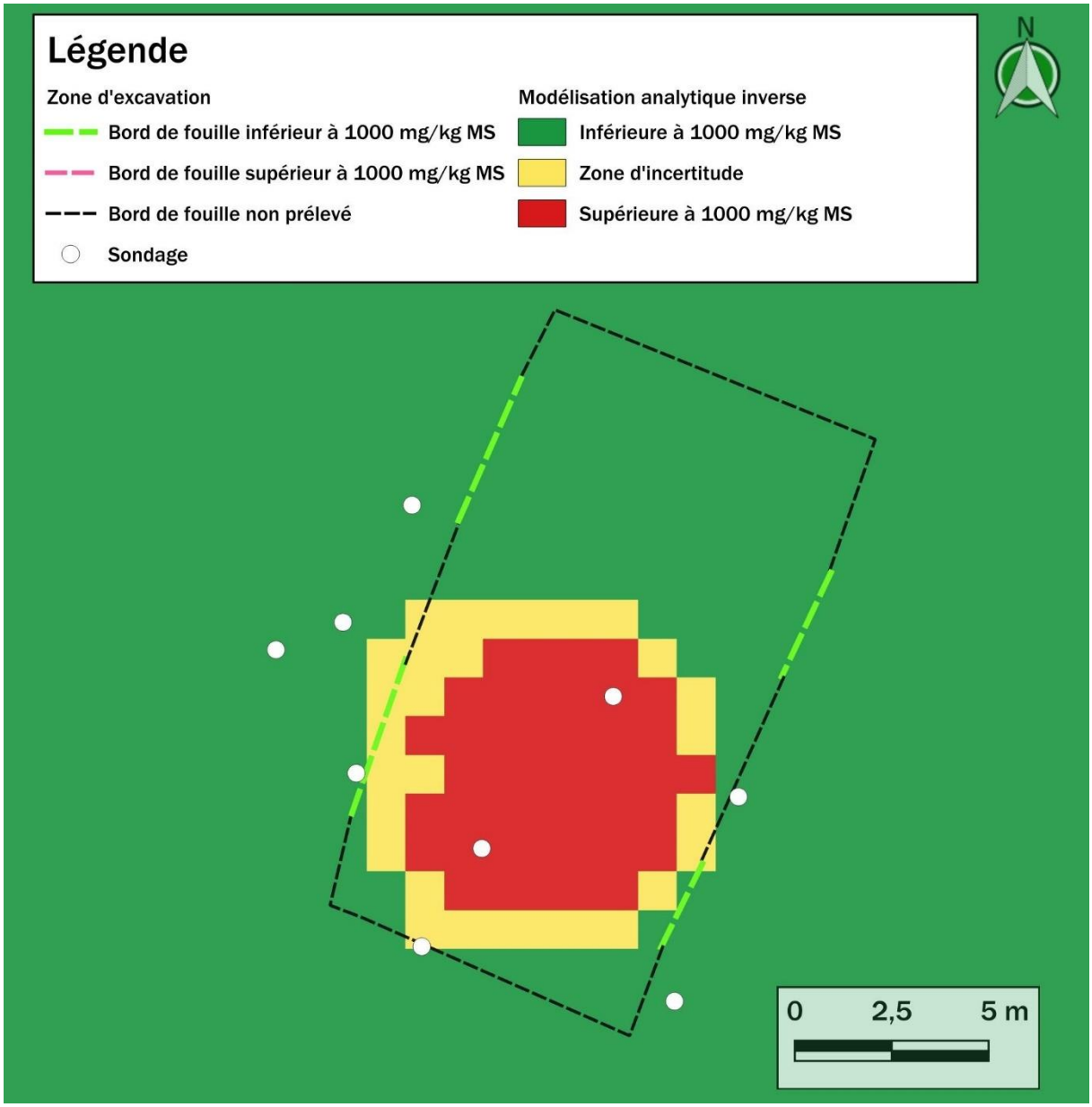

Légende
Zone d'excavation
Bord de fouille inférieur à 2500 mg/kg MS
Bord de fouille supérieur à 2500 mg/kg MS
Sondage
Modélisation analytique inverse
Inférieure à 2500 mg/kg MS
Zone d'incertitude
Supérieure à 2500 mg/kg MS
N
0
5
10 m

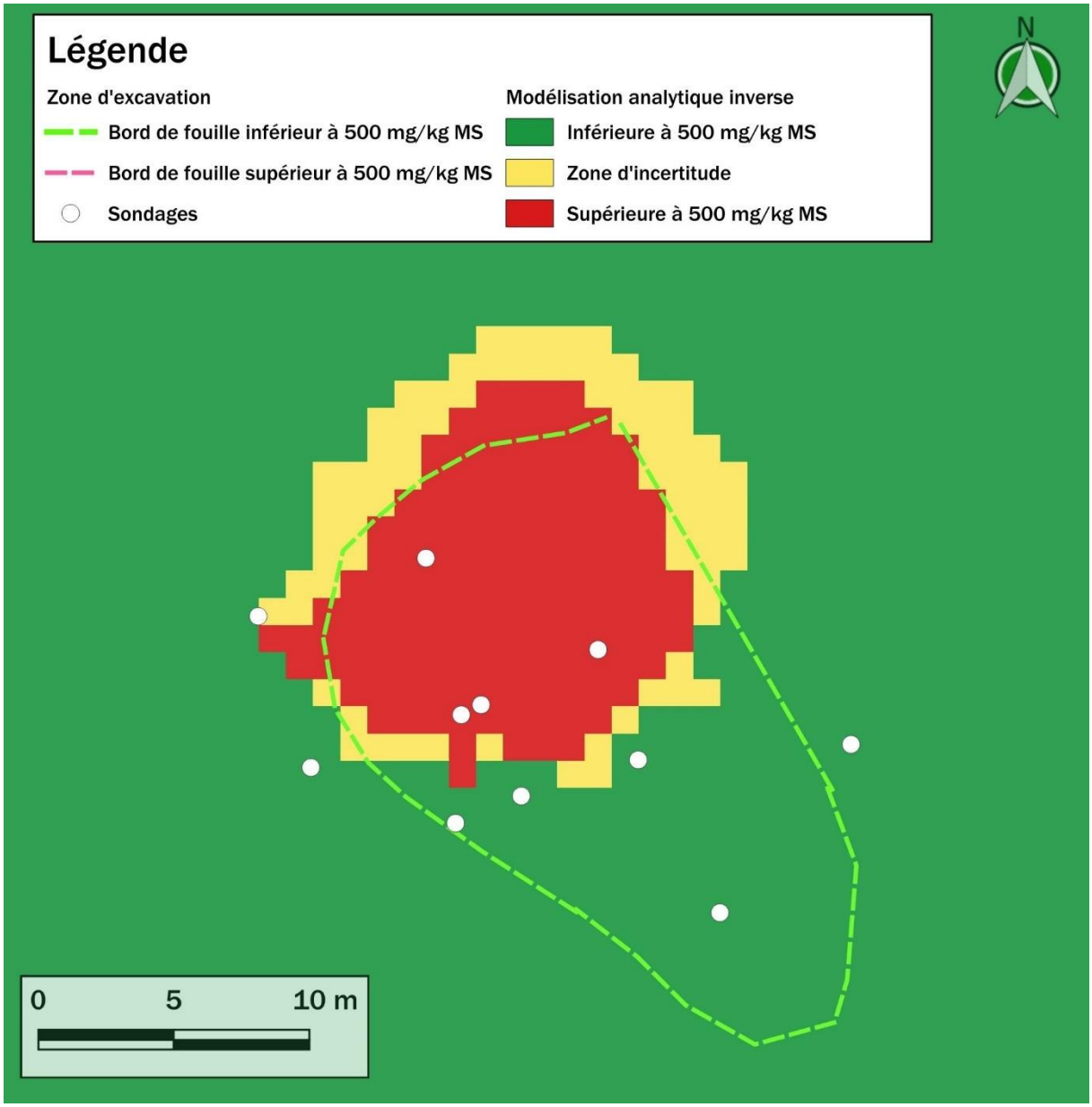

Légende
Zone d'excavation
Bord de fouille inférieur à 1000 mg/kg MS
Bord de fouille supérieur à 1000 mg/kg MS
Bord de fouille non prélevé
Sondage
Modélisation analytique inverse
Inférieure à 1000 mg/kg MS
Zone d'incertitude
Supérieure à 1000 mg/kg MS
N
0
2,5
5 m

