



HAL
open science

Géostatistique et cartographie du fond pédo-géochimique

Cécile Le Guern, Baptiste Sauvaget, Chantal De Fouquet

► **To cite this version:**

Cécile Le Guern, Baptiste Sauvaget, Chantal De Fouquet. Géostatistique et cartographie du fond pédo-géochimique. 5e Rencontres Nationales de la Recherche sur les sites et sols pollués, ADEME, Feb 2025, Paris, France. hal-04850552

HAL Id: hal-04850552

<https://hal.science/hal-04850552v1>

Submitted on 20 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Géostatistique et cartographie du fond pédo-géochimique

Test d'application du filtrage géostatistique sur le jeu de données METOTRASS

Cécile Le GUERN¹, Baptiste SAUVAGET², Chantal de FOUQUET³

¹ : BRGM, Direction régionale Pays de la Loire, 1 rue des Saumonières, 44323 NANTES Cedex 3, France, c.leguern@brgm.fr

² : eOde Sarl, 7 chemin de Mont-Riant, 2000 NEUCHÂTEL, Suisse, baptistesauvaget@eode.ch

³ : Mines Paris, PSL - Centre de Géosciences, 35 rue Saint-Honoré, 77305 FONTAINEBLEAU Cedex, France, chantal.de_fouquet@minesparis.psl.eu

Résumé

La méthode nationale de gestion des sites et sols pollués, de même que les guides de gestion des terres excavées se réfèrent aux fonds pédo-géochimiques comme valeurs de référence pour la qualité des sols. Or, du fait de l'absence d'échantillonnage spécifique pour la détermination de ces valeurs, une méthode de cartographie géostatistique par filtrage par analyse krigéante, visant à séparer les anomalies du fond, a été développée (Sauvaget et al, 2022). D'abord testée à l'échelle d'un quartier de centre urbain, elle est ici expérimentée à l'échelle d'un département via le jeu de données METOTRASS. Les résultats obtenus montrent que la méthode est applicable directement pour le zinc (en tenant compte de la relation avec l'altitude) ou plus indirectement pour l'arsenic, le plomb et le cuivre en partitionnant le territoire. Enfin le dégroupement appliqué aux données apparaît comme primordial dans le traitement statistique de données à la répartition spatiale hétérogène comme c'est souvent le cas pour les données en milieu urbain.

Introduction

Les fonds pédo-géochimiques représentent des référentiels importants pour la gestion des sols, de l'échelle du site à celle du territoire. La méthode nationale de gestion des sites et sols pollués s'y réfère, de même que les guides de gestion des terres excavées. Ils sont importants pour l'interprétation des analyses chimiques des sols, qui peuvent être menées dans le cadre de diagnostics de pollution, de valorisation de terres excavées ou de mise en culture alimentaire de sols. Ces problématiques sont particulièrement prégnantes en milieu urbain. Les travaux menés à l'échelle nationale dans le cadre du projet FGU3 ont permis de s'accorder sur une méthode de traitement statistique des données (Brunet et al, 2024), en prenant en compte les difficultés liées aux données anomaliques et aux regroupements de données. Compte-tenu de l'hétérogénéité des sols urbains (Penížek et Rohořková, 2006 ; Greinert, 2015)), le fond pédogéochimique est loin d'être homogène à l'échelle d'un territoire. Il apparaît ainsi nécessaire de cartographier sa variabilité pour disposer d'un référentiel plus fin. Néanmoins, la difficulté réside dans la gestion des données anomaliques, très présentes à l'échelle nationale dans la mesure où une grande partie des données d'analyse chimique de sols sont liées à des diagnostics de pollution. Pour résoudre cette difficulté, B. Sauvaget (2022) a développé une méthode de filtrage géostatistique pour cartographier le fond pédogéochimique.

Le filtrage géostatistique consiste en une analyse krigéante, qui est la décomposition de la teneur en plusieurs composantes, selon leur échelle de variabilité : sur le domaine d'étude, les composantes de « petite portée » sont interprétées comme les « anomalies », et celles de plus grande portée comme le « fond pédo-géochimique ». Cette décomposition présente l'avantage de dépendre de l'étendue du domaine d'étude.

