

Constitution de référentiels géochimiques locaux pour les sols et proches sous-sols urbains : de la base de données à l'interprétation géostatistique

Baptiste Sauvaget

► **To cite this version:**

Baptiste Sauvaget. Constitution de référentiels géochimiques locaux pour les sols et proches sous-sols urbains : de la base de données à l'interprétation géostatistique. Géochimie. Université Paris sciences et lettres, 2019. Français. NNT : 2019PSLEM014 . tel-02310465

HAL Id: tel-02310465

<https://pastel.archives-ouvertes.fr/tel-02310465>

Submitted on 10 Oct 2019

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.



THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à MINES ParisTech

Dans le cadre d'une collaboration avec le BRGM et l'Ademe

**Constitution de référentiels géochimiques locaux
pour les sols et proches sous-sols urbains :**
de la base de données à l'interprétation géostatistique

**Constitution of local geochemical references
for urban soils and subsoils :**
from database to geostatistical interpretation

Soutenue par

Baptiste SAUVAGET

Le 28 mars 2019

Ecole doctorale n° 398

**Géosciences, Ressources
Naturelles et Environnement**

Spécialité

Géosciences et géoingénierie

Composition du jury :

Catherine KELLER Professeur, CEREGE	<i>Rapporteur</i>
Fayçal REJIBA Professeur, Université de Rouen Normandie	<i>Rapporteur</i>
Philippe NEGREL Directeur de laboratoire, BRGM	<i>Examineur</i>
Nicolas SABY Chef de projet, INRA	<i>Examineur</i>
Olivier ROUSTANT Professeur, Mines Saint-Etienne	<i>Président du jury</i>
Hélène ROUSSEL Chef de projet, ADEME	<i>Examinatrice</i>
Cécile LE GUERN Chef de projet, BRGM	<i>Co-encadrante</i>
Chantal de FOUQUET Directrice de recherche, Mines ParisTech	<i>Directrice de thèse</i>
Jean-François BRUNET Ingénieur de recherche, BRGM	<i>Membre invité</i>



THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL



Géosciences pour une Terre durable

brgm

ADEME



Agence de l'Environnement
et de la Maîtrise de l'Énergie

REMERCIEMENTS

Avant de présenter ces travaux de recherches, je tiens à exprimer ma gratitude auprès de tous ceux grâce à qui cette thèse a aboutie.

Je tiens à remercier en premier lieu Chantal de Fouquet et Cécile Le Guern, ma directrice et mon encadrante de thèse, pour m'avoir accueilli dans leurs laboratoires respectifs, pour leur patience et leurs conseils ainsi que pour avoir guidé ces travaux tout au long de ces trois ans.

Je remercie également les personnes qui ont contribué à cette thèse : Hélène Roussel, Jacques Rivoirard, Béatrice Béchet, Jean-François Brunet, Christophe Schwartz, Nicolas Saby, Christophe Ducommun et Francis Douay, pour le temps qu'ils m'ont consacré, pour tous leurs précieux conseils et pour les différentes discussions que nous avons eues.

Je remercie chaleureusement Catherine Keller et Fayçal Rejiba pour avoir accepté de consacrer une partie de leur temps à évaluer ces travaux en tant que rapporteurs, et Philippe Négrel et Olivier Roustant pour avoir accepté de faire partie de ce jury de thèse.

Mes remerciements s'adressent aussi aux partenaires financiers sans lesquels ces travaux n'auraient pas vu le jour, l'Ademe et le BRGM, l'IRSTV, Pollusols et la région Pays de la Loire. Mais aussi la SAMOA et NMA car, sans la possibilité d'application sur leurs données, ces travaux n'auraient pas été possibles.

Merci également à Pierre-François Berrier, Bastien Collet, Antoine Billard, Hélène Demougeot-Renard, Hubert Leprond, Coline Eychène, Claire Fauchoux, Anne Blanchart, Aurélie Cambou et tous les autres avec qui j'ai pu collaborer au sein du Groupe de Travail du « Fond pédo-géochimique anthropisé », et des projets SUPRA et GéoBaPa.

Je tiens à remercier chaleureusement l'ensemble des équipes de Fontainebleau et de Nantes pour les bons moments lors des discussions de couloirs, des repas ensemble et des pauses cafés. Merci pour votre accueil, vos conseils et votre bienveillance de tous les jours. Merci à Didier et à Fabien pour leur aide précieuse sur R et en informatique sur des ordinateurs parfois très capricieux. Merci également à Sandrine, Nathalie, Véronique, Chloé et Régine pour leur aide dans l'organisation des nombreux déplacements de cette thèse.

Je voudrais aussi remercier ceux qui m'ont soutenu sans trop comprendre dans quoi je m'engageais ni ce que je faisais, ma famille, ainsi que mes amis.

Enfin, je terminerai ces remerciements par celle qui me soutient depuis bientôt 10 ans, qui a toujours compris l'emploi du temps souvent chargé d'une thèse et qui m'a épaulé même quand j'étais loin. Je suis chanceux de t'avoir à mes côtés car sans toi je n'aurais pas été si loin, merci Céleste.

TABLE DES MATIERES

INTRODUCTION.....	15
Partie I : ETAT DE L'ART ET DONNEES DISPONIBLES.....	21
I.A. Etat de l'art.....	21
I.A.1. Définitions	21
I.A.1.1. Sol et Sous-Sol.....	21
I.A.1.2. Contamination / Pollution.....	22
I.A.1.3. Fond pédo-géochimique anthropisé.....	23
I.A.2. Cartographie géochimique	26
I.A.2.1. Contexte historique	26
I.A.2.2. Bases de données	28
I.A.2.3. Méthodes de cartographie et intérêt de la géostatistique	31
I.A.3. Echantillonnage et analyse	33
I.A.3.1. Stratégie d'échantillonnage.....	33
I.A.3.2. Mode d'échantillonnage	36
I.B. Présentation des données utilisées.....	38
I.B.1. Les secteurs tests	38
I.B.2. Données de géochimie	41
Partie II : METHODES STATISTIQUES DE DETERMINATION D'UNE VALEUR SEUIL DE FOND PEDO-GEOCHIMIQUE ET ETUDE DE SENSIBILITE.....	47
II.A. Méthodes usuelles de détermination d'une valeur seuil de fond pédo-géochimique	47
II.A.1. Méthodes graphiques	48
II.A.2. Calculs statistiques	49
II.B. Sensibilité des calculs statistiques du seuil de fond géochimique	52
II.B.1. Influence de la distribution.....	53
II.B.2. Mélange de lois	58
II.B.3. Effectif et limite de quantification	62
II.B.4. Hétérogénéité d'échantillonnage.....	66
II.B.5. Discussion	70
Partie III : PRISE EN COMPTE DE L'HETEROGENEITE DES SOLS URBAINS	73
III.A. Déterminants de la géochimie des sols urbains.....	73
III.A.1. Nature des matériaux	73
III.A.2. Pressions anthropiques.....	75

III.A.3. Influence croisée des déterminants sur la géochimie : hétérogénéité des sols urbains	76
III.A.4. Données descriptives courantes de la nature et de l'occupation des sols : échelle de la ville	76
III.A.5. Données descriptives spécifiques au cas d'étude à l'échelle du quartier	78
III.B. Proposition de méthodologies pour la prise en compte des déterminants de la géochimie urbaine	80
III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques.....	82
III.C.1. Principe général et échelle de travail	82
III.C.2. Application à la ville et au quartier (Nantes)	87
III.C.3. Résultats cartographiques du croisement des descripteurs	91
III.C.4. Comparaison aux données réelles	94
III.C.5. Calculs statistiques de valeurs seuil	98
III.D. Regroupement des échantillons en classes homogènes.....	102
III.D.1. Objectif et principe général	102
III.D.2. La classification ascendante hiérarchique.....	103
III.D.3. Application aux données.....	106
III.D.4. Résultats et discussion	107
III.D.5. Test sur un autre quartier	112
III.E. Conclusions sur la prise en compte de l'hétérogénéité du milieu et les indicateurs possibles	115
Partie IV : SEPARER FOND ET ANOMALIES PAR FILTRAGE GEOSTATISTIQUE ..	117
IV.A. Le filtrage par analyse krigéante	120
IV.A.1. Une autre définition d'une anomalie	120
IV.A.2. Principe de l'analyse krigéante.....	122
a) Cas monovariéable	122
b) Cas multivariéable.....	123
IV.B. Application à un site.....	126
IV.B.1. Sélection du site et des données.....	126
IV.B.2. Résultats et discussion	133
IV.C. Application à un quartier	140
IV.C.1. Différences et limites par rapport à l'application à un site	140
IV.C.2. Résultats et discussion	145
IV.D. Prise en compte d'échantillons de supports variables.....	155

IV.E. Conclusion et perspectives	159
CONCLUSION	163
REFERENCES	169
Sites internet	185
TABLE DES ABREVIATIONS	186
Annexe 1 : Présentation du variogramme	187
Annexe 2 : Exemple de sondage et échantillonnage d'un diagnostic de pollution	189
Annexe 3 : Liste des laboratoires et principaux protocoles d'analyse recensés dans la base de données locale	190
Modalités d'analyse chimique	190
Données disponibles :	192
Annexe 4 : Résultats des sondages complémentaires et comparaisons (pXRF / ICP).....	194
Annexe 5 : Résultats de METOTRASS utilisés pour la notation du potentiel d'anomalie géochimique de la ville de Nantes.....	198
Annexe 6 : Tableau de synthèse des choix de notes de potentiel d'anomalie géochimique de Corine Land Cover et copernicus.....	201
Annexe 7 : Exemple des matrices de l'Inventaire Historique Urbain utilisé et principe de fonctionnement.....	204
Annexe 8 : Résultats d'analyses en cuivre suivant les découpages du territoire par analyse multi-critères	206
Annexe 9 : Représentation des échantillons du site sélectionnés pour l'essai de filtrage.....	207

TABLE DES FIGURES

Figure 1 : Schéma d'aménagement d'un ancien quartier industriel de centre urbain, exemple de l'île de Nantes, quartier de la création (SAMOA / Mg Design 2014).	15
Figure 2 : Principe de la démarche de valorisation des terres excavées à trois niveaux. (Coussy et al.,2017).	17
Figure 3: Schéma conceptuel de l'organisation du manuscrit.....	19
Figure 4 : Schéma conceptuel du fond pédo-géochimique anthropisé. Extrait du guide Ademe 2018.....	24
Figure 5: Schéma conceptuel du Fond Pédo-Géochimique naturel et urbain (modifié d'après Le Guern, communication personnelle).	25
Figure 6: Extrait de l'atlas de Webb,1978, sur la géochimie du Royaume-Uni.	27
Figure 7: Estimation des concentrations en plomb des sols de la ville de Pékin (Chine). Interpolation par krigeage ordinaire sur des échantillons de sols de surface (0-20cm). (Zheng, 2008)	28
Figure 8: Cartographie des teneurs en plomb dans les sols. Données RMQS (GisSol).....	30
Figure 9 :Carte géochimique de la région de Denver présentant la concentration en Pb d'échantillons de sols prélevés dans les 15 premiers centimètres (Smith et al., 2011).....	31
Figure 10 : Estimation des teneurs en arsenic (à gauche) et de l'écart-type de krigeage (à droite) dans des échantillons de sols d'une région de Calabre (Italie) (Guagliardi et al., 2012)	32
Figure 11 : Plan d'échantillonnage du Réseau de Mesure de la Qualité des Sols (RMQS) en France et mise en évidence de l'échantillonnage régulier dans la région parisienne (Saby, 2006).	34
Figure 12 : Exemple de l'influence de la densité d'échantillonnage. Une même zone hypothétique est échantillonnée suivant 3 densités différentes : a) 500m, b) 1000m et c) 1500m. Les teneurs en un métal Xm et la carte associée aux résultats est représentée en fon. (Demetriades et Birke, 2015).	34
Figure 13: Adaptation du maillage selon l'hétérogénéité géochimique, mise en place dans le cadre du projet METOTRASS afin d'optimiser le nombre d'échantillons de sols à prélever pour déterminer le fond géochimique (d'après Le Guern et al, 2013).	35
Figure 14: Schéma du prélèvement d'un échantillon composite. 5 sous-échantillons (points rouges) sont prélevés aux sommets et au centre d'un carré de côté « d ».....	36
Figure 15: Extrait de la carte géologique harmonisée de la Loire-Atlantique à 1/50 000 (Béchenec, 2007)	38
Figure 16 : Extrait des sites industriels et de services potentiellement polluants recensés sur Nantes dans la base de données BASIAS (www.georisques.gouv.fr)(2015).....	39

Figure 17 : Carte des sondages collectés en fonction du type de document d'origine (diagnostic de pollution ou étude géotechnique).....	41
Figure 19 : Deux protocoles d'échantillonnage présents dans les données collectées, sur un sondage. Les zones rouges et grises représentent des couches différentes du sous-sol. Le cas a) échantillonne selon une longueur constante (926 échantillons), alors que le cas b) échantillonne chaque couche(3704 échantillons).....	42
Figure 18 : Longueur et profondeur des échantillons de diagnostic de pollution collectés.	42
Figure 20 : Exemple de résultats d'ACP mettant en évidence les trois familles identifiées sur toutes les données brutes (pour un même effectif de 1853 échantillons). Rouge : hydrocarbures, bleu : métaux/ métalloïdes influencés par les pressions anthropiques, vert : métaux/métalloïdes d'origine géogénique	44
Figure 21 : Carte d'implantation des données collectées suivant leur origine.....	46
Figure 22 : Méthode de Lepeltier. Les fréquences cumulées sont exprimées en fonction des concentrations en mg/kg sur une échelle logarithmique. La flèche représente le passage au deuxième mode de distribution (Matschullat, 2000).	48
Figure 23 : Exemple de fonction de distribution théorique. Les deux lignes représentent les distributions originales (cercles) et théorique calculée (carrés). Le jeu de données utilisé est le même que celui de la figure 25 (Matschullat, 2000).	48
Figure 24 : Schéma d'une boîte de dispersion.	50
Figure 25 : Loi uniforme de minimum "m" et maximum "M"	53
Figure 26 : Exemple de loi normale de moyenne μ	55
Figure 27 : Exemple de loi log-normale de moyenne μ	56
Figure 28 : Mélange de lois uniformes.....	58
Figure 29 : Valeur seuil sur des mélanges de lois uniformes à différentes proportions α de fond.....	60
Figure 30 : Valeur seuil sur des mélanges de lois log-normales à différentes proportions α de fond, calculées par calcul direct.	61
Figure 31 : Résultats minima et maxima des seuils calculés sur 1000 itérations, sur un effectif croissant. Rouge : Q95, Vert : Vibrisse supérieure, Bleu : Seuil MAD. L'histogramme des données simulées est indiqué en ordonnée.Cas lognormal (moyenne = 1 et écart-type = 2.5).	62
Figure 32 : Minima et maxima des seuils calculés sur 1000 itérations, considérant une part d'échantillons sous la limite de quantification croissante. Remplacement des valeurs inférieures à la LQ par la LQ (en haut) et par 0 (en bas). Rouge : Q95, Vert : Vibrisse supérieure, Bleu : Seuil MAD. L'histogramme des données simulées est reporté en ordonnée. Cas lognormal (moyenne = 1 et écart-type = 2.5).....	63
Figure 33 : Résultats minimums et maximums des calculs de la distance interquartile (vert) et du terme MAD (bleu) sur 1000 itérations, considérant une part d'échantillons sous la limite de quantification croissante et un remplacement des valeurs censurées par la LQ.	64

Figure 34 : Résultats de calcul de seuils sur les échantillons du cas d'étude pour les 3 méthodes. Vert : Valeurs du fond pédo-géochimique naturel (calculé sur des alluvions).....	66
Figure 36 : Carte des teneurs en plomb dans une partie des échantillons collectés.	67
Figure 35 : Résultats de calcul de seuils sur les échantillons du cas d'étude pour les 3 méthodes, sans pondération, ou en pondérant par l'inverse du nombre d'échantillons dans des mailles de 25 ou 100 m de côté. Analyse de plomb, cuivre et zinc en mg/kg de MS.....	67
Figure 37: Diagrammes quantiles-quantiles des concentrations en cuivre, plomb et zinc, en mg/kg de MS, avec ou sans pondération par l'inverse du nombre d'échantillons dans une maille de 100m de côté.....	68
Figure 38 : Résultats de calculs de seuils sur les échantillons du cas d'étude pour les méthodes de la vibrisse supérieure et du 90 ^{ème} centile, en fonction de la typologie des remblais.(Le Guern et al., 2016).....	69
Figure 39 : Exemple de carte d'affleurement à 1 m de profondeur réalisée à partir du modèle 3D du sous-sol suivant différents types de remblais.	78
Figure 40 : Degré de potentialité aux hydrocarbures lié aux sources potentielles de pollutions localisées au sein des sites industriels et activités de service recensés – exemple de représentation possible (cas de l'île de Nantes) (Le Guern et al, 2017).....	79
Figure 41 : Photo-interprétation de l'historique d'occupation des sols par des activités maraîchères sur un des quartiers nantais en années.	79
Figure 42: Schéma conceptuel de la constitution de valeurs de fond pédo-géochimique présentant l'importance de la prise en compte d'une typologie de sols urbains, associée à la gestion des anomalies.....	81
Figure 43: Schéma global de l'analyse multi-critères empirique, dans le cas d'un seul descripteur de la nature des sols et de deux descripteurs de la pression anthropique.....	86
Figure 44: Croisement des descripteurs retenus suivant leur pertinence vis-à-vis de l'objectif de cartographie de potentiel d'anomalies géochimiques et de l'échelle.	87
Figure 45 : Occupation des sols fournie par Corine Land Cover (2016) sur la ville de Nantes.	90
Figure 46 : Potentiel de contamination en plomb, cuivre et zinc, suivant les notes intermédiaires liées à l'occupation des sols fournie par Corine Land Cover sur la ville de Nantes.....	90
Figure 47 : Carte du potentiel de présence d'anomalie en cuivre, plomb et zinc à l'échelle de la ville, obtenue par le croisement de Corine Land Cover et de la carte géologique au 1/50 000 ^{ème}	91
Figure 48 : Carte du potentiel de présence d'anomalie en cuivre, plomb et zinc à l'échelle de la ville obtenue par le croisement de Copernicus et de la carte géologique au 1/50 000 ^{ème}	92
Figure 49: Superposition des zones d'influence des sites Basias sur la carte du potentiel de présence d'anomalie en cuivre, plomb et zinc à l'échelle de la ville obtenue par le croisement de Copernicus et de la carte géologique au 1/50 000 ^{ème}	93

Figure 50: Carte du potentiel de présence d'anomalie en cuivre, plomb et zinc à l'échelle d'un quartier obtenue par le croisement de l'Inventaire Historique Urbain, de Copernicus et du modèle géologique 3D du sous-sol.	94
Figure 51 : Teneurs en plomb et zinc des échantillons collectés dans la ville. Les couleurs représentent le potentiel de présence d'anomalie suivant la carte réalisée à partir de Corine Land Cover (figure 45). Les moyennes par classe sont représentées par le trait noir.....	95
Figure 52 : Teneurs en plomb et zinc des échantillons collectés dans la ville. Les couleurs représentent le potentiel de présence d'anomalie suivant la carte réalisée à partir de Copernicus (figure 46). Les moyennes par classe sont représentées par le trait noir.....	96
Figure 53: Effectifs des échantillons collectés dans les 2 premiers mètres de profondeur suivant les classes d'anomalies de l'analyse multi-critère sur le quartier.....	96
Figure 54: Histogramme des teneurs en plomb et zinc des échantillons collectés sur un quartier de la ville. Les couleurs représentent le potentiel de pollution de la figure 50.....	97
Figure 55 : Schéma conceptuel de la constitution de valeurs de fond pédo-géochimique présentant l'importance de la prise en compte d'une typologie et d'un découpage du territoire par Analyse Multi-Critères.....	101
Figure 56: Schéma conceptuel des différentes étapes des deux approches pour définir une typologie des sols et proches sous-sols urbains adaptée au fond pédo-géochimique.	102
Figure 57: Distance euclidienne dans un repère formé par les paramètres X et Y, pour deux individus i et j.....	104
Figure 58: Exemple de dendrogramme.	104
Figure 59: Résultats sur un même jeu de données de a : K-moyennes, b : classification hiérarchique,.....	105
Figure 60 : Résultats de classification sur les données brutes des échantillons de moins de 2m de profondeur. Distance euclidienne, sans spatialisation.....	107
Figure 61 : Résultats de classification sur les données brutes des échantillons de remblais les moins profond de chaque sondage. Distance euclidienne, sans spatialisation.....	107
Figure 62: Résultats de classification sur les données centrées-réduites des échantillons de moins de 2 m de profondeur. Distance euclidienne, sans spatialisation.	108
Figure 63 : Résultats de classification sur les données log-translatées des échantillons de moins de 2 m de profondeur. Distance euclidienne, sans spatialisation.	109
Figure 64 : Résultats de classification sur les données log-translatées des échantillons de moins de 2 m de profondeur. Distance par la matrice de corrélation, sans spatialisation.....	110
Figure 65 : Résultats de classification sur les données log-translatées des échantillons de moins de 2 m de profondeur. Distance euclidienne, spatialisation par triangulation de Delaunay.....	111
Figure 66 : Dendrogrammes des classifications sur les teneurs en cuivre, plomb et zinc brutes et log-translatées du quartier des Gohards. Les points de couleurs représentent les trois classes distinguées.....	112

Figure 67: Nuages de corrélation entre les teneurs en cuivre, plomb et zinc suivant les trois classes identifiées par la classification sur données log-translatées du quartier des Gohards.	113
Figure 68: Nuage de corrélation des teneurs en cuivre et zinc suivant les classes identifiées par la classification sur les teneurs log-translatées du quartier des Gohards, avec une distance calculée via la matrice de corrélation.	113
Figure 69: Schéma conceptuel de l'évolution de la méthodologie de détermination du fond pédo-géochimique anthropisé intégrant la géostatistique pour l'estimation d'une gamme de valeurs de fond.	119
Figure 70: Schéma conceptuel de l'anomalie (en noir) et du fond (en gris).	120
Figure 71: Séparation des anomalies (en rouge) du fond (pointillés noirs) d'une variable Z le long d'un linéaire x, selon l'échelle de variabilité spatiale.	121
Figure 72: Plan d'implantation des sondages sur lesquels ont été prélevés les 208 échantillons. Les deux opérations sont indiquées en bleu et rouge.	126
Figure 73: Nuage de corrélation entre les profondeurs des échantillons du site et les typologies de matériaux définies par (Le Guern et al. 2016). Le trait plein relie les moyennes des profondeurs par classe.	127
Figure 74: Nuages de corrélation des concentrations en cuivre, hydrocarbures et chrome dans les échantillons du site en fonction de la profondeur et du type de remblais (rouge : remblais suspects, marron : remblais divers, orange : remblais assimilés naturels et vert foncé : alluvions).	128
Figure 75: Longueur des 208 échantillons du site. En gris les 86 échantillons de longueur comprise entre 50 et 70 cm, sélectionnés pour la suite de l'analyse exploratoire.	128
Figure 76 : Profondeur moyenne des 86 échantillons sélectionnés précédemment. En gris les 53 échantillons à moins de 2 m de profondeur.	129
Figure 77: Implantation des 53 échantillons sélectionnés (en noir) par rapport aux 208 d'origine (en rouge).	129
Figure 78 : Variogrammes expérimentaux simples des indicatrices de remblais. Le modèle ajusté est représenté en trait plein. Les données sont les 53 échantillons sélectionnés pour leur longueur comprise entre 50 et 70cm et leur profondeur inférieure à 2 m.	131
Figure 79 : Variogrammes expérimentaux simples et croisés des teneurs en cuivre, plomb et zinc. Le modèle ajusté est représenté en trait plein. Les données sont les 53 échantillons sélectionnés avec une longueur comprise entre 50 et 70cm et une profondeur inférieure à 2m.	131
Figure 80 : Variogrammes expérimentaux croisés des teneurs en cuivre, plomb et zinc avec les indicatrices de remblais divers et suspects. Le modèle ajusté est représenté en trait plein.	132
Figure 81 : Estimation des teneurs en zinc par cokrigage (en haut) et estimation du fond pédo-géochimique de la même substance par filtrage (en bas).	135

Figure 82 : Estimation des teneurs en plomb par cokrigeage (en haut) et estimation du fond pédo-géochimique de la même substance par filtrage (en bas).....	135
Figure 83: Teneurs en zinc (mg/kg_MS) dans les échantillons du site (en haut) et dans les estimations (gauche : cokrigeage, droite : filtrage). L’histogramme du cokrigeage est tronqué et ne présente pas les teneurs supérieures à 300 mg/kg_MS.....	136
Figure 84:Diagrammes de dispersion des échantillons de départ et des résultats de cokrigeage et de filtrage du zinc en mg/kg_MS. La moyenne est représentée par le x. Les outliers supérieurs à 250 mg/kg_MS dans les diagrammes de dispersion des échantillons et du cokrigeage ne sont pas représentés.....	137
Figure 85: Teneurs en plomb (mg/kg_MS) dans les échantillons du site (en haut) et dans les estimations (gauche : cokrigeage, droite : filtrage).L’histogramme des échantillons et du cokrigeage est tronqué et ne présente pas les teneurs supérieures à 400 mg/kg_Ms.....	138
Figure 86: Nuage de corrélation entre les teneurs en cuivre, plomb et zinc et la longueur de l’échantillon et leur moyenne par classe (de 20 échantillons) (trait plein).	141
Figure 87: Fréquence des pourcentages représentatifs suivant le type de remblais.....	142
Figure 88:Nuage de corrélation entre les teneurs en cuivre, plomb et zinc et le pourcentage représentatif et leur moyenne par classe (de 20 échantillons).....	142
Figure 89 : Variogrammes expérimentaux simples et croisés des teneurs en cuivre, plomb et zinc. Le modèle ajusté est représenté en trait plein. L’histogramme des paires de calcul est reporté en rouge.....	144
Figure 90 : Estimation des teneurs en cuivre par cokrigeage (en haut) et du fond pédo-géochimique de la même substance par filtrage (en bas).....	146
Figure 91: Estimation des teneurs en plomb par cokrigeage (en haut) et du fond pédo-géochimique de la même substance par filtrage (en bas).....	147
Figure 92 : Estimation des teneurs en zinc par cokrigeage (en haut) et du fond pédo-géochimique de la même substance par filtrage (en bas).....	148
Figure 93: Teneurs en cuivre (mg/kg_MS) dans les échantillons des remblais assimilables naturels du quartier (en haut) et dans les estimations (gauche : cokrigeage, droite : filtrage). Les flèches noires indiquent des données supérieures qui ne sont pas représentées.....	149
Figure 94: Teneurs en plomb (mg/kg_MS) dans les échantillons des remblais assimilables naturels du quartier (en haut) et dans les estimations (gauche : cokrigeage, droite : filtrage).Les flèches noires indiquent des données supérieures qui ne sont pas représentées.	150
Figure 95:Teneurs en zinc (mg/kg_MS) dans les échantillons des remblais assimilables naturels du quartier (en haut) et dans les estimations (gauche : cokrigeage, droite : filtrage).Les flèches noires indiquent des données supérieures qui ne sont pas représentées.	150
Figure 96: Carte d’affleurement des remblais à 1m de profondeur, selon le modèle géologique 3D réalisé à partir des sondages et de données historiques.....	153
Figure 97: Estimation du fond pédo-géochimique anthropisé dans les remblais divers et assimilables naturels à une profondeur de 1m.	154

Figure 98 : Nuages de corrélation des estimations en cuivre, plomb et zinc obtenues par filtrage avec sélection de données (ordonnée) et en discrétisant le support en ndisc points (abscisse).	157
Figure 99 : Nuages de corrélation des écarts-type d'estimation du cuivre, plomb et zinc obtenues par filtrage avec sélection de données (ordonnée) et en discrétisant le support en ndisc points (abscisse).	158
Figure 100: Résultats de cartographie des zones de valorisation d'un lot de terres excavées dans la tranche 0-1m de profondeur. Les teneurs utilisées pour simuler le lot sont :	160
Figure 101: Schéma conceptuel des traitements à appliquer pour la constitution de valeurs de référence de la qualité géochimique des sols urbains, à l'échelle de la ville en cas d'un effectif réduit de données disponibles.	167
Figure 102: Schéma conceptuel des traitements à appliquer pour la constitution de valeurs de référence de la qualité géochimique des sols urbains, à l'échelle d'un quartier en cas d'un effectif important de données disponibles.	168
Figure 103: Exemple de variogramme illustrant le palier, la portée et l'effet de pépité.	187
Figure 104: Exemple de résultats de sondage extrait d'un diagnostic de pollution. Les échantillons sont indiqués comme « T2-1 » et « T2-2 ». Figure extrait d'un rapport fourni par la SAMOA.	189
Figure 105 : Exemple de diagramme des étapes de la préparation des analyses de sols urbains. (Allen et Cave, 2011)	190
Figure 106: Diagramme de la répartition de toutes les données collectées sur les quartiers suivant le laboratoire d'analyse.	192
Figure 107: Récapitulatif des méthodes d'analyse principale renseignées dans la base de données suivant la fraction et la famille analysée, ainsi que le pourcentage de données ayant une méthode d'analyse renseignée.	193
Figure 108 : Position des sondages complémentaires	194
Figure 109: Nuage de corrélation et droite de regression des analyses en nickel par pXRF et ICP.	195
Figure 110 : Nuage de corrélation et droite de regression des analyses en plomb par pXRF et ICP.	195
Figure 111: Nuage de corrélation et droite de regression des analyses en cuivre par pXRF et ICP.	196
Figure 112: Nuage de corrélation des analyses en antimoine par pXRF et ICP	196
Figure 113: Nuages de corrélation et droite de regression des analyses en arsenic, baryum, chrome et zinc par pXRF et ICP.	197
Figure 114 : Carte géologique harmonisé (d'après Bechenec, 2007) à 1/50 000 de la ville de Nantes.	200
Figure 115 : Potentiel d'anomalie géochimique en cuivre, plomb et zinc suivant la géologie	200

Figure 116 : Potentiel d'anomalie géochimique en cuivre, plomb et zinc suivant l'occupation des sols de la cartographie Copernicus.	203
Figure 117: Cartographie de synthèse du potentiel de présence de contamination en plomb à partir des résultats d'IHU (modifiée d'après Desse-Engrand, 2015).	204
Figure 118 : Potentiel d'anomalie géochimique en cuivre, plomb et zinc lié aux activités industrielles et de service potentiellement polluantes de 1850 à nos jours.	205
Figure 119: Teneurs en cuivre des échantillons collectés dans la ville. Les couleurs représentent le potentiel de présence d'anomalie suivant la carte réalisée à partir de Corine Land Cover (Gauche) et celle réalisée à partir de Copernicus (Droite). Les moyennes par classe sont représentées par le trait noir.	206
Figure 120 : Echantillons du site sélectionné pour l'essai de filtrage le long de leur sondage en fonction de la profondeur. Le code couleur correspond aux typologies de remblais (Le Guern et al, 2016).	207