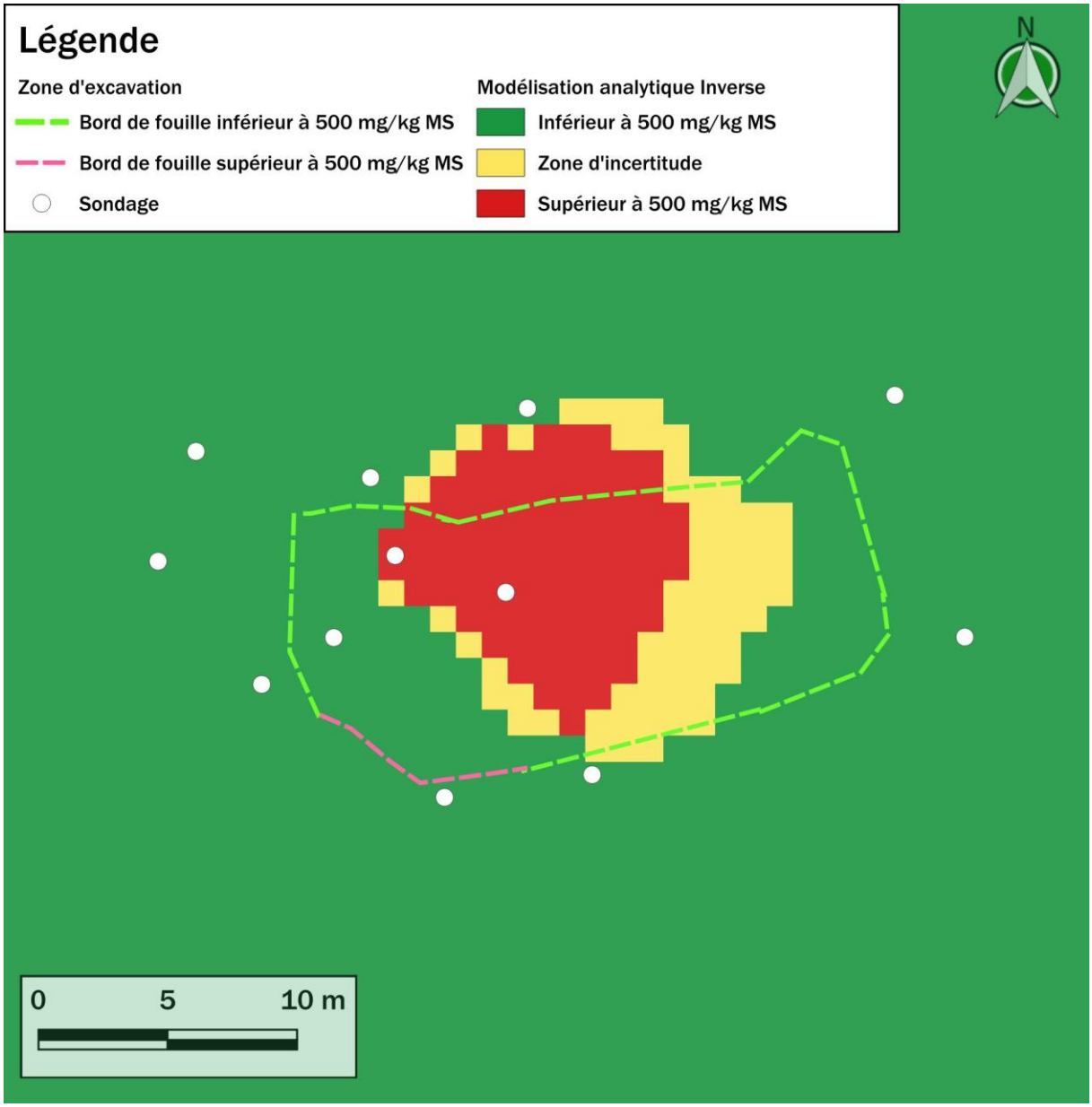

Légende
Zone d'excavation
Bord de fouille inférieur à 500 mg/kg MS
Bord de fouille supérieur à 500 mg/kg MS
Bord de fouille non prélevé
Sondage
Modélisation analytique inverse
Inférieure à 500 mg/kg MS
Zone d'incertitude
Supérieure à 500 mg/kg MS
N
0
5
10 m

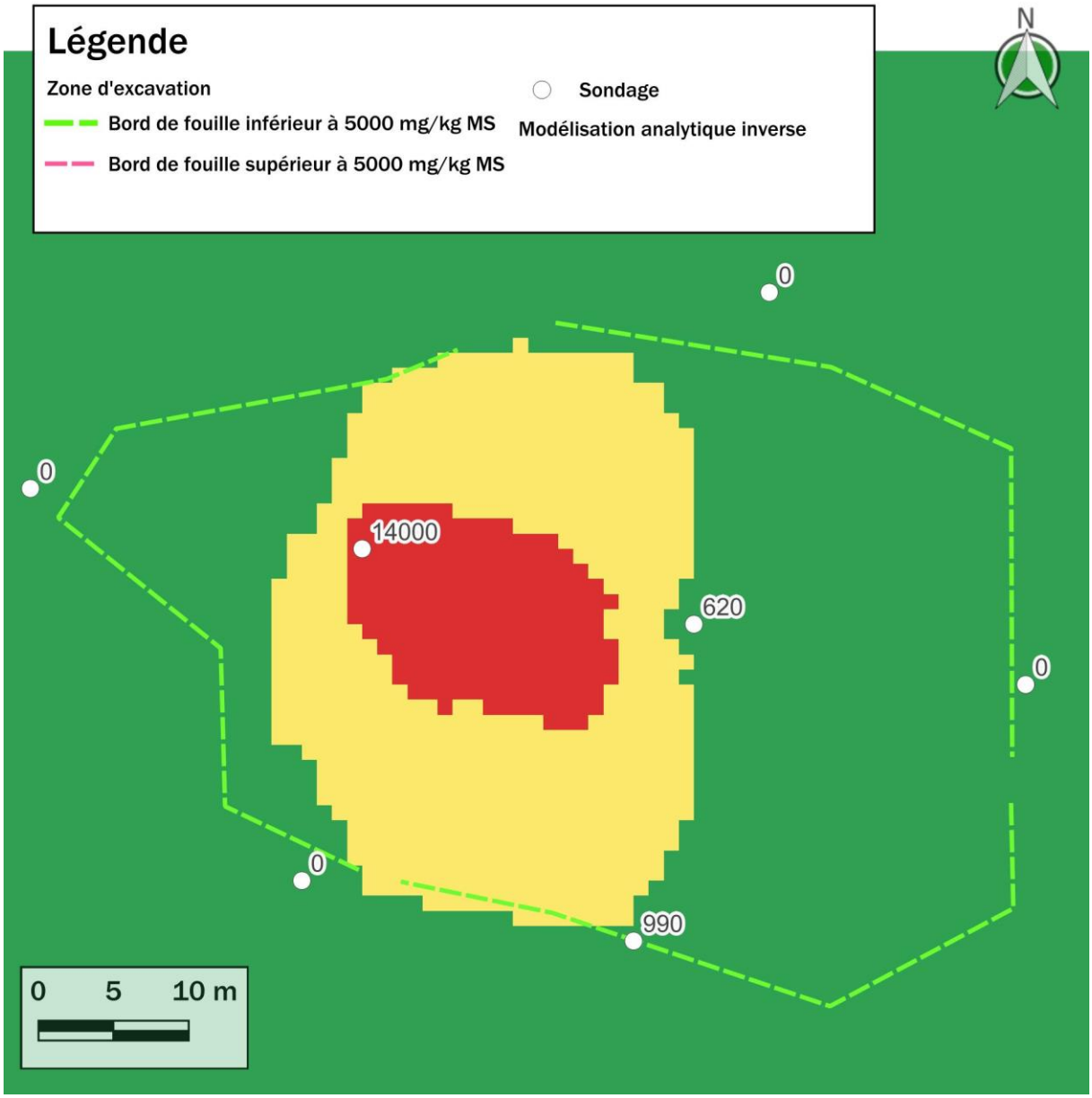

Légende
Zone d'excation
Bord de fouille inférieur à 2000 mg/kg MS
Bord de fouille supérieur à 2000 mg/kg MS
Sondages
Modélisation analytique inverse
Inférieure à 500 mg/kg MS
Zone d'incertitude
Supérieure à 500 mg/kg MS
N
0
5
10 m

