

Apport de la géostatistique au diagnostic de sites et sols pollués : prolongement d'un cas d'étude de démonstration GeoSiPol

Jean-Baptiste MATHIEU^{1*}, Sébastien KASKASSIAN² et Michel H. GARCIA³

¹:KIDOVA, 155 avenue Roger Salengro, 92370 Chaville, jb.mathieu@kidova.com

²:BURGEAP, 19, rue de la Villette, 69425 LYON Cedex03, s.kaskassian@burgeap.fr

³:KIDOVA, 155 avenue Roger Salengro, 92370 Chaville, michel.garcia@kidova.com

Résumé

L'étude des sites potentiellement pollués comprend généralement deux grandes étapes : 1) l'analyse de l'état de pollution des sols pour juger de son impact sur les milieux et décider de la nécessité ou non d'intervenir selon l'usage prévu du site, 2) si nécessaire, la délimitation et la quantification de la pollution des sols pour cibler et dimensionner au mieux le traitement et donc optimiser les coûts de réhabilitation [1][2].

Cet article présente et compare, pour chacun de ces objectifs, les résultats d'une approche dite « usuelle » et d'une approche « géostatistique » sur un ancien site de stockage de produits chimiques. Il dresse un bilan coût/avantage chiffré de l'approche géostatistique selon les données disponibles, l'objectif et le temps consacré à l'étude.

Ce travail s'appuie sur une des études de démonstration [3] de l'application de la géostatistiques aux sites et sols pollués (SSP), réalisées, dans le cadre du groupe de travail GeoSiPol avec un financement de l'ADEME, et en complète les résultats. L'étude porte sur les données du site 1bis du projet de recherche « ATTENA » [4] financé également par l'ADEME.

Introduction

A la demande de ses membres et grâce à un financement de l'ADEME, le groupe de travail GeoSiPol a mené des études de démonstration de l'application de la géostatistique aux SSP afin d'en illustrer l'intérêt [3]. A partir de cas d'étude réels, l'un des objectifs de ces études était de dresser le bilan critique des coûts et des avantages d'une approche « géostatistique » par rapport aux pratiques « usuelles » de délimitation d'une pollution des sols. Sans pouvoir facilement impliquer de bureau d'étude dans les délais impartis à ces études, ces aspects n'avaient finalement pas pu être abordés. Ils font l'objet de cet article en s'intéressant aux données du site 1bis ATTENA particulièrement bien adaptées à ce type d'exercice.

La comparaison entre approches géostatistique et usuelle vise à dresser un bilan coût/avantage chiffré qui permette de mieux évaluer l'apport et la pertinence de géostatistique en prenant en compte :

- la taille du jeu de données disponible (diagnostic standard ou diagnostic détaillé),
- l'objectif recherché (estimation du volume ou localisation d'une zone source),
- la répartition des coûts entre l'étude et la réhabilitation du site selon les volumes estimés.

Matériel et méthodes

Le site 1bis ATTENA est un ancien site de stockage, conditionnement et expédition de produits chimiques. Les sols y sont pollués par des solvants chlorés, des hydrocarbures (dont des solvants mono-aromatiques) et des composés organo-solubles (alcools, cétones). Dans ce cadre, l'étude de démonstration s'est intéressée plus particulièrement à l'état de pollution des sols en Perchloroéthylène (PCE), Trichloroéthylène (TCE) et Cis-Dichloroéthylène (cis-DCE abrégé DCE dans la suite) sur les 6 premiers mètres de profondeur. Ce choix a été motivé par la présence dominante de ces trois polluants qui forment ensemble une source de pollution en phase liquide non aqueuse. Le domaine étudié représente un volume total de 37 000 m³.

Le jeu de données total comprend des analyses en laboratoire sur 287 échantillons de sols provenant de 46 sondages et 802 mesures de PID qui ont servi principalement à sélectionner les échantillons analysés en laboratoire. Ce jeu de données particulièrement conséquent résulte d'exigences scientifiques du projet de recherche ATTENA et ne correspond pas aux standards habituels des études SSP. Il a ainsi été demandé à un ingénieur d'étude de BURGEAP étranger au projet d'en extraire un jeu de donnée dit « restreint », constitué de 89 mesures PID et 49 échantillons de sol analysés en provenance de 13 sondages. Les données restantes ont servi de données de validation des modèles d'estimation fondés sur le jeu de données restreint puis de données complémentaires pour améliorer la précision des estimations. Dans ces conditions, l'objectif de l'étude de démonstration était d'estimer et délimiter la zone source où les trois polluants étudiés forment ensemble un phase liquide (organique) non aqueuse (NAPL en anglais) [5].