L'objectif est de vérifier l'applicabilité du filtrage géostatistique pour cartographier à l'échelle territoriale le fonds pédo-géochimique à partir d'un jeu de données comportant des nombreuses anomalies. Le jeu de données METOTRASS (Le Guern et al, 2013, 2019) relatifs aux teneurs usuelles en As, Cu, Pb et Zn dans les sols de Loire-Atlantique en lien avec le substrat géologique, présente l'intérêt de couvrir un large territoire de manière assez complète et de comporter des anomalies d'origine naturelle. Une question centrale concerne les co-variables à utiliser pour faciliter le filtrage géostatistique.

Matériel et méthodes

Données utilisées

Les données de géochimie utilisées proviennent du jeu de données (METOTRASS, 479 données), du RMQS (26 données) et de la BDETM (230 données, mais aucune sur l'arsenic). Les données de géologie correspondent aux entités géochimiques cohérentes définies par Le Guern et al (2013), aux accidents tectoniques et aux unités géologiques issues de la carte géologique harmonisée de Loire-Atlantique (Béchenec, 2007). Des données de topographie (MNT IGN au pas de 25 m) ont été utilisées également. Le jeu de données METOTRASS se divise en un jeu de données principal acquis selon une reconnaissance aléatoire stratifiée pour la détermination des fonds pédo-géochimiques, et un jeu de données complémentaire à la recherche d'anomalies géochimiques (As et Pb principalement). Les teneurs en As, Cu, Pb et Zn mesurées au laboratoire par pXRF ont été corrigées à partir de corrélations

obtenues avec des teneurs obtenues par ICP sur un lot d'échantillons représentatifs. Elles sont ainsi supposées comparables aux teneurs des autres jeux de données qui ont été obtenues par ICP.

Source	Code	As	Cu, Pb, Zn
Metotrass	1	352	352
	2	21	21
	3	107	107
RMQS	0	26	26
BDETM	0	0	230
Total		506	736

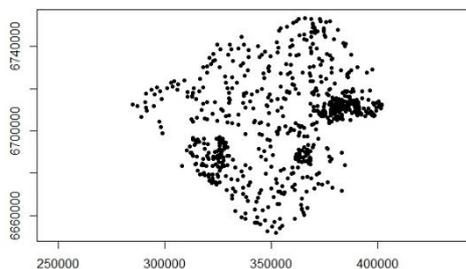


Tableau 1 : Nombre de données d'analyses de sol par source et localisation des échantillons sur le département de Loire-Atlantique

Traitement des données

Les traitements, réalisés sous R, Excel stat et Isatis, se divisent en trois étapes successives. La première est l'analyse exploratoire des données, cherchant à mettre en évidence les relations entre les données et à guider la suite des traitements en testant en particulier les liens avec des paramètres. Elle comprend des statistiques classiques et des ACP. La seconde étape est l'analyse variographique, étudiant la variabilité spatiale des données et l'influence des covariables sélectionnées. Le but est de fournir un modèle variographique adapté qui pourra être utilisé pour l'estimation des teneurs et du fond. La troisième étape regroupe l'estimation par krigeage (ou cokrigeage si des covariables pertinentes ont été identifiées) des teneurs au sein de la zone d'étude, ainsi que l'estimation par filtrage (analyse krigeante) du fond pédo-géochimique.

L'analyse variographique a démarré par le calcul des variogrammes expérimentaux simples et croisés omnidirectionnels. Pour essayer d'obtenir de meilleures structurations des données, des sélections dans les jeux de données d'une part, et la prise en compte d'une anisotropie d'autre part ont été effectuées. La recherche de covariables a fait l'objet d'une approche itérative. Dans ce cadre, l'altitude et la pente ont été retenues car elles sont susceptibles d'avoir une influence, en particulier sur les teneurs en zinc (Leenaers et al, 1990). Les domaines géologiques ont été testés sur la base de la répartition des teneurs. Les distances aux accidents tectoniques (cisaillements, grandes failles) et la aux plutons granitiques ont été considérées pour l'arsenic, car les anomalies s'expliquent par les grands accidents tectoniques (ex. cisaillement sud-armoricain) et la présence de plutons granitiques (via notamment des remontées hydrothermales).

Résultats et discussion

Analyse exploratoire des données

L'analyse exploratoire ne montre pas de corrélation marquée entre les 4 éléments étudiés (As, Cu, Pb, Zn). Une légère corrélation entre As et Pb ressort néanmoins. Elle peut s'expliquer partiellement par une interférence au niveau des spectres pXRF. Une corrélation intéressante du Zn avec l'altitude ressort. Elle peut s'expliquer par la mobilité plus importante de cet élément qui engendre des teneurs plus importantes en altitude plus basse. Aucune corrélation avec la pente ne ressort en revanche.