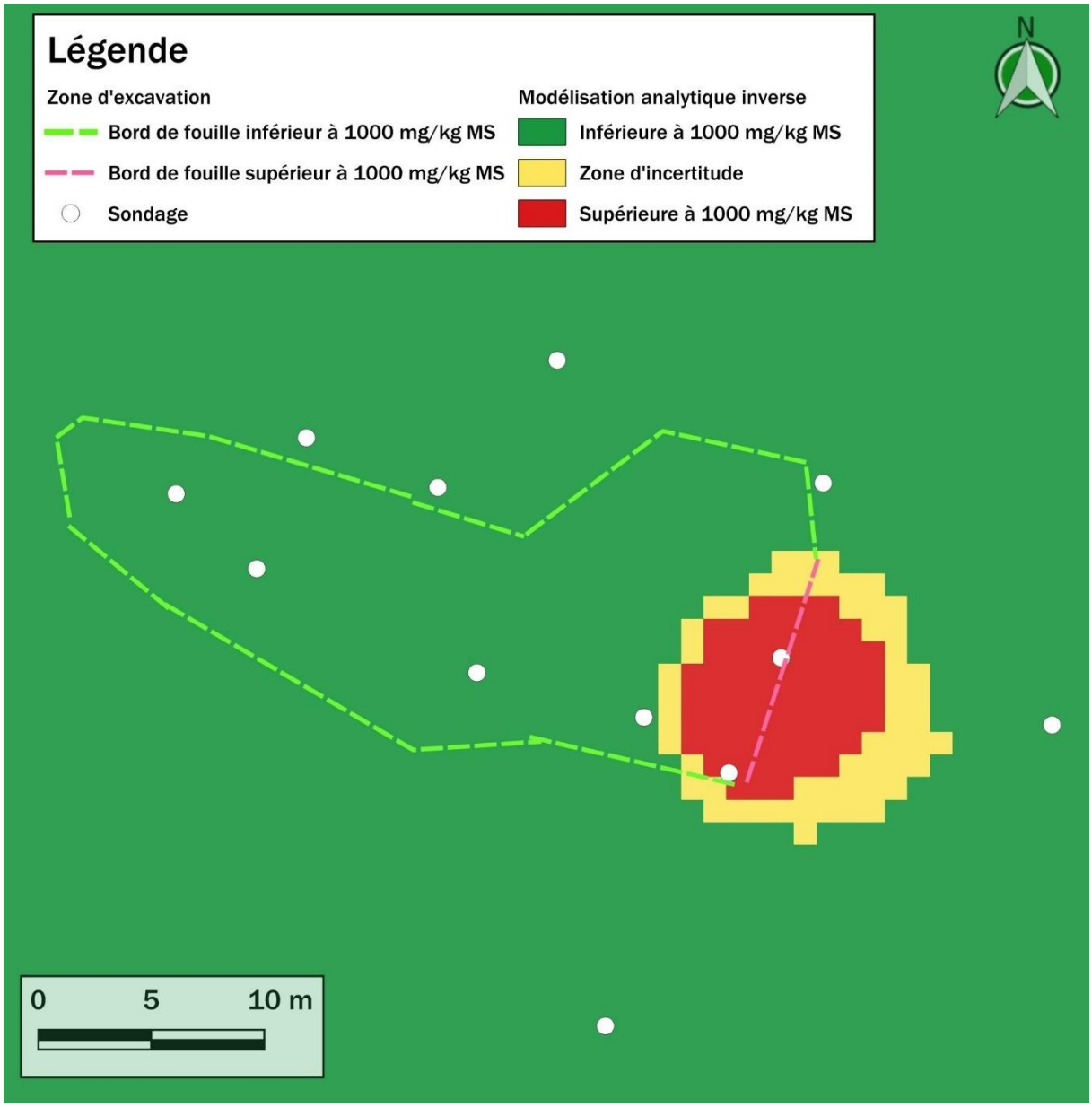

Légende
Zone d'excavation
Bord de fouille inférieur à 1000 mg/kg MS
Bord de fouille supérieur à 1000 mg/kg MS
Bord de fouille non prélevé
Sondage
Modélisation analytique inverse
Inférieure à 1000 mg/kg MS
Zone d'incertitude
Supérieure à 1000 mg/kg MS
N
0
5
10 m

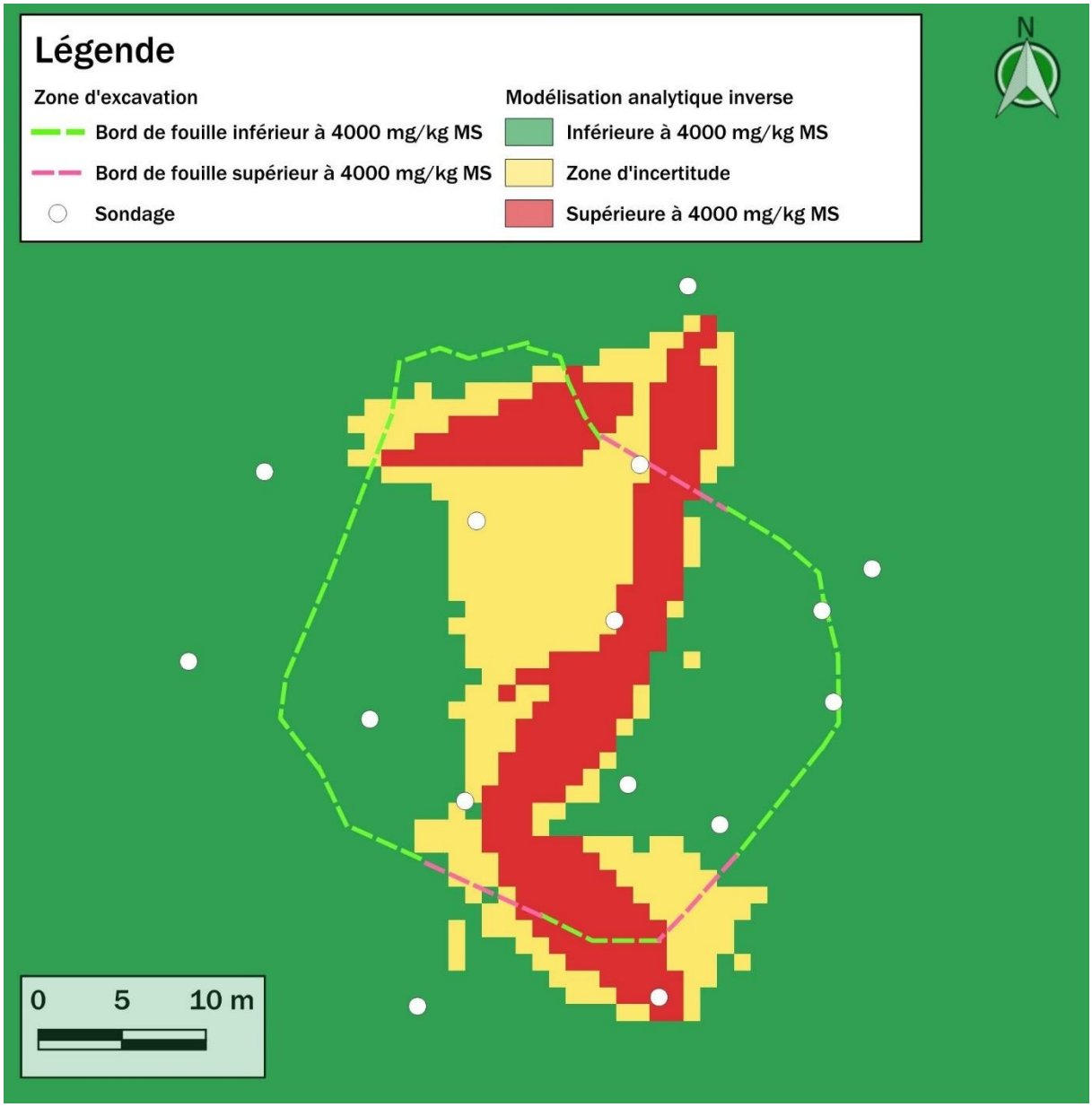

Légende
Zone d'excavation
Bord de fouille inférieur à 1000 mg/kg MS
Bord de fouille supérieur à 1000 mg/kg MS
Bord de fouille non prélevé
Sondage
Modélisation analytique inverse
Inférieure à 1000 mg/kg MS
Zone d'incertitude
Supérieure à 1000 mg/kg MS
N
0
2,5
5 m

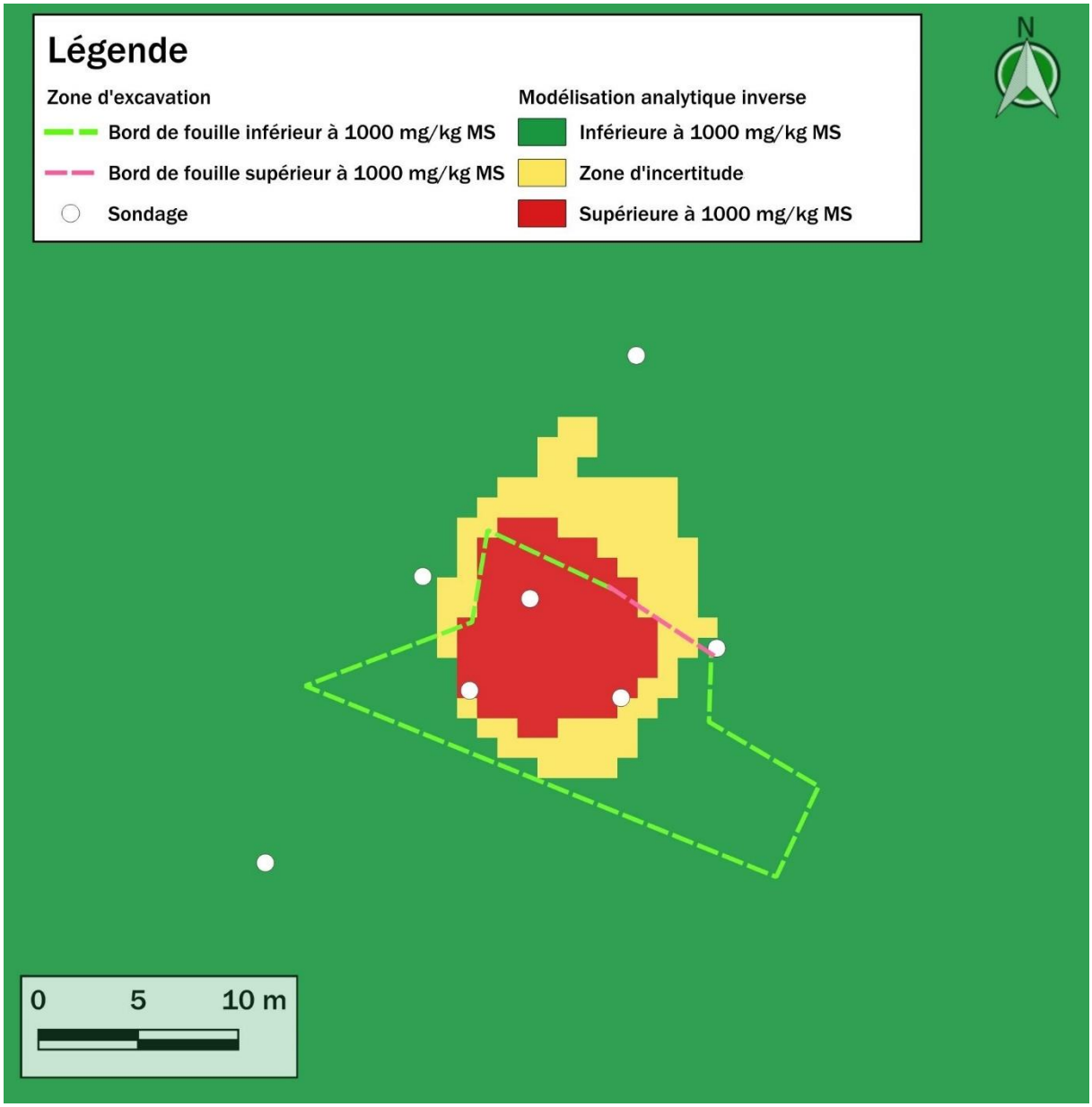

Légende
Zone d'excavation
Bord de fouille inférieur à 500 mg/kg MS
Bord de fouille supérieur à 500 mg/kg MS
Sondages
Modélisation analytique inverse
Inférieure à 500 mg/kg MS
Zone d'incertitude
Supérieure à 500 mg/kg MS
N
0
5
10 m