TABLE DES TABLEAUX

Tableau 1: Effectifs des 14 paramètres considérés dans la suite des traitements.	43
Tableau 2 : Matrice de corrélation mettant en évidence les trois familles identifiées sur les données brutes (pour un même effectif de 1853 échantillons).	44
Tableau 3 : Rapport seuil/ e^{μ} en fonction de σ , pour le calcul direct dans le cas d'une loi log-normale.	56
Tableau 4 : Différenciation des types de remblais. Les remblais assimilables naturels sont caractérisés par l'absence de tous les termes cités dans le tableau (Le Guern et al., 2016).	74
Tableau 5: Récapitulatif des descripteurs disponibles, de leur échelle de validité et de leur disponibilité.	83
Tableau 6 : Choix des notes intermédiaires si deux descripteurs existent pour la pression anthropique ou la nature des sols ou sous-sols. Etape renseignée comme « synthèse par type de descripteur » à la figure 43.	85
Tableau 7: Notes de potentiel d'anomalies géochimique attribuées aux codes de Corine Land Cover et de Copernicus dans le cas de métaux liés aux pressions anthropiques (cuivre, plomb et zinc). Les notes sont fournies dans la colonne « Pression ». La note 999 correspond aux surfaces immergées non prises en compte dans le modèle.	88
Tableau 8 : Notes de potentiel d'anomalie géochimique suivant les combinaisons possibles des notes intermédiaires (de 0 pour un potentiel nul à 2 pour un potentiel fort) des descripteurs de pression anthropique et de nature des sols et sous-sols dans le cas de métaux marqués par une origine anthropique (Cu, Pb, Zn) dans les deux premiers mètres de profondeur.	89
Tableau 9 : Résultats des calculs de seuils statistiques du fond pédo-géochimique anthropisé en plomb et zinc (en mg/kg_MS) par classe de potentiel de présence d'anomalie géochimique d'un quartier urbain sur la tranche 0-2 m.	99

Tableau 10 : Résultats des calculs de seuils par type de remblais du quartier de l’Ile de Nantes (recalculé pour les échantillons de moins de 2 m de profondeur à partir du jeu de données de Le Guern et al., 2016).....	100
Tableau 11 : Coefficients de corrélation régionalisés du cuivre, du plomb, du zinc et des remblais divers et suspect pour l’effet de pépite du modèle linéaire de corrégalisation...	133
Tableau 12 : Coefficients de corrélation régionalisés du cuivre, du plomb, du zinc et des remblais divers et suspect pour la composante sphérique du modèle linéaire de corrégalisation.	134
Tableau 13 : Coefficients de corrélation régionalisés du cuivre, du plomb, du zinc et des remblais divers et suspect pour la composante linéaire du modèle linéaire de corrégalisation.	134
Tableau 14: Quantiles calculés sur les échantillons et les estimations par cokrigeage et filtrage des teneurs en zinc du site.	137
Tableau 15: Quantiles calculés sur les échantillons et les estimations par cokrigeage et filtrage des teneurs en plomb du site.	138
Tableau 16 : Coefficients de corrélation régionalisés du cuivre, du plomb et du zinc pour l’effet de pépite du modèle linéaire de corégionalisation, à l’échelle du quartier.....	145
Tableau 17 : Coefficients de corrélation régionalisés du cuivre, du plomb et du zinc pour la composante sphérique de portée 200 m du modèle linéaire de corégionalisation, à l’échelle du quartier.	145
Tableau 18 : Coefficients de corrélation régionalisés du cuivre, du plomb et du zinc pour la composante sphérique de portée 1000 m du modèle linéaire de corégionalisation, à l’échelle du quartier.	146
Tableau 19: Quantiles calculés sur les échantillons et les estimations par cokrigeage et filtrage des teneurs en cuivre des remblais assimilables naturels du quartier, et leur représentation sur l’histogramme des teneurs estimées par filtrage (Q95 : rouge, Vibrisse supérieure : vert, seuil MAD : bleu).	151
Tableau 20: Quantiles calculés sur les échantillons et les estimations par cokrigeage et filtrage des teneurs en plomb des remblais assimilables naturels du quartier, et leur représentation sur l’histogramme des teneurs estimées par filtrage (Q95 : rouge, Vibrisse supérieure : vert, seuil MAD : bleu).	152
Tableau 21: Quantiles calculés sur les échantillons et les estimations par cokrigeage et filtrage des teneurs en zinc des remblais assimilables naturels du quartier, et leur représentation sur l’histogramme des teneurs estimées par filtrage (Q95 : rouge, Vibrisse supérieure : vert, seuil MAD : bleu).	152
Tableau 22 : Potentiel d’anomalie géochimique des faciès géologiques présents sur la ville de Nantes d’après la base de données METOTRASS (Le Guern et al, 2013) et interprétés pour l’analyse multi-critères du potentiel de présence d’anomalies géochimiques.	199
Tableau 23 : Détail des choix de note de potentiel d’anomalie en cuivre, plomb et zinc attribuées à Corine Land Cover.....	202

Tableau 24: Extrait de la table d'attribution des notes intermédiaires pour l'analyse multi-critère suivant les résultats des matrices polluant de l'Inventaire Historique Urbain de Nantes.
..... 205

INTRODUCTION

A partir des années 1980, après les « Trente Glorieuses », la France comme d'autres pays fortement industrialisés, a connu de grands changements socio-économiques. L'intensification de la pression démographique en ville, liée à un fort exode rural couplé à une importante désindustrialisation, pousse les aménageurs à repenser et à réaménager les villes. Dans un contexte de hausse des prix de l'immobilier puis de politique de limitation de l'étalement urbain, la libération de vastes emprises industrielles, souvent à proximité du centre-ville, constitue une aubaine pour de nombreuses collectivités. Les projets d'aménagement visent alors à redensifier et à reconstruire les villes sur elles-mêmes, notamment en reconvertissant les anciennes friches industrielles.

Des villes comme Lyon, Bordeaux, Le Havre ou Nantes (Figure 1) et aujourd'hui Paris (<http://www.grandparisamenagement.fr/>), ont ainsi investi dans d'ambitieux projets de renouvellement urbain, visant à répondre à la demande en logements et à renforcer l'attractivité de leur agglomération.

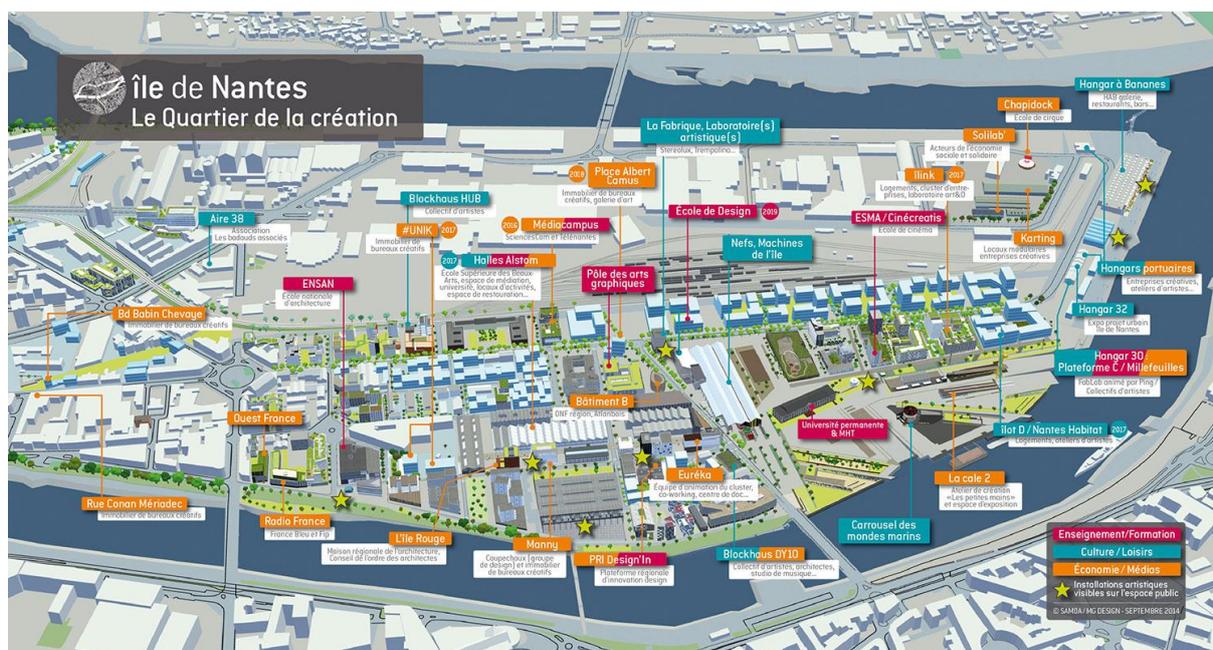


Figure 1 : Schéma d'aménagement d'un ancien quartier industriel de centre urbain, exemple de l'île de Nantes, quartier de la création (SAMOA / Mg Design 2014).

Une difficulté lors du réaménagement de friches industrielles est liée aux pollutions des sols. Si les pollutions ponctuelles problématiques rentrent dans le cadre de la méthodologie nationale de gestion des sites et sols pollués (cf. circulaires de 1993 et 2007), une partie importante de ces friches mais aussi du tissu urbain est impactée par des pollutions diffuses. Or, le réaménagement de friches urbaines génère d'importants volumes de terres excavées.

INTRODUCTION

Cependant, selon l'Ordonnance n° 2010-1579 du 17 décembre 2010, les terres excavées (naturelles ou non) qui sortent du site dont elles sont extraites ont un statut de déchet¹. Ainsi, toute gestion ou valorisation des terres excavées en dehors de leur site d'origine doit se faire suivant la réglementation des déchets². Ce statut de déchet apparaît comme un facteur limitant pour leur valorisation.

Toutefois, pour éviter la mise en installation de stockage de ces terres et du fait de la tension sur les ressources en matériaux de construction, le Ministère en charge de l'environnement a souhaité faciliter leur valorisation en aménagement et en technique routière, tout en préservant les enjeux environnementaux et sanitaires. Pour cela, il a édité un guide permettant d'appliquer aux terres modérément contaminées une logique d'économie circulaire (Blanc et al. 2012, Coussy et al. 2017). Cette gestion devrait engendrer des gains économiques et environnementaux importants (Jeannot et al, 2014). Elle doit s'assurer néanmoins de limiter l'impact sur l'environnement tout en vérifiant la compatibilité entre la qualité des milieux et les usages dans les aménagements futurs. Dans le cas d'une valorisation initialement impossible, les terres peuvent faire l'objet de traitements (in-situ ou en centre de traitement) afin de les rendre compatibles avec leur usage futur.

Pour permettre la valorisation des terres excavées en aménagement et technique routière, le guide de gestion de terres modérément contaminées (Coussy et al. 2017) impose que trois critères soient vérifiés : a) le maintien de la qualité des sols, b) la préservation des ressources en eau et c) la compatibilité sanitaire avec l'usage futur (Figure 2). Le fond pédogéochimique anthropisé intervient dans le premier critère (a).

Les valeurs seuils libératoires pour la gestion des terres excavées à l'échelle nationale du niveau 1 (Figure 2) s'appuient sur des données Aspitet (Apports d'une Stratification Pédologique à l'Interprétation des Teneurs en Eléments Traces) (Baize, 2008). Si les terres à valoriser ne respectent pas ces seuils, il est possible de s'appuyer sur un fond géochimique local. Cependant, il manque en France, pour établir le fond pédogéochimique local, tant des données de référence sur les sols urbains que des méthodologies spécifiques. La démarche de mise en place de données de référence est en cours, via le projet ADEME-BRGM « Fond pédogéochimique urbain - FGU » et la base de données BDSolU³.

¹ Les déchets sont définis comme « toute substance ou tout objet, ou plus généralement tout bien meuble, dont le détenteur se défait ou dont il a l'intention ou l'obligation de se défaire » (article L. 541-1-1 du code de l'environnement).

² selon l'article L541-2 du Code de l'Environnement, « tout producteur ou détenteur de déchets est tenu d'en assurer ou d'en faire assurer la gestion, conformément aux dispositions du présent chapitre. Tout producteur ou détenteur de déchets est responsable de la gestion de ces déchets jusqu'à leur élimination ou valorisation finale, même lorsque le déchet est transféré à des fins de traitement à un tiers. Tout producteur ou détenteur de déchets s'assure que la personne à qui il les remet est autorisée à les prendre en charge ».

³ <http://www.BDSolU.fr>

INTRODUCTION

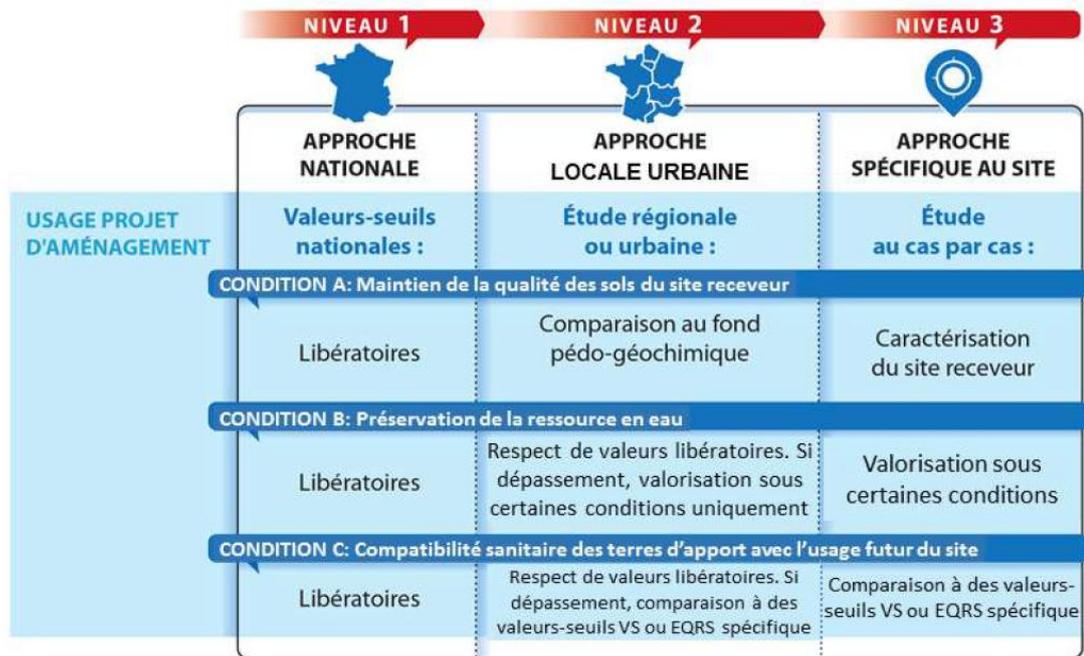


Figure 2 : Principe de la démarche de valorisation des terres excavées à trois niveaux. (Coussy et al., 2017).

Faute de données de référence sur les sols urbains disponibles à l'échelle nationale⁴, il est donc actuellement nécessaire d'utiliser d'autres données pour élaborer des référentiels de fond géochimique en milieu urbain. La thèse s'intéresse plus particulièrement à l'exploitabilité des données issues de diagnostics de pollution qui représentent la majorité des sondages effectués en milieu urbain. Elle contribue aux développements méthodologiques en vue de proposer des référentiels de qualité géochimique des sols urbains, dans une perspective de valorisation des terres excavées à l'échelle du quartier ou de la ville.

Réalisée en collaboration avec l'Ademe, le BRGM et le centre de géosciences de Mines ParisTech, cette thèse s'inscrit dans un contexte d'études sur le fond pédo-géochimique et sur les sols urbains, en relation avec le Groupe de Travail sur les valeurs de Fond pédo-géochimique anthropisé, piloté par l'Ademe. Ce groupe avait pour objectif la rédaction d'un guide national de bonnes pratiques (Ademe, 2018). La thèse s'intègre également dans le cadre du projet SUPRA (Sols Urbains en Projet d'Aménagements), co-financé par le programme Ademe GRAINE. Ce projet vise notamment à éclairer par une approche novatrice et pluridisciplinaire (géosciences, pédologie, agronomie urbaine, cartographie, urbanisme,

⁴Les campagnes d'échantillonnage et d'analyse nécessaires pour la détermination de valeurs de fond pédo-géochimique sont plus aisées en milieux naturels ou ruraux car les variations spatiales de la géochimie sont plus faibles. Ainsi un échantillon peut être représentatif d'une surface importante. En revanche, en milieu urbain où la géochimie est plus largement influencée par les activités humaines, obtenir un échantillonnage représentatif est plus complexe. Cela nécessite une connaissance poussée du sol/sous-sol et des anciennes activités ayant pu impacter la géochimie, ainsi qu'un grand nombre d'échantillons, notamment dans des zones difficiles d'accès. C'est une des raisons pour lesquelles des campagnes d'échantillonnage spécifiques au fond pédo-géochimique en milieu urbain ne sont pas envisagées à l'heure actuelle en France.

INTRODUCTION

aménagement du territoire) la connaissance des sols urbains et la définition d'un référentiel commun d'indicateurs de qualité de ces sols, en lien avec les services écosystémiques attendus. Ce projet prévoit également la mise en place d'un réseau de connaissance et de gestion des sols urbains notamment par une approche des services rendus par les sols. Il intègre deux autres thèses cofinancées par l'Ademe :

- Anne Blanchart : « Vers une prise en compte des potentialités des sols dans la planification et l'urbanisme opérationnel » (Blanchart, 2018).
- Aurélie Cambou : « Évaluation du stock et de la stabilité du carbone organique dans les sols urbains » (Cambou, 2018).

La thèse a été également soutenue par le programme POLLUSOLS, cofinancé par la région des Pays de la Loire. POLLUSOLS étudie les pollutions diffuses dans le continuum terre-mer, en s'appuyant sur le bassin versant aval de la Loire. Un des nombreux objectifs du programme est de structurer la recherche régionale dans le domaine en favorisant les échanges pluridisciplinaires.

Elle a également contribué aux discussions au sein du projet GéoBaPa. Celui-ci a pour but de fournir un outil de cartographie du fond pédo-géochimique à l'échelle régionale pour une meilleure valorisation des terres excavées sur le bassin de la Seine, de Paris au Havre. Cette zone d'étude se situe dans les régions Ile de France et Normandie.

Dans ce contexte pluridisciplinaire, la thèse fait appel à trois disciplines : la géochimie, la géomatique et la géostatistique. Elle aborde les problèmes liés au contexte français de disponibilité des données de géochimie des sols urbains. Ne disposant pas, contrairement à d'autres pays (Royaume-Unis, Suède, Allemagne), de données de qualité géochimique des sols en milieux urbains acquises pour la détermination du fond pédo-géochimique anthropisé selon un maillage régulier, mais principalement de données issues de diagnostics de pollution, de nombreuses questions se posent sur l'exploitation de ces données pour déterminer le fond pédo-géochimique anthropisé. Les méthodes statistiques de détermination d'une valeur seuil sont-elles pertinentes ? Quels outils de réflexion pour la mise en place de référentiels (Fond Pédo-géochimique) à une échelle locale peuvent être mis en place ? Comment tenir compte d'une typologie de sols urbains ? Quels sont les apports de la géostatistique à cette réflexion ?

Pour répondre à ces questionnements, le manuscrit présente en première partie un état de l'art des différentes notions et méthodologies actuelles pour la détermination de valeurs seuils du fond pédo-géochimique, ainsi que les données disponibles qui ont pu être collectées et qui seront utilisées pour la suite des travaux. La deuxième partie propose une étude critique des critères statistiques usuels et une étude des paramètres susceptibles d'influencer les résultats fournis par ces outils statistiques. Pour cela, leurs applications à des cas théoriques et pratiques seront étudiées. Deux méthodes de segmentation du territoire s'appuyant sur l'intégration des données descriptives et d'une typologie de sols urbains sont discutées dans la troisième partie. La première méthode vise à fournir une segmentation *a priori* du territoire suivant des données descriptives interprétées via les résultats de recherches bibliographiques

INTRODUCTION

et de retours d'expérience ; elle correspond à l'approche privilégiée dans le guide de détermination du fond pédo-géochimique anthropisé (Ademe, 2018). La seconde méthode présente une classification des échantillons collectés suivant les teneurs mesurées de façon à réaliser des classes homogènes, ensuite mises en relation avec des données descriptives cartographiques. Enfin, la quatrième partie de la thèse discute des apports de la géostatistique, et notamment du filtrage pour la séparation des anomalies et l'estimation du fond pédo-géochimique anthropisé sur un critère d'échelle de variabilité spatiale.

L'ensemble de la démarche est synthétisée à la Figure 3.

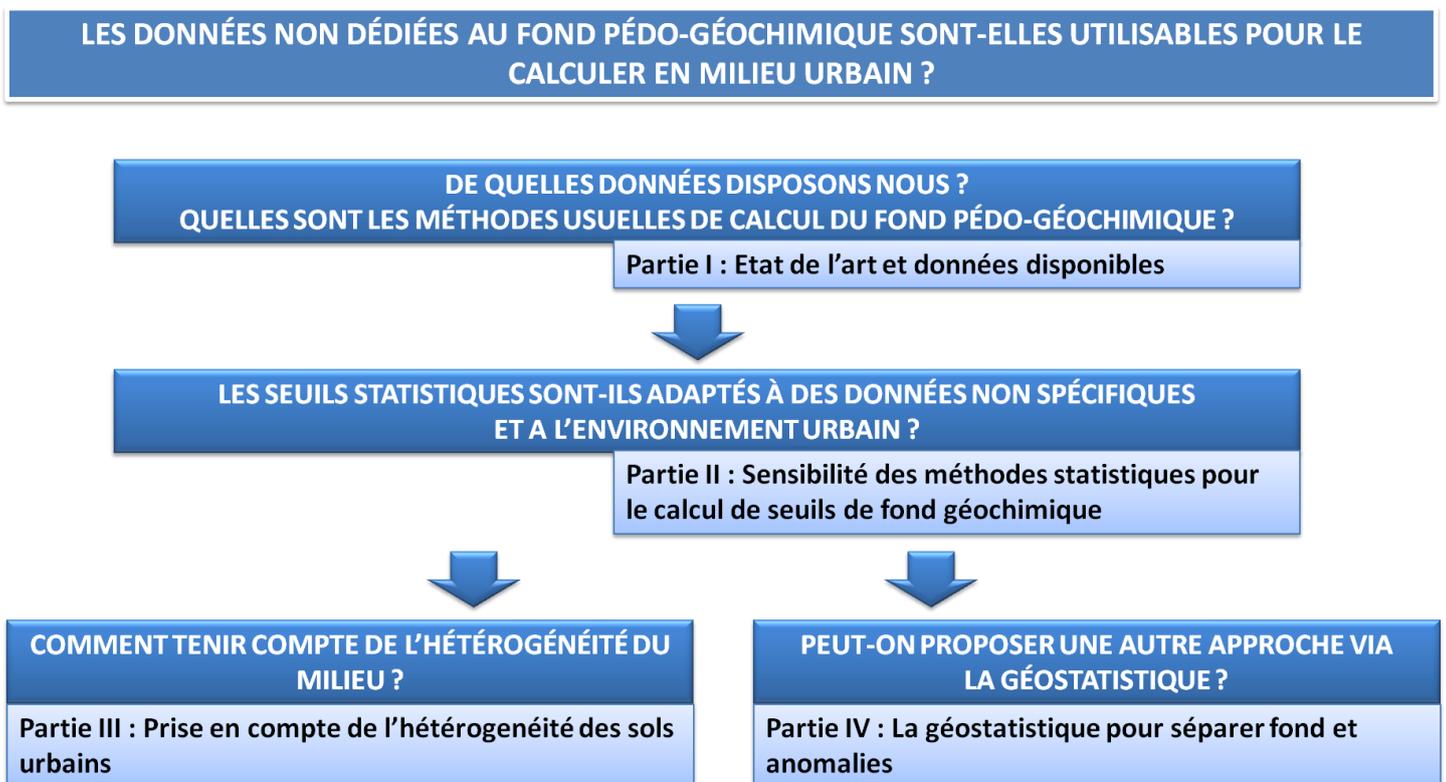


Figure 3: Schéma conceptuel de l'organisation du manuscrit

Partie I : ETAT DE L'ART ET DONNEES DISPONIBLES

Il existe différentes définitions du fond pédo-géochimique qui ont pu évoluer au cours du temps et selon les pays. On cherche ici à le caractériser afin de l'utiliser ensuite comme référentiel pour la valorisation des terres excavées. On revient alors sur les définitions des objets d'étude, l'évolution de cartographie de la géochimie et les données couramment utilisées pour y parvenir. En parallèle, les spécificités de l'approche française et de cette thèse sont rappelées.

I.A. Etat de l'art

I.A.1. Définitions

I.A.1.1. Sol et Sous-Sol

La définition du sol est primordiale car elle concerne l'objet d'étude. Une définition relativement générale le décrit comme « *faisant partie intégrante des écosystèmes terrestres et constituant l'interface entre la surface de la terre et le socle rocheux. Il se subdivise en couches horizontales successives aux caractéristiques physiques, chimiques et biologiques spécifiques. Il a également différentes fonctions* » (Council of Europe, 1990).

Or il apparaît que cette définition dépend du domaine d'expertise.

En agronomie, le sol est défini comme un **support des plantes**, cultivées ou non. C'est la zone exploitée par les racines. Il présente donc un intérêt comme ressource pour la vie végétale.

En pédologie il est décrit comme « *la formation naturelle de surface, à structure meuble et d'épaisseur variable, résultant de la transformation de la roche mère sous-jacente sous l'influence de divers processus, physiques, chimiques et biologiques, au contact de l'atmosphère et des êtres vivants* » (Demolon, 1952). Il n'y a donc pas de notion d'épaisseur limite à proprement parler ; dans certains cas, celle-ci peut atteindre plusieurs mètres. Toutefois, la majorité des études se concentrent sur les 30 premiers centimètres, en se concentrant sur les applications agronomiques.

En archéologie urbaine, il est défini comme « *le volume de sol produit par les activités humaines de leurs origines à nos jours. Caractérisé par une épaisseur importante généralement comprise entre deux et douze mètres, il est une source de l'histoire des villes et de leur communauté d'habitants* » (Laurent et Fondrillon, 2010, Laurent, 2007 ; Fondrillon, 2007).

Ainsi la définition varie selon les enjeux de l'étude. Le sol peut apparaître comme un **support de la vie**, une **zone d'échange** entre différents milieux (hydrosphère, atmosphère, biosphère), une **archive du passé**, un **bien foncier** ou encore une **ressource**.

Nous le considérerons ici davantage comme un ensemble de **matériaux potentiellement mobilisé lors de travaux d'aménagements** (et d'excavation). Nous avons donc besoin d'une définition plus large, telle que celle fournie par la base de référence mondiale sur les sols (World Reference Base for Soils), où les sols sont décrits comme « *tout objet faisant partie de l'épiderme de la terre* » (Sokolov, 1997; Nachtergaele, 2005). Dans cette définition le sol comprend « *tout matériau dans les deux premiers mètres de la surface de la Terre qui est en contact avec l'atmosphère, à l'exclusion des organismes vivants, des zones couvertes par un continuum de glace qui n'est pas recouverte par un autre matériau et des masses d'eau d'une profondeur supérieure à deux mètres* » (WRB, 2015).

Néanmoins, dans le contexte de la réutilisation de terres excavées, nous considérerons cette dernière définition des sols en considérant la profondeur possible d'excavation (une couche de plusieurs mètres d'épaisseur). Les remblais, généralement présents en zone urbaines sont également pris en compte dans notre définition. Dans la suite le terme de sol, intègre donc une partie du sous-sol.

Ce point de vue est proche de la définition proposée dans la norme ISO 11074 décrivant le sol comme la « *couche supérieure de la croûte terrestre transformée par des processus climatiques, physico-chimiques et biologiques et composée de particules minérales, de matière organique, d'eau, d'air et d'organismes vivants, organisée en horizons de sols génériques. Dans une acception plus large relevant du génie civil, le terme « sol » inclut l'horizon superficiel et le sous-sol; les dépôts tels que les argiles, limons, sables, graviers, gravillons, pierres, ainsi que la matière organique et les dépôts tels que la tourbe; les matériaux d'origine anthropique tels que les déchets; les gaz et l'humidité du sol; et les organismes vivants* ».

I.A.1.2. Contamination / Pollution

Si les termes de contamination et de pollution sont couramment considérés comme synonymes, ces deux termes présentent des concepts différents qu'il convient de rappeler. Ces concepts varient selon les pays.

En France, la présence d'une substance potentiellement dangereuse sur un site n'est pas nécessairement synonyme de pollution. Une distinction est à faire entre a) la contamination qui est l'introduction « *d'une substance potentiellement dangereuse, **quelle que soit sa teneur*** » (Jeannot, 2001), dans les milieux naturels de transfert et d'exposition que sont le sol, l'air et l'eau, et b) la pollution, qui est présente dès lors que « *la teneur est **potentiellement dangereuse** ou qu'elle atteint des valeurs limites fixées par les normes* » (Jeannot, 2001).

Un site pollué est alors « *un site présentant un risque pérenne, réel ou potentiel, pour la santé humaine ou l'environnement du fait d'une pollution résultant d'une activité actuelle ou ancienne sur ce site* » (Commissariat Général au Développement Durable, 2012).

On parlera ici de contamination diffuse pour une contamination « *qui affecte tous les sols, plus ou moins faiblement, mais de manière **généralisée**. Ce type de contamination **ne peut être totalement évité** : il est la conséquence d'apports involontaires (émissions atmosphériques industrielles par exemple) ou d'impuretés indésirables dans des apports volontaires (métaux dans les engrais minéraux par exemple)* » (SPAQuE, 2010). On en distingue deux sortes :

- Contaminations atmosphériques : les retombées généralisées affectent l'ensemble d'un territoire, les sources sont multiples et difficilement identifiables. On y trouve les rejets industriels transportés sur de longues distances, mais aussi les rejets automobiles ainsi que les émissions liées au chauffage collectif ou individuel.
- Contaminations culturelles, principalement dues aux épandages sur les parcelles agricoles.

Les contaminations locales ou ponctuelles sont définies comme une « *contamination spatialement **localisée**, sur un **territoire restreint** : sources ponctuelles (ex : les activités industrielles pouvant avoir des émissions dans les divers milieux environnementaux)* » (RECORD, 2010).

Ainsi la différence majeure entre une contamination locale ou diffuse s'effectue d'après des critères spatiaux de localisation et d'étendue.

Bien qu'une distinction puisse être effectuée entre contamination et pollution, celle-ci n'est pas toujours faite dans le langage courant ainsi que dans certains termes utilisés par les acteurs de la valorisation des terres excavées. Les termes de « pollutions diffuses et ponctuelles » seront utilisés dans la suite du manuscrit bien que celles-ci se rapportent généralement à des contaminations.

I.A.1.3. Fond pédo-géochimique anthropisé

La notion de fond pédo-géochimique anthropisé, intègre de nombreuses notions telles que la source de la contamination (ponctuelle/diffuse) ou le type de substance (naturelle / anthropique). Une partie de ces notions est résumée à la Figure 4.