L'approche par domaine géologique apporte des nuances intéressantes. Les absences de corrélations entre éléments, et la légère corrélation entre As et Pb se retrouvent. Les teneurs en As sont plus élevées dans les domaines géologiques 2 et 3 (Tableau 2). Les teneurs en As sont en moyenne plus basses et moins dispersées dans les domaines géologiques 1 et 4. Les teneurs les plus élevées se situent à proximité de plutons granitiques dans les domaines géologiques 3 et 4.

As (mg/kg)		Jeu de données complet		METOTRASS 1 (principal)	
		sans pondération	avec pondération	sans pondération	avec pondération
Domaine géologique	Nb éch.				
1	44	27	26	25	27
2	113	34	37	35	37
3	156	81	49	49	39
4	150	23	23	23	22
1234		33	31	30	29,5

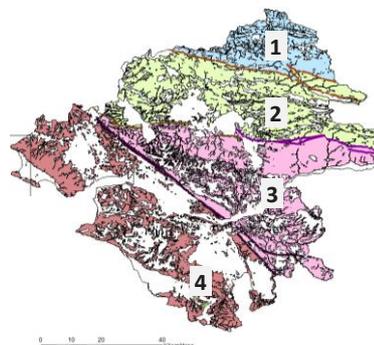


Tableau 2. Vibrisses interne supérieure calculées à partir des teneurs en As par domaine géologique et sur l'ensemble des domaines, avec et sans pondération, selon les jeux de données considérés. Pondération par l'inverse du nombre de voisins dans un cercle centré sur chaque donnée (de rayon 1 km pour le jeu de données complet, 1.5 km pour le jeu de données principal). Localisation des domaines géologiques sur la droite.

La distance aux accidents tectoniques et aux plutons granitiques explique une partie des teneurs, mais de nombreuses teneurs faibles sont également constatées dans ou à proximité. Ils n'apparaissent ainsi pas comme des co-variables pertinentes.

Le declustering corrige pour partie le suréchantillonnage. La pondération appliquée est proportionnelle à l'inverse du nombre de données situées dans un disque de rayon donné centré sur chaque donnée successivement. Le rayon est défini à partir de l'analyse statistique des distances au plus proche voisin des données. Si le rayon du disque reste partiellement arbitraire, cette méthode évite de fixer une origine arbitraire, comme c'est le cas lorsqu'une grille est utilisée pour déterminer la densité locale de l'échantillonnage. Une étude de sensibilité au rayon choisi pour le calcul de la pondération peut être utile au cas par cas. La pondération appliquée sur le jeu de données complet aboutit à une vibrisse équivalente à celle obtenue avec le jeu de données principal METOTRASS sans pondération (Tableau 2), ce qui semble très cohérent.

Analyse variographique

Les variogrammes simples par élément obtenus à partir du jeu de données complet sur l'ensemble du territoire apparaissent peu structurés sauf pour le Zn. Le croisement du variogramme du Zn avec l'altitude montre une anti-corrélation cohérente avec l'hypothèse d'une surconcentration en bas de pente. L'altitude peut donc servir de covariable pour l'estimation spatialisée des teneurs en zinc. Les cartes variographiques de chaque paramètre sur jeu de données complet ou partiel montrent une anisotropie globalement sud-est/nord-ouest, cohérente avec le cisaillement sud armoricain qui sépare les domaines géologiques 3 et 4. Les variations en arsenic apparaissent plus rapides le long du cisaillement que dans le reste du département. Ceci est cohérent avec les connaissances actuelles des faciès riches en arsenic (plus fortes anomalies le long du cisaillement). Le cuivre, le plomb et le zinc semblent plus homogènes.

L'analyse variographique omnidirectionnelle par domaine géologique montre par exemple pour l'arsenic un effet de pépite lié à la présence de forts contrastes de teneurs à petites distance. Le variogramme du Pb apparaît mieux structuré, de même que le variogramme croisé As-Pb. Le Pb peut donc être utilisé comme covariable pour la cartographie de l'arsenic.