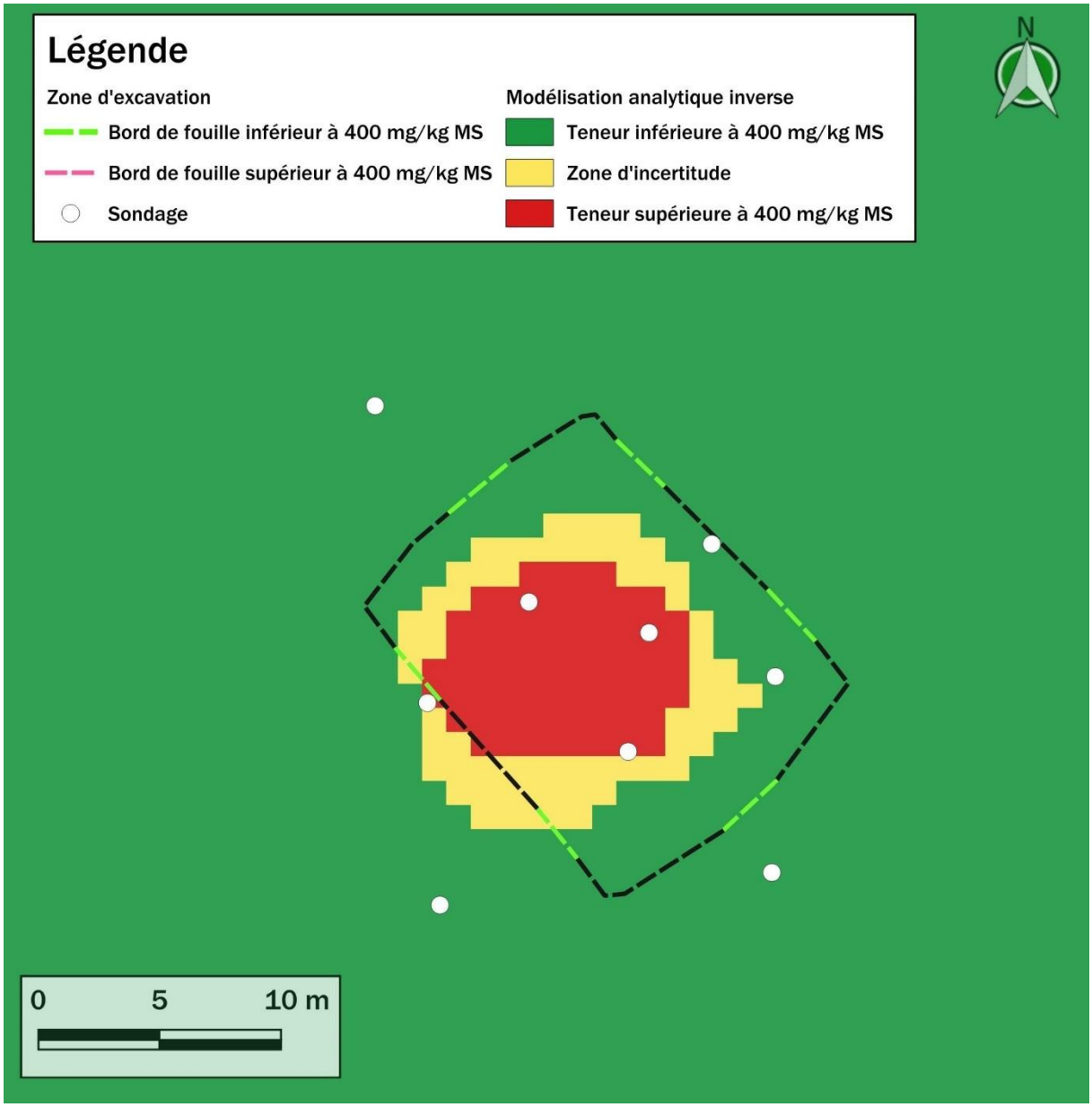

Légende
Zone d'excavation
Bord de fouille inférieur à 500 mg/kg MS
Bord de fouille supérieur à 500 mg/kg MS
Sondage
Modélisation analytique Inverse
Inférieur à 500 mg/kg MS
Zone d'incertitude
Supérieur à 500 mg/kg MS
N
0
5
10 m

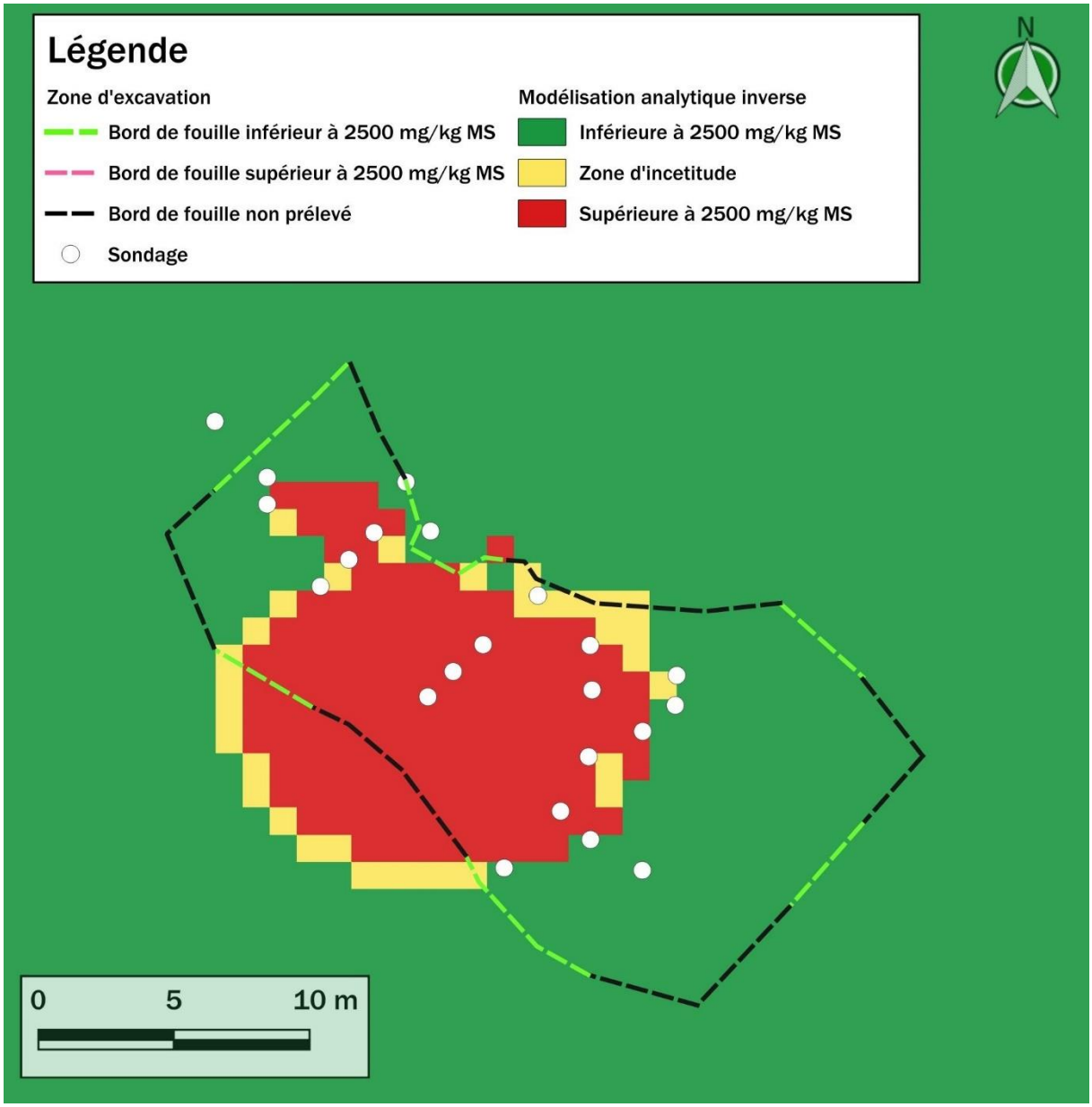

Légende
Zone d'excavation
Bord de fouille inférieur à 5000 mg/kg MS
Bord de fouille supérieur à 5000 mg/kg MS
Sondage
Modélisation analytique inverse
N
0
0
14000
620
0
0
990
0 5 10 m