Suivant ce même objectif, l'approche géostatistique (résultat de l'étude de démonstration) et celle dite « usuelle » (objet des travaux complémentaires) ont été répétées sur le jeu de donnée « restreint » et sur l'ensemble des données afin de quantifier l'impact du nombre de données sur l'estimation du volume total de la zone source, de sa localisation et de la masse de polluant qu'elle représente.

Les principales étapes de ces approches sont rappelées ci-après.

Commun à l'approche usuelle et géostatistique

- *Mise en forme des données pour leur traitement numérique.*
Compilation des données dans des fichiers et import dans un programme adapté pour les étudier (SIG, modèleur géologique, autre). Dans le cas présent, les données ont été chargées et exploitées dans le modèleur géologique GOCAD-SKUA[®] à l'aide d'outils de la SoilRemediation[®] Suite.
- *Estimation d'un seuil de teneur pour les polluants d'intérêt.*
Calcul de ce seuil, sur la base d'une EQRS ou par référence à une valeur de bruit de fond. Dans cette étude, le logiciel OREOS[6] a été utilisé et la teneur totale révélatrice de NAPL est estimée à 400 ppm.

Spécifique à l'approche usuelle

- *Calcul du volume représentatif de chaque échantillon.* (Figure 1)
Calcul par plus proche voisinage fonction de la répartition des données (maillage dit de Voronoi).
- *Calcul du volume de la zone source et de la masse de polluant associée*
Somme des volumes associés aux échantillons dont la teneur totale en polluants est supérieure au seuil considéré. Bien qu'inhabituel, l'incertitude statistique sur ce volume peut être quantifiée par une méthode de ré-échantillonnage aléatoire des données [7]. Utilisé ici à titre de comparaison avec la géostatistique.
- *Positionnement de la zone source*
Zone source définie par l'ensemble des mailles (plus proches voisinages) associées aux échantillons où la teneur totale mesurée est supérieure au seuil considéré.

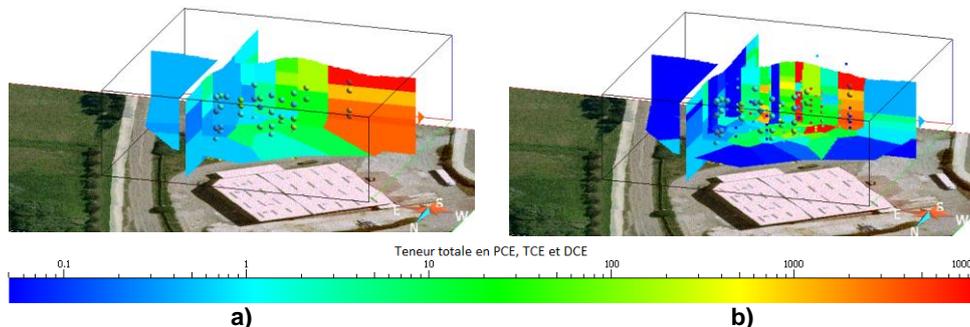


Figure 1: Décomposition du site en blocs de plus proche voisinage associés aux données pour le jeu de données a) restreint, b) complet. Les teneurs des blocs sont celles des données associées.

Spécifique à l'approche géostatistique [3]

- *Analyses statistique et variographique des teneurs.*
Analyse et modélisation des distributions (histogrammes) des teneurs pour les différents polluants, des relations ou corrélations entre polluants et des corrélations spatiales.
- *Simulation des teneurs.*
Génération d'un grand nombre de simulations (100) des teneurs en polluants sur une grille 3D par une approche géostatistique adaptée (production d'images statistiquement réalistes de la pollution).
- *Quantification de l'incertitude sur le volume de la zone source et de la masse de polluant associée*
Calcul sur chaque simulation du volume total des cellules où les teneurs simulées correspondent à une zone source et calcul des masses associées.
- *Quantification de l'incertitude locale sur la position de la zone source*
Classification de chaque cellule comme « zone source », « hors zone source » ou « incertaine » en fonction de la proportion (p) de simulations prédisant la cellule dans une zone source. Ici, la cellule est « zone source » si $p \geq 80\%$, « hors zone source » si $p \leq 20\%$, sinon « incertaine ».