Cartographie du fond pédo-géochimique

Les variogrammes expérimentaux suffisamment structurés du Zn ont permis l'application du krigeage et du filtrage sur le jeu de données complet, sans tri. Les variogrammes expérimentaux simples et croisés du zinc et de l'altitude ont servi à établir le modèle de corrélation par ajustement automatique avec trois structures de bases : une composante pépitique, une composante sphérique de portée 3,3 km et une autre composante sphérique de portée 30,5 km. L'anisotropie étant relativement faible dans le cas du zinc, elle n'est pas prise en compte.

L'estimation réalisée par cokrigeage suivant une grille régulière de 1 km de côté (Figure 1) montre une forte influence des composantes de courtes portées dans l'estimation. Cela fournit un signal relativement bruité qui ne semble pas cohérent avec le fond pédo-géochimique. Le filtrage par analyse krigéante s'appuyant sur la troisième composante permet une meilleure estimation du fond pédo-géochimique en zinc, avec une médiane de l'ordre de 40 mg/kg_MS. Le fond apparaît légèrement plus élevé dans le nord du département par rapport au sud.

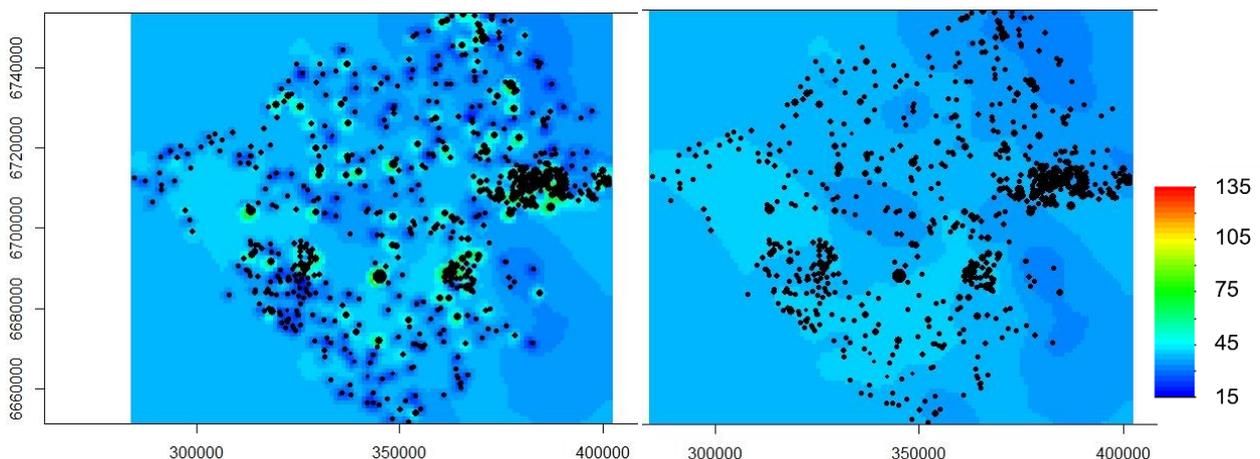


Figure 1 : Estimation des teneurs en zinc (mg/kg) dans les sols de la Loire-Atlantique par cokrigeage avec l'altitude à gauche, et par analyse krigéante à droite

Pour le filtrage, en l'absence d'indication supplémentaire, il y a une indétermination sur la « moyenne », à répartir entre « fond » et « anomalies ». Dans le cas présent, l'espérance (la moyenne probabiliste) des anomalies est arbitrairement supposée nulle.

L'indétermination est la suivante : Z désigne la teneur, dont le variogramme $\gamma(h)$ se décompose en une composante de petite portée $\gamma_1(h)$ et une composante « de grande portée » ou non bornée $\gamma_2(h)$. Dans le cas de deux composantes stationnaires d'ordre deux (donc dont le variogramme admet une portée), soient respectivement Y_1 et Y_2 deux Fonctions Aléatoires d'espérance nulle, et de variogramme respectif $\gamma_1(h)$ et $\gamma_2(h)$. m désignant l'espérance (la moyenne spatiale) de la teneur Z , alors toutes les décompositions $Z(x) = Z_1(x) + Z_2(x)$ telles que $Z_1(x) = m_1 + Y_1(x)$ et $Z_2(x) = m_2 + Y_2(x)$ avec $m_1 + m_2 = m$ sont admissibles ($m_2 > 0$ si la

teneur « de fond » est supposée positive). Le choix de m_2 (ou de m_1) détermine la moyenne de l'autre composante : $m_j = m - m_i$. Des hypothèses supplémentaires sont nécessaires pour déterminer m_1 et m_2 .