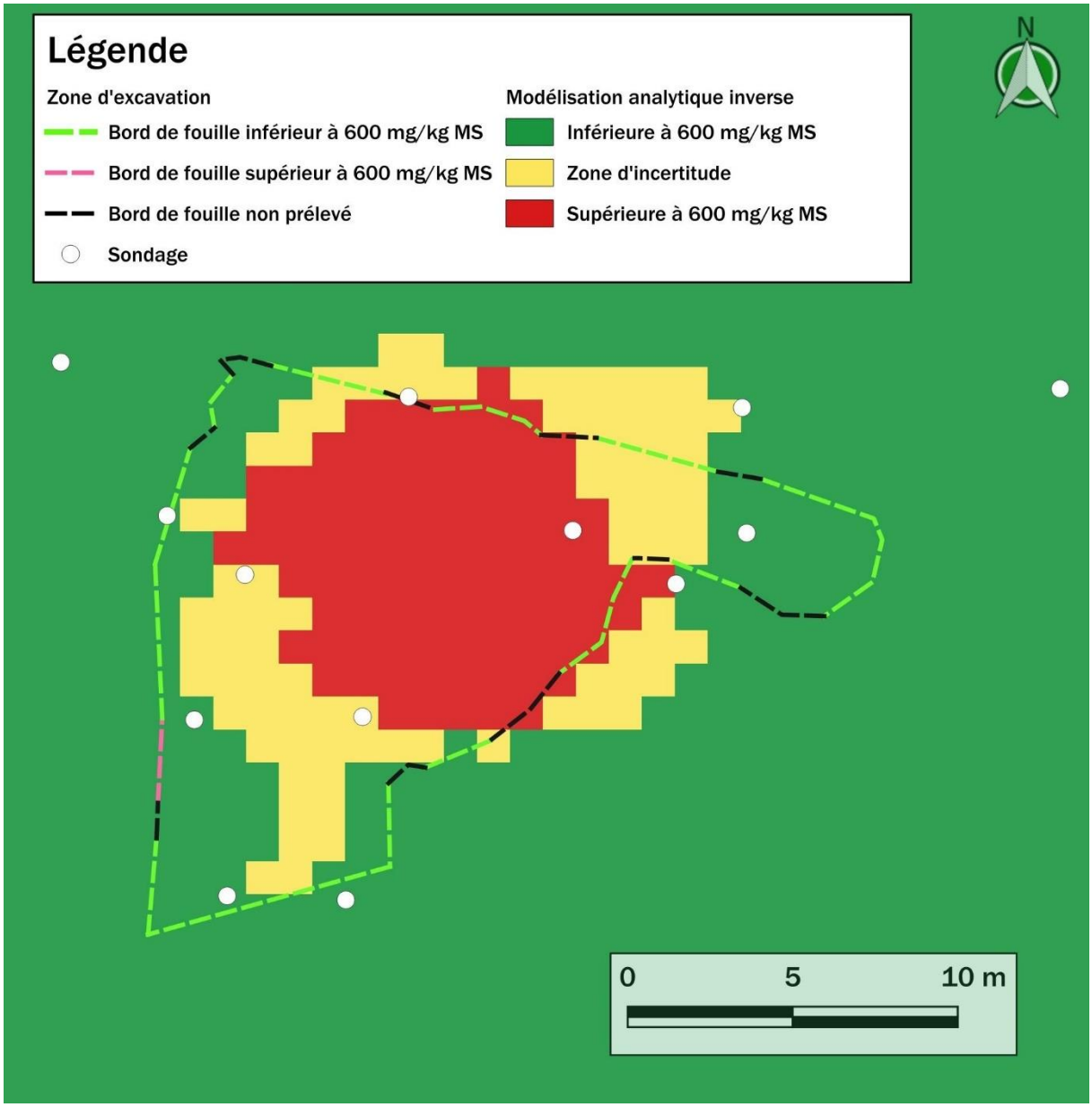

Légende
Zone d'excavation
Bord de fouille inférieur à 1000 mg/kg MS
Bord de fouille supérieur à 1000 mg/kg MS
Sondage
Modélisation analytique inverse
Inférieure à 1000 mg/kg MS
Zone d'incertitude
Supérieure à 1000 mg/kg MS
N
0
5
10 m

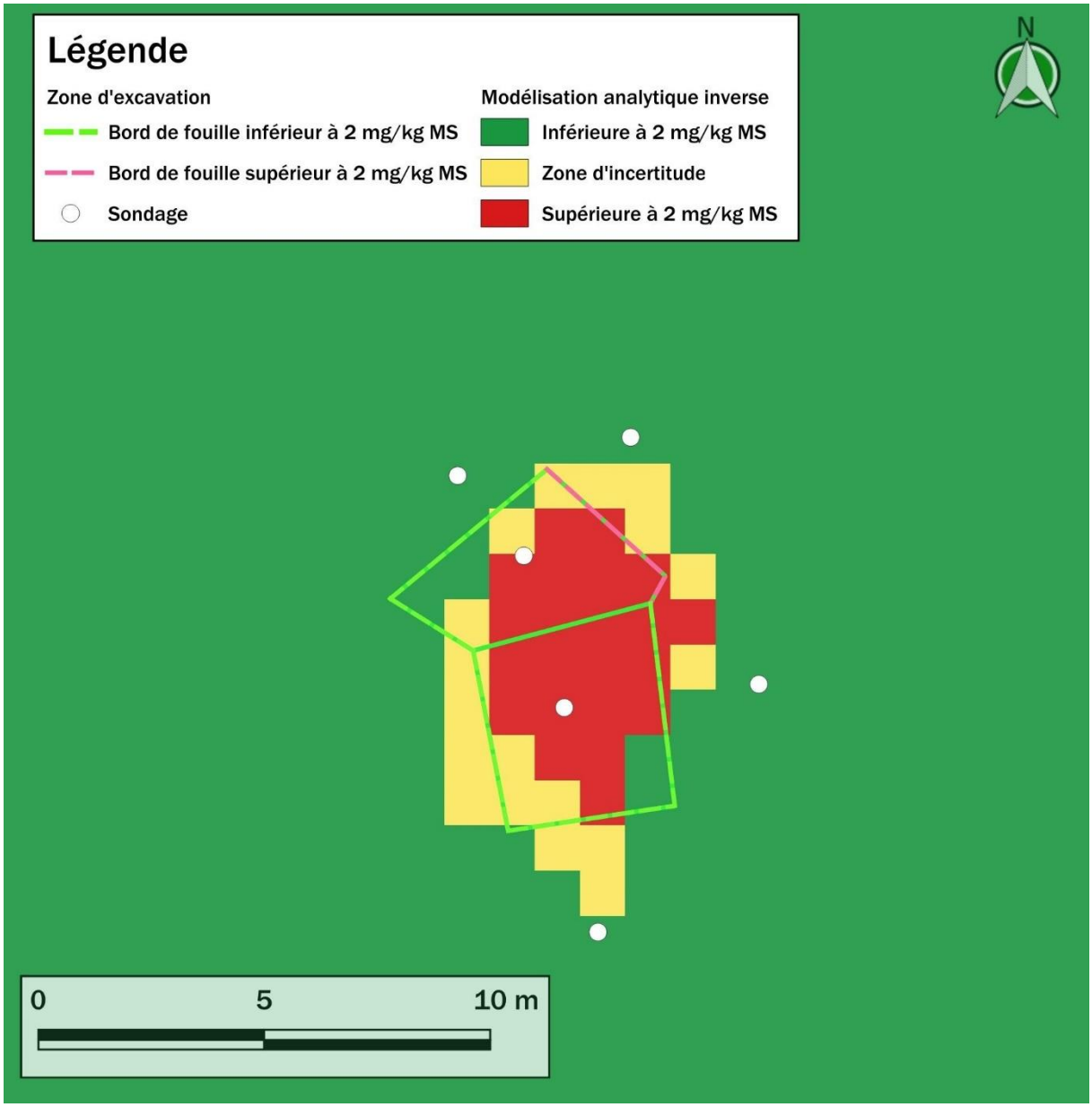

Légende
Zone d'excavation
Bord de fouille inférieur à 4000 mg/kg MS
Bord de fouille supérieur à 4000 mg/kg MS
Sondage
Modélisation analytique inverse
Inférieure à 4000 mg/kg MS
Zone d'incertitude
Supérieure à 4000 mg/kg MS
N
0
5
10 m