Résultats et discussion

Le Tableau 1 résume les statistiques sur le volume de la zone source, tel qu'il peut être estimé à partir de teneurs à l'échelle des échantillons, en fonction de l'approche suivie et du jeu de données restreint ou complet. Les intervalles de confiance établis avec l'approche usuelle proviennent de ré-échantillonnages aléatoires des données (méthode de *bootstrap*) [7]. Ces résultats appellent les commentaires suivants.

- Comme attendu, l'augmentation de la taille du jeu de données améliore les estimations quel que soit l'approche considérée (réduction de l'intervalle de confiance), le volume moyen étant par ailleurs réduit.
- Si entre les deux approches les écarts d'estimation diminuent avec le nombre de données croissant, des différences continuent d'exister tant sur le volume moyen (simple au double) que sur l'incertitude.
- Contrairement à l'approche géostatistique, les volumes estimés avec celle usuelle varient grandement d'un jeu de données à l'autre, confirmant le besoin d'avoir beaucoup de données pour pouvoir l'appliquer. En ce sens, l'approche géostatistique est certainement plus robuste (moins sensible aux données).

Tableau 1. Statistiques sur le volume de la zone source estimé à l'échelle des échantillons.

Approche	jeu restreint		jeu complet	
	« usuelle »	« géostatistique »	« usuelle »	« géostatistique »
Volume moyen	8 117 m ³	3 422 m ³	1 172 m ³	2 161 m ³
Intervalle de confiance à 95%	[1 864 m ³ ; 14 931 m ³]	[1 642 m ³ ; 6 121 m ³]	[677 m ³ ; 1 780 m ³]	[1 439 m ³ ; 3 168 m ³]

La Figure 2 montre les sols classés comme situés dans la zone source avec les deux approches et pour les deux jeux de données. Dans le cas de l'approche géostatistique, les sols dont la classification comme zone source est incertaine (entre 20 et 80% de chances d'être en zone source) sont délimités par des surfaces transparentes. Ces résultats graphiques illustrent et complètent les observations précédentes.

- Avec le jeu de données restreint, l'approche usuelle est incapable de délimiter convenablement la zone source a beaucoup d'endroits (Figure 2.a), expliquant ainsi la surestimation du volume moyen.
- Dans l'approche géostatistique à utiliser comme élément d'aide à la décision pour prendre en compte les incertitudes, les sols incertains seraient à traiter, sauf à décider de données complémentaires aux endroits incertains. Le jeu de données complet permet ainsi de réduire l'étendue des sols incertains tout en identifiant davantage de sols situés (très certainement) en zone source.
- En augmentant le nombre de données, l'approche usuelle tend à mieux délimiter la zone source en se rapprochant davantage des résultats de l'approche géostatistique (sols en zone source et incertains).
- L'impact du jeu de données sur les résultats est spectaculaire avec l'approche usuelle peu robuste. Il se traduit par des prédictions plus mauvaises de cette approche par rapport à l'autre. Les 238 données complémentaires prédites à partir du jeu de données restreint sont mal classées à 29% avec l'approche usuelle contre 8% avec l'approche géostatistique.