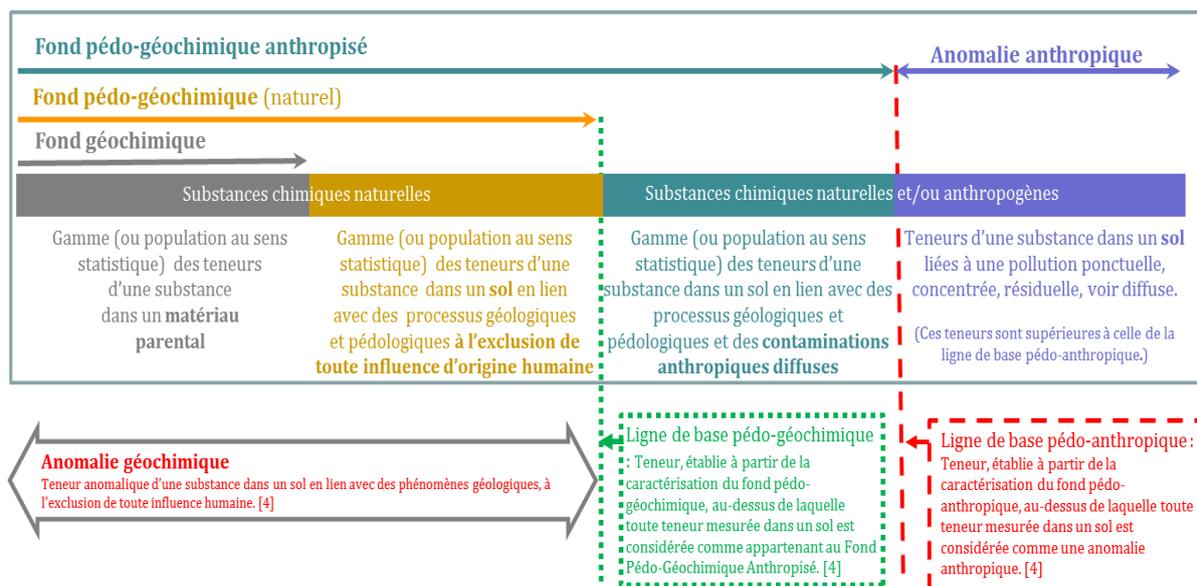


Figure 4 : Schéma conceptuel du fond pédogéochimique anthropisé. Extrait du guide Ademe 2018

Le fond pédogéochimique (naturel) (FPG), est alors défini comme « *la concentration chimique dans un sol résultant des évolutions naturelles, géologiques et pédologiques en dehors de tout apport d'origine humaine* » (Baize, 2008). En France, compte-tenu de la grande diversité géologique, aucun seuil général et unique de pollution ne pourrait être défini. De plus, « *le fond pédogéochimique n'a de sens pratique que rapporté à un territoire. Plus celui-ci est grand, plus la variabilité des roches et sols est importante, et plus les fourchettes de concentrations en certains éléments s'élargissent, d'où moins elles peuvent être utiles dans l'interprétation* » (Baize, 2008). Cette remarque montre la nécessité d'un découpage réfléchi du territoire afin de tenir compte des hétérogénéités spatiales de la géochimie.

La composante anthropique est liée aux impacts des activités humaines. En ville, celles-ci ont pu s'exercer sur des périodes de temps longues. Le fond pédogéochimique urbain ou anthropisé (FPGA) est défini, selon le décret sol belge de 2008, mis à jour en 2018, comme une « *concentration ambiante d'un polluant dans le sol [ou l'eau souterraine] ; les concentrations ambiantes peuvent indiquer des variations géologiques naturelles ou l'influence d'une activité agricole, industrielle ou urbaine généralisée* ». En France, ils sont définis comme « *des niveaux de fond normaux dans le sens où ils font référence à une pollution mineure uniquement associée aux activités quotidiennes de la vie moderne urbaine et rurale* » (Daniau, 2009). Celui-ci prend en compte à la fois le fond pédogéochimique naturel (géogénique) et l'impact diffus de l'activité humaine (anthropogénique) dans le but de les différencier des contaminations ponctuelles d'extension limitée. Cette notion nécessite toutefois une interprétation de données d'analyses géochimiques de façon à différencier les contaminations diffuses des contaminations ponctuelles (Figure 5).

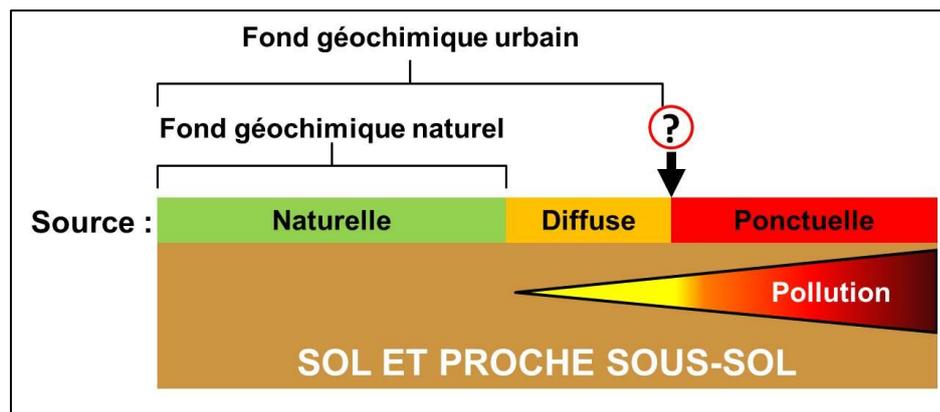


Figure 5: Schéma conceptuel du Fond Pédogéochimique naturel et urbain (modifié d'après Le Guern, communication personnelle).

Le fond pédogéochimique (anthropisé) est donc une gamme de valeurs souvent exprimée par sa **borne supérieure** au-delà de laquelle les concentrations ne peuvent plus être considérées comme reflétant l'environnement général (Reimann, 2005). A noter qu'aucune des définitions précédentes du fond pédogéochimique ne prend en compte de paramètres sanitaires. Elles ne se fondent que sur les concentrations en éléments analysés dans les sols de façon à refléter une teneur de référence dans une zone donnée et non sur les risques potentiels sur les populations.

Enfin, le préfixe « pédo » faisant référence au sol nous l'associons ici au sol défini précédemment (ensemble de la couche potentiellement excavées). Dans la suite, le terme de « fond pédogéochimique anthropisé » (FPGA) est utilisé dans le sens défini par le groupe de travail national sur le fond pédogéochimique (Ademe, 2018) : « *Gamme (ou population au sens statistique) des teneurs d'une substance dans un sol en lien avec des processus géologiques et pédologiques et des contaminations anthropiques diffuses.* »

Les valeurs seuils se rapportent au concept de « ligne de base pédo-anthropique » du guide, elle-même définie comme une « *teneur, établie à partir de la caractérisation du fond pédo-anthropique, au-dessus de laquelle toute teneur mesurée dans un sol est considérée comme une anomalie anthropique* » (Ademe 2018).

L'application de ces seuils pose néanmoins la question de la représentativité des données utilisées, à plus forte raison lorsque celles-ci sont collectées auprès de différents bureaux d'études comme c'est le cas pour cette thèse. L'étude de la stratégie d'échantillonnage mise en place est donc importante.

A noter, enfin, qu'aucune de ces définitions ne tient, compte du support des données (volume de l'échantillon analysé). Implicitement, le support est donc supposé ponctuel alors que ses variations de volume ont une influence sur la variance et donc sur les résultats des traitements statistiques ou géostatistiques appliqués aux données ainsi que sur les critères statistiques (hormis la moyenne).

I.A.2. Cartographie géochimique

La cartographie de la géochimie, qu'elle concerne les milieux urbains ou non, a largement évolué tant dans les objectifs que dans les outils utilisés au cours du temps (Garrett et al., 2008 ; Teng et al., 2014). Elle est ainsi partie de considérations économiques avec la recherche de minerai et a évolué ensuite vers des objectifs plus environnementaux via l'étude de la qualité environnementale et sanitaire des sols.

I.A.2.1. Contexte historique

La cartographie géochimique apparaît dans les années 1930 et 1950 d'abord en Union Soviétique puis à travers le monde (Fersman, 1935 ; Lovering, 1950). Ces premiers questionnements présentent des enjeux principalement économiques à travers la recherche de ressources exploitables par le secteur minier. Ces travaux se généralisant, les premiers atlas nationaux voient le jour dans les années 1960 et 1970 (Webb, 1978) (Figure 6). En France, l'Inventaire Minier National est réalisé à partir de 1970, par le BRGM. Toutefois, du fait de l'impossibilité d'exploiter des ressources minérales dans les régions urbaines, les villes sont alors peu étudiées. On remarque leur absence dans les représentations cartographiques de la géochimie comme dans l'atlas de Webb, 1978 (Londres entouré en rouge à la Figure 6).

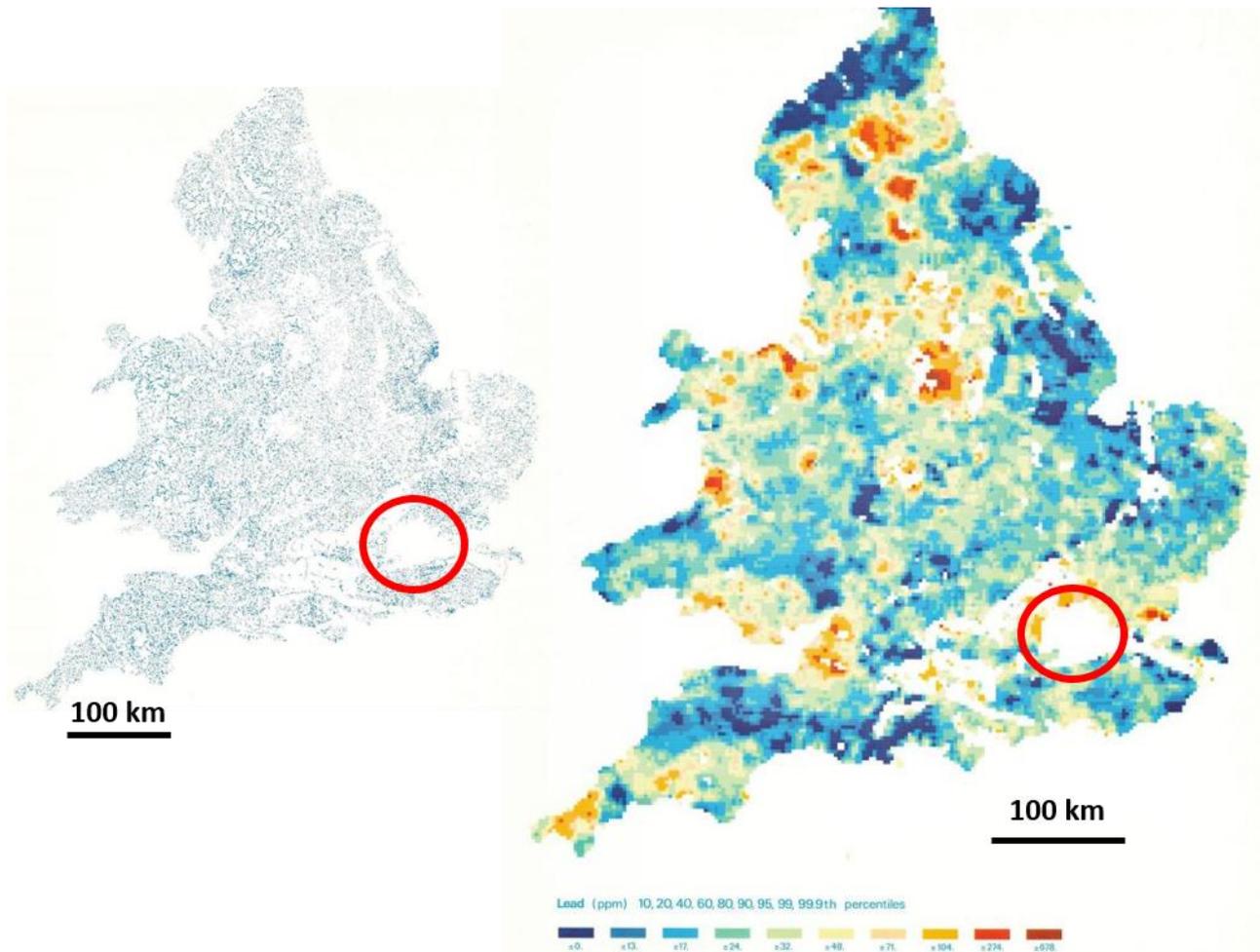


Figure 6: Extrait de l'atlas de Webb, 1978, sur la géochimie du Royaume-Uni.

La ville de Londres est entourée en rouge

A gauche : Carte des 50 000 échantillons de sols et de sédiments collectés

A droite : Cartographie des teneurs en plomb

Au début des années 1970, les problématiques écologiques et sanitaires sont prises en compte dans le cadre de la cartographie géochimique (Thorton et Webb, 1979). D'abord délaissées, les zones urbaines vont progressivement prendre de l'importance dans les études du sol et proche sous-sol du fait des enjeux économiques et sanitaires importants (Johnson et Ander, 2008 ; Zheng, 2008) (Figure 7). Des cartographies de la géochimie urbaine sont entreprises à une échelle nationale en Grande-Bretagne dès 1992 (Fordyce, 2005).

Cette prise de conscience de la nécessité de la connaissance des sols et sous-sols urbains est renforcée par l'urbanisation et l'exode rural observés depuis des décennies dans la plupart des pays (Données de la Banque Mondiale). Ainsi en France, près de 80% de la population peut être qualifiée d'urbaine (Clanché et Rascol, 2011).

I.A. Etat de l'art

Les enjeux sanitaires sont importants (Le Guern et Sauvaget, 2017 ; Le Guern, 2017), et représentent le principal sujet d'étude de la géochimie urbaine. Dans ce cadre la méthode nationale des sites et sols pollués intègre la notion de risque sanitaire et environnementale à l'échelle du site et fait référence au fond pédo-géochimique local obtenu à l'aide de quelques échantillons de référence.

Pour la gestion des terres excavées, une échelle plus vaste, du quartier ou de la ville, est nécessaire.

Il est intéressant de se pencher alors sur les grandes bases de données.

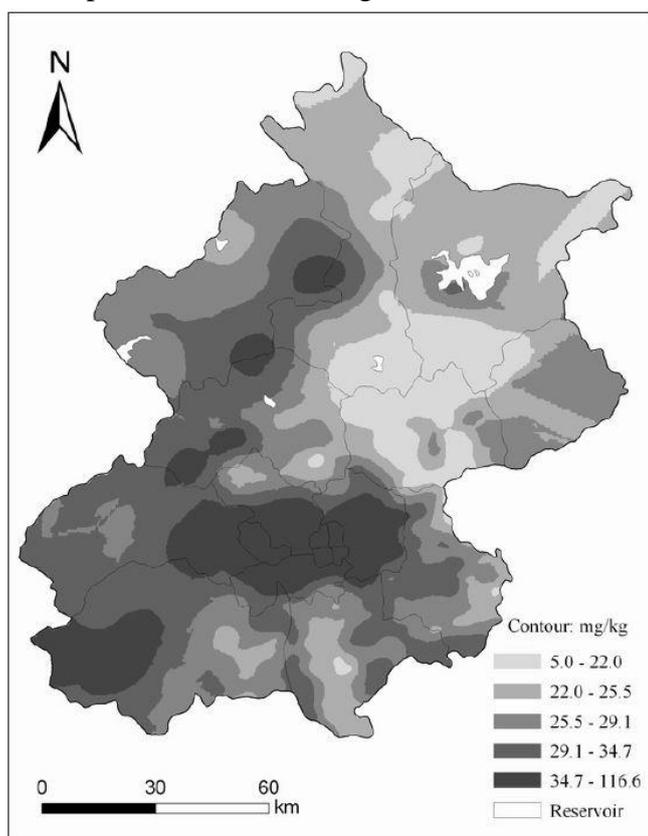


Figure 7: Estimation des concentrations en plomb des sols de la ville de Pékin (Chine). Interpolation par krigeage ordinaire sur des échantillons de sols de surface (0-20cm). (Zheng, 2008)

I.A.2.2. Bases de données

La cartographie géochimique est fondée sur des bases de données géoréférencées. Afin de mieux les stocker et de faciliter leur consultation, les données géologiques et géochimiques ont fait l'objet de bancarisations à des échelles nationales.

A l'échelle nationale, on peut ainsi citer la G-Base du British Geological Survey (BGS) fournissant une cartographie de la géochimie des sols, pour l'ensemble de la Grande Bretagne. Outre les données cartographiques, une partie des données d'analyses sont également téléchargeables librement (lien internet : <http://www.bgs.ac.uk/products/geochemistry/gBaseUK.html>).

Les Etats-Unis fournissent également une cartographie détaillée de la géochimie à l'échelle des états. Des statistiques descriptives sont fournies, toutefois les données de base (échantillons, analyses, ...) ne sont pas accessibles depuis le site (lien internet : <https://mrdata.usgs.gov/soilgeochemistry/#/periodictable>).

En France, le Gis Sol présente un ensemble de cartes thématiques sur la géochimie des sols de surface dont : la BDETM (Base de Données de Eléments Traces Métalliques), l'IGCS (Inventaire, Conservation et Gestion des Sols) et la BDAT (Base de Données d'Analyses des Terres). Les cartes sont accessibles librement ainsi que des données sous forme de tableaux téléchargeables et de web services permettant l'intégration et l'utilisation dans un SIG (Système d'Informations Géographiques) (lien internet : <https://www.gissol.fr/donnees>). Dans cartes d'évolution sont également disponibles.

En particulier, le Réseau de Mesure de la Qualité des Sols (RMQS) du GisSol (Figure 8), cartographie depuis l'an 2000 l'évolution de la qualité des sols français. A partir de l'échantillonnage de 2240 sites répartis sur l'ensemble du territoire (métropole et outre-mer) avec une maille de 16 km de côté et recommencée tous les 15 ans. Les échantillons composites sont prélevés dans les 30 premiers centimètres de sol, dans des zones éloignées de pollution directes importantes. L'échantillonnage du RMQS ne concerne pas les zones urbaines.

A plus large échelle il est possible de citer les travaux du GEMAS (Geochemical Mapping of Agricultural and Grazing land Soil). Ce projet mené par l'EuroGeoSurveys a pour but de cartographier les teneurs de fond en 60 éléments chimiques à travers 33 pays européens. Pour cela plus de 4 000 échantillons de sols de surface (0-20 cm) ont été collectés dans des terrains agricoles ou naturels éloignés de contaminations directes. Le projet cherche à définir des valeurs de référence pour les sols naturels ou faiblement anthropisés à l'échelle d'un continent (lien internet : <http://www.eurogeosurveys.org/projects/gemas/>).

Antérieur au projet GEMAS, on peut aussi présenter le programme Forum of European Geological Surveys (FOREGS) qui vise lui aussi à fournir des valeurs de fond pédo-géochimique à l'échelle de l'Europe pour plus de 60 substances. Contrairement au projet GEMAS, les substrats prélevés sont plus diversifiés avec des prélèvements d'alluvions en plus des sols de surface. Là encore, ces échantillonnages ne ciblent que des sols peu ou pas anthropisés (lien internet : <http://weppi.gtk.fi/publ/foregsatlas/index.php>).

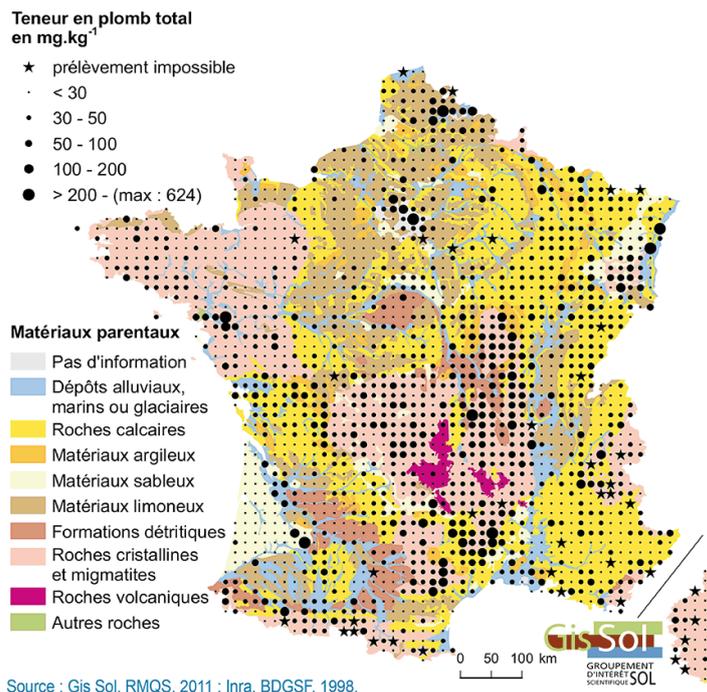


Figure 8: Cartographie des teneurs en plomb dans les sols. Données RMQS (GisSol)

Les bases de données précédentes sont nationales ou internationales et ne concernent (sauf pour une partie de la G-Base) pas les zones urbanisées. C'est pourquoi, en France, une nouvelle base de données est mise en place actuellement par l'Ademe et le BRGM. Cette dernière, appelée BDSolU, banarise des données géochimiques sur les sols urbains provenant de diverses sources. Ces données ont pour but de servir de référence pour le fond pédogéochimique anthropisé (données brutes ou élaboration de référentiels). La base accueillera des données issues de projets auxquels le BRGM participe mais aussi des données liées à des projets d'aménagements fournies par des aménageurs ou des collectivités (pour l'essentiel, des données issues des diagnostics de pollutions...). En échange d'un apport de données, et selon les termes des conventions passées avec le BRGM, les fournisseurs peuvent bénéficier du contenu de BDSolU ainsi que d'une expertise sur le fond pédogéochimique urbain sur leur territoire (en lien avec le guide national, Ademe 2018) ([lien internet : http://www.bdsolu.fr/](http://www.bdsolu.fr/)).

Ainsi, bien que des initiatives soient en cours pour banariser des données de géochimie urbaine (BDSolU), la majorité des données collectées sur les sols en ville le sont par des bureaux d'études pour la caractérisation de la structure ou de la qualité locale des sols. Ces données représentent des volumes très importants d'informations mais leur collecte, leur banarisation et leur exploitation restent complexes à mettre en place à grande échelle et représentent un des enjeux de cette thèse.

Pour représenter ces résultats d'analyses, différentes méthodes de cartographie sont disponibles.

I.A.2.3. Méthodes de cartographie et intérêt de la géostatistique

Deux démarches sont utilisées pour la cartographie des teneurs en géochimie. Dans la première, seules sont représentées les teneurs aux points où elles ont été mesurées (Argyraki et Kelepertzis, 2014 ; Smith et al., 2011). Les cartes obtenues sont donc un ensemble de points dont la taille ou la couleur est liée à la teneur mesurée sur l'échantillon. (Figure 9)

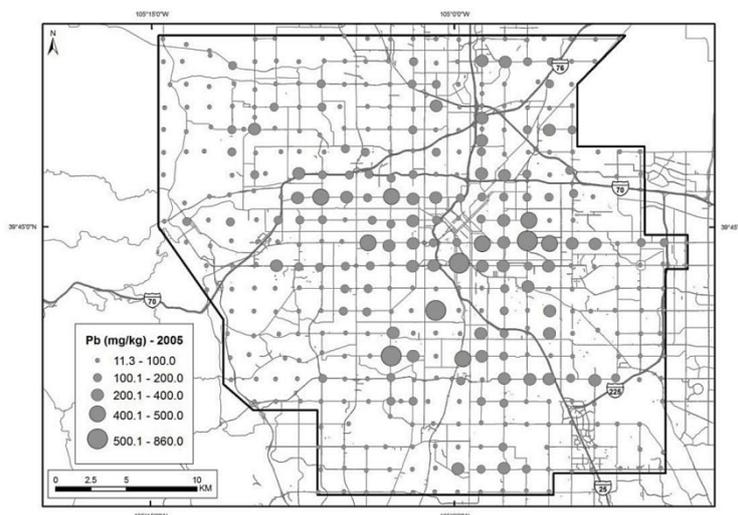


Figure 9 : Carte géochimique de la région de Denver présentant la concentration en Pb d'échantillons de sols prélevés dans les 15 premiers centimètres (Smith et al., 2011).

Cette approche présente l'avantage de ne présenter que des données réellement mesurées. Néanmoins, la cartographie des teneurs ne couvre qu'une partie réduite du territoire (uniquement la position des échantillons) et présente donc un intérêt limité pour des outils d'aide à la décision pour lesquels une cartographie de l'ensemble du territoire est généralement nécessaire.

Afin de fournir des valeurs pour l'ensemble du territoire, la seconde approche consiste à estimer les teneurs entre les échantillons. Ce ne sont donc plus des données réelles mesurées qui sont représentées mais les résultats d'un modèle reflétant la réalité.

Différentes méthodes d'estimation sont disponibles comme les polygones d'influence (polygones de Voronoï ou de Thiessen) ou l'estimation par l'inverse du carré de la distance. Ces deux méthodes, couramment utilisées pour les estimations en cartographie (Acosta et al., 2011 ; Appleton et Adlam, 2012 ; Argyraki et Kelepertzis, 2014 ; Guillen et al., 2011), représentent une vision très simplifiée de la variation des variables dans l'espace. En effet la variabilité spatiale des données est souvent complexe et des méthodes simples d'estimation ne sont pas toujours pertinentes.

La géostatistique a été développée pour améliorer les estimations. D'abord utilisée en contexte minier pour l'estimation des gisements, elle est maintenant appliquée à de nombreux domaines dont la géochimie (Cheng et al., 2009 ; de Fouquet, 2006 ; 2011 ; Donati et de Fouquet,

2018 ; Ha et al., 2014 ; Hani et Pazira, 2011 ; Li et Feng, 2012 ; Lv et al., 2015 ; Poggio et Vrščaj, 2009, Saby et al., 2006 ; Yuan et al., 2013).

En géostatistique, les teneurs dans les sols sont interprétées comme des variables régionalisées : ces variables varient dans l'espace de façon structurée. Ainsi, des teneurs mesurées sur des échantillons proches sont plus susceptibles de se ressembler que deux mesures sur des échantillons éloignés. La relation entre l'écart quadratique des données et leur distance dans l'espace reflète la variabilité spatiale. Cette relation est modélisée par le variogramme (voir annexe 1).

Ce variogramme représente le modèle décrivant la variabilité des teneurs dans l'espace ; il est utilisé pour l'estimation à l'échelle d'un territoire par krigeage (Figure 10).

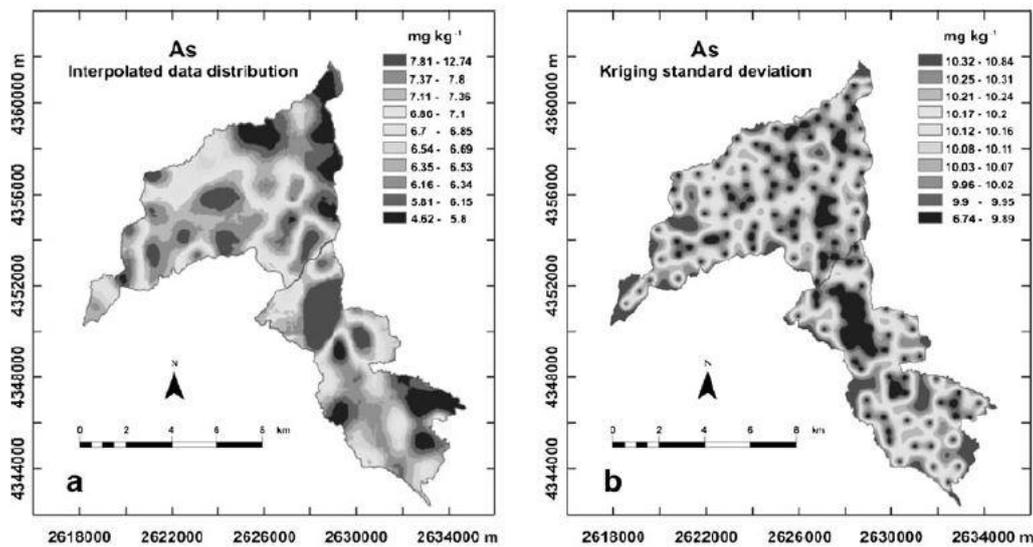


Figure 10 : Estimation des teneurs en arsenic (à gauche) et de l'écart-type de krigeage (à droite) dans des échantillons de sols d'une région de Calabre (Italie) (Guagliardi et al., 2012)

Le krigeage est le meilleur estimateur linéaire, au sens de la minimisation de la variance de l'erreur d'estimation. Il fait intervenir le variogramme modélisé. Toutefois, le krigeage reste une estimation et donc une interprétation de la réalité. Il présente néanmoins l'avantage de fournir une cartographie de l'écart-type d'estimation qui mesure la précision de l'estimation réalisée.

La géostatistique permet également de tenir compte de plusieurs covariables, corrélées, à la variable à estimer. Dans une première étape, des variables des analyses en composantes principales (ACP) s'avèrent intéressantes (Ha et al., 2014 ; Li et Feng, 2012 ; Schaefer et al., 2010, Yuan et al., 2013). Ce traitement transforme des variables corrélées en des facteurs (ou composantes principales) point à point décorrélés fournissant des informations complémentaires. Cela permet de réduire le nombre de variables et de les rendre moins redondantes. Dans une deuxième étape, la modélisation du variogramme multi-variable permet de tenir compte des covariables dans l'estimation par cokrigeage.

Parmi les co-variables, il est possible d'intégrer des variables quantitatives ou qualitatives (Jeannée et de Fouquet, 2003), permettant, par exemple, la prise en compte de typologies de matériau ou d'occupation des sols.

L'intérêt de la géostatistique réside notamment dans la compréhension des variations spatiales des teneurs, notamment par l'interprétation du variogramme. En effet, si le but initial de ces travaux était de mettre en évidence des anomalies géochimiques en prospection minière (Matheron, 1982 ; Wackernagel, 1993), le parallèle avec les contaminations locales que l'on cherche à éliminer pour calculer le fond pédo-géochimique est évident. Ainsi, l'analyse krigéante apparaît intéressante pour l'estimation du fond pédo-géochimique anthropisé.

L'analyse krigéante (Bourennane et al, 2003 ; Goovaerts, 1992 ; Saby et al., 2009; Wackernagel, 1993) permet de décomposer une variable en composantes spatiales, associées aux différentes échelles de variabilité mises en évidence par le variogramme expérimental.

I.A.3. Echantillonnage et analyse

I.A.3.1. Stratégie d'échantillonnage

Les données bancarisées dans les bases de données (G-Base, RMQS, BDSolU, ...) sont issues d'études avec différentes stratégies d'échantillonnage.

Deux classes de plans d'échantillonnage peuvent être distinguées suivant leur objectif (Johnson et Ander, 2008). Les premiers sont dits « réguliers » et concernent des sondages avec une répartition spatiale homogène afin d'établir des cartes ou des atlas. Les autres, dits « orientés » montrent une répartition plus hétérogène et cherchent à mettre en évidence un objet précis du sous-sol, le plus souvent une anomalie.

Les échantillonnages réguliers suivent généralement une grille plus ou moins resserrée (Figure 11). Cette méthode est communément utilisée à travers le monde pour réaliser des cartographies géochimiques de zones relativement grandes (pays, région ou ville) (**Chine** : Gong (2010), Li (2015), Lv (2015); **Russie** : Gavrilenko (2013) ; **Royaume-Unis** : Ander (2013), Fordyce (2005), Lark (2013) ; **Italie** : Poggio (2009) ; **Espagne** : Acosta (2011), Guillén (2011) ; **Slovénie** : Bavec (2014), Šajn (2011) ; **France** : Saby (2006 et 2009) ; **Grèce** : Argyraki (2014)). Cette méthode présente l'avantage d'une bonne répartition des données pour la cartographie et d'un schéma aisément reproductible. Une tolérance est généralement admise sur l'écart entre l'implantation théorique et l'implantation réelle des échantillons afin d'éviter les sites non compatibles (contamination avérée, difficulté d'accès, ...).