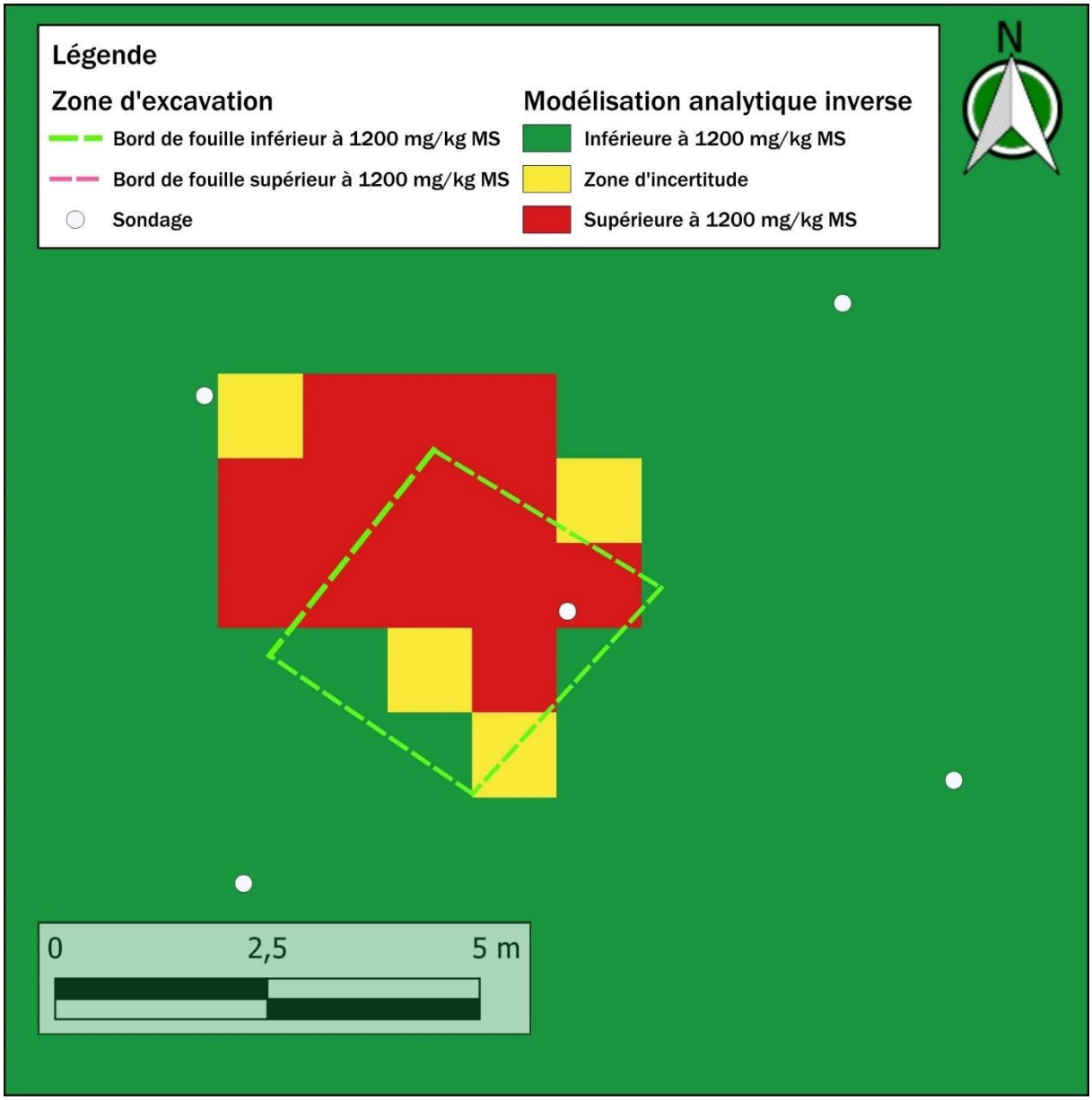

Légende
Zone d'excavation
Bord de fouille inférieur à 1000 mg/kg MS
Bord de fouille supérieur à 1000 mg/kg MS
Sondage
Modélisation analytique inverse
Inférieure à 1000 mg/kg MS
Zone d'incertitude
Supérieure à 1000 mg/kg MS
N
0
5
10 m

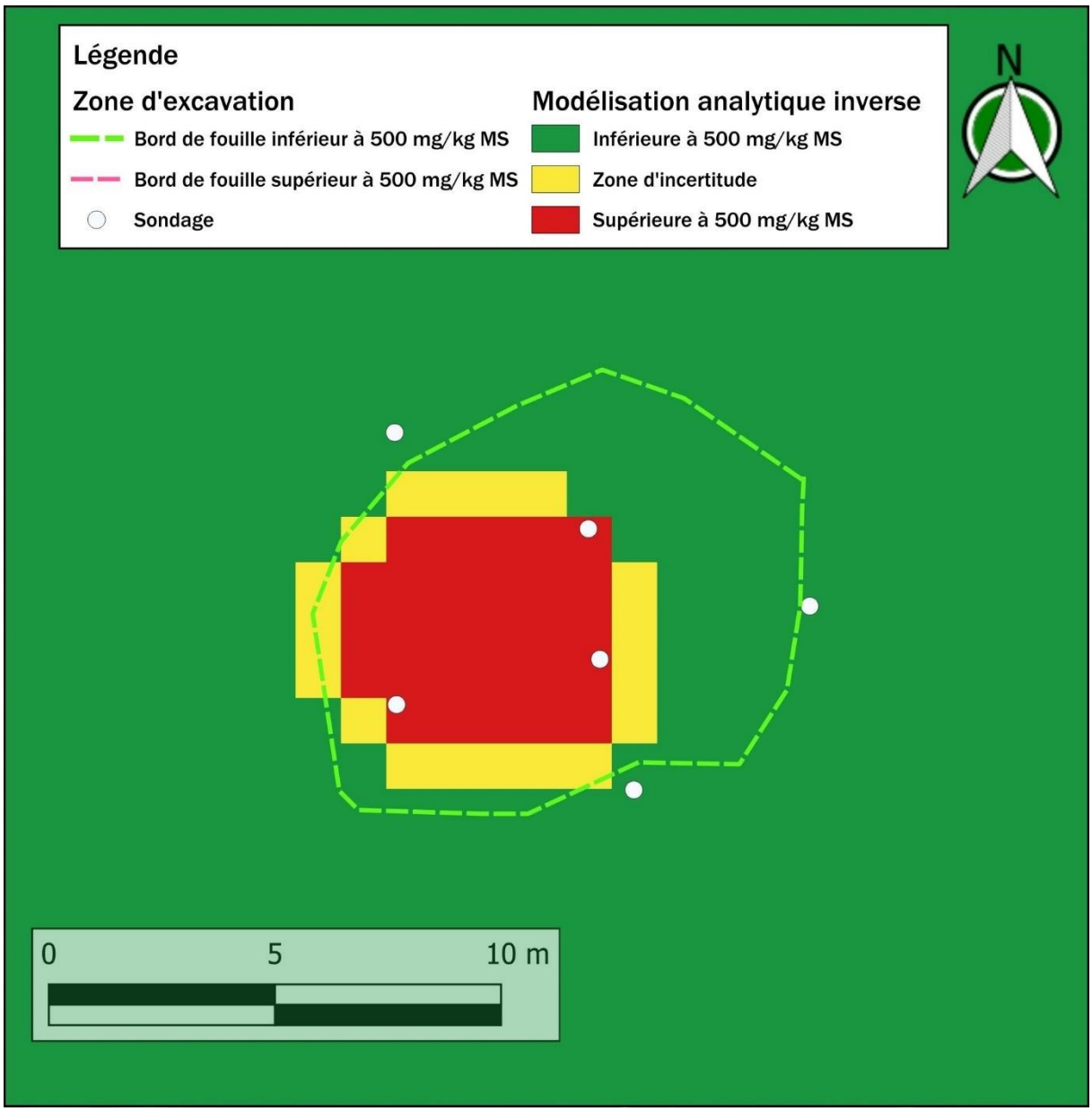

Légende
Zone d'excavation
Bord de fouille inférieur à 400 mg/kg MS
Bord de fouille supérieur à 400 mg/kg MS
Sondage
Modélisation analytique inverse
Teneur inférieure à 400 mg/kg MS
Zone d'incertitude
Teneur supérieure à 400 mg/kg MS
N
0
5
10 m

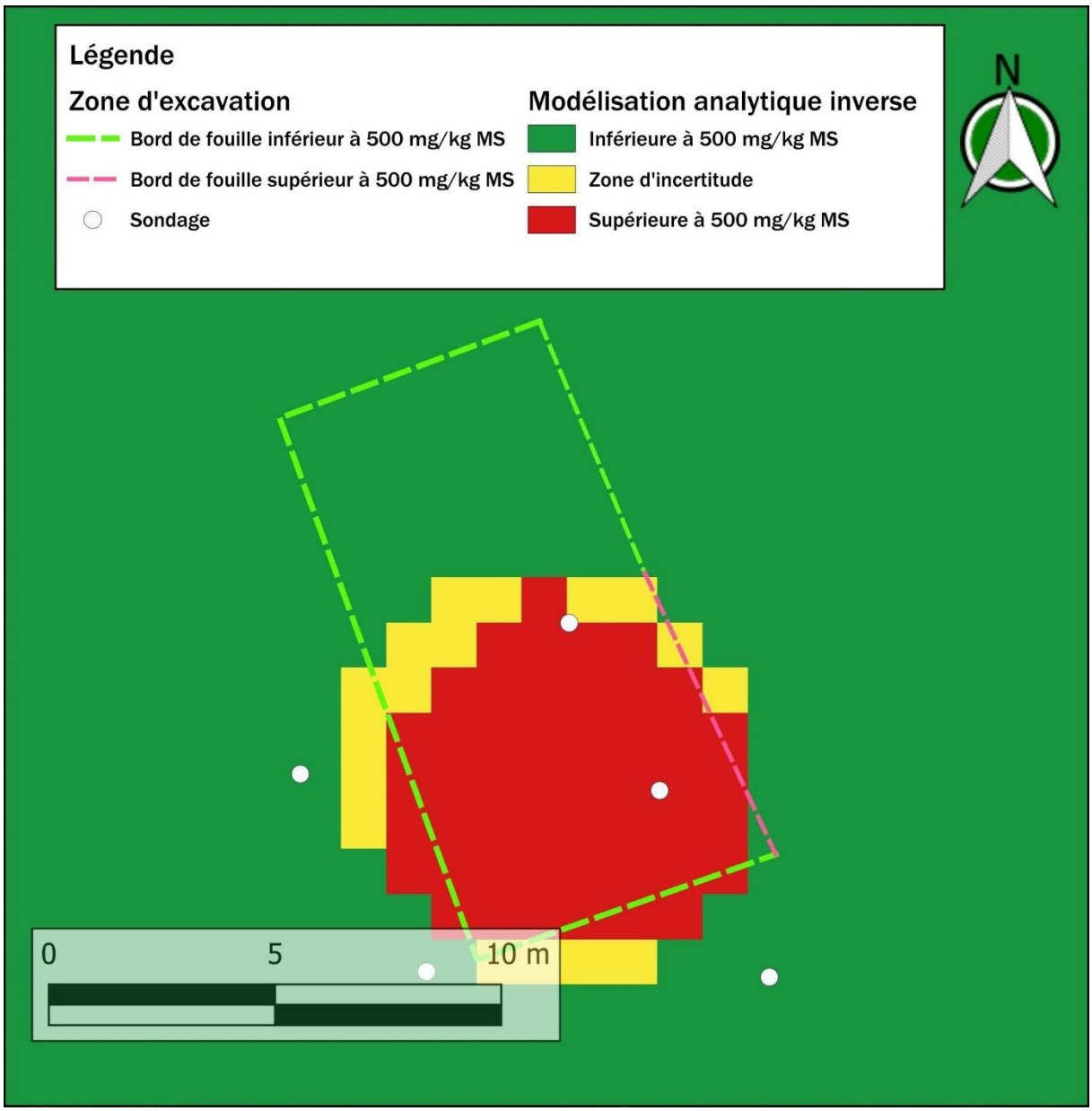

Légende
Zone d'excavation
Bord de fouille inférieur à 2500 mg/kg MS
Bord de fouille supérieur à 2500 mg/kg MS
Bord de fouille non prélevé
Sondage
Modélisation analytique inverse
Inférieure à 2500 mg/kg MS
Zone d'incetitude
Supérieure à 2500 mg/kg MS
N
0
5
10 m