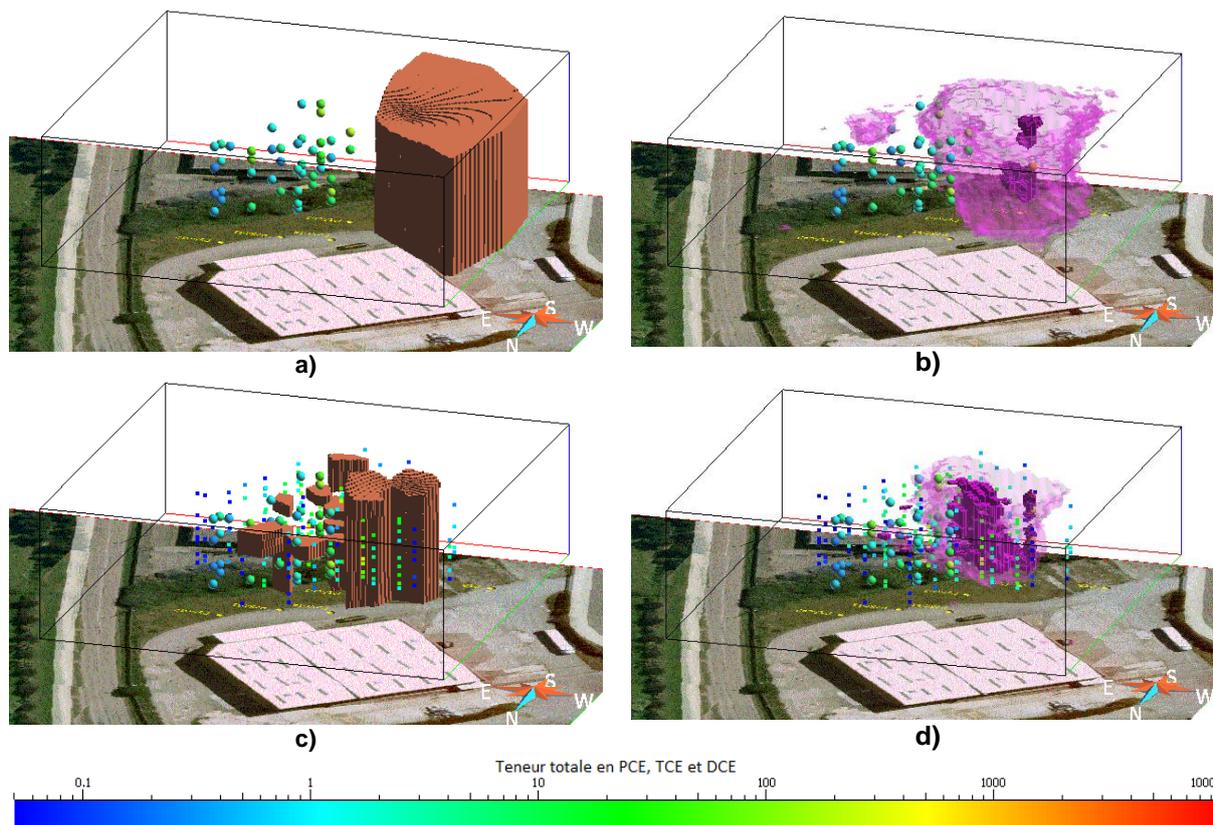


Figure 2: Délimitation des zones sources selon l'approche, usuelle (surfaces marrons à gauche) ou géostatistique (surfaces violettes opaques à droite), et le jeu de données considéré, restreint (en haut) et complet (en bas). Les surfaces transparentes à droite délimitent les sols incertains.

A partir de la répartition spatiale des sols classés comme situés en zone source ou incertains (cas de l'approche géostatistique), il a été délimité, pour chacun des quatre modèles d'estimation, le volume de sols à excaver et à traiter en prenant en compte les contraintes habituelles de chantier sur les excavations. Les volumes correspondants se trouvent dans le Tableau 2 ainsi que la superficie des terrains à excaver et des statistiques sur la masse de polluants en place estimée. Ces résultats ont servi à calculer les coûts d'études et de dépollution présentés dans le Tableau 3 selon l'approche et le jeu de donnée utilisés. Ces coûts s'apparentent à ce que serait une solution générique d'excavation et de traitement sur site. Les fourchettes indiquées intègrent l'incertitude sur les masses mais aussi différents scénarios : mode d'excavation avec talutage ou blindage, mode de suivi et tri des terres, traitement par déstructuration sous tente complétée ou non par chaulage, quantité de charbon actif à utiliser.

Tableau 2. Estimation des éléments de dimensionnement de la réhabilitation.

Approche	Jeu restreint		Jeu complet	
	Usuelle	Géostatistique	Usuelle	Géostatistique
Volume à excaver	8 100 m ³	7 500 m ³	4 800 m ³	4 500 m ³
Volume à traiter	8 100 m ³	3 400 m ³	1 200 m ³	2 200 m ³
Masse de polluants estimée en zone source ou incertaine	48 t	27 t	24 t	17 t
Intervalle de confiance à 95% sur masse	18 à 78 t	4 à 62 t	9 à 34 t	10 à 27 t
Superficie de la zone d'excavation	1 500 m ²	1 500 m ²	830 m ²	920 m ²

Tableau 3. Estimation des coûts de l'étude et de la réhabilitation [k€ HT].