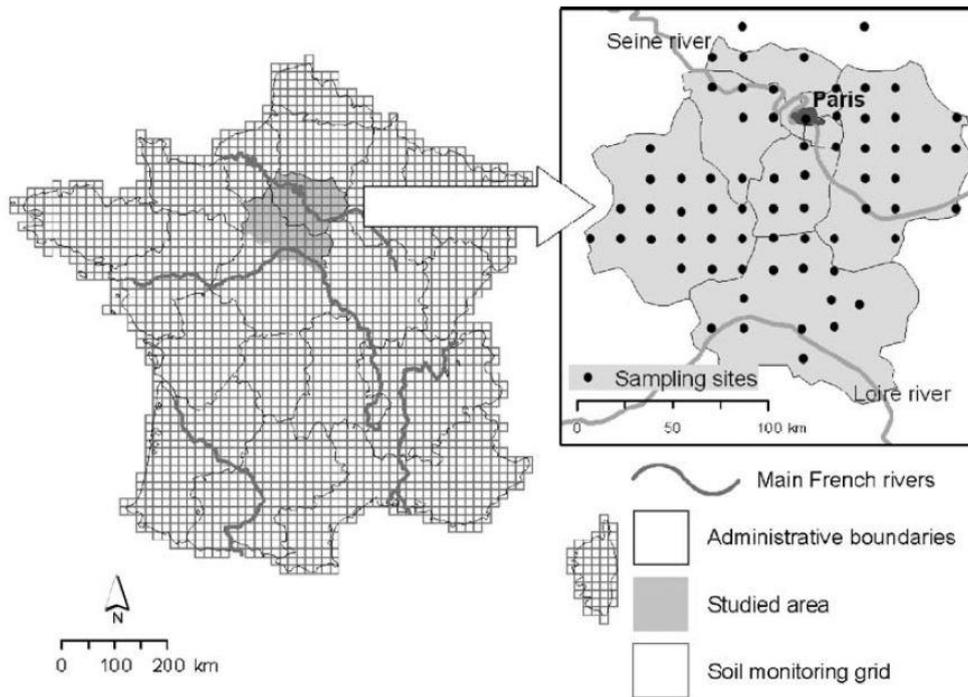


Figure 11 : Plan d'échantillonnage du Réseau de Mesure de la Qualité des Sols (RMQS) en France et mise en évidence de l'échantillonnage régulier dans la région parisienne (Saby, 2006).

A propos de la maille de reconnaissance, des recommandations ont été émises dans le cadre du projet URGE 2 de l'Euro Geo Surveys en 2015 (Demetriades et Birke, 2015) pour la comparaison des résultats d'une ville à l'autre en recourant à un échantillonnage régulier dont le maillage varie selon la densité du tissu urbain. Afin de tenir compte de la forte hétérogénéité du milieu urbain, un maillage resserré est nécessaire (Figure 12). Les centres-villes et les parties les plus urbanisées sont échantillonnées selon une maille de 500 m de côté et les zones périurbaines tous les 1000 m.

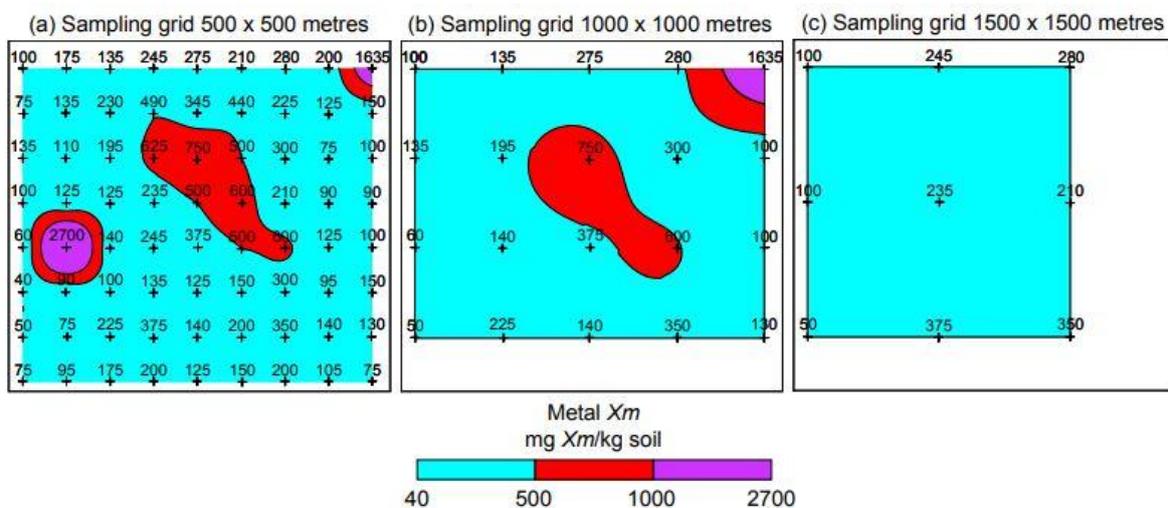


Figure 12 : Exemple de l'influence de la densité d'échantillonnage. Une même zone hypothétique est échantillonnée suivant 3 densités différentes : a) 500m, b) 1000m et c) 1500m. Les teneurs en un métal Xm et la carte associée aux résultats est représentée en fon. (Demetriades et Birke, 2015).

Toutefois une telle stratégie d'échantillonnage nécessite, pour caractériser le milieu urbain, des moyens conséquents, qui ne sont pas toujours possibles. D'autres plans d'échantillonnage sont parfois utilisés afin de limiter ces coûts.

La stratégie d'échantillonnage développée dans le projet METOTRASS (Le Guern et al, 2013), s'appuie à la fois sur un découpage du territoire en entités géochimiquement cohérentes et sur une densité d'échantillonnage variant selon le degré d'hétérogénéité de ces entités définies. Cette approche permet d'optimiser le nombre de prélèvements en adaptant le maillage à l'hétérogénéité supposée des entités comme le montre la Figure 13. En milieu urbain, la difficulté provient de la forte hétérogénéité et de la méconnaissance globale des sols urbains.

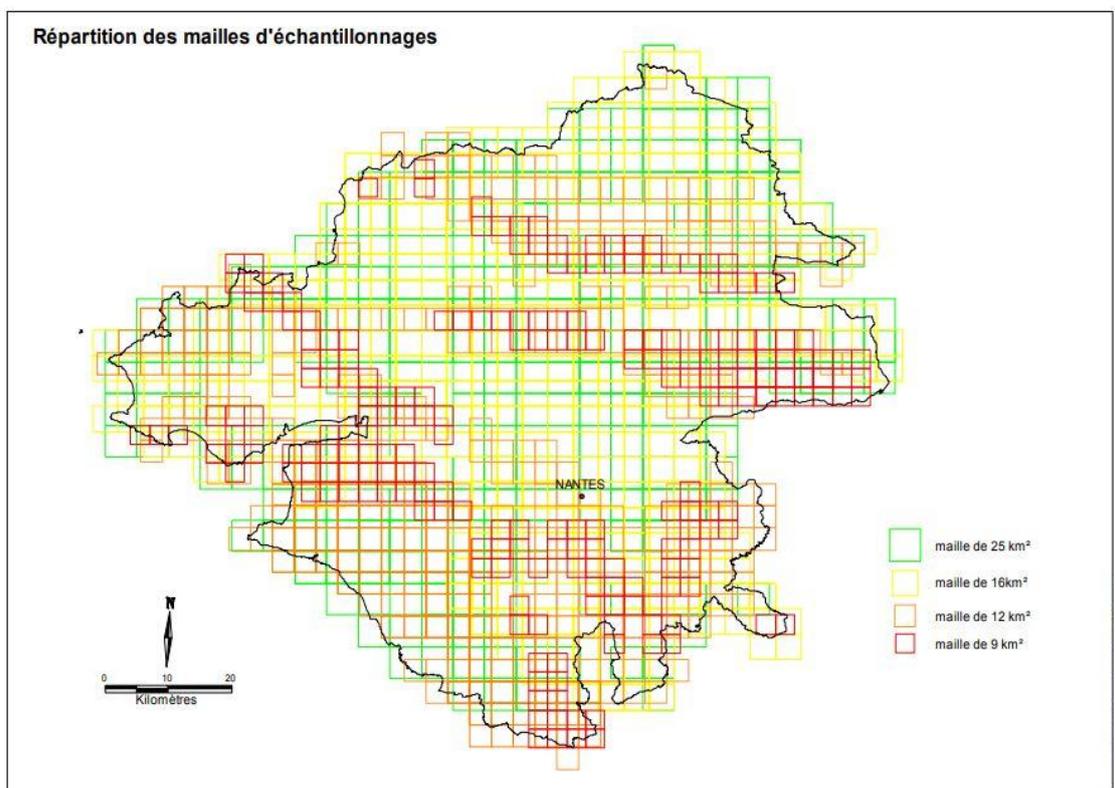


Figure 13: Adaptation du maillage selon l'hétérogénéité géochimique, mise en place dans le cadre du projet METOTRASS afin d'optimiser le nombre d'échantillons de sols à prélever pour déterminer le fond géochimique (d'après Le Guern et al, 2013).

Les plans d'échantillonnage orientés ciblent un phénomène ou un objet, par exemple pour définir des valeurs de référence pour un type de sol ou délimiter une zone contaminée. Ce type de stratégie d'échantillonnage est effectuée pour les diagnostics de pollution réalisés par les bureaux d'étude afin d'identifier et de quantifier les sources de pollution sur un site. Il est également utilisé à travers le monde entier pour les études de la géochimie. Parmi les études utilisant ce type de stratégie d'échantillonnage, on peut citer des études de sites pollués (Demougeot-Renard, 2008 ; Ramos-Miras, 2011), de jardins urbains (Douay, 2005 ; Fetzer, 1998),

de contamination des sols urbains (Malik, 2010 ; Manta, 2002) et enfin de définition de valeurs seuils ou de référentiel de qualité des sols nationaux (Reimann, 2005) ou urbains (Puskás, 2009 ; Jarva, 2014). L'échantillonnage est alors orienté dans le sens où il cible un type de sol ou d'environnement précis (parcs ou jardins le plus souvent). Les échantillons sont implantés de façon relativement homogène sur la zone d'étude.

Pour calculer des valeurs de référence pour la qualité géochimique des sols, un échantillonnage régulier ou aléatoire stratifié reste la meilleure option puisque cela garantit une répartition homogène sur l'aire d'étude et évite les biais (sur ou sous-échantillonnage d'un type de sol). Néanmoins cela nécessite des campagnes d'échantillonnages spécifiques qui peuvent s'avérer très coûteuses, notamment dans le cas d'échantillonnages en profondeur. Ne disposant pas de telles campagnes en France, pour le fond pédo-géochimique anthropisé, nous utilisons les données issues d'échantillonnages orientés et devons donc tenir compte des possibles biais.

I.A.3.2. Mode d'échantillonnage

Une fois le plan d'échantillonnage établi, le mode d'échantillonnage doit également être pris en compte.

Dans la majorité des études du fond pédo-géochimique, l'échantillonnage se concentre sur les sols de surface (20 à 30 premiers centimètres de profondeur), les plus concernés par les enjeux sanitaires du fait des risques d'exposition par ingestion ou inhalation. Ils sont relativement simples à prélever dans le cas de campagnes de grande ampleur (nationale ou régionale). De façon à disposer d'un volume de matériau suffisant pour les analyses et afin d'être représentatifs d'une zone ou d'une parcelle, ces échantillons sont généralement composites (Figure 14), c'est à dire des mélanges homogénéisés de plusieurs prélèvements (Saby, 2006 ; Albanese, 2007 ; Karim, 2015 ; Lark, 2013).

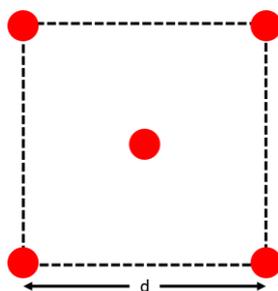


Figure 14: Schéma du prélèvement d'un échantillon composite. 5 sous-échantillons (points rouges) sont prélevés aux sommets et au centre d'un carré de côté « d ».

Dans son guide, URGE 2 préconise des échantillons prélevés sur un carré homogénéisé de sol de 25 cm de côté et de 10 cm de profondeur, afin de tenir compte du maximum de contamination anthropique sans qu'elle ne soit diluée par du matériel

géogénique. Toutefois les échantillons ne doivent être prélevés que dans des zones sans contamination directe (parc, jardins ...).

Dans le cas du calcul du fond pédo-géochimique anthropisé pour la valorisation des terres excavées, un tel échantillonnage de surface n'apparaît cependant pas représentatif d'une tranche de sols de plusieurs mètres d'épaisseur sur l'ensemble de la ville (et non simplement les parcs et jardins).

Dans le cas des diagnostics de pollution, et plus généralement des études géochimiques sur des épaisseurs de sol importantes (plusieurs mètres), les échantillons sont généralement prélevés sur des sondages carottés (voir annexe 2). L'échantillon est alors caractérisé par une profondeur de début et de fin le long du sondage. Ces profondeurs peuvent dépendre des horizons (ou lithologies) ou être fixées à l'avance. Pour être représentatif de la couche prélevée, l'échantillon est homogénéisé.

Un parallèle peut être fait avec les mesures de qualité de l'air où une typologie des stations de mesures a été établie. L'INERIS a défini trois types d'influences des sources d'émissions de polluants : fond, industrielle ou trafic. Les mesures sont classées de fond *« lorsque les niveaux de concentration ne sont pas influencés de manière significative par une source particulière (ex: émetteur industriel, voirie...) mais plutôt par la contribution intégrée de multiples sources. Ce type de mesure est destiné au suivi de l'exposition moyenne de la population et des écosystèmes aux phénomènes de pollution atmosphérique qui affectent la zone de surveillance sur de larges distances (plusieurs kilomètres voire plusieurs dizaines ou centaines de kilomètres) »* (LCSQA, 2015).

Ce n'est donc plus la mesure en elle-même qui définit le fond mais la position de la station et de plus, pour une même station, certains paramètres analysés pourront correspondre au fond alors que d'autres subiront l'influence industrielle ou du trafic. Cette approche par typologie sera utilisée à partir de la segmentation du territoire dans la partie III de ce manuscrit.

I.B. Présentation des données utilisées

I.B.1. Les secteurs tests

Contexte géologique

La ville de Nantes se situe sur la bordure orientale du « Sillon de Bretagne », un massif de roches très dures s'étendant de Montaigu à la Pointe du Raz (Le Guern et al., 2001). D'après la carte géologique au 1/50 000 (Nantes n°481) (Figure 15), la ville repose principalement sur des micaschistes et des altérites indifférenciées, la partie sud-ouest de la ville étant occupée par des massifs de granite et de mylonites. La partie de la ville bordant la Loire est occupée par des remblais anthropiques reposant sur des alluvions dont l'épaisseur varie de quelques mètres à plusieurs dizaines de mètres (jusqu'à 30 m).

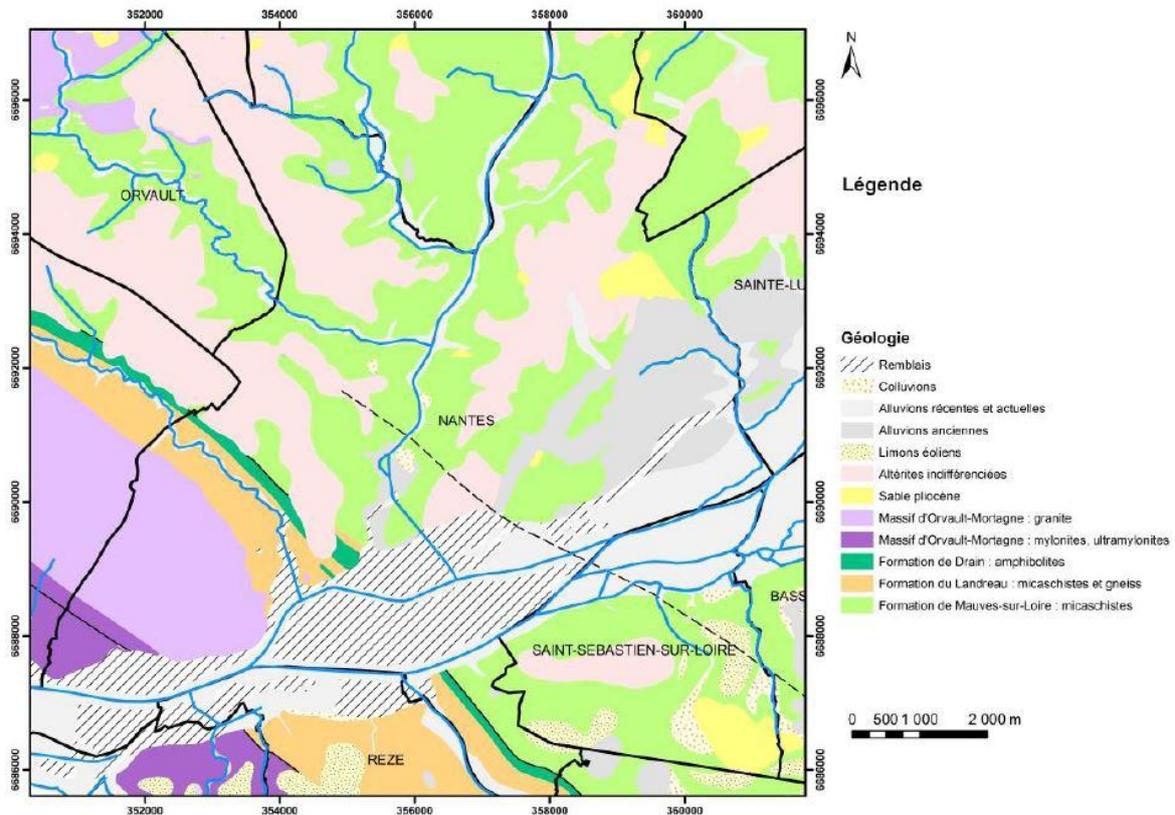


Figure 15: Extrait de la carte géologique harmonisée de la Loire-Atlantique à 1/50 000 (Béchenec, 2007)

Activités industrielles et de service passées

Comme beaucoup de grandes villes françaises, Nantes a été fortement transformée par le développement de l'industrie. On retrouve d'anciens sites industriels liés à l'activité portuaire, au travail du bois ou du métal. Sur la Figure 16, les sites BASIAS indiquant d'anciennes activités industrielles ou de service potentiellement polluantes (de 1850 à nos jours) apparaissent principalement regroupées dans le centre-ville et le long de la Loire. Nombre de ces sites sont des stations-services des années 1970 présentant des dépôts de liquides inflammables (DLI). Enfin, parmi les activités recensées en bord de Loire, se trouvent des fonderies ou des chantiers navals, mais la fabrication d'engrais et le traitement de déchets.

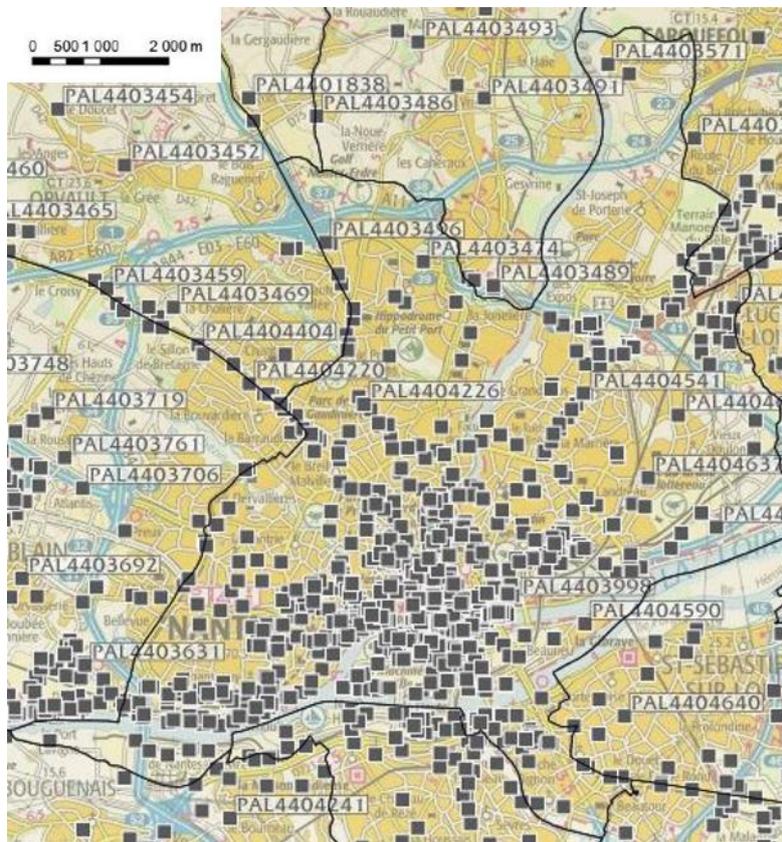


Figure 16 : Extrait des sites industriels et de services potentiellement polluants recensés sur Nantes dans la base de données BASIAS (www.georisques.gouv.fr)(2015).

Cas particulier de l'Île de Nantes

Les données collectées n'ayant pas une répartition homogène sur la ville (cf. Figure 21), il a été choisi de retenir plusieurs quartiers présentant un fort enjeu lié à d'importants travaux de réaménagements actuels ou futurs. Nous nous focaliserons dans la suite sur l'un de ces quartiers, l'Île de Nantes.

Ce quartier du centre-ville est situé sur la Loire. L'île mesure 4,9 km de long pour 1 km de large avec une superficie de 337 hectares. Le quartier comprenait en 2011, 18 000 habitants, soit environ 6 % de la population de la ville.

Son contexte géologique est principalement marqué par une importante couche de remblais de différentes natures et avec une épaisseur comprise principalement entre 2 et 5 m. En dessous, se trouvent des alluvions de la Loire (d'épaisseur pouvant aller jusqu'à 20 m), reposant sur un socle de micaschiste.

L'activité industrielle passée était principalement influencée par le port et les chantiers navals à l'ouest de l'île. Cette activité ayant progressivement décliné jusqu'à la fermeture des derniers chantiers navals en 1987, un réaménagement du quartier est, depuis, mis en œuvre. Ce réaménagement a été confié à la Société d'Aménagement de la Métropole Ouest Atlantique (SAMOA) en 1990. Cette dernière estime le volume de terres excavées générées à environ 100 000 tonnes par an entre 2015 et 2025.

I.B.2. Données de géochimie

Données issues de diagnostics de pollution

Comme évoqué dans l'introduction, il n'existe pas de campagne nationale d'échantillonnage dédié au fond pédo-géochimique anthropisé en France. La base BDSolU étant en début d'alimentation, elle n'était pas suffisamment fournie pour être utilisée. Nous avons ainsi travaillé avec la base de données locale mise en place par le BRGM sur les quartiers nantais dans le cadre d'études avec les aménageurs locaux (Le Guern et al, 2016, Le Guern et al, 2017). La Figure 17 présente à titre d'exemple la répartition des sondages collectés sur l'île de Nantes.

Le BRGM a travaillé en partenariat avec la Société d'Aménagement de la Métropole Ouest Atlantique (SAMOA) et Nantes Métropole Aménagement (NMA) sur la collecte et la bancarisation de données sur un ensemble de quatre quartiers de la ville de Nantes. Ce travail a permis la bancarisation de 3570 sondages et 4630 échantillons (visant principalement des sondages de diagnostics de pollution). Ces sondages sont géoréférencés dans une base de données sous Access 2010 incluant également les descriptions des diverses couches rencontrées, les profondeurs de début et fin des échantillons collectés ainsi que les résultats d'analyses de ces échantillons. Dans le cadre de ma thèse, j'ai contribué à alimenter cette base à hauteur de 861 sondages et 14300 paramètres géochimiques analysés.

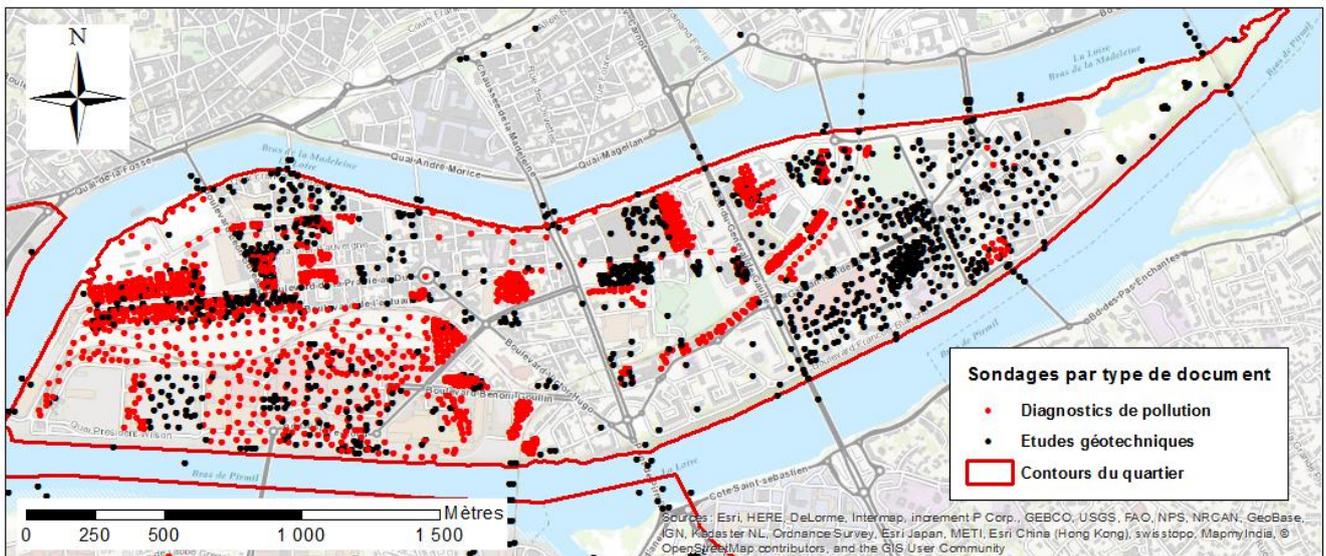


Figure 17 : Carte des sondages collectés en fonction du type de document d'origine (diagnostic de pollution ou étude géotechnique)

Les données de diagnostics de pollution ont été collectées auprès de différents bureaux d'études, avec des résultats d'analyses géochimiques d'un grand nombre d'échantillons, mais ceux-ci présentent des profondeurs et des volumes d'échantillons très variables du fait de protocoles d'échantillonnage différents (Figure 18). Dans ce cadre, deux principaux

I.B. Présentation des données utilisées

protocoles sont à distinguer. Le premier (Figure 19, cas a) consiste à réaliser des échantillons de volume (longueur) constant quelles que soient les couches rencontrées pour le sondage (ex : de 0 à 1 m puis de 1 à 2 m). Le second protocole (Figure 19, cas b) consiste à échantillonner séparément les couches rencontrées. On échantillonne ainsi la première couche, puis la seconde. Selon les protocoles retenus, les longueurs et les volumes échantillonnés sont donc très variables.

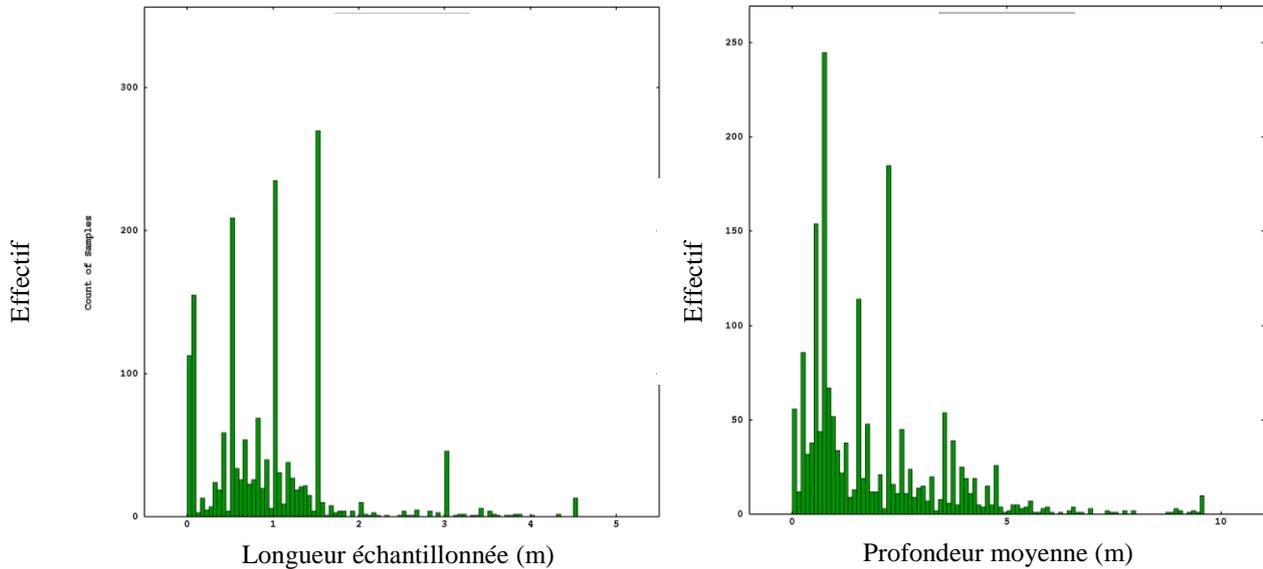


Figure 18 : Longueur et profondeur des échantillons de diagnostic de pollution collectés.

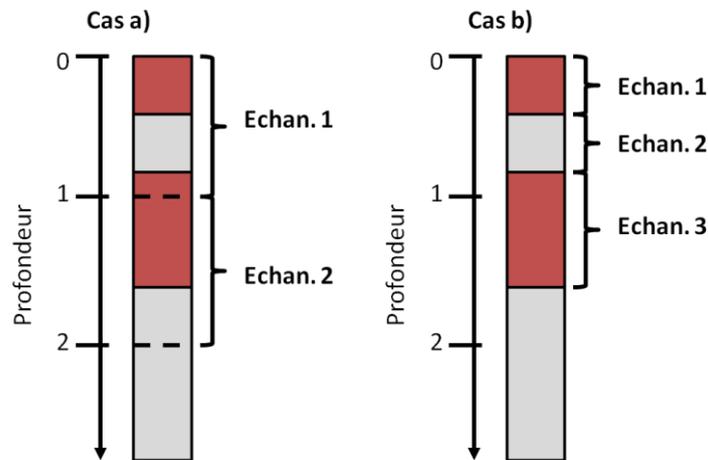


Figure 19 : Deux protocoles d'échantillonnage présents dans les données collectées, sur un sondage. Les zones rouges et grises représentent des couches différentes du sous-sol. Le cas a) échantillonne selon une longueur constante (926 échantillons), alors que le cas b) échantillonne chaque couche (3704 échantillons).