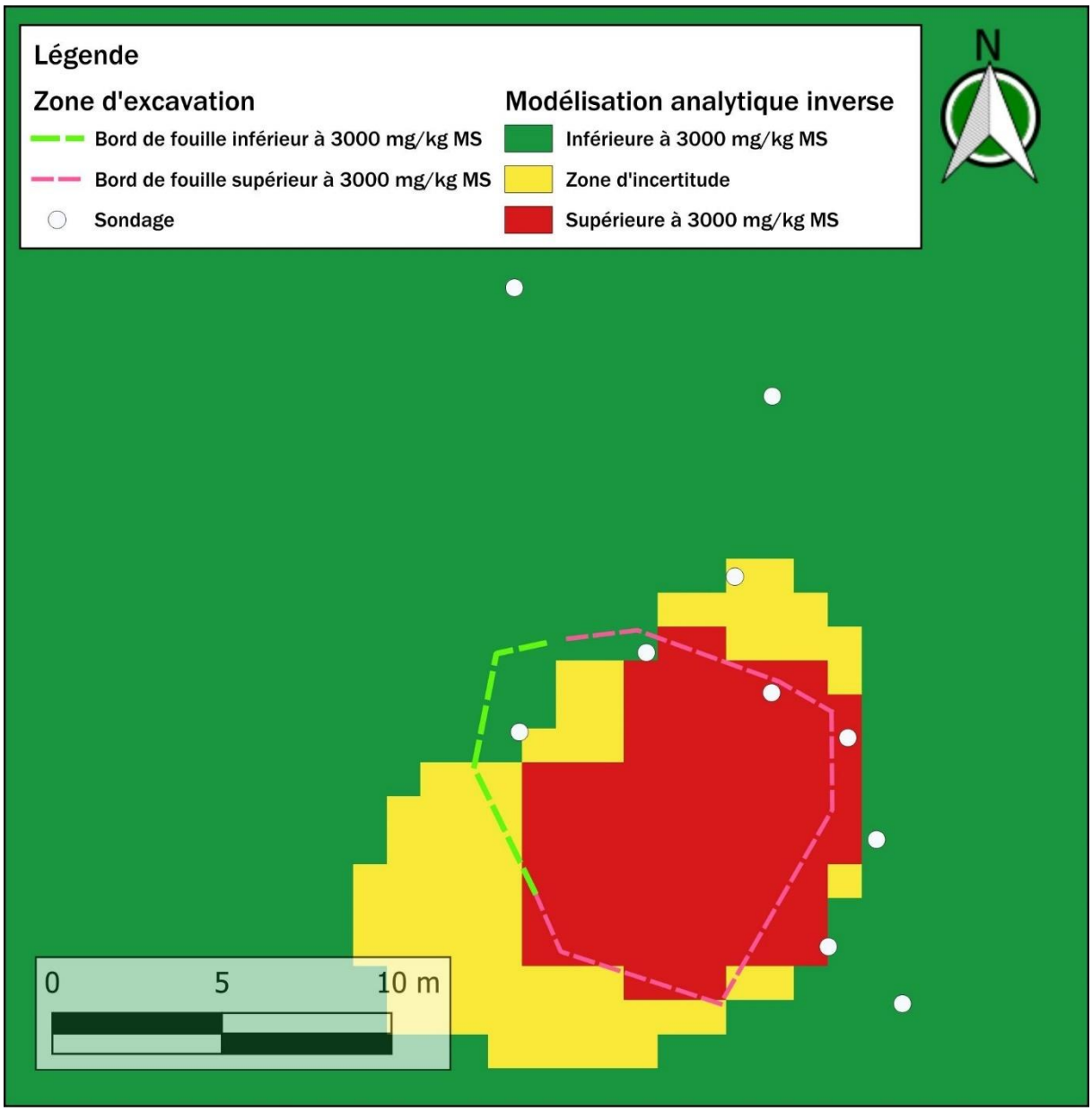

Légende
Zone d'excavation
Bord de fouille inférieur à 600 mg/kg MS
Bord de fouille supérieur à 600 mg/kg MS
Bord de fouille non prélevé
Sondage
Modélisation analytique inverse
Inférieure à 600 mg/kg MS
Zone d'incertitude
Supérieure à 600 mg/kg MS
N
0
5
10 m

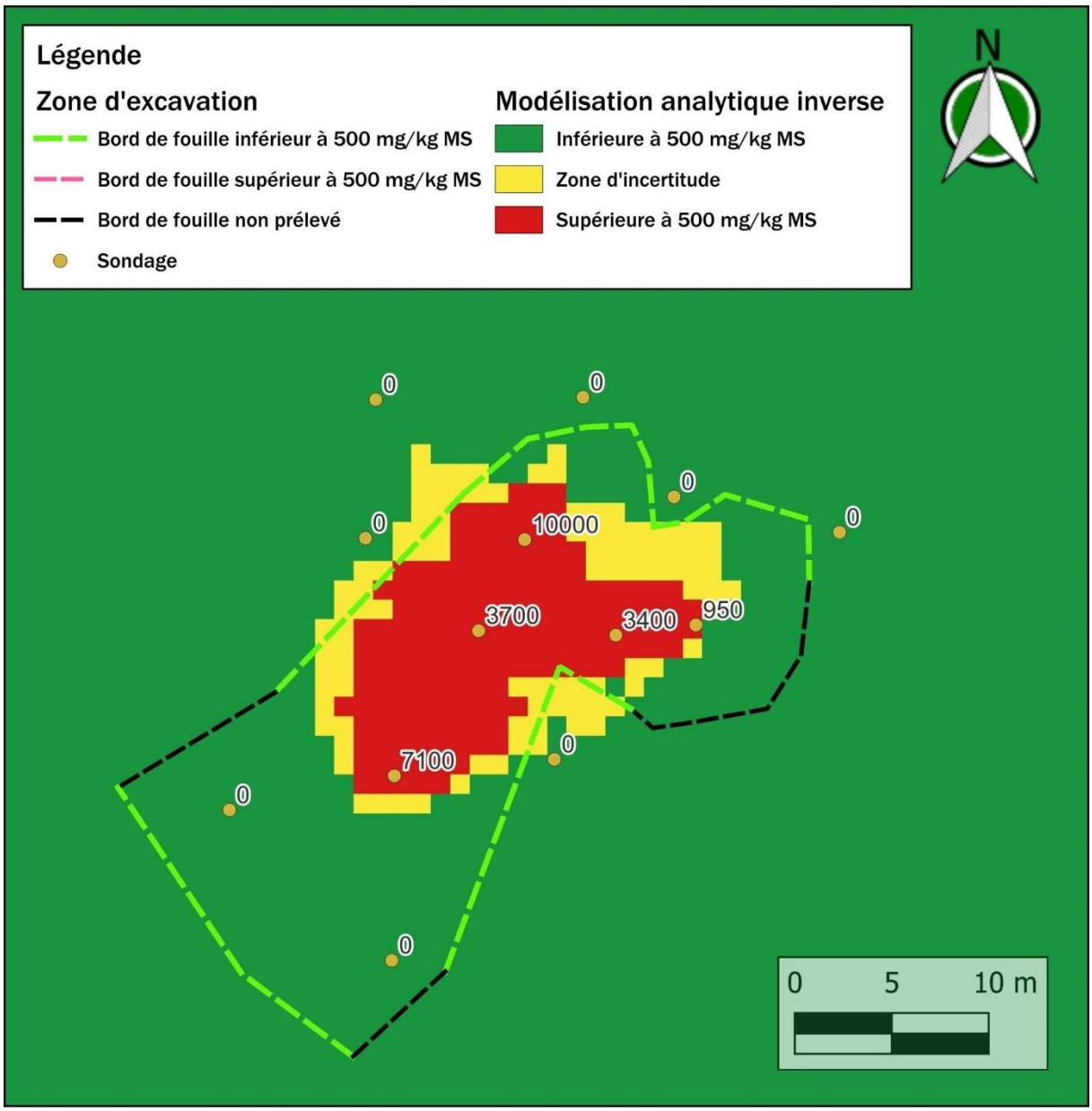

Légende
Zone d'excavation
Bord de fouille inférieur à 2 mg/kg MS
Bord de fouille supérieur à 2 mg/kg MS
Sondage
Modélisation analytique inverse
Inférieure à 2 mg/kg MS
Zone d'incertitude
Supérieure à 2 mg/kg MS
N
0
5
10 m