On montre (de Fouquet, 2022) que cette indétermination se retrouve dans les équations du filtrage. Dans la version du filtrage utilisée, $m_1 = 0$ et par suite, $m = m_2$. La teneur totale et le « fond » ont même espérance, ce qui n'apparaît pas réaliste. Le résultat de l'estimation est plutôt à interpréter comme la délimitation d'anomalies « négatives », « faibles » ou « fortes », ces classes étant donc définies à une constante inconnue près.

Conclusions et perspectives

La mise en œuvre du filtrage géostatistique (ou analyse krigeante) a bien fonctionné pour le zinc avec l'ensemble du jeu de données (METOTRASS, RMQS, BDETM) et la pente comme covariables. Elle apparaît plus compliquée pour l'arsenic, le plomb et le cuivre. Néanmoins, le découpage du territoire selon les domaines géologiques permet de dégager des corrélations spécifiques à chaque domaine. L'analyse exploratoire et variographique nécessite néanmoins dans ce cadre une forte expertise. Les résultats obtenus donnent bon espoir de réussir à estimer spatialement les teneurs en arsenic (mais aussi plomb et cuivre) par domaine.

Les difficultés rencontrées s'expliquent en partie par la nature du jeu de données, et en particulier par l'effet de pépité généré par les anomalies en arsenic. Des données plus nombreuses pour établir les variations à courte distance auraient été utiles.

La méthode de déclustering mise en œuvre présente un grand intérêt pour exploiter statistiquement des jeux de données présentant des biais, comme c'est le cas fréquemment dans les jeux de données en milieu urbain qui sont fréquemment acquises dans le cadre de diagnostics de pollution, avec des concentrations de données sur les sites étudiés.

Références

Béchenec, F. (2007) Carte géologique harmonisée du département de Loire-Atlantique. Rapport BRGM/RP55703-FR, 369 p.

Brunet et al (2024) FGU – Fonds Pédogéochimiques Urbains, Rapport final, Convention ADEME 2018-2022

Greinert, A. (2015) The heterogeneity of urban soils in the light of their properties. *J Soils Sediments* 15, 1725–1737 (2015). <https://doi.org/10.1007/s11368-014-1054-6>

Leenaers, H., Okx, J. and Burrough, P. (1990) Employing elevation data for efficient mapping of soil pollution on floodplains. *Soil Use and Management*, 6: 105-114. <https://doi.org/10.1111/j.1475-2743.1990.tb00818.x>

Le Guern C, Baudouin V, Conil P, Courtais B avec la collaboration de Houel M, Coffinet A, Latourte C, Gourmelen F (2013) METOTRASS : méthodologie optimisée pour l'évaluation des teneurs en éléments trace (As, Pb, Cu, Zn) dans les sols en domaine de socle : test sur le département de la Loire-Atlantique. Rapport final. ADEME. Angers, France Convention 0972C0030. Rapport BRGM/RP-63998-FR. 112 p. <https://infoterre.brgm.fr/rapports/RP-63998-FR.pdf>

Le Guern C., V. Baudouin, B. Clozel, P. Conil, H. Roussel (2019) Stratégie d'échantillonnage optimisée pour le fond pédo-géochimique de territoires hétérogènes, Quatrièmes Rencontres nationales de la Recherche sur les sites et sols pollués 26 et 27 novembre 2019, Le Beffroi de Montrouge (Portes de Paris)

Penížek V. & Rohošková M. (2006). Urban Soils: A Part of Man's Environment. In *Environmental Health in Central and Eastern Europe*, Eds. K. C. Donnelly and Leslie H. Cizmas, Springer Netherlands, p. 213-220.

Sauvaget B., de Fouquet C., Le Guern C., Renard D., Roussel H., 2022. Geostatistical filtering to map a 3D anthropogenic pedo-geochemical background for excavated soil reuse, *Journal of Geochemical Exploration*, vol. 240. <https://doi.org/10.1016/j.gexplo.2022.107031>

Remerciements

Nous remercions l'ADEME pour le soutien financier du projet FGU3 dans lequel ces travaux se sont inscrits (convention 1772C0050).