I.B. Présentation des données utilisées

Si la méthode d'échantillonnage illustrée par le cas « a » présente l'avantage d'une épaisseur constante et facilite donc la cartographie des résultats, elle mélange néanmoins différents matériaux, pouvant avoir des propriétés géochimiques très différentes. Elle introduit de ce fait une dégradation d'information. Pour la détermination de valeurs de fond pédogéochimique, il est préférable de différencier les types de matériaux et donc les couches échantillonnées, en privilégiant la méthode « b ». Celle-ci permet alors d'adapter les valeurs selon le type de matériau, avec à terme une précision plus importante des résultats, notamment en séparant une éventuelle anomalie liée à un type de matériau précis (exemple : des mâchefers). C'est d'ailleurs ce que le BRGM a préconisé à la SAMOA pour les futures acquisitions de données (Le Guern et al, 2016). Pour la cartographie géostatistique, il est possible de travailler en puissance-accumulation ou d'adapter le système de krigeage à des données de support variable.

Les protocoles d'analyse peuvent également varier selon les laboratoires (liste des protocoles et des laboratoires en annexe 3), selon les paramètres, le bureau d'étude chargé du diagnostic et la date de l'analyse. Nous avons choisi de retenir les substances les plus analysées pour garantir un effectif suffisant. Nous utilisons les résultats d'analyses pseudo-totaux (eau régale) des métaux et métalloïdes quantifiés par ICP/MS, ainsi que les résultats d'analyses des hydrocarbures C10-C40 et des 16 HAP indicateurs (spectrométrie de masse (GS MS), et extraction par acétone/hexane). Les substances suivantes sont retenues (Tableau 1).

Famille	Paramètre	Effectif
Métaux-Métalloïdes	Antimoine(Sb)	2215
	Arsenic (As)	2444
	Baryum (Ba)	1988
	Cadmium (Cd)	2304
	Chrome (Cr)	2299
	Cuivre (Cu)	2401
	Mercure (Hg)	2450
	Molybdène (Mo)	1991
	Nickel (Ni)	2333
	Plomb (Pb)	2414
	Sélénium (Se)	1986
Zinc (Zn)	2372	
HAP	Somme des HAP (16)	1325
Hydrocarbures	Total C10-C40	1668

Tableau 1: Effectifs des 14 paramètres considérés dans la suite des traitements.

Disposant d'un nombre important de substances, nous avons cherché à identifier des familles de substances fortement corrélées, présentant différents aspects de la géochimie urbaine. Pour identifier ces familles, nous examinons la matrice de corrélation (Tableau 2) (en

I.B. Présentation des données utilisées

vérifiant la pertinence des résultats par les nuages de corrélation) et avons réalisé des analyses en composantes principales (ACP) sur les données brutes et log-translatées (Figure 20).

	Cu	Pb	Zn	Ba	Cr	Ni	C10-C40	HAP 16
Cu	1							
Pb	0,89	1						
Zn	0,89	0,82	1					
Ba	0,66	0,65	0,77	1				
Cr	0,46	0,43	0,6	0,74	1			
Ni	0,53	0,47	0,68	0,8	0,85	1		
C10-C40	0,42	0,37	0,38	0,21	0,04	0,09	1	
HAP16	0,54	0,51	0,53	0,35	0,06	0,17	0,74	1

Tableau 2 : Matrice de corrélation mettant en évidence les trois familles identifiées sur les données brutes (pour un même effectif de 1853 échantillons).

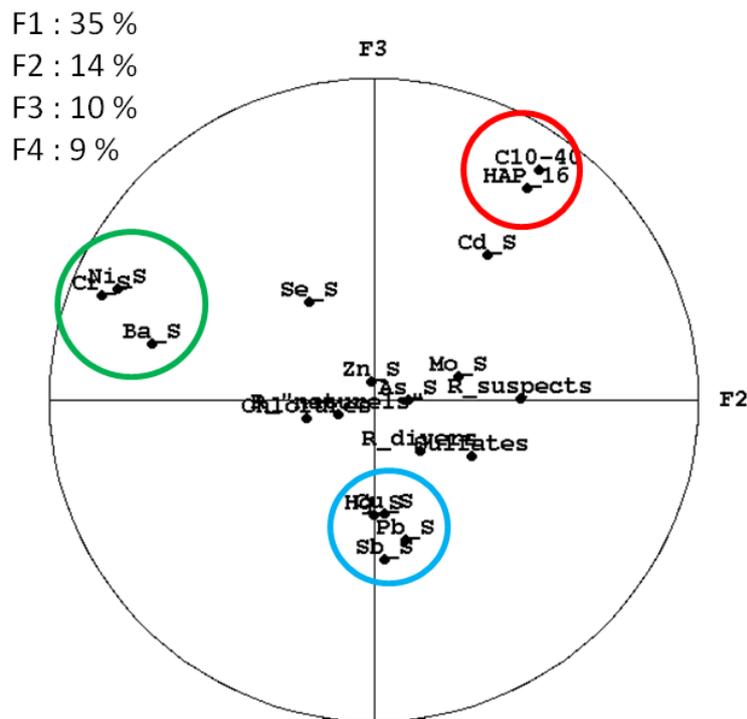


Figure 20 : Exemple de résultats d'ACP mettant en évidence les trois familles identifiées sur toutes les données brutes (pour un même effectif de 1853 échantillons). Rouge : hydrocarbures, bleu : métaux/ métalloïdes influencés par les pressions anthropiques, vert : métaux/métalloïdes d'origine géogénique

Trois familles ont été identifiées. La première concerne les métaux/métalloïdes dont les teneurs sont fortement influencées par l'activité humaine (en bleu dans la figure 20)

I.B. Présentation des données utilisées

(cuivre, plomb, zinc) (Alloway, 1990; Birke, 2000 ; Birke et Rauch, 1997) (le zinc n'est pas associé au plomb ou au cuivre dans l'exemple de l'ACP, il a été ajouté pour sa forte corrélation aux métaux de cette famille (cf. Tableau 2). La seconde famille regroupe, à l'inverse, des métaux/métalloïdes plutôt d'origine naturelle (en vert dans la figure 20) (baryum, chrome, nickel) dans le secteur d'étude (Le Guern et al, 2016). Enfin, la troisième famille regroupe les hydrocarbures C10-C40 et les 16 HAP indicateurs (en rouge dans la figure 20). Ces derniers sont retenus comme second marqueur d'anthropisation.

Il est donc possible d'obtenir un nombre important de données géologiques et géochimiques sur une aire urbaine via la collecte de diagnostics de pollution. Toutefois, il convient de tenir compte des limites liées aux différents protocoles d'échantillonnage et d'analyses avant d'utiliser les données pour calculer des valeurs de référence pour la qualité géochimique des sols et sous-sols urbains. Cela se fait par une sélection des données pertinentes et comparables entre-elles.

Données à l'échelle de la ville

Afin de compléter les données de diagnostics de pollution, d'autres données géochimiques ont été étudiées.

Les premières sont des données issues de campagnes d'échantillonnage de sols de jardins nantais. Ces données ont été acquises en partenariat entre le BRGM, l'IFSTTAR, l'Université de Nantes, l'IRSTV et la ville de Nantes dans le cadre de projets de recherche et développement (dont l'ANR Jassur).

Les données sont issues d'échantillons de sols principalement analysés au spectroscope à fluorescence X portable (pXRF), avec des analyses complémentaires à l'ICP pour une sélection d'échantillons représentatifs des gammes de teneurs observées en métaux et métalloïdes (Pb, As, Zn, Cu notamment).

J'ai complété la base de données locale rassemblant les analyses issues des diagnostics de pollution avec les données d'analyse ICP des échantillons géoréférencés (ponctuel ou composite de parcelle). Disposant déjà d'un nombre conséquent de données sur la majorité des quartiers, les données pXRF seules n'ont pas été prises en compte à ce stade afin de ne pas introduire de variabilité liée à cette méthode d'analyse, bien que des corrélations aient été proposées avec les données d'ICP.

Nous nous sommes également intéressés aux données du projet METOTRASS. Néanmoins, aucune donnée n'étant disponible pour les aires urbaines, et plus précisément à Nantes, ces données n'ont pu servir que de comparaison, notamment pour les analyses sur les alluvions et les sols naturels.

La Figure 21 illustre la répartition des données collectées.

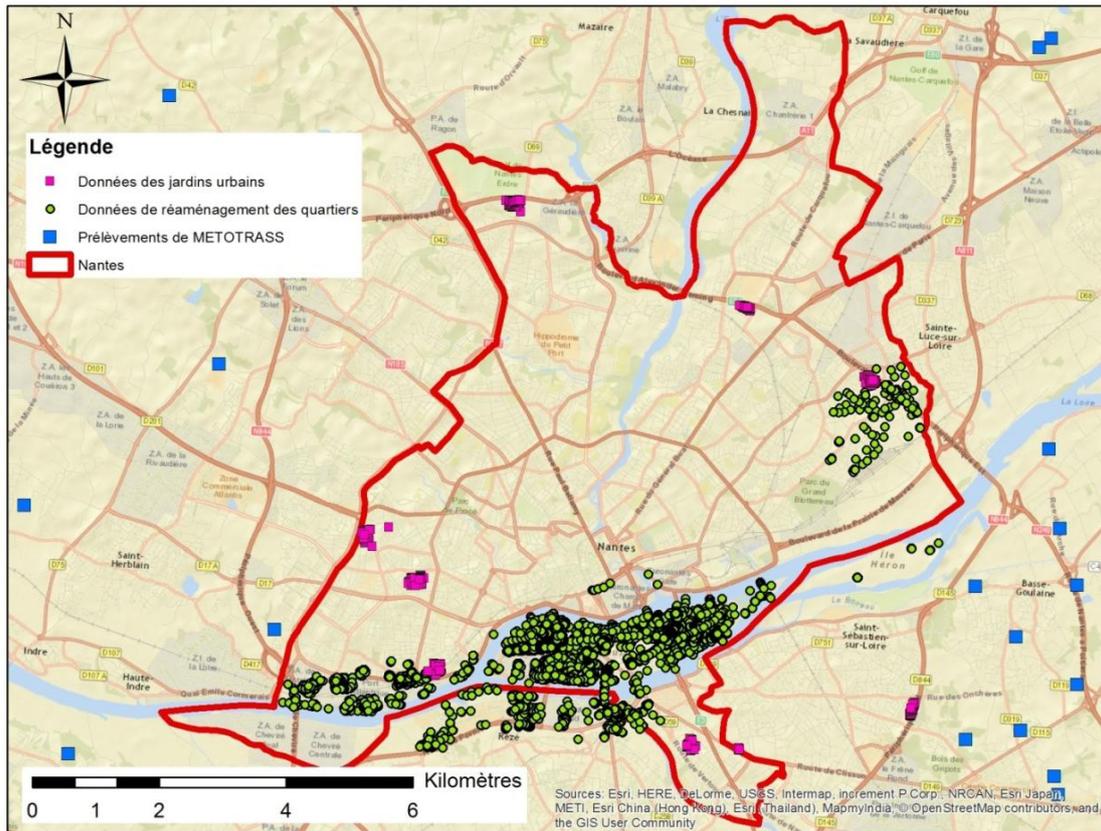


Figure 21 : Carte d'implantation des données collectées suivant leur origine.

Vert : diagnostics de pollution, Violet : jardins, Bleu : METOTRASS

Acquisition de données complémentaires

Enfin, des données complémentaires ont été acquises dans un quartier disposant de peu de données (Doulon – Gohards). Ce quartier péri-urbain avec un passé maraîcher très marqué, présente peu d'anciens sites industriels et par conséquent les données de diagnostics de pollution y sont peu nombreuses.

Nous avons prélevé 43 échantillons sur 24 sondages d'environ 1,5 m de profondeur réalisés à la tarière à main sur l'ensemble du quartier. Les anciennes parcelles de maraîchages, les espaces verts et les terrains en friches ont été principalement ciblés pour des raisons d'accès. Les échantillons (ponctuels) ont été prélevés en veillant à ne pas mélanger les différentes couches rencontrées, mis en sachet plastique hermétique et placés en glacière. Au laboratoire les prélèvements ont été homogénéisés et séchés à l'air libre, puis tamisés à 2 mm. La fraction tamisée a été mesurée par pXRF (NITON XL3t Goldd).

Nous avons également cherché à calibrer les mesures par pXRF par des analyses par ICP MS sur une sélection d'échantillons représentatifs des gammes de teneurs observées pour les éléments trace présentant les gammes les plus étendues (Pb, Zn, As, Cu...). Les résultats sont fournis en annexe 4.

Partie II : METHODES STATISTIQUES DE DETERMINATION D'UNE VALEUR SEUIL DE FOND PEDO-GEOCHIMIQUE ET ETUDE DE SENSIBILITE

II.A. Méthodes usuelles de détermination d'une valeur seuil de fond pédo-géochimique

Nous avons présenté les différentes définitions associées au fond pédo-géochimique ainsi que les spécificités des données disponibles. Il convient maintenant de s'intéresser aux méthodologies de calculs des valeurs de références du fond pédo-géochimique anthropisé.

Qu'elles proviennent de la collecte de données existantes ou de l'analyse de nouveaux échantillons, les informations obtenues doivent être interprétées afin d'en déduire une valeur de fond pédo-géochimique.

Parmi les nombreux protocoles visant à fournir des valeurs seuils de fond géochimique, deux grandes démarches sont à distinguer. La première se fonde sur la géochimie pour calculer des valeurs moyennes ou médianes d'échantillons de référence pour de grandes zones géographiques. Par exemple, des prélèvements de sédiments océaniques sont effectués ou les concentrations moyennes de la croûte continentale sont calculées. Cela fournit un aperçu de la géochimie qui ne tient pas compte des variations spatiales à l'intérieur de la zone d'étude.

Cette démarche est fortement influencée par les connaissances des experts, pour le choix des échantillons et leur interprétation. Les résultats peuvent alors être biaisés par des décisions subjectives et il est parfois complexe d'effectuer des comparaisons fiables entre différentes études. De plus, ne pouvant être appliqué à un secteur urbain du fait de la forte hétérogénéité du milieu, elle ne sera pas abordée davantage.

La seconde démarche est basée sur une approche statistique. Les protocoles se concentrent sur des calculs ou des diagrammes et sont donc moins susceptibles d'être biaisés par une décision subjective de l'opérateur (en dehors du choix de la méthode), l'influence du jeu de données reste toutefois à prendre en compte. Ces protocoles cherchent à déterminer une valeur seuil séparant les teneurs liées au fond pédo-géochimique des anomalies (naturelles ou anthropiques). Pour cela il existe des méthodes graphiques et des calculs statistiques de seuils.

II.A.1. Méthodes graphiques

Les méthodes graphiques se basent sur l'observation et l'interprétation de diagrammes. Fondée sur les travaux de Lepeltier (1969), la première méthode consiste à tracer un graphique des fréquences cumulées comprenant l'axe des abscisses et des ordonnées en logarithme (voir Figure 22). Le premier changement de pente indique un changement de mode de distribution et, dans notre cas, le passage à des valeurs anomales.

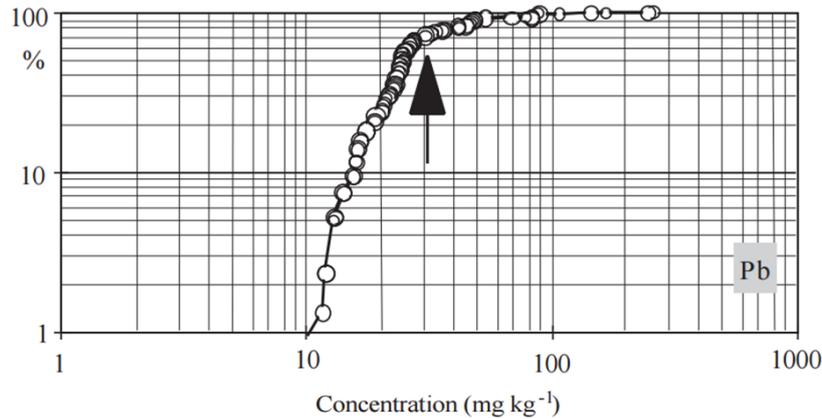


Figure 22 : Méthode de Lepeltier. Les fréquences cumulées sont exprimées en fonction des concentrations en mg/kg sur une échelle logarithmique. La flèche représente le passage au deuxième mode de distribution (Matschullat, 2000).

Il est également possible de calculer une distribution théorique des données en l'absence d'anomalies supposées. Les auteurs se fondent alors sur l'hypothèse qu'il n'y a pas de données anomales entre le minimum et la médiane. Ils extraient de cette façon la moitié inférieure des données et les reproduisent symétriquement par rapport à la médiane d'origine (voir Figure 23). Supposant la distribution symétrique, ils obtiennent ainsi une population sur laquelle il est possible de calculer une valeur limite du fond géochimique (Matschullat, 2000).

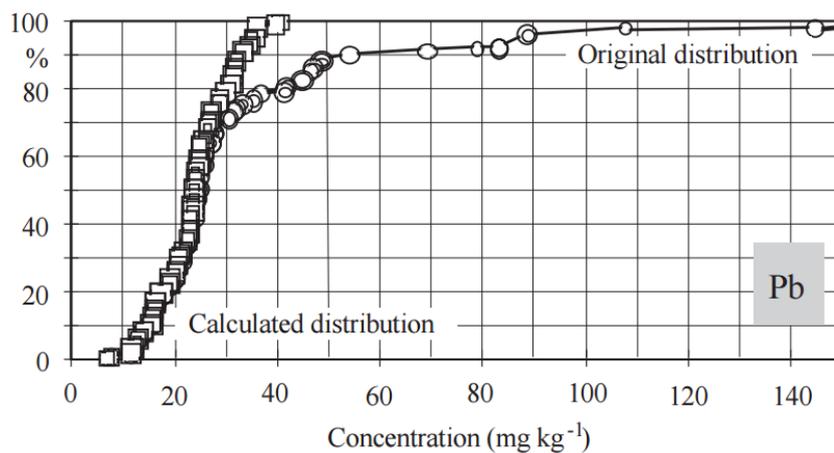


Figure 23 : Exemple de fonction de distribution théorique. Les deux lignes représentent les distributions originales (cercles) et théorique calculée (carrés). Le jeu de données utilisé est le même que celui de la figure 25 (Matschullat, 2000).

II.A. Méthodes usuelles de détermination d'une valeur seuil de fond pédo-géochimique

Ces méthodes peuvent être biaisées lors du choix du changement de pente ou de la valeur de coupure. En effet cette valeur peut être influencée par l'appréciation de l'opérateur.

Afin d'obtenir des valeurs fiables et comparables avec ce type de méthodes, il est nécessaire de s'assurer de la représentativité des données utilisées, de disposer d'un effectif important et de conserver le même opérateur pour toutes les interprétations.

L'application de ces méthodes pour la détermination de valeurs de référence du fond pédo-géochimique anthropisé à l'échelle des grandes aires urbaines françaises n'est pas encouragée, du fait des données collectées (dont la représentativité est à vérifier) et d'une volonté de limiter les biais liés à des décisions de l'opérateur.

II.A.2. Calculs statistiques

Les méthodes basées sur des calculs statistiques cherchent à déterminer une valeur seuil ou ligne de base, séparant les anomalies du fond pédo-géochimique.

Historiquement, l'une des premières méthodes utilisée pour déterminer le fond géochimique par un calcul statistique a été la moyenne plus ou moins deux fois l'écart type ($\text{moy} \pm 2\sigma$) (Hawkes et Webb, 1962 ; Liebens, 2012 ; Matschullat, 2000 ; Reimann, 2005). Cette méthode se base sur l'hypothèse d'une distribution normale des données et fournit alors un intervalle comprenant approximativement 95% des données. Toutefois il ne doit être effectué que sur un jeu de données unimodal (Reimann, 2005) et la moyenne peut-être influencée par les valeurs extrêmes (de Fouquet, 2012).

Une méthode, dérivée de la première, consiste à calculer sur les données la gamme de valeurs correspondant à la moyenne plus ou moins deux fois l'écart type et à enlever les valeurs en dehors de cet intervalle. On répète ensuite l'opération autant de fois que nécessaire afin que l'ensemble de valeurs soient comprises dans l'intervalle. Les valeurs restantes sont alors considérées comme reflétant le fond géochimique de la zone (Erhardt, 1998).

Les jeux de données résultant d'analyses géochimiques contiennent généralement de nombreuses valeurs « extrêmes ». Celles-ci comprennent notamment des valeurs parfois très fortes liées aux anomalies naturelles ou aux contaminations. La moyenne utilisée dans les calculs précédents y est très sensible et les résultats sont influencés par ces valeurs. C'est pour cela qu'une nouvelle méthode fondée sur la médiane (insensible aux valeurs extrêmes), communément appelée « Seuil MAD » (Médian Absolute Deviation) (Tukey, 1977), a été mise en place et est couramment utilisée (Reimann, 2005 ; Rothwell, 2015). Le principe en est le suivant :

Soit X la variable étudiée (une teneur en élément chimique dans les sols). Notons Q_{50} sa médiane ($P(X < Q_{50}) = 1/2$). On a alors :

II.A. Méthodes usuelles de détermination d'une valeur seuil de fond pédo-géochimique

- seuil associé au MAD, par la méthode de la médiane (Leys, 2013 ; Reimann et al., 2018 ; Tukey, 1977) :

$$(1) \quad S_{MAD} = Q_{50} + (2 \text{ MAD}) \text{ avec } MAD = \text{médiane} (|Z - Q_{50}|)$$

Exemple : Pour un jeu de données (1, 1, 2, 2, 4, 6, 9), la médiane est 2, la distance absolue à la médiane est donc (1, 1, 0, 0, 2, 4, 7), qui une fois classé donne (0, 0, 1, 1, 2, 4, 7). Ainsi le terme MAD est 1 et le seuil S_{MAD} est égal à 3.

L'expression de ce seuil comprend un quantile (médiane) et un terme de distance, le MAD. Le résultat est insensible aux valeurs extrêmes. Ce seuil n'est, à l'heure actuelle, pas utilisé dans des normes et n'est généralement évoqué que pour la comparaison à d'autres seuils.

Une autre solution pour le calcul de valeurs limites est basée sur les box-plot ou boîtes de dispersion (Figure 24). Cette méthode dépend des quartiles et s'exprime de la façon suivante :

Soient respectivement Q_{25} et Q_{75} les deux autres quartiles ($P(Z < Q_{25}) = 1/4$ et $P(Z < Q_{75}) = 3/4$)

- seuil associé à la vibrisse, par le centile de Tukey (Jarva, 2014 ; Leys, 2013 Reimann, 2005, 2017, 2018 ; Rothwell, 2015 ; Tarvainen, 2011) défini comme :

$$(2) \quad S_{\text{vibrisse}} = Q_{75} + a (Q_{75} - Q_{25}) \text{ avec généralement } a = 3/2$$



Figure 24 : Schéma d'une boîte de dispersion.

Cette méthode, plus récente, est fréquemment utilisée et figure dans la méthodologie du programme national finlandais (Jarva, 2010) et du Réseau de Mesure de Qualité des Sols (RMQS) français. Elle ne pose pas d'hypothèse sur la loi des valeurs et ne nécessite donc pas de transformation ni de prétraitement. Tout comme le seuil MAD, le seuil obtenu avec la vibrisse supérieure fournit une valeur théorique ne correspondant pas à une valeur réelle mesurée. Ainsi la valeur retenue peut être la valeur théorique calculée ou la valeur mesurée inférieure la plus proche. Dans la suite nous considérons uniquement le cas de la valeur calculée, jugée plus indépendante de la variabilité des valeurs fortes.

Enfin, un autre protocole, mis en place au Royaume-Uni par le British Geological Survey (BGS) (Ander, 2013), fournit une valeur limite du « normal background

II.A. Méthodes usuelles de détermination d'une valeur seuil de fond pédo-géochimique

concentration » (NBC) pouvant être assimilé au bruit de fond géochimique. Cette valeur cherche à mettre en évidence des sites qui « *si rien n'est fait, présenteraient des risques significatifs de danger tels que la mort, la maladie ou des blessures graves* » (Ander, 2013). Cette valeur limite a été fixée au 95^{ème} centile. Une variante faisant intervenir le 90^{ème} centile est également décrite dans la norme européenne (ISO Standard 19258, 2005).

- seuil associé au 95^{ème} ou 90^{ème} centile (Ander, 2013 ; Cave, 2012 ; McIlwaine, 2014 ; Rothwell, 2015) :

(3) $S_{95}=Q_{95}$ et $S_{90}=Q_{90}$

Contrairement aux deux autres, cette valeur seuil du fond pédo-géochimique n'associe plus un terme de distance et correspond simplement à un quantile prédéfini.

II.B. Sensibilité des calculs statistiques du seuil de fond géochimique

Compte-tenu des différences entre les trois principales méthodes de calcul statistiques de seuils de fond pédo-géochimique, la question se pose de leur comparaison dans les cas de lois usuelles, mais aussi dans les cas spécifiques de données non dédiées au fond pédo-géochimique.

Trois lois élémentaires sont examinées. Nous étudierons notamment s'il existe une relation d'ordre systématique entre les valeurs seuil selon ces lois. En particulier, l'un des trois seuils apparaît-il plus conservatif (fournissant systématiquement la valeur minimum) quelle que soit la loi ?

Nous examinons ensuite le cas d'un mélange de lois, pour vérifier si les seuils peuvent différencier les deux modes de la distribution dans le cas d'une population statistiquement hétérogène.

L'influence du nombre de données est ensuite examinée. En effet, certaines statistiques, comme les quantiles d'ordre faible ou élevé, sont très sensibles à l'effectif (voir par exemple, Bernard-Michel et al., 2005 & 2006, en contexte environnemental). La question se pose également de l'influence de la limite de quantification (LQ), présente dans les données géochimiques "du fond". Plusieurs ouvrages soulignent cette influence (Reimann, 2005 ; Rothwell, 2015 ; Sancho, 2016), sans toutefois la quantifier. On examinera par simulation l'effet de la proportion de données inférieures à la Limite de Quantification ainsi que celle de la valeur de substitution choisie. Enfin, l'influence de l'hétérogénéité spatiale des échantillons est examinée sur des données réelles bancarisées par le BRGM.

Les calculs sont d'abord effectués sans prise en compte de l'implantation des données. Les calculs sont ensuite repris en introduisant une pondération des données, afin de corriger l'influence de l'irrégularité spatiale de l'échantillonnage. Enfin, la typologie de remblais est aussi prise en compte (Le Guern et al, 2016).

La majorité des calculs a été effectuée avec le logiciel libre R. Les logiciels Access et Excel de la suite Microsoft office, sont utilisables pour ces calculs.

II.B.1. Influence de la distribution

Les distributions empiriques de données géochimiques sur un site, une agglomération ou une région sont variées (Karim et al., 2015 ; Reimann, 2005).

Afin d'étudier l'influence des lois de distribution sur le seuil de fond pédo-géochimique calculé, trois lois élémentaires sont examinées : la loi uniforme, la loi normale et la loi log-normale.

Pour toutes les lois (théoriques ou empiriques), $S_{90} \leq S_{95}$.

a) Lois symétriques :

Notons Q_{50} la médiane de la distribution Z . La loi étant symétrique, $Q_{75}-Q_{50}=Q_{50}-Q_{25}$.

Par suite,

$$\begin{aligned} \text{MAD} &= \text{médiane } (|Z-Q_{50}|) \\ &= Q_{75}-Q_{50} \end{aligned}$$

D'où :

$$\begin{aligned} S_{\text{MAD}} &= Q_{50} + 2\text{MAD} \\ &= Q_{50} + Q_{75} - Q_{25} \\ &= Q_{75} + (Q_{50} - Q_{25}) \end{aligned}$$

Comme $S_{\text{vibrise}} = Q_{75} + \alpha (Q_{75} - Q_{25})$, en dehors du cas particulier où $Q_{75}=Q_{50}$, nécessairement

$$\forall \alpha \geq 1, S_{\text{MAD}} < S_{\text{vibrise}}$$

b) Loi uniforme (Figure 25).

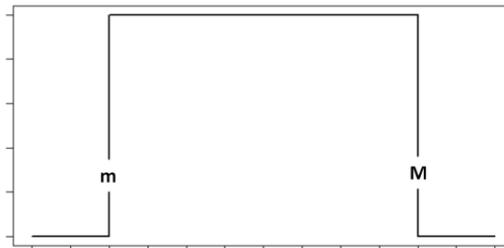


Figure 25 : Loi uniforme de minimum "m" et maximum "M".

La loi uniforme sur l'intervalle $[Z_{\min}, Z_{\max}]$, est un cas peu probable en géochimie mais permet de poser les bases de la réflexion. L'expression d'un quantile est :

$$P(Z < z) = (z - Z_{\min}) / (Z_{\max} - Z_{\min}), \quad Z_{\min} \leq z \leq Z_{\max}$$

D'où l'expression suivante du S_{95} :

II.B. Sensibilité des calculs statistiques du seuil de fond géochimique

$$S_{95} = Q_{95} \\ = Z_{\min} + 0,95 [Z_{\max} - Z_{\min}]$$

Dans le cas de la vibrisse supérieure :

$$\begin{aligned} S_{\text{vibrisse}} &= Q_{75} + 1,5 (Q_{75} - Q_{25}) \\ \text{Avec } (Q_{75} - Q_{25}) &= Z_{\min} + 0,75 [Z_{\max} - Z_{\min}] - Z_{\min} - 0,25 [Z_{\max} - Z_{\min}] \\ &= 0,5 [Z_{\max} - Z_{\min}] \end{aligned}$$

$$S_{\text{vibrisse}} = Z_{\min} + 1,5 [Z_{\max} - Z_{\min}]$$

Pour S_{MAD} , la loi uniforme étant symétrique :

$$\begin{aligned} \text{MAD} &= \text{médiane } (|Z - Q_{50}|) \\ &= Q_{75} - Q_{50} \\ &= Q_{50} - Q_{25} \end{aligned}$$

d'où :

$$\begin{aligned} S_{\text{MAD}} &= Q_{50} + 2\text{MAD} \\ &= Q_{50} + 2 (Q_{75} - Q_{25}) \\ &= 2 Q_{75} - Q_{50} \\ &= 2 Z_{\min} + 2 \cdot \frac{3}{4} [Z_{\max} - Z_{\min}] - Z_{\min} - \frac{1}{2} [Z_{\max} - Z_{\min}] \\ S_{\text{MAD}} &= Z_{\min} + [Z_{\max} - Z_{\min}] \\ &= Z_{\max} \end{aligned}$$

Les seuils vérifient les inégalités suivantes :

$$S_{90} < S_{95} < S_{\text{MAD}} < S_{\text{vibrisse}}$$

Choisir les quantiles comme seuils revient à considérer que les 10% ou les 5% supérieurs des données ne font pas partie du fond pédo-géochimique. Or, rien ne justifie a priori cette différenciation. A l'inverse, la vibrisse supérieure permettrait ici d'intégrer des données supérieures à Z_{\max} : une donnée supplémentaire supérieure à Z_{\max} ne serait donc pas nécessairement considérée comme une anomalie.

c) loi normale (gaussienne) (Figure 26).