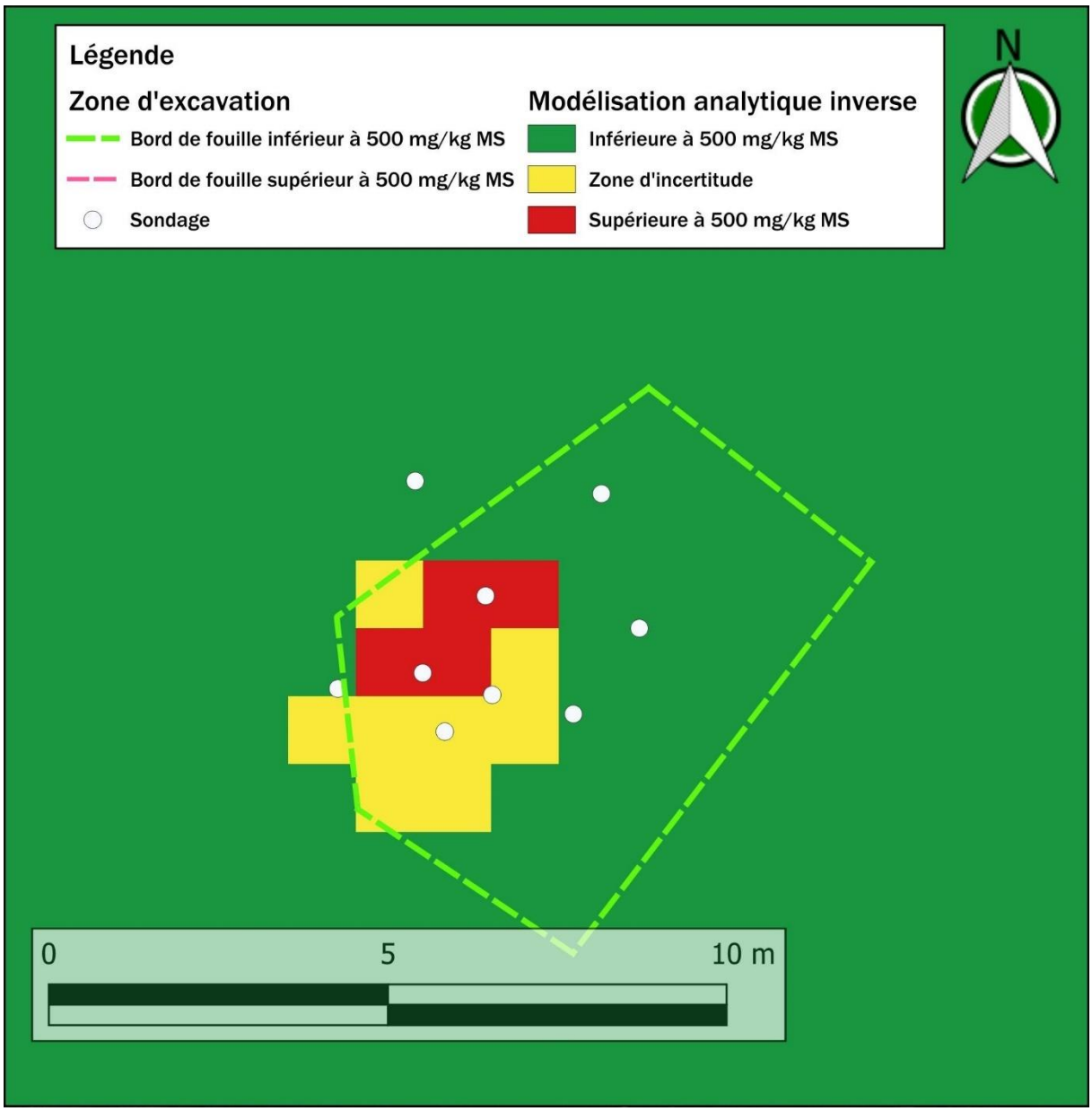

Légende
Zone d'excavation
Bord de fouille inférieur à 1200 mg/kg MS
Bord de fouille supérieur à 1200 mg/kg MS
Sondage
Modélisation analytique inverse
Inférieure à 1200 mg/kg MS
Zone d'incertitude
Supérieure à 1200 mg/kg MS
N
0
2,5
5 m

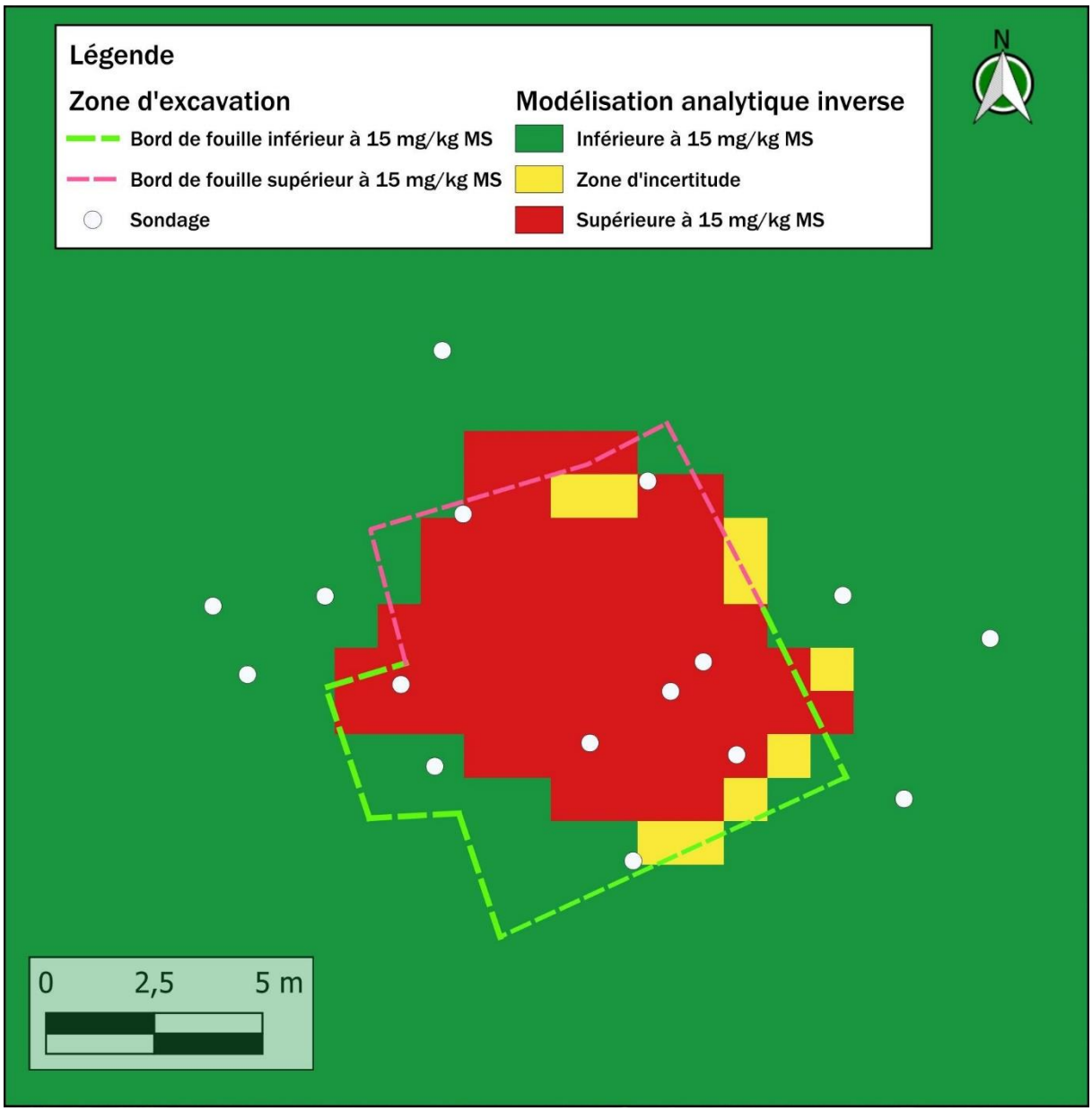

Légende
Zone d'excavation
Bord de fouille inférieur à 500 mg/kg MS
Bord de fouille supérieur à 500 mg/kg MS
Sondage
Modélisation analytique inverse
Inférieure à 500 mg/kg MS
Zone d'incertitude
Supérieure à 500 mg/kg MS
N
0
5
10 m


Légende
Zone d'excavation
Bord de fouille inférieur à 500 mg/kg MS
Bord de fouille supérieur à 500 mg/kg MS
Sondage
Modélisation analytique inverse
Inférieure à 500 mg/kg MS
Zone d'incertitude
Supérieure à 500 mg/kg MS
N
0
5
10 m


Légende
Zone d'excavation
Bord de fouille inférieur à 3000 mg/kg MS
Bord de fouille supérieur à 3000 mg/kg MS
Sondage
Modélisation analytique inverse
Inférieure à 3000 mg/kg MS
Zone d'incertitude
Supérieure à 3000 mg/kg MS
N
0
5
10 m


Légende
Zone d'excavation
Bord de fouille inférieur à 500 mg/kg MS
Bord de fouille supérieur à 500 mg/kg MS
Bord de fouille non prélevé
Sondage
Modélisation analytique inverse
Inférieure à 500 mg/kg MS
Zone d'incertitude
Supérieure à 500 mg/kg MS
N
0
0
0
0
0
10000
3700
3400
950
0
7100
0
0
0 5 10 m


Légende
Zone d'excavation
Bord de fouille inférieur à 500 mg/kg MS
Bord de fouille supérieur à 500 mg/kg MS
Sondage
Modélisation analytique inverse
Inférieure à 500 mg/kg MS
Zone d'incertitude
Supérieure à 500 mg/kg MS
N
0
5
10 m


Légende
Zone d'excavation
Bord de fouille inférieur à 15 mg/kg MS
Bord de fouille supérieur à 15 mg/kg MS
Sondage
Modélisation analytique inverse
Inférieure à 15 mg/kg MS
Zone d'incertitude
Supérieure à 15 mg/kg MS
N
0
2,5
5 m