Approche	Jeu restreint		Jeu complet	
	Usuelle	Géostatistique	Usuelle	Géostatistique
Acquisition des données, mise en forme et compte-rendu	13 à 17		48 = 13+35 à 65 = 17+48	
Interprétation et rapport	3 à 4	7 à 10	6 à 10	7 à 10
Coût moyen de réhabilitation AMO et contrôle	1 200 [645 à 1 765]	840 [515 à 1 240]	670 [490 à 825]	655 [545 à 800]
Total	1 220 [660 à 1 790]	865 [535 à 1 270]	735 [545 à 900]	720 [600 à 875]

Dans ces évaluations de coûts, le passage du jeu de données restreint au jeu de données complet est considéré comme résultant d'une campagne de reconnaissance complémentaire. La position des échantillons complémentaires n'est cependant pas optimale car issue de considérations scientifiques dictées par le projet ATTENA et non pas d'un objectif de réduction de l'incertitude spatiale. Sur les 238 données nouvelles du jeu complet, seules 69 sont situées à des endroits suffisamment incertains. Les 169 autres sont redondantes et apportent peu d'informations complémentaires par rapport au jeu de données restreint. Dans le cas d'une approche géostatistique, la campagne de reconnaissance complémentaire pourrait donc se résumer à 16 sondages et 69 analyses sur échantillons pour un coût entre 17 et 20 k€ au lieu de 35 à 48 k€.

Conclusions et perspectives

Ces résultats montrent que l'approche géostatistique permet de produire des estimations cohérentes et fiables, assorties d'incertitudes sur les volumes de sols pollués, leurs positions et les masses de polluants en place, que le jeu de données soit restreint ou complet. La quantification de l'incertitude permet aussi de juger du manque d'information quand les données sont insuffisantes. A l'inverse, l'approche usuelle ne peut s'appliquer qu'à la condition d'avoir de nombreuses données pour délimiter sans trop d'erreurs les sols pollués. Si cette condition est vérifiée, l'approche usuelle conduit à des estimations et des délimitations qui tendent vers celles obtenues avec l'approche géostatistique. Elle ne permet cependant pas de gagner sur les coûts d'étude. Dans le cas d'étude présenté, cette différence est en moyenne de 15 k€ (735 - 720 k€) à l'avantage de l'approche géostatistique. Dans tous les cas, les coûts de l'étude géostatistique proprement dite, par rapport à une approche usuelle (coûts d'interprétation et rapport dans le Tableau 3), apparaissent insignifiants comparés aux coûts totaux de réhabilitation d'un site (quelques milliers d'euros par rapport à plusieurs centaines de milliers d'euros).

Références

- [1] EPA (2004). *Improving Sampling, Analysis, and Data Management for Site Investigation and Cleanup*, EPA-542-F-04-001a, <http://www.epa.gov/tio/download/char/2004triadfactsheeta.pdf>.
- [2] MEDAD (2007). *Diagnostic du site*, Ministère de l'Ecologie, www.sites-pollues.ecologie.gouv.fr/OutilsMethodologiques/Diagnostics_du_site_V0-022007.pdf.
- [3] GeoSiPol (2012). *Etudes de démonstration de l'intérêt de la géostatistique dans le domaine des sites et sols pollués*, www.geosipol.org.
- [4] Kaskassian S., Gleize T., Chastanet J., Côme J.-M. (2013). *Projet ATTENA – Phase 2, Tâche 3.1.2 : mise en œuvre des guides méthodologiques MACAOH par BURGEAP sur le Site 1bis (solvants chlorés)*, Rapport Final, 3 vol., p.296, www.attena.org
- [5] Côme, J.-M., Kaskassian, S., Ropars, M., Razakarisoa, O., Quintard, M., Schäfer, G., Haeseler, F. (2007). *Organo-chlorés aliphatiques. Caractérisation d'une zone source dans les aquifères. Programme R&D MACAOH*, ADEME, Collection Connaître pour Agir, Guides et cahiers techniques. <http://www2.ademe.fr/servlet/getDoc?cid=96&m=3&id=51385>
- [6] Chastanet J., Kaskassian S., Côme J.M. (2012). *User's guide – OREOS v1*, Burgeap Report RREDMS00014, 37p. www.oreos-software.com.
- [7] Garcia M.H., Mathieu J.-B., Garcia V. (2014). *Methodological and practical aspects of geostatistical bootstrap for quantifying global and local soil contamination uncertainty*. [présenté à geoENV 2014, Paris, juillet 2014](http://www.geoENV2014.org).

Remerciements

Nous remercions l'ADEME pour son soutien sans lequel les études de démonstration et le projet ATTENA n'auraient pas été possibles. Nous adressons également nos remerciements aux membres du consortium ATTENA d'avoir accepté de mettre à la disposition de GeoSiPol leurs données.