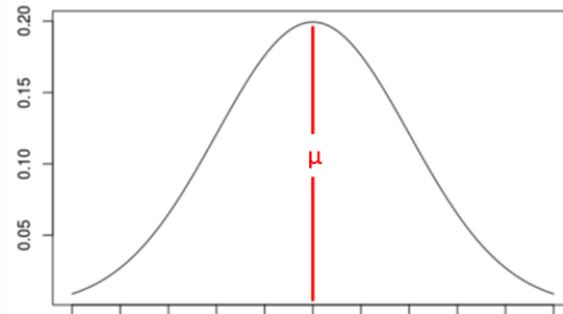


Figure 26 : Exemple de loi normale de moyenne μ .

Rappelons que pour une gaussienne réduite, $Q_{75} = 0.6745$, $Q_{90}=1.282$ et $Q_{95} =1.645$. Les seuils associés à la variable normale, $Z = \mu + \sigma Y$ où Y est une gaussienne réduite, sont :

$$\begin{aligned} \text{d'où} \quad S_{90} &= \mu + \sigma Q_{90} \\ S_{90} &= \mu + 1.282\sigma \end{aligned}$$

$$\begin{aligned} \text{d'où} \quad S_{95} &= \mu + \sigma Q_{95} \\ S_{95} &= \mu + 1.645\sigma \end{aligned}$$

La loi normale étant symétrique,

$$\begin{aligned} S_{\text{vibrise}} &= \mu + \sigma [Q_{75} + 1.5 (Q_{75} - Q_{25})] \\ &= \mu + (1 + 2 * 1.5) \sigma Q_{75} \\ &= \mu + 2.698\sigma \end{aligned}$$

$$\begin{aligned} S_{\text{MAD}} &= \mu + 2 * \text{médiane}_i (IZ_i - \mu I) \\ &= \mu + 2\sigma * \text{médiane}_i (IY_i I) \\ &= \mu + 2\sigma Q_{75} = \mu + 1.349\sigma \end{aligned}$$

L'ordre des seuils est le suivant :

$$S_{90} < S_{\text{MAD}} < S_{95} < S_{\text{vibrise}}$$

Il diffère donc de celui du cas uniforme ; en particulier l'ordre des seuils du 95^{ème} centile et de la méthode MAD sont inversés.

Les données géochimiques suivent néanmoins rarement des lois de distribution symétrique. Il est alors intéressant d'appliquer ces seuils à une **loi log-normale** (Figure 27), plus proche des données géochimiques habituelles.

d) Loi Log-normale

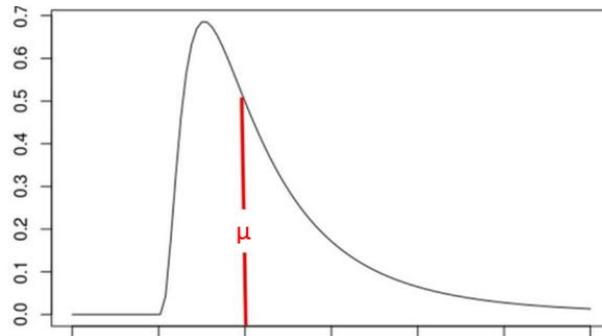


Figure 27 : Exemple de loi log-normale de moyenne μ .

Dans le cas d'une loi log-normale, $Z = e^{\mu + \sigma Y}$ où Y est une gaussienne réduite. Deux modes de calcul sont possibles : on peut exponentier les seuils calculés pour la loi normale, ce qui donne

$$\begin{aligned} S_{90} &= e^{\mu + 1.282\sigma} \\ S_{95} &= e^{\mu + 1.645\sigma} \\ S_{\text{Vibrisse}} &= e^{\mu + 2.698\sigma} \\ S_{\text{MAD}} &= e^{\mu + 1.349\sigma} \end{aligned}$$

L'ordre reste identique au cas précédent puisque la fonction exponentielle est croissante :

$$S_{90} < S_{\text{MAD}} < S_{95} < S_{\text{Vibrisse}}$$

On peut aussi effectuer le calcul directement, sur la loi log-normale ; les formules donnent alors :

$$S_{90} = e^{\mu + 1.282\sigma} \text{ et } S_{95} = e^{\mu + 1.645\sigma}$$

Résultats identiques à ceux du calcul précédent.

$$\begin{aligned} S_{\text{Vibrisse}} &= e^{\mu + \sigma Q_{75}} + 1.5 e^{\mu} (e^{\sigma Q_{75}} - e^{-\sigma Q_{75}}) \\ D'où \quad S_{\text{Vibrisse}} &= e^{\mu + \sigma Q_{75}} (5/2 - e^{-2\sigma Q_{75}}) \\ S_{\text{MAD}} &= e^{\mu} + 2 e^{\mu} \text{médiane} (| e^{\sigma Y} - 1 |) \end{aligned}$$

Ces seuils s'expriment en fonction de σ sous la forme du rapport seuil/ e^{μ} (cf. Tableau 3) :

σ	0.01	0.1	0.25	0.5	0.75	1	2	3	5
S_{90}/e^{μ}	1.0129	1.137	1.378	1.898	2.616	3.604	12.988	46.805	607.603
S_{95}/e^{μ}	1.017	1.179	1.509	2.276	3.434	5.180	26.835	139.012	3730.000
$S_{\text{Vibrisse}}/e^{\mu}$	1.524	1.740	2.114	2.789	3.543	4.398	9.374	18.780	72.844
S_{MAD}/e^{μ}	1.0135	1.134	1.333	1.654	1.944	2.200	2.790	2.961	2.999

Tableau 3 : Rapport seuil/ e^{μ} en fonction de σ , pour le calcul direct dans le cas d'une loi log-normale.

II.B. Sensibilité des calculs statistiques du seuil de fond géochimique

Les résultats diffèrent de ceux du premier calcul, l'ordre des seuils variant ici selon σ :

- lorsque $\sigma \leq 0,01$, $S_{90} < S_{MAD} < S_{95} < S_{Vibrise}$
- lorsque $\sigma \in]0.01 ; 0.75]$, $S_{MAD} < S_{90} < S_{95} < S_{Vibrise}$
- lorsque $\sigma \in]0.75 ; 1]$, $S_{MAD} < S_{90} < S_{Vibrise} < S_{95}$
- lorsque $\sigma > 1$, $S_{MAD} < S_{Vibrise} < S_{90} < S_{95}$

Ainsi, la loi de probabilité, mais aussi, dans le cas log-normal, la méthode de calcul influencent les seuils obtenus. Dans la majorité des cas, la méthode MAD fournit le seuil le plus bas et tend donc à être la plus conservative pour l'environnement. La valeur absolue $|Z - Q_{50}|$ fait jouer le même rôle aux valeurs respectivement inférieures ou supérieures à la médiane. Or généralement, les valeurs faibles étant moins dispersées que les valeurs fortes, cet écart est plus réduit que l'écart interquartiles $Q_{75} - Q_{25}$.

II.B.2. Mélange de lois

Considérons maintenant le mélange de deux lois de même type, et de paramètres différents. Nous admettrons que le premier mode correspond au fond et le second aux anomalies.

a) Mélange de lois uniformes

Nous commençons avec le cas élémentaire du mélange de lois uniformes. Soient (Figure 28) :

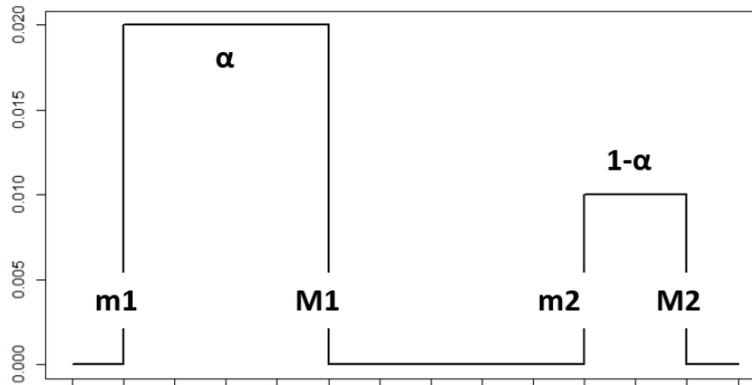


Figure 28 : Mélange de lois uniformes

- m1 et m2 : les minimums du premier et second mode
- M1 et M2 : les maximums du premier et second mode (avec $m_1 < M_1 < m_2 < M_2$)
- α : proportion du premier mode dans le mélange

Les quantiles s'expriment comme :

$$Q_z = m_1 + z [(M_1 - m_1) / \alpha] \quad \text{si } z < \alpha$$

$$Q_z = m_2 + (z - \alpha) [(M_2 - m_2) / (1 - \alpha)] \quad \text{si } z > \alpha$$

Nous cherchons ici les conditions dans lesquelles un seuil « S » sépare les deux modes, c'est-à-dire, est tel que $M_1 < S < m_2$.

Les formules ayant pour premier ou unique terme un quantile (95^{ème} pour le S_{95} , 90^{ème} pour le S_{90} , 75^{ème} pour S_{vibrisse} , et le 50^{ème} pour S_{MAD}), ces derniers représentent la première contrainte d'inégalité. Ainsi, le 90^{ème} centile ne pourra se situer entre M_1 et m_2 que si α est égal à 0.9. Et 0.95 pour le 95^{ème} centile. Cette condition est rarement vérifiée dans un cas réel.

Dans le cas de la vibrisse supérieure, un terme positif est ajouté au 3^{ème} quartile ; pour que le seuil se trouve dans les conditions recherchées, ce quartile doit appartenir au premier mode et donc α doit être supérieur à 0,75. Cette hypothèse apparaît plausible, car cela sous-entend que les anomalies représentent moins de 25% des données totales. Une proportion du

II.B. Sensibilité des calculs statistiques du seuil de fond géochimique

second mode, supérieure à 0,25 peut en effet remettre en question son caractère d'anomalie. La deuxième condition est que le seuil doit être inférieur au second mode.

En s'appuyant sur la relation suivante :

$$\begin{aligned}(Q_{75}-Q_{25}) &= m_1 + 0.75 [(M_1-m_1)/\alpha] - m_1 + 0.25 [(M_1-m_1)/\alpha] \\ &= 0.5 [(M_1-m_1)/\alpha]\end{aligned}$$

Et comme :

$$S_{\text{Vibrisse}} = m_1 + 0.75[(M_1-m_1)/\alpha] + 1.5 (0.5 [(M_1-m_1)/\alpha])$$

Alors :

$$S_{\text{Vibrisse}} = m_1 + 1,5 (M_1-m_1)/\alpha < m_2$$

Or si S_{Vibrisse} est inférieur à m_2 , alors les conditions posées sur le seuil sont respectées. Ces conditions restent restrictives mais plus plausibles que dans le cas des quantiles.

Enfin, dans le cas de la méthode MAD, la première contrainte est posée par la médiane. Ainsi la première condition est que α soit supérieur à 0.5. Ce qui implique que les données du fond doivent nécessairement être plus nombreuses que les données anormales. Par la suite nous étudions les deux cas, $\alpha \geq 0.75$ et $0.5 \leq \alpha < 0.75$.

- Si α est supérieur ou égal à 0.75, alors le seuil s'exprime de la façon suivante :

Comme la loi uniforme est symétrique, alors si $\alpha \geq 0.75$:

$$\begin{aligned}MAD = Q_{50}-Q_{25} &= m_1 + 0.5 [(M_1-m_1)/\alpha] - m_1 + 0.25 [(M_1-m_1)/\alpha] \\ &= 0.25 [(M_1-m_1)/\alpha]\end{aligned}$$

$$S_{MAD} = m_1 + 0.5 [(M_1-m_1)/\alpha] + 2 (0.25 [(M_1-m_1)/\alpha])$$

$$(4) \quad S_{MAD} = m_1 + ((M_1-m_1)/\alpha)$$

Comme $\alpha < 1$ alors $((M_1-m_1)/\alpha) > (M_1-m_1)$ et $S_{MAD} > M_1$. Si $S_{MAD} < m_2$, dans le cas où $\alpha > 0.75$, le seuil peut séparer les deux modes.

- Enfin, si α est compris entre 0.5 et 0.75 alors le seuil peut s'exprimer de la façon suivante :

$0.5 < \alpha < 0.75$ et sous la condition supplémentaire $(Q_{50}-m_1) < (m_2-Q_{50})$ soit $2Q_{50} < m_2-m_1$

$$MAD = Q_{50} - Q_{(\alpha-0.5)} = (1-\alpha) [(M_1-m_1)/\alpha]$$

$$S_{MAD} = m_1 + 0.5 [(M_1-m_1)/\alpha] + 2 [(1-\alpha) [(M_1-m_1)/\alpha]]$$

$$(5) \quad S_{MAD} = m_1 + 2.5 ((M_1-m_1)/\alpha) + 2\alpha ((M_1-m_1)/\alpha)$$

Comme $\alpha < 0.75$, alors $2.5 - 2\alpha > 1$ et $S_{MAD} > M_1$

II.B. Sensibilité des calculs statistiques du seuil de fond géochimique

Ainsi l'algorithme peut séparer les deux modes si α est compris entre 0,5 et 0,75 et si le résultat de l'équation (5) est inférieur à m_2 . Toutefois, cela nécessite que les anomalies représentent plus de 25% de la population, ce qui n'est réellement cohérent que si les anomalies sont sur-échantillonnées par rapport au fond.

Les résultats précédents sont illustrés à la figure suivante :

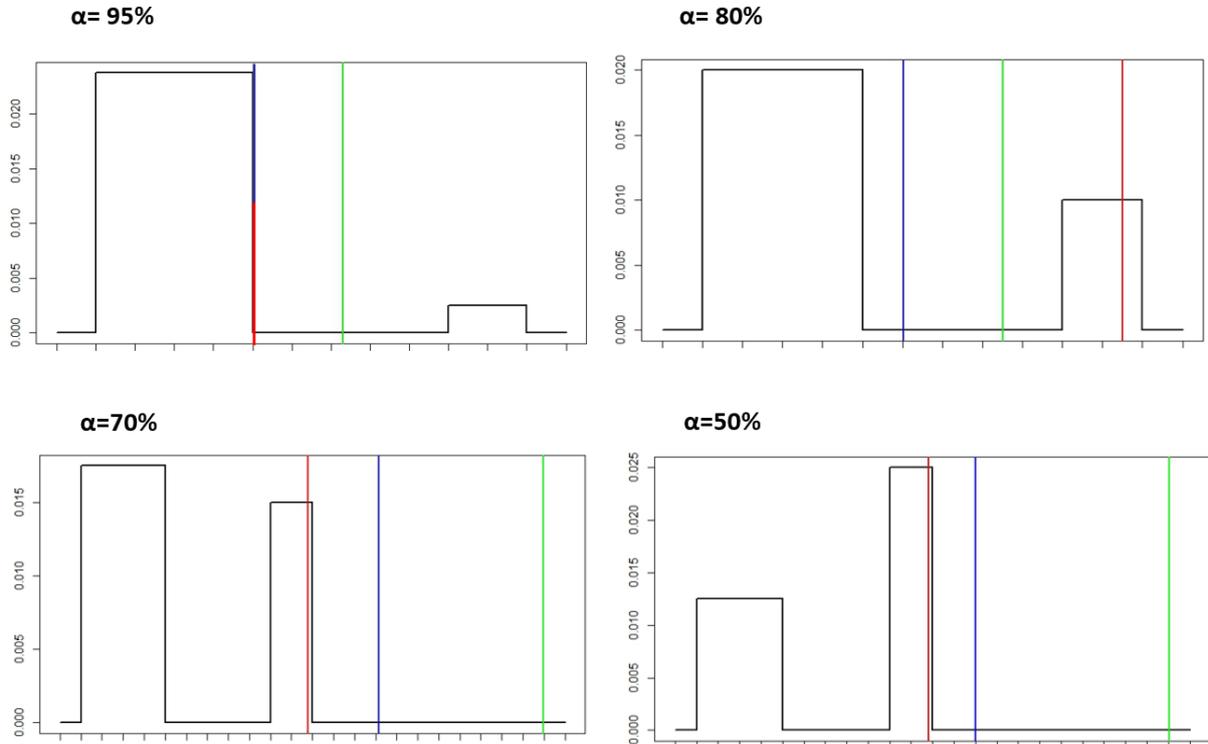


Figure 29 : Valeur seuil sur des mélanges de lois uniformes à différentes proportions α de fond.

Le mélange de deux lois uniformes montre les contraintes nécessaires pour que les seuils fournis par les trois algorithmes soient cohérents avec la limite du fond pédogéochimique supposée. Ces conditions très restrictives ne seraient évidemment que rarement vérifiées en pratique (en supposant que les données correspondent à un mélange uniforme). Nous venons de montrer que le résultat d'un algorithme ne correspond pas nécessairement au concept recherché.

b) Mélange de lois log-normales

Comme pour le cas unimodal, l'application à des lois uniformes permet de simplifier la réflexion. Toutefois les données géochimiques sont généralement plus proches de lois log-normales. Nous appliquons donc la même démarche à un mélange de loi log-normales.

Nous simulons un mélange de lois log-normales de moyennes respectives 1 et 10 et d'écart-type 1 et 1.5 afin que les deux modes soient clairement séparés. Plusieurs proportions α de fond dans le mélange sont testées (95, 80, 70 et 50%).

Un échantillonnage aléatoire d'effectif 3000 est ensuite réalisé sur ce mélange et les seuils sont calculés et représentés graphiquement (Figure 30).

II.B. Sensibilité des calculs statistiques du seuil de fond géochimique

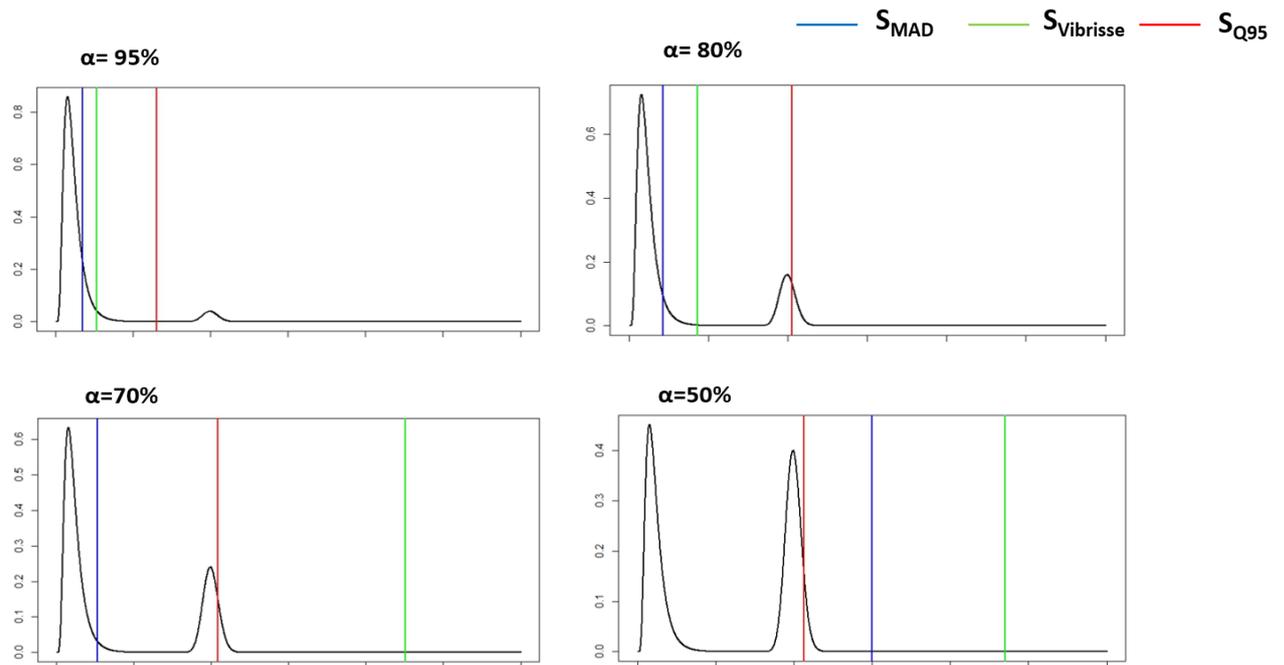


Figure 30 : Valeur seuil sur des mélanges de lois log-normales à différentes proportions α de fond, calculées par calcul direct.

Les résultats sont similaires à ceux obtenus avec le mélange de lois uniformes. La proportion α de fond dans le mélange étant le principal paramètre définissant la cohérence du résultat des seuils vis-à-vis de la définition. Ces résultats analogues dans le cas d'un mélange de lois normales, ne sont pas présentés ici.

c) Discussion

Le fait que la loi de distribution des données influence l'ordre des seuils calculés ou que ceux-ci ne correspondent pas nécessairement à la définition choisie du fond pédogéochimique, dans le cas de populations non-homogènes, montre la difficulté d'implémentation du calcul de limite supérieure de fond pédogéochimique suivant un critère statistique unique, qui ne tient pas compte du contexte. En effet, dans le cadre d'une campagne de prélèvements et d'analyses pour le diagnostic d'un site pollué (origine la plus fréquente des données disponibles), les données représentatives du fond pédogéochimique peuvent être beaucoup moins nombreuses que les "anomalies", et ceci même avec une reconnaissance à maille régulière. Ce peut aussi être le cas dans un quartier avec un passé industriel marqué, où les différents sites industriels avec des sols pollués couvrent la majorité du territoire. La notion de fond dépend donc clairement de l'échelle de travail. La zone étudiée doit être suffisamment grande pour que les "anomalies" restent minoritaires par rapport à ce que l'on considère comme le "fond" local. Il est donc préférable de faire intervenir la connaissance (notamment historique) du milieu pour délimiter la zone d'étude.

Dans le cas de données implantées irrégulièrement ou dans celui d'un sur-échantillonnage préférentiel, il est aussi nécessaire de tenir compte de l'irrégularité de l'échantillonnage. Toutefois, cette correction n'est pas toujours suffisante, notamment lorsque la reconnaissance préférentielle n'échantillonne pas suffisamment le mode des valeurs basses.

II.B.3. Effectif et limite de quantification

Examinons maintenant la sensibilité des trois seuils au nombre de données ainsi qu'à la limite de quantification. Cette étude est effectuée par simulation dans le cas log-normal (par calcul empirique sur les tirages) sur une loi de moyenne 1 et écart-type 2.5.

a) Influence de l'effectif

Pour examiner l'influence de l'effectif, les trois seuils sont calculés sur des ensembles d'effectifs croissants : 20, 30, 50, 100, 200, 300, 500, 750, 1000 données. Les calculs sont répétés 1000 fois et les résultats minimums et maximums pour chaque sous-groupe d'échantillons sont représentés graphiquement (voir Figure 31), afin de montrer l'amplitude des fluctuations.

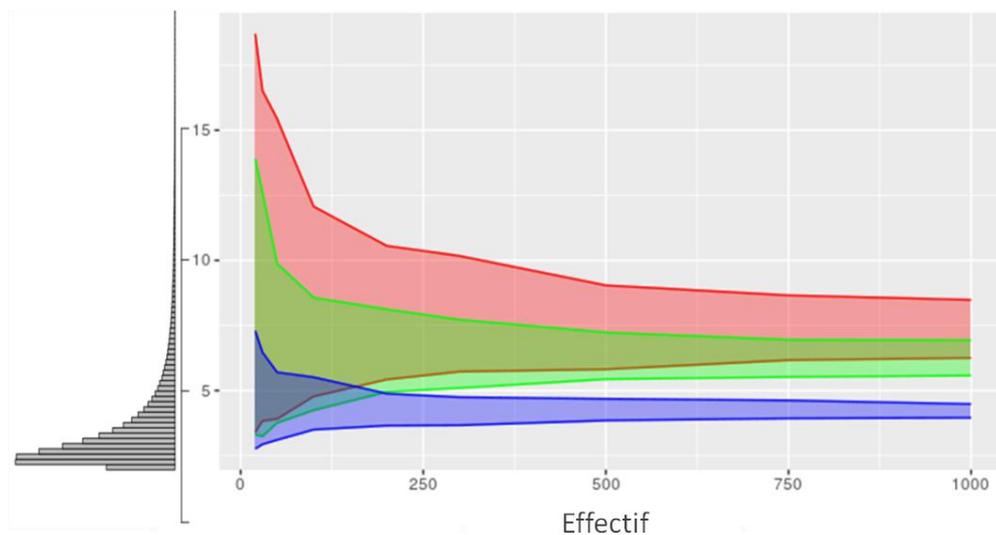


Figure 31 : Résultats minima et maxima des seuils calculés sur 1000 itérations, sur un effectif croissant. Rouge : Q95, Vert : Vibrisse supérieure, Bleu : Seuil MAD. L'histogramme des données simulées est indiqué en ordonnée. Cas lognormal (moyenne = 1 et écart-type = 2.5).

L'écart entre le maximum et le minimum obtenus par les trois algorithmes sur les 1000 itérations de la simulation représente l'amplitude des fluctuations pour les valeurs de seuil calculées. Les résultats (cf. Figure 31) montrent logiquement une diminution des fluctuations à mesure que l'effectif augmente. Pour les trois seuils les fluctuations diminuent rapidement jusqu'à 100 échantillons. Elle s'atténue ensuite pour atteindre une certaine stabilité au-delà de 200 échantillons pour le seuil MAD et 500 pour les deux autres seuils. Là encore, le seuil MAD apparaît le plus conservatif.

Le calcul des seuils à partir d'effectifs trop réduits (inférieurs à 50) est clairement à éviter. Le résultat, trop instable varierait avec la mise à disposition de données supplémentaires. Ce cas de figure reste toutefois courant dans les études environnementales. Néanmoins le seuil MAD semble présenter la plus faible amplitude de fluctuation et serait donc la meilleure option dans l'hypothèse d'un calcul sur un effectif réduit. La valeur fournie doit toutefois rester purement informative du fait de sa forte incertitude.

b) Influence de la limite de quantification

L'influence de la proportion de valeurs inférieures à la limite de quantification est examinée à partir d'un même ensemble de valeurs, les valeurs inférieures à cette limite étant supposées non quantifiées par les techniques analytiques. La proportion de valeurs inférieures à la limite de quantification est obtenue en prenant comme limite de quantification des quantiles croissants de la distribution empirique.

La Figure 32 présente les résultats obtenus par simulation en intégrant une proportion croissante d'échantillons inférieurs à la limite de quantification dans le cas d'un remplacement par cette limite ou par 0.

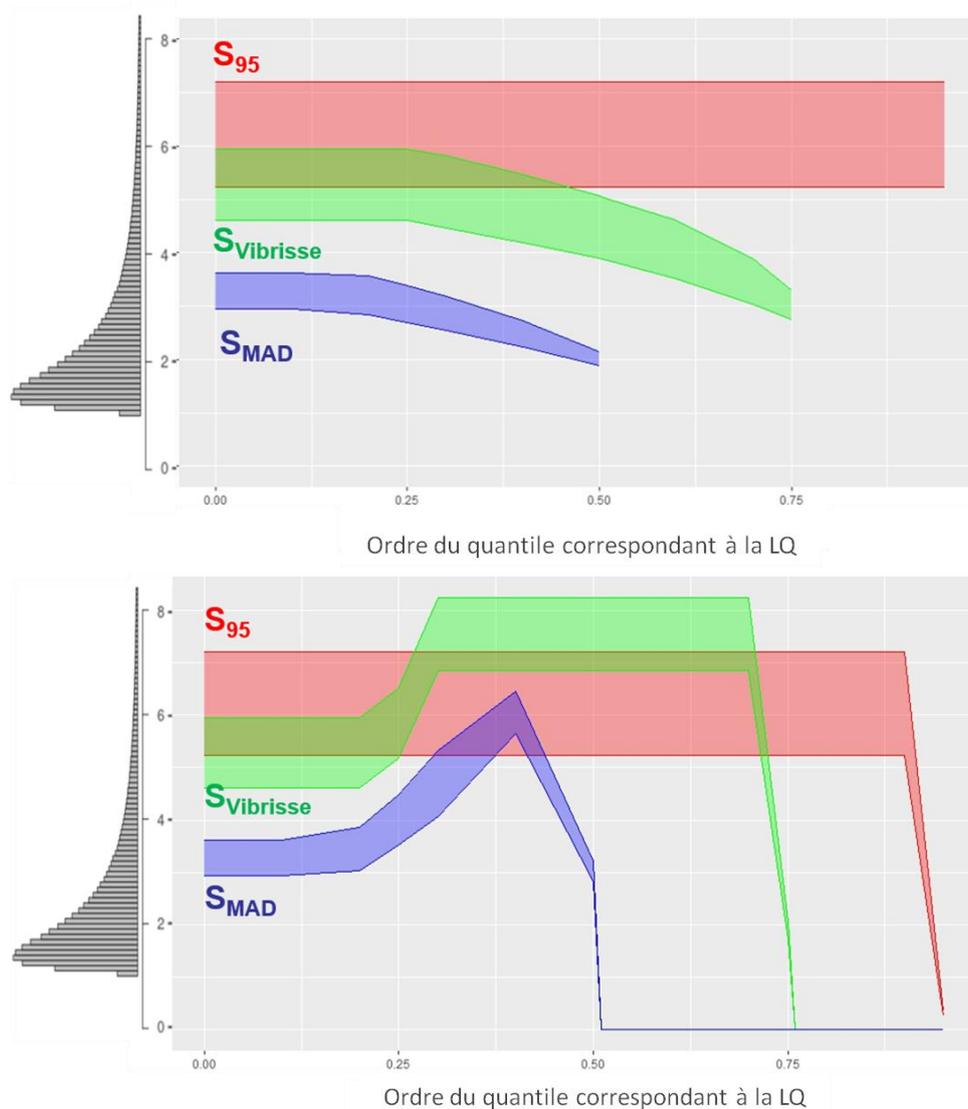


Figure 32 : Minima et maxima des seuils calculés sur 1000 itérations, considérant une part d'échantillons sous la limite de quantification croissante. Remplacement des valeurs inférieures à la LQ par la LQ (en haut) et par 0 (en bas). Rouge : Q₉₅, Vert : Vibrisse supérieure, Bleu : Seuil MAD. L'histogramme des données simulées est reporté en ordonnée. Cas lognormal (moyenne = 1 et écart-type = 2.5).

II.B. Sensibilité des calculs statistiques du seuil de fond géochimique

Tant que la limite de quantification reste inférieure au quantile Q95, le seuil S_{95} reste invariable à la valeur de remplacement : 0, LQ/2 ou LQ. De même pour le seuil S_{90} .

Les seuils MAD et la vibrisse supérieure sont tout d'abord insensibles à la limite de quantification pour un pourcentage respectif inférieur à 20% et 25% de données inférieures à la limite de quantification, et décroît au-delà. Le seuil MAD décroît jusqu'à 50% et la vibrisse supérieure jusqu'à 75%. Cela résulte du terme de distance dans ces deux seuils qui décroissent jusqu'à zéro pour les pourcentages cités précédemment (cf. Figure 33). L'augmentation des seuils est ensuite uniquement liée à la valeur de la limite de quantification.

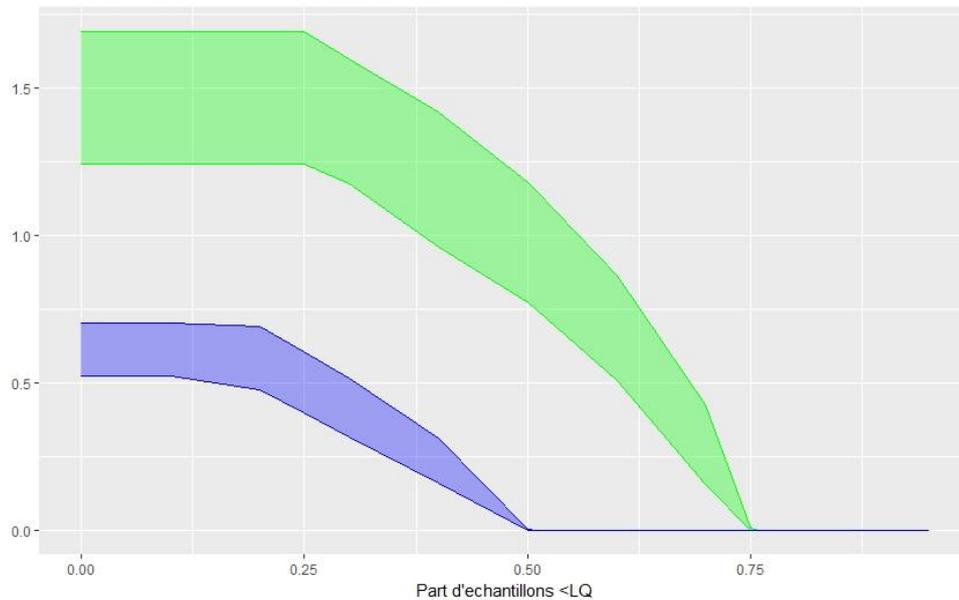


Figure 33 : Résultats minimums et maximums des calculs de la distance interquartile (vert) et du terme MAD (bleu) sur 1000 itérations, considérant une part d'échantillons sous la limite de quantification croissante et un remplacement des valeurs censurées par la LQ.

Dans le cas du remplacement des données inférieures à la limite de quantification par 0, on observe comme précédemment que le 95^{ème} centile n'est pas affecté avec moins de 95% des données inférieures à cette limite de quantification. Il tombe ensuite à 0. Le comportement des seuils MAD et de la vibrisse supérieure est moins homogène que précédemment. Les seuils ne sont tout d'abord pas affectés avec moins de 20% de données inférieures à la limite de quantification. On observe ensuite une croissance rapide des seuils pour la vibrisse supérieure et plus lente pour le seuil MAD. Cette croissance s'explique par le terme de distance de ces seuils qui prend en compte l'écart entre la valeur de remplacement (ici 0) et la valeur de la limite de quantification, augmentant ainsi la distance calculée de la différence entre la limite de quantification et la valeur de remplacement. Par la suite la vibrisse supérieure montre une stabilité entre 30 et 70%. Cela est expliqué par le fait que les valeurs du 1^{er} quartile (0) et du 3^{ème} quartile ne varient pas. On observe ensuite une décroissance des seuils, très rapide pour la vibrisse et plus lente pour le seuil MAD, les deux atteignant 0 à 50% pour le seuil MAD et 75% pour la vibrisse.

II.B. Sensibilité des calculs statistiques du seuil de fond géochimique

Ces résultats sont applicables à tous les cas de substitution par une valeur inférieure à la limite de quantification, comme par exemple par $LQ/2$, cas fréquent en géochimie.

Ainsi, les résultats de ces algorithmes dépendent de la proportion de valeurs inférieures à la limite de quantification et de la valeur de substitution. Bien que le choix de substituer par 0 semble plus conservatif (puisque l'on diminue des valeurs), il faut tenir compte de l'influence du calcul des distances (MAD et distance interquartile) dans les seuils MAD et la vibrisse supérieure.

Enfin, à partir d'un certain pourcentage de valeurs inférieures à la limite de quantification, les trois seuils fournissent des résultats correspondant à la valeur de substitution des concentrations inférieures à la limite de quantification et deviennent donc indicatifs. Ces pourcentages, sont respectivement : 50% pour la méthode MAD, 75% pour la vibrisse supérieure et 95% pour le 95^{ème} centile.

c) Discussion

Ces résultats montrent l'importance de la prise en compte de l'effectif et de la limite de quantification dans l'établissement de valeurs de fond géochimique. L'effectif est un paramètre très limitant car un faible nombre d'échantillons induit une forte incertitude sur les seuils calculés. Or la majorité des études urbaines se heurte à des échantillonnages difficiles et coûteux et ne peuvent généralement garantir un nombre suffisant d'échantillons (>50). Il est donc conseillé d'accompagner les seuils calculés de l'effectif afin que la pertinence de ces seuils puisse être évaluée par le lecteur.

De nombreux paramètres géochimiques montrent une forte censure (données inférieures à la limite de quantification), or il a été démontré ci-dessus que celle-ci influence fortement les seuils calculés. Il est donc important de tenir compte du pourcentage de valeurs censurées. Enfin, la valeur de substitution des données censurées est à examiner et, afin de ne pas surestimer les seuils calculés avec les méthodes MAD ou de la vibrisse supérieure, il est conseillé de choisir la valeur de la limite de quantification comme valeur de substitution. Ne présentant aucune difficulté, l'étude de sensibilité du résultat à la valeur de substitution (des données inférieures à la limite de quantification) est préconisée.

II.B.4. Hétérogénéité d'échantillonnage

L'application à un cas pratique permet d'étudier l'influence d'un échantillonnage préférentiel et donc hétérogène. Le cas d'étude utilisé est le quartier de l'île de Nantes.

a) Pondération des données

Les trois algorithmes sont d'abord appliqués sans prise en compte de l'implantation des données (Figure 34). Là encore le seuil MAD fournit les résultats les plus conservatifs. Les résultats des trois algorithmes sont cohérents avec cet environnement urbain à travers des enrichissements marqués pour l'arsenic, le cuivre, le plomb ou le zinc influencés par l'activité anthropique, et des seuils inférieurs au fond (pédo-)géochimique naturel de la zone pour le baryum, le chrome ou le nickel, ici d'origine géogénique alors que les échantillons concernent plutôt des matériaux anthropiques.

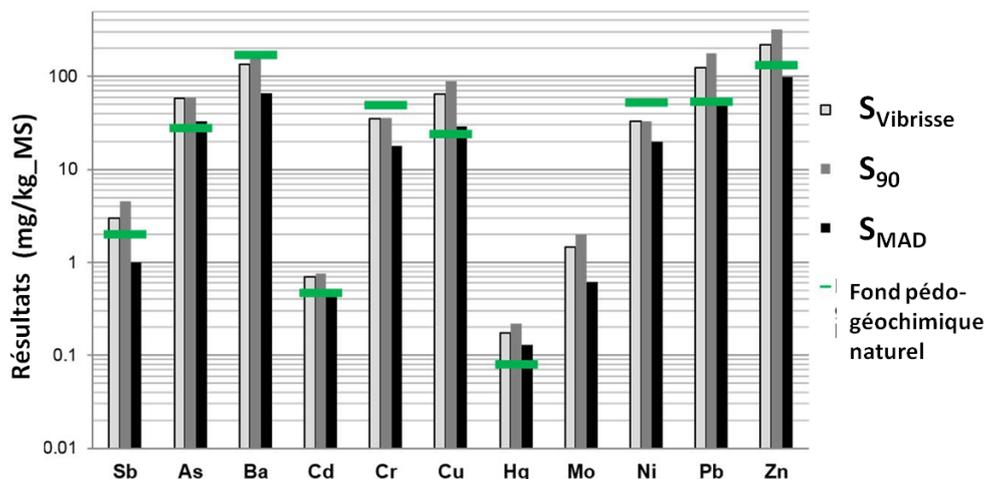


Figure 34 : Résultats de calcul de seuils sur les échantillons du cas d'étude pour les 3 méthodes. Vert : Valeurs du fond pédo-géochimique naturel (calculé sur des alluvions).

Nous avons ensuite étudié l'influence de la pondération pour corriger l'irrégularité spatiale de l'échantillonnage. Celui-ci est ici orienté de façon à étudier les anomalies et tend donc à sur échantillonner les zone d'anomalies potentielles. Les résultats des analyses ont été pondérés par l'inverse du nombre d'échantillons présents dans une maille de 25 ou 100m de côté, ce qui diminue l'influence des échantillons resserrés.

Les résultats (cf. Figure 35) montrent que, quelle que soit la pondération des échantillons, l'ordre des seuils reste identique et le seuil MAD est le plus conservatif. On remarque ensuite une croissance des seuils calculés quand la taille de la maille utilisée pour la pondération augmente. D'après les diagrammes de la Figure 37, cela s'explique par une proportion plus importante de valeurs fortes après pondération. Cette augmentation est due au grand nombre de valeurs faibles mais géographiquement proches dans le jeu de données (Figure 36), qui perdent alors de leur influence avec la pondération. La plus grande proportion de valeurs fortes, dans les données pondérées, influe sur les quantiles qui rehaussent les valeurs seuils calculées. On observe donc l'inverse de ce qui était attendu en supposant que

II.B. Sensibilité des calculs statistiques du seuil de fond géochimique

les anomalies étaient sur-échantillonnées. Enfin, le seuil MAD, fondé sur la médiane et les valeurs proches, apparaît peu sensible à la pondération.

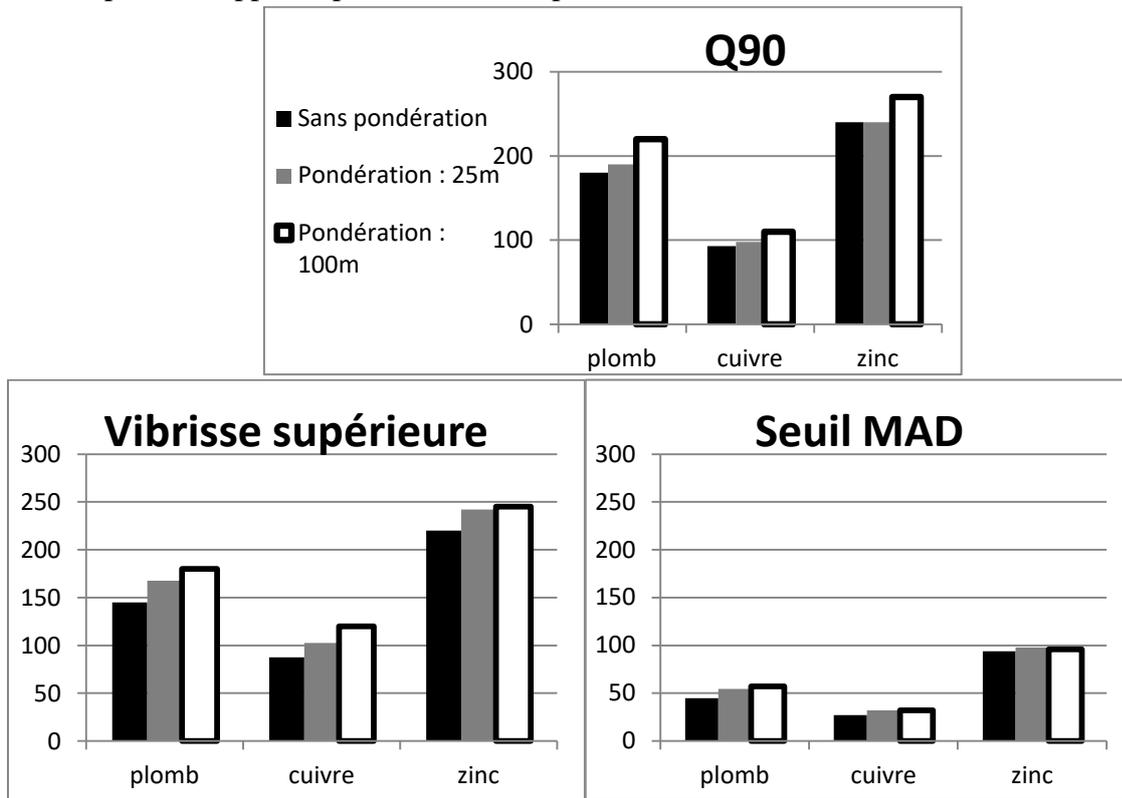


Figure 35 : Résultats de calcul de seuils sur les échantillons du cas d'étude pour les 3 méthodes, sans pondération, ou en pondérant par l'inverse du nombre d'échantillons dans des mailles de 25 ou 100 m de côté. Analyse de plomb, cuivre et zinc en mg/kg de MS.

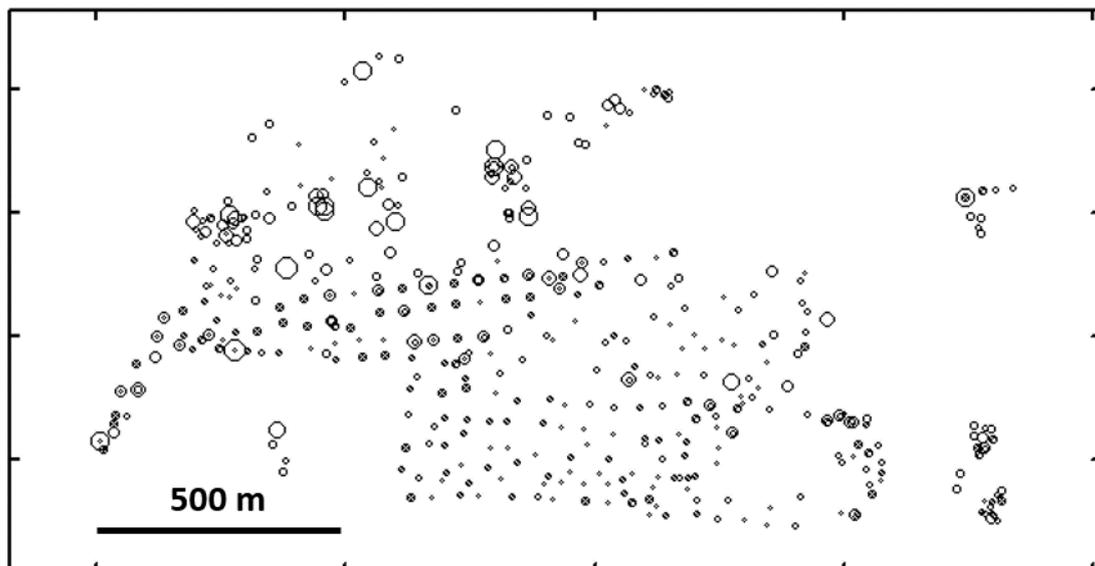


Figure 36 : Carte des teneurs en plomb dans une partie des échantillons collectés.

II.B. Sensibilité des calculs statistiques du seuil de fond géochimique

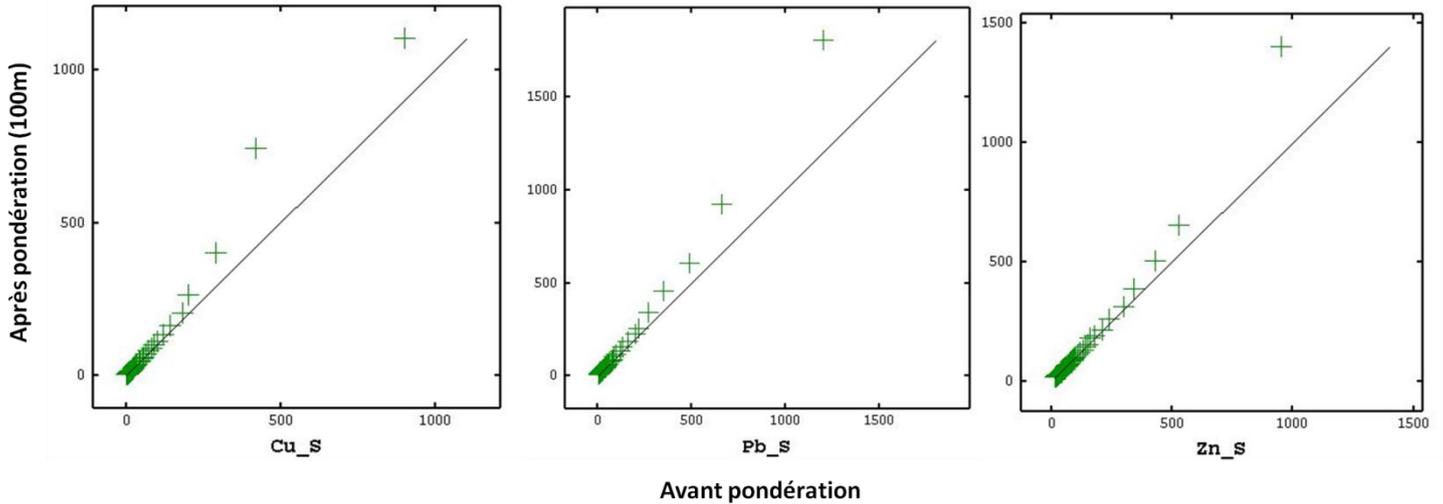


Figure 37: Diagrammes quantiles-quantiles des concentrations en cuivre, plomb et zinc, en mg/kg de MS, avec ou sans pondération par l'inverse du nombre d'échantillons dans une maille de 100m de côté.

Ainsi, dans ce cas, la pondération des échantillons suivant leur proximité géographique, augmente les seuils calculés. Cela atténue l'effet des données proches géographiquement et permet de mieux tenir compte de la géochimie réelle de la zone. Enfin, une sous-estimation des seuils en l'absence de pondération réduit également les possibilités de réutilisation des terres excavées.

b) Regroupement par typologie de sols

La géochimie urbaine résulte de différents déterminants concernant la nature des sols et sous-sols (données géologiques, pédologiques) ou dépendant de l'occupation actuelle ou passée de ces sols. Afin de mieux appréhender la géochimie urbaine et donc le fond pédogéochimique urbain, il est intéressant de tenir compte de ces déterminants comme co-variables. Sur ce principe, ont été développées des typologies de remblais (Le Guern et al., 2016). Les échantillons ont été regroupés selon trois typologies : remblais assimilables naturels, divers ou suspects et alluvions. Les résultats des algorithmes sont illustrés à la Figure 38.

II.B. Sensibilité des calculs statistiques du seuil de fond géochimique

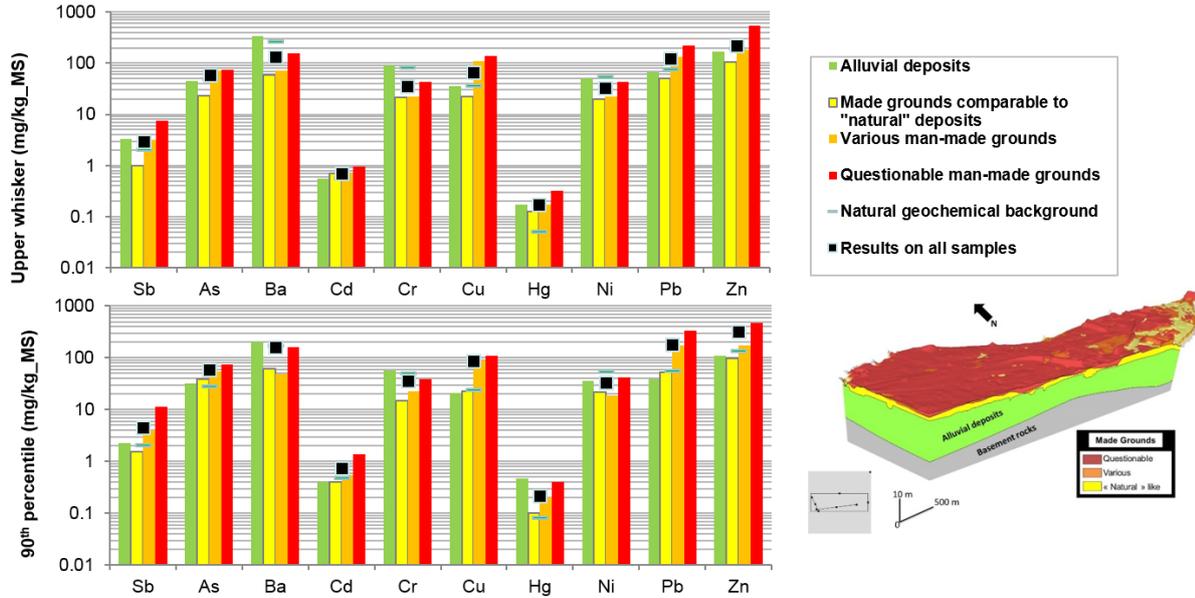


Figure 38 : Résultats de calculs de seuils sur les échantillons du cas d'étude pour les méthodes de la vibrisse supérieure et du 90^{ème} centile, en fonction de la typologie des remblais. (Le Guern et al., 2016)

Trait vert : Valeurs du fond pédo-géochimique naturel (calculé sur des échantillons d'alluvions d'une zone naturelle en amont),

carré noir : seuil calculé sur tous les échantillons.

En bas à droite : Vue 3D du modèle géologique du quartier comprenant la typologie utilisée.

On retrouve des tendances évoquées précédemment, à savoir des seuils plus élevés en baryum, chrome et nickel dans les alluvions (traits verts verticaux) que dans les remblais (traits jaunes, oranges et rouges verticaux). Inversement les seuils pour le plomb, le cuivre et le zinc montrent bien un enrichissement des remblais vis-à-vis des alluvions. On note toutefois des seuils plus faibles dans les remblais assimilables naturels que dans les alluvions. Cela s'explique ici par la nature de ces remblais, composés de sables de Loire triés lors de leur mise en place et présentant moins de particules fines, riches en métaux, que les alluvions.

Ainsi la prise en compte des typologies de sols aux propriétés géochimiques différentes permet de préciser des fond pédo-géochimiques différents. En revanche, l'intégration de ces données descriptives partitionne les données, ce qui diminue le nombre d'échantillons des groupes sur lesquels sont calculés les seuils, avec les conséquences vues précédemment. Plutôt que de limiter l'intégration de données descriptives pourtant cruciales dans l'établissement d'un fond pédo-géochimique urbain pertinent, il apparaît préférable d'examiner le regroupement des classes.

II.B.5. Discussion

Le calcul de seuils statistiques pour la définition du fond pédo-géochimique nécessite la prise en compte de différents paramètres pouvant influencer le résultat. Une étude préalable des données, comprenant notamment l'implantation des échantillons, leur nombre et la part de données non quantifiées, est donc nécessaire.

Le 95^{ème} centile considère par définition que le fond représente 95% des données. Cette définition correspond plutôt à la prise en compte d'une marge d'erreur pour un échantillonnage orienté évitant les zones contaminées. Cela correspond notamment bien au type d'échantillonnage mis en place par le BGS pour la réalisation de la G-Base. De plus, son insensibilité aux valeurs inférieures à la limite de quantification peut permettre une première approche dans des cas spécifiques de forte censure (données inférieures à la LQ). Il reste toutefois déconseillé dans des études urbaines où l'échantillonnage hors anomalie est complexe et plus incertain. Le quantile 90 est, lui aussi déconseillé, car si la marge d'erreur est ici plus large (10%) elle reste néanmoins purement arbitraire et potentiellement insuffisante dans le cas d'une collecte de données urbaines où l'anomalie peut représenter, comme le montre l'exemple de Nantes, un pourcentage plus important des données.

La vibrisse supérieure permet de considérer des anomalies pouvant représenter jusqu'à 25% des données totales, ce pourcentage correspond plus à un échantillonnage non orienté dans des zones peu impactées ou à un échantillonnage orienté vers des zones présumées sans anomalies en milieu urbain (en gardant une marge d'erreur conséquente). Cette méthode est cohérente avec les échantillonnages réalisés dans le cadre du RMQS (Réseau de Mesure de la Qualité des Sols) ou du programme de fond géochimique Finlandais (Jarva, 2010). Les seuils relativement hauts favorisent une approche économique du fond pédo-géochimique visant à la réutilisation de volumes plus importants de terres excavées. Enfin, la limite de quantification est à étudier du fait de la sensibilité du seuil à partir de 25% de censure. Le remplacement des données par une valeur inférieure à la limite de quantification est à proscrire (ou du moins à valider par une étude de sensibilité) du fait de l'impact sur la distance interquartile. Cela conduit en effet dans certains cas à une surestimation du fond, qui n'est pas conservatif pour l'environnement.

Enfin, le seuil MAD peut différencier fond et anomalie, dans certaines conditions, jusqu'à 50% d'anomalies. Il permet d'intégrer des données issues de campagnes d'échantillonnage plus orientées vers des sites contaminés comme les diagnostics de pollution. Il apporte donc une solution intéressante pour la valorisation de ces données. Les seuils calculés étant relativement bas (par rapport aux quantiles ou à la vibrisse), il convient plutôt à une approche conservative au détriment de la réutilisation des terres. Comme la vibrisse supérieure, il est relativement sensible aux données inférieures à la limite de quantification (à partir de 20% de données censurées), et la valeur de remplacement ne doit pas être inférieure à cette limite car cela conduit à une surestimation du fond.

II.B. Sensibilité des calculs statistiques du seuil de fond géochimique

Ainsi, le choix de ces méthodes statistiques repose en partie sur l'évaluation de la proportion des anomalies et du fond. Dans ce sens, nos conclusions rejoignent celles de Reimann (2005). Il démontre en effet que pour une faible proportion d'anomalies, la vibrisse supérieure peut fournir des valeurs seuils adéquates. En revanche, tout comme nous l'avons également présenté, une proportion relativement importante d'anomalies ne pourra être séparée que par le critère MAD.

Cette proportion d'anomalies dépend alors du type de terrain étudié et des caractéristiques de l'échantillonnage. Dans notre cas l'échantillonnage orienté vers les zones potentiellement polluées concerne la grande majorité des données, ce qui présente l'avantage de mieux localiser et identifier les anomalies, mais est moins représentatif de l'ensemble de la zone étudiée et doit être pris en compte. Un échantillonnage complémentaire pourrait intervenir pour quelques échantillons en complément des données collectées. Un échantillonnage aléatoire géographiquement stratifié par exemple, permettrait de garantir une meilleure représentativité de secteurs moins échantillonnés (hors friches industrielles notamment).

Enfin, la notion d'anomalie et les processus ou activités qui y sont associés varient suivant l'échelle de travail. En effet une zone industrielle peut apparaître comme une anomalie à l'échelle de la ville mais faire partie du fond à l'échelle d'un quartier. La notion d'anomalie se rapporte donc à l'étendue de la zone étudiée. Ne peuvent être considérés comme anomalies que des phénomènes ayant une emprise limitée (par exemple inférieure à 25% de la zone). L'échantillonnage peut modifier (sur ou sous-représentation) ce rapport dans les données finales.

Recommandations

La première étape du calcul d'une valeur de fond pédo-géochimique anthropisé à partir de données collectées, consiste à vérifier la représentativité et la distribution de ces données. Pour cela, l'étude de la structure du sol est déterminante et peut être réalisée via la collecte de sondages géotechniques, l'apport de données cartographique (carte géologique, pédologique) et la prise en compte des descriptions de carottages. L'occupation des sols actuelle et passée est une source d'informations importante pour la représentativité des données. Elle renseigne sur le degré d'hétérogénéité structurelle ou géochimique du sol et du sous-sol.

Des zones homogènes peuvent éventuellement être identifiées. La représentativité des échantillons à l'intérieur de ces zones dépend alors de la pertinence de l'échantillonnage (échantillonnage du fond ? Surestimation des anomalies ? ...). Si les échantillons sont jugés représentatifs, alors les seuils peuvent être calculés, en précisant l'effectif et l'emprise de la zone concernée par la valeur. Dans les cas d'effectifs faibles (moins de 30 échantillons) il est conseillé d'insister sur le caractère purement indicatif de la valeur fournie et sur son incertitude.

II.B. Sensibilité des calculs statistiques du seuil de fond géochimique

Lorsque la représentativité des données ne peut être assurée, il est préférable de ne pas fournir de valeurs seuils puisque celles-ci ne pourraient être associées à une emprise géographique claire.

Enfin l'utilisation de ces méthodes statistiques de détermination d'une valeur seuil du fond pédo-géochimique repose sur l'hypothèse que le fond représente une proportion bien plus importante des données que les anomalies. Cette hypothèse n'est pas toujours vérifiée suivant l'échelle de travail et le type d'échantillonnage utilisé (cf. exemple de l'Ile de Nantes).

Partie III : PRISE EN COMPTE DE L'HETEROGENEITE DES SOLS URBAINS

Dans la partie précédente il a été démontré que si les critères statistiques ne garantissent pas, dans tous les cas, une valeur seuil correspondant à la borne supérieure ou à un majorant du fond pédo-géochimique, ils présentent néanmoins des valeurs simples à obtenir susceptibles d'être utilisés comme référence. Or chaque critère ne fournit qu'une seule valeur seuil par paramètre pour l'ensemble des données traitées. Dans le cas où toutes les données d'une ville sont traitées ensemble, afin de garantir la conservation de la qualité des sols, il est préférable de favoriser les seuils les plus conservatifs. Cela limite toutefois les possibilités de valorisation des terres excavées. La solution proposée précédemment est de séparer les données suivant une typologie des sols ayant un sens géochimique. Cette typologie est basée sur les déterminants de la géochimie en milieu urbain. De cette façon, plusieurs niveaux de seuils peuvent-être fournis et rattachés à des entités géographiques cohérentes.

Dans un premier temps il est donc important de bien définir ces déterminants de la géochimie urbaine et les données cartographiques afin d'en tenir compte dans les développements méthodologiques.

III.A. Déterminants de la géochimie des sols urbains

La géochimie des sols en milieu urbain dépend principalement de deux types de déterminants :

- la **nature des matériaux** constituant le sol ou le sous-sol (Argyraki et Kelepertzis, 2014 ; Birke et Rauch, 2000 ; Rodrigues et al., 2009). Cela comprend la roche mère, le type de sol mais aussi la nature des remblais éventuellement mis en place (Le Guern et al., 2016 ; 2017).

- les **pressions anthropiques** exercées sur les sols (Cicchella, 2008 ; Madrid et al., 2006 ; Pouyat et al., 2007 ; Rodrigues et al., 2009 ; Le Guern et al., 2016 ; 2018), incluant les diverses activités actuelles ou passées (industrielles, agricoles, domestiques, transport...) ayant pu se développer sur l'aire urbanisée.

III.A.1. Nature des matériaux

La géochimie des sols varie selon l'origine et la nature des matériaux qui le constituent (Appleton et al., 2012, 2013 ; Argyraki et Kelepertzis, 2014 ; Négrel et al., 2018).

Une distinction peut être faite entre les matériaux **naturels** et **anthropiques**. Cette différenciation concerne l'origine du matériau et sa mise en place. Ainsi, trois catégories sont distinguées (Ford, 2010) :

III.A. Déterminants de la géochimie des sols urbains

- les matériaux d'origine naturelle et déposés par des processus naturels (ex : alluvions) ;
- les matériaux d'origine naturelle mais déplacés par l'Homme (ex : sables utilisés en remblais) ;
- les matériaux d'origine anthropique mis en place par l'Homme (ex : résidus de fonderie).

Si la distinction de l'origine des matériaux est relativement aisée, il n'en est pas toujours de même pour le processus de mise en place.

Ainsi, dans le cas de la présence de matériaux anthropiques au sein de matériaux d'origine naturelle (exemple : des morceaux de briques dans du sable), le pourcentage volumique d'artéfacts anthropiques est généralement utilisé pour faire la distinction entre les types de sols (WRB, 2015 ; Girard et Baize, 2009). Ce pourcentage n'est cependant pas suffisant pour évaluer son potentiel polluant. C'est pourquoi Le Guern et al. (2016) ont proposé un ensemble de termes de description des artéfacts destinés à évaluer le potentiel polluant intrinsèque des remblais. Ces termes peuvent être appliqués plus largement aux sols anthropisés (Tableau 4). Cette typologie des sols est utilisée dans la suite des travaux.

Remblais Suspects	Remblais divers
pyrite	brique
mâchefer	dalle
fonderie	pavé
scorie	verre
asphalt	grave
bitume	béton
goudron	gravat
cencre	maçonnerie
enrobé	bloc
violet	remblais
suie	
charbon	
noir	

Tableau 4 : Différenciation des types de remblais. Les remblais assimilables naturels sont caractérisés par l'absence de tous les termes cités dans le tableau (Le Guern et al., 2016).

III.A.2. Pressions anthropiques

La qualité géochimique des sols urbains est largement déterminée par le cumul des pressions anthropiques (Lark et Scheib, 2013 ; Li et al., 2015 ; Lu et Bai, 2010 ; Thackway et Specht, 2015). Ces pressions regroupent l'ensemble des impacts de l'homme sur la géochimie, ne comprenant pas la modification physique ou morphologique du matériau. Cela intègre en particulier l'apport en contaminants.

Ainsi différents types d'activités : industrie, domestique, agricole, transport, ... sont susceptibles de contaminer les sols. L'industrie est susceptible de contaminer les sols à travers les processus de fabrication, de nettoyage ou de refroidissement (Li et al., 2004). Les contaminations associées aux activités domestiques sont, quant à elles, plutôt liées aux matériaux (ruissèlement sur les toitures) ou au chauffage domestique (Göbel et al., 2007). Les transports génèrent également des gaz et des particules qui s'accumule ensuite dans les sols proches (Gasperi, 2006 ; Shaheen, 1975). Enfin, l'agriculture peut-elle aussi générée des contaminations à travers l'utilisation d'engrais ou d'amendement, ou de fongicides tels que la bouillie bordelaise utilisée pour la vigne (Baize, 2000).

Les types de contaminants générés par l'ensemble de ces activités sont très variés. On peut citer quelques exemples parmi les activités industrielles et de service. Les dépôts de liquides inflammables (DLI) sont susceptibles de générer des contaminations en hydrocarbures dans les sols. Les ateliers de métallurgie et les fonderies peuvent générer des contaminations en métaux comme le cuivre, le plomb ou le zinc (Lark et Scheib, 2013 ; Li et al., 2015 ; Lu et Bai, 2010). Il faut également prendre en compte les usines de fabrication d'engrais, de textile, de savon, mais aussi les décharges et les sites de stockages de minerais ou de résidus de fonderies.

Les contaminations liées aux pressions anthropiques peuvent être relativement ponctuelles. C'est le cas en particulier pour nombre d'activités industrielles où les sources de contamination sont localisées sur des emprises spatiales limitées (ex. dépôt de carburant) (Lark et Scheib, 2013 ; Li et al., 2015 ; Lu et Bai, 2010). Les contaminations engendrées peuvent aussi être diffuses. C'est le cas notamment suite à des rejets atmosphériques (usine, transport, chauffage urbain...) ou d'épandages sur de grandes surfaces (ex. en agriculture) (Baize, 2000).

Comme pour la nature des sols et proches sous-sols, les contaminations générées dépendent de plusieurs paramètres :

- Le type d'activité : un dépôt d'hydrocarbures n'aura pas les mêmes impacts géochimiques qu'une fonderie ou qu'un atelier.
- L'intensité de la pression. Pour reprendre l'exemple de la cuve d'hydrocarbures, celle-ci n'aura pas le même impact suivant sa dimension, sa fréquence d'utilisation ou son entretien.

III.A. Déterminants de la géochimie des sols urbains

- La durée. Plus l'activité potentiellement polluante durera et plus l'impact sur la géochimie des sols risque d'être important, que ce soit en termes de concentration ou d'étendue.

L'historique d'occupation des sols permet de retrouver les successions d'activités qui se sont exercées et apportent des éléments de connaissances sur les pressions anthropiques exercées (Franck-Neel, 2015 ; Le Guern et al, 2016, 2011 ; Le Guern et al, en préparation). Ainsi, la présence des polluants potentiels associés aux activités industrielles et de services qui se sont succédées dans le temps est décrite dans certains inventaires historiques urbains (IHU), par exemple sur Saint Nazaire et sur les quartiers nantais (Le Guern et al., 2011, 2016, 2017). La matrice activités-polluants récemment publiée (2018) (<http://ssp-infoterre.brgm.fr/bd-activipoll>) permet de faire le lien entre les activités exercées et les contaminants potentiels.

L'occupation des sols peut également être déterminée par photo-interprétation (Ban et al., 2015 ; Li et al., 2015), cela afin d'obtenir rapidement des informations sur de grandes aires urbaines. Son application nécessite toutefois des images aériennes ou satellites de qualité qu'il est parfois difficile d'obtenir, et est indisponible pour les occupations passées.

III.A.3. Influence croisée des déterminants sur la géochimie : hétérogénéité des sols urbains

Ces deux déterminants (nature des matériaux et pressions anthropiques) ont une importance variable suivant la substance considérée et la profondeur. En effet, la présence dans les sols de métaux fortement marqués par l'activité anthropique tels que le cuivre ou le plomb est ainsi plutôt déterminée par l'occupation des sols (bien que la présence de matériaux anthropiques tels que des remblais industriels ait aussi un impact), alors que la présence de métaux d'origine plutôt géogénique (comme le nickel ou le chrome sur l'Ile de Nantes) est plus influencée par la nature des sols en lien avec le matériau parental.

Cela contribue à l'hétérogénéité des sols urbains (Béchet et al, 2009 ; Rodrigues et al., 2009) à différentes échelles (ville, quartier, îlot).

III.A.4. Données descriptives courantes de la nature et de l'occupation des sols : échelle de la ville

La géochimie des sols urbains étant influencée par la nature des sols et leur occupation par les activités humaines, nous avons fait un point sur les différents jeux de données existants. Dans un premier temps nous avons choisi de nous intéresser aux données facilement accessibles à l'échelle de la ville.

Les sols naturels dérivant des roches sous-jacentes, l'information géologique devrait

III.A. Déterminants de la géochimie des sols urbains

permettre d'approcher le potentiel d'anomalie et d'hétérogénéité géochimique des sols naturels avant impact et pression anthropique, comme l'a montré Metotrass (Le Guern et al, 2013). La carte géologique informe de la présence de grandes zones de remblais et est disponible à l'échelle du 1/50 000 sur l'ensemble du territoire national. Une version numérique vectorisée de la carte de Nantes a été mise à notre disposition par le BRGM. La carte géologique actuelle ne différencie cependant pas les types de remblais. La carte pédologique peut aussi être une source d'informations importante. Toutefois il n'existe pas de méthodologie de cartographie des sols urbains et ce genre de donnée n'est pas disponible pour les villes (Béchet et al, 2017). Des cartographies ou des extrapolations peuvent exister dans les zones périurbaines, comme c'est le cas pour la ville de Nantes. La carte pédologique de Loire-Atlantique, carte IGCS 1/250 000, intègre ainsi des extrapolations sur les zones périurbaines.

L'occupation des sols actuelle et récente est fournie au format vecteur par Corine Land Cover (CLC) à l'échelle du 1/100 000. Ces données sont gratuites et en libre-service sur internet. Toutefois, l'emprise minimale d'un polygone de Corine Land Cover est de 25 ha, ce qui est peu adapté à un traitement précis des zones urbaines. La cartographie du projet Copernicus fournit un niveau de détail supérieur pour un ensemble d'aires urbaines européennes. L'échelle est de 1/10 000 et, là encore les données sont gratuites et en libre-service sur internet. Ces cartographies sont obtenues par photo-interprétation d'images aériennes ou satellites. Elles sont structurées en un ensemble de polygones comprenant différents niveaux de codes numériques correspondent à différents types d'occupation de sols allant du plus large au plus précis. Un niveau de code supplémentaire permet à Copernicus un gain de précision par rapport à Corine Land Cover.

Les deux cartographies d'occupation des sols précédentes ne tiennent toutefois compte que de l'occupation actuelle ou récente. La version accessible la plus ancienne de CLC, accessible, date de 1990. Pour les pressions anthropiques passées, la base de données Basias apporte des informations ponctuelles sur les activités industrielles et de services potentiellement polluantes de 1850 à nos jours. Ces données sont également accessibles sur internet et un export nous a été fourni par le BRGM de Nantes. Ces données ponctuelles (adresse ou centre du site concerné) sont liées à une table de données regroupant les activités potentiellement polluantes et les dates correspondantes ; elles comprennent également différents codes, dont les codes NAF permettant de lier l'activité à des polluants potentiels (Matrice activité-polluant, 2018).

Ces données présentent toutes l'avantage d'être gratuites et accessibles. Toutefois, l'échelle du 1/50 000 ou du 1/100 000 est mieux adaptée à de larges territoires. Dans le cadre de travaux plus précis, sur l'emprise d'un quartier par exemple, cette échelle de validité présente des limites et d'autres données d'échelle plus fine doivent être acquises ou traitées.

III.A.5. Données descriptives spécifiques au cas d'étude à l'échelle du quartier

Dans un deuxième temps, nous nous sommes intéressés à des données descriptives plus précises, mais moins fréquemment disponibles.

Le premier jeu de données est un modèle géologique en trois dimensions du sous-sol urbain constitué à partir de descriptions des sondages collectés (Le Guern et al, 2016). Celui-ci permet de tenir compte de la géologie à différentes profondeurs et d'intégrer une typologie de remblais lors de sa construction. Les données du modèle 3D peuvent ensuite être utilisées sous forme de cartes d'affleurement à différentes profondeurs avec une échelle de validité de l'ordre de 1/5 000 (cf. Figure 39).

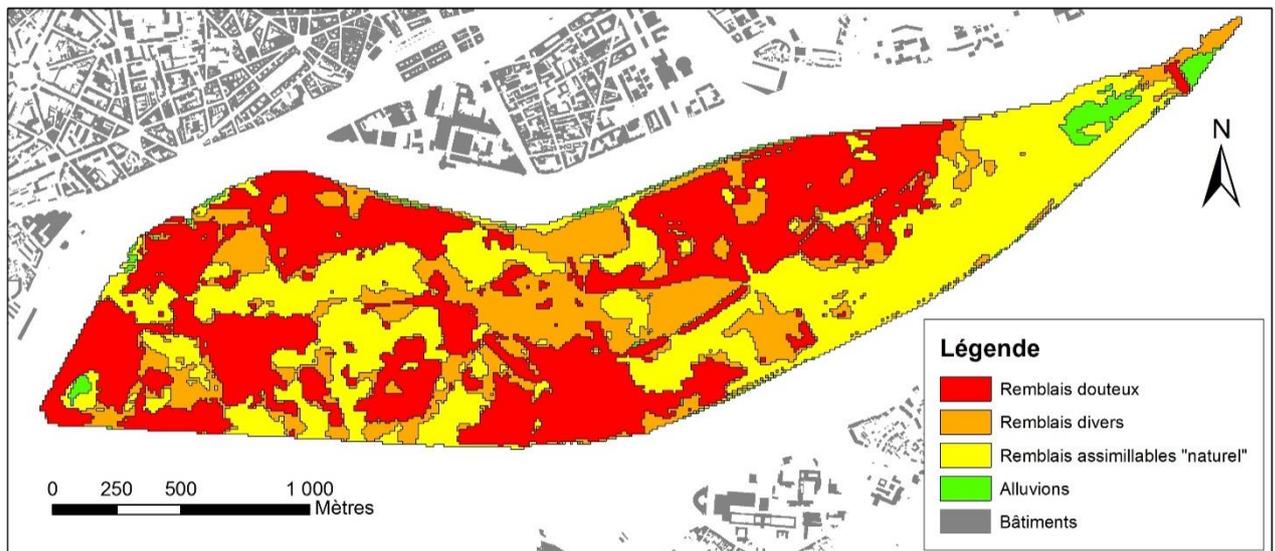


Figure 39 : Exemple de carte d'affleurement à 1 m de profondeur réalisée à partir du modèle 3D du sous-sol suivant différents types de remblais.

Afin de tenir compte avec plus de précision de l'occupation des sols passée, un Inventaire Historique Urbain (IHU) peut également être exploité. Celui-ci se base sur le dépouillement d'archives afin de localiser précisément les activités industrielles et de services potentiellement polluantes passées (de 1850 à nos jours). Ce travail, mis en place par le BRGM sur plusieurs quartiers nantais (Le Guern et al, 2016, 2017), permet une cartographie de l'emprise des sites identifiés mais aussi des intrasites (lieux privilégiés de contamination ponctuelle à l'intérieur des sites) et de les relier à des polluants potentiels, là aussi avec une précision de l'ordre du 1/5 000 (cf. Figure 40).

III.A. Déterminants de la géochimie des sols urbains

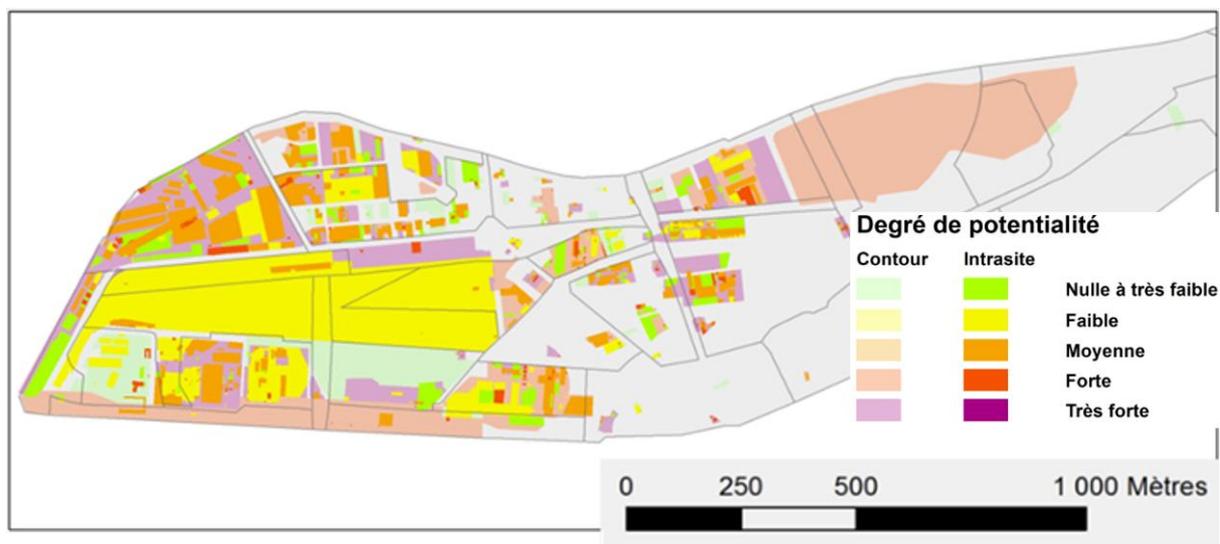


Figure 40 : Degré de potentialité aux hydrocarbures lié aux sources potentielles de pollutions localisées au sein des sites industriels et activités de service recensés – exemple de représentation possible (cas de l’Ile de Nantes) (Le Guern et al, 2017)

Toutefois, l’Inventaire Historique Urbain ne tient compte que des activités industrielles et de services passées les plus potentiellement polluantes. Or, dans le cas de quartiers péri-urbains, d’autres activités potentiellement polluantes comme l’agriculture doivent être considérées. Pour cela, un traitement de photos aériennes anciennes est nécessaire. Cette photo-interprétation permet de tenir compte des types de culture et de leur durée (cf. Figure 41). Je l’ai mené en parallèle à l’acquisition des données de terrain (cf § I.B.2 et Annexe 4).

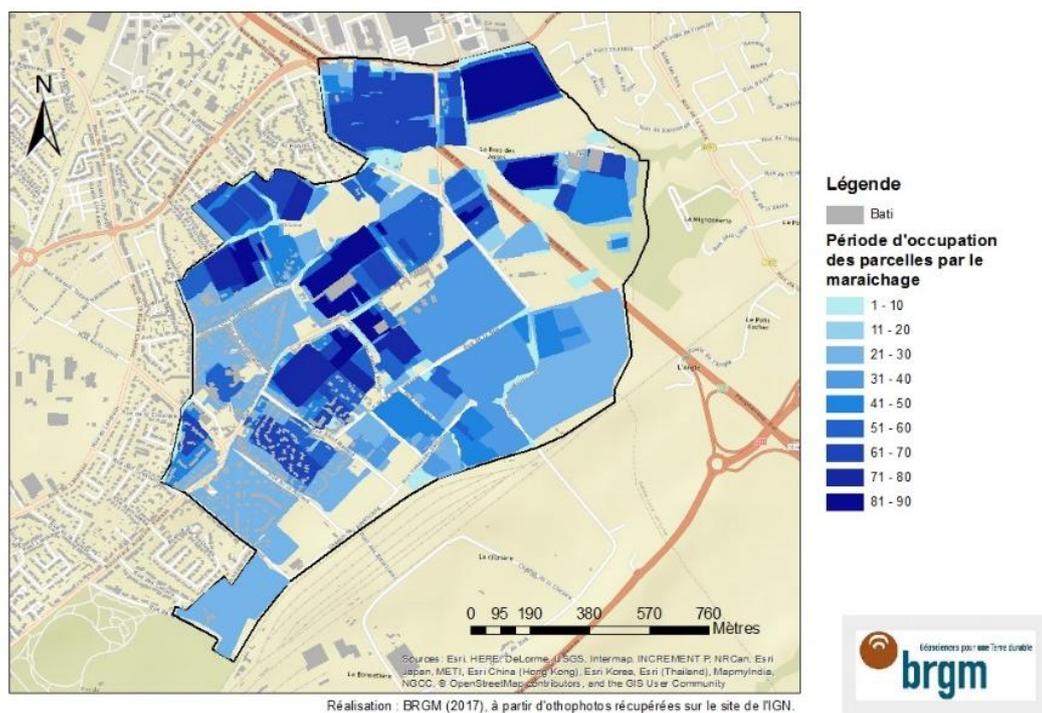


Figure 41 : Photo-interprétation de l’historique d’occupation des sols par des activités maraîchères sur un des quartiers nantais en années.

III.B. Proposition de méthodologies pour la prise en compte des déterminants de la géochimie urbaine

Nous avons démontré précédemment que l'effectif et le plan d'échantillonnage peuvent avoir une influence majeure sur les seuils statistiques calculés pour déterminer le fond pédo-géochimique. Les données géochimiques sont toutefois difficiles à obtenir et il est nécessaire de trouver d'autres données plus commodes d'accès et pouvant fournir des informations complémentaires.

Parmi ces données figurent les déterminants de la géochimie des sols urbains, liés à la nature des matériaux ainsi qu'aux pressions anthropiques. Suivant cette démarche, Le Guern et al. (2016) ont mis en évidence la relation entre une typologie de remblais et des données géochimiques d'un quartier de centre urbain et montré l'intérêt de la prise en compte de la typologie pour le calcul de valeurs de fond pédo-géochimique en vue de la réutilisation des terres excavées. Cela conduit à autant de valeurs de fond pédo-géochimique que de typologies de référence retenues, et ce pour chaque substance. Les terres appartenant aux fonds les plus bas peuvent être réutilisées sur des zones où les valeurs de fond sont plus élevées mais pas l'inverse. Cela permet de préserver la qualité géochimique des sols en place tout en favorisant la réutilisation des terres excavées. Dans une optique opérationnelle, il apparaît utile de rattacher ces différentes valeurs de fond à des zones suffisamment grandes pour faciliter la valorisation des terres excavées.

Dans la mesure où le fond pédo-géochimique doit être associé à une emprise géographique claire (Baize, 2008), nous cherchons à définir des entités géochimiquement cohérentes. Comme cela a été proposé dans Métotrass (Le Guern et al, 2013), il s'agit de prendre en considération également le degré d'hétérogénéité de ces entités. Dans le cas des aires urbaines, les matériaux sont fortement remaniés et le degré d'hétérogénéité peut être important (Pickett et Cadenasso, 2009 ; Greinert, 2015).

On se propose alors de développer une méthodologie permettant de découper le territoire urbain en entités géographiques cohérentes afin de tenir compte des fortes hétérogénéités du milieu et éviter de regrouper des données de zones de géochimie trop contrastée.

Nous envisageons **deux méthodes de découpage** du territoire suivant deux hypothèses différentes (Figure 42).

La première méthode s'appuie sur l'hypothèse d'un découpage du territoire en ne tenant compte que des données descriptives et de retours d'expérience. On parlera alors d'un **découpage suivant le potentiel de présence d'anomalie géochimique**. Cela permettrait d'avoir une cartographie a priori du territoire et d'identifier des zones potentielles pour des échantillonnages complémentaires aux données collectées. Cette approche suit la logique des développements actuels menés dans le cadre du groupe de travail national et des différents

III.B. Proposition de méthodologies pour la prise en compte des déterminants de la géochimie urbaine

projets français sur le fond pédo-géochimiques urbains. En effet, l'objectif est alors de fournir une cartographie a priori du potentiel de contamination des sols avant même la collecte d'échantillons ou de données géochimiques. Cette approche se base sur des recherches bibliographiques et des retours d'expérience afin de définir les notes et pondérations appliquées lors de la segmentation du territoire à l'échelle d'une région (Bastian, 2000 ; Recatalá et Sacristán, 2014 ; Kheiret al., 2014 ; Reimann et al., 2019) ou d'une ville (Effland et Pouyat, 1997 ; Schindelbeck et al., 2008 ; Sobocká et Fulajtár, 2009).

La seconde méthode se fonde sur l'hypothèse inverse. Elle part du principe que les résultats d'analyses géochimiques collectés sont les informations principales et que l'on doit donc organiser le découpage du territoire en fonction du classement de ces données. Le découpage final est alors le résultat d'une **classification des échantillons en classes géochimiques homogènes**.

Le détail de chaque méthodologie et les résultats correspondants sont décrits dans les sous-chapitres suivants.

Enfin, dans un souci opérationnel, il est important de vérifier si le découpage utilisé pour définir les entités géographiques peut s'appuyer sur des critères physiques visibles sur le terrain. En effet, en vue de la valorisation de terres excavées, un site ou une partie de celui-ci doit pouvoir être clairement lié à une typologie d'entités géographiques afin de recevoir des terres adaptées. La cartographie obtenue doit donc correspondre à des objets identifiables sur le terrain (en lien avec l'occupation des sols actuelle ou la nature des sols).

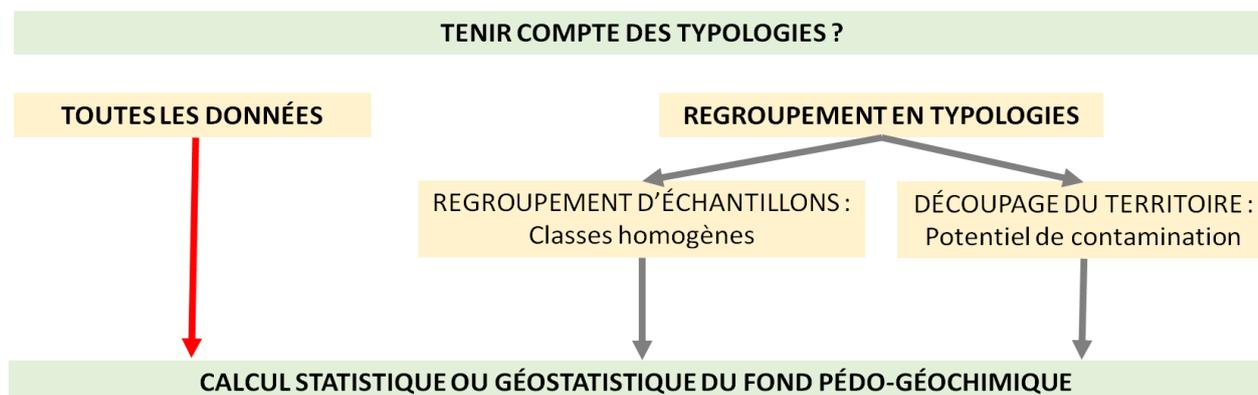


Figure 42: Schéma conceptuel de la constitution de valeurs de fond pédo-géochimique présentant l'importance de la prise en compte d'une typologie de sols urbains, associée à la gestion des anomalies

III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques

III.C.1. Principe général et échelle de travail

Si la notion d'entité géochimique cohérente mise en place dans le cadre du projet METOTRASS (Le Guern et al, 2013) s'applique en milieu urbain, la méthodologie de construction nécessite d'être adaptée à la spécificité et à l'hétérogénéité des sols urbains. En effet, METOTRASS ne prend en considération que l'influence du substrat géologique. En milieu urbain, il faut prendre en compte également les autres déterminants de la géochimie, à savoir pour la nature des matériaux en place a) le mécanisme de mise en place (naturel/anthropique) et b) l'origine des matériaux (naturels/anthropiques).

Le découpage du territoire en entités géochimiques cohérentes s'appuie sur la recherche d'ensembles géographiques classés selon leur potentiel de présence d'anomalies géochimiques. La méthode développée s'inspire des travaux réalisés sur la perturbation des sols (Franck-Néel, 2015 ; Zeilhofer et Topanotti, 2008) et de ceux d'exposition des nappes d'eau souterraine (méthode DRASTIC) (Ahmed, 2015 ; Krishna, 2014). Ce type de découpage est également appliqué à d'autres domaines tels que l'imperméabilisation des sols (Artmann, 2015).

Deux échelles de travail

Nous avons choisi de travailler à deux échelles différentes. La première est la ville, afin d'avoir un aperçu des grandes tendances. La représentation sera simplifiée du fait de l'utilisation des données disponibles. Elle vise en outre à identifier les zones selon leur potentiel d'anomalies géochimiques. Par exemple, les zones à fort potentiel d'anomalies sont susceptibles de générer d'importants volumes de terres excavées de mauvaise qualité.

La deuxième échelle est celle du quartier en ayant recours à des données d'échelles plus fines. Elles ne sont pas disponibles de manière systématique et nécessitent un travail préalable pour leur constitution. Plus fidèle à la réalité par un niveau de détail géographique plus fin (la simplification liée à l'échelle est moins importante pour des échelles plus fines), cette échelle est mieux adaptée à des projets d'aménagement.

Données source

La méthode consiste à sélectionner des descripteurs (données cartographiques) cohérents vis-à-vis du sujet et de l'échelle de travail. On utilisera alors des données les moins redondantes possibles et avec une échelle suffisante pour le niveau de détail souhaité. Pour plus de cohérence, il est préférable de travailler avec des données d'échelle similaire.

III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques

Le choix des descripteurs s'appuie sur les données spatialisées disponibles pour les deux déterminants identifiés. Ces données sont décrites au § III.A.4.

Le Tableau 5 récapitule l'ensemble des données géographiques retenues et leur échelle de validité.

TYPE DE DESCRIPTEUR		DESCRIPTEUR	ECHELLE	DISPONIBILITE
PRESSION	Actuelle	Corine Land Cover	1 / 100 000	Nationale
		Copernicus	1 / 10 000	Grandes agglomérations
	Passée	BASIAS	(Ponctuelle) 1 / 25 000	Nationale
		Inventaire Historique Urbain	1 / 5 000	Quelques quartiers
NATURE		Carte géologique	1 / 50 000	Nationale
		Modèle 3D	1 / 5 000	Quelques quartiers

Tableau 5: Récapitulatif des descripteurs disponibles, de leur échelle de validité et de leur disponibilité

Méthode de croisement

La méthodologie de découpage des entités s'appuie sur une analyse multi-critères. Celle-ci consiste à noter et pondérer les données sources suivant l'indicateur que l'on cherche à cartographier. Elle représente alors une interprétation des données d'origines afin de leur donner un sens (ici géochimique).

Cette méthode d'analyse multi-critères se développe actuellement dans de nombreux domaines grâce aux systèmes d'informations géographique (SIG) et aux outils de gestion de données (Huang, 2010 ; Malczewski, 2006, 2010 ; Romano et al., 2015 ; Velasquez et Hester, 2013). La méthodologie se décompose en 3 étapes : la sélection des données descriptives, leur interprétation (notes) afin de leur donner un sens suivant l'objectif visé et le croisement en un indicateur.

Interprétation des descripteurs en fonction du potentiel de contamination

Une fois les descripteurs choisis, il est nécessaire de les interpréter en termes de géochimie. Dans ce cadre, nous avons choisi de leur attribuer des notes suivant le potentiel de présence d'une anomalie géochimique (anomalie naturelle ou contamination anthropique). Les données étant organisées selon des codes (type de substrat géologique, type d'occupation des sols), nous attribuons par code de chaque descripteur une note de 0 à 2 selon le potentiel d'anomalie croissant (**0 : nul, 1 : moyen, 2 : fort**).

L'interprétation des descripteurs est réalisée à partir de recherches bibliographiques et de retours d'expériences sur les liens entre les descripteurs et la présence d'anomalie. Ce volet est détaillé au chapitre III.B.2 et aux annexes 5, 6 et 7.

Dans le cas du plomb par exemple, le code « zone industrielle » de Corine Land Cover se verra attribuer la note de 2, et les « espaces verts urbains » la note 0.

A cette étape nous pouvons donc établir des cartes décrivant le potentiel de présence d'anomalie géochimique, liée à la nature du sol ou aux pressions anthropiques.

Croisement des descripteurs

Pour croiser les descripteurs urbains et obtenir des indices de potentiel de présence d'anomalies géochimiques, nous nous appuyons sur une Analyse Multi-Critères (AMC) (Malczewski, 2006 ; Karnatak et al., 2007 ; Ersoy et Bulut, 2009 ; Franck-Néel, 2015). Cette méthode itérative permet des regroupements de données cartographiques descriptives à partir des notes intermédiaires attribuées aux descripteurs, puis des notes finales par regroupement (Figure 43). Pour simplifier la présentation, seule la dernière itération de la cartographie du potentiel de présence d'anomalies en cuivre, plomb et zinc sera présentée.

Plusieurs descripteurs étant utilisés pour la nature ou pour les pressions, un regroupement de ces informations est nécessaire : il faut donc déterminer des résultats intermédiaires pour les deux types de descripteurs identifiés (Nature / Pression). Nous avons choisi de traduire les résultats intermédiaires sous forme de notes comprises entre 0 et 2 en fonction des combinaisons possibles des notes liées aux descripteurs (Tableau 6). Ainsi, la note 0 est attribuée si les 2 descripteurs ont une note nulle ; la note 1 si la somme des 2 descripteurs vaut 1 ou 2 ; et la note 2 si la somme des descripteurs est égale à 3 ou 4 (Tableau 6).

III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques

Note descripteur A	Note descripteur B	Note intermédiaire
2	2	2
2	1	
1	2	
2	0	1
0	2	
1	1	
1	0	
0	1	
0	0	0

Tableau 6 : Choix des notes intermédiaires si deux descripteurs existent pour la pression anthropique ou la nature des sols ou sous-sols. Etape renseignée comme « synthèse par type de descripteur » à la figure 43.

La carte d'indices de potentiel de présence d'anomalies géochimiques s'appuie d'une part sur le croisement cartographique des indicateurs qui permet de déterminer les polygones d'intersection entre les différentes couches, et d'autre part sur la note finale obtenue à partir d'une requête croisant les notes intermédiaires des descripteurs.

Le croisement géographique est réalisé sous *ArcGis 10.2.2* à partir de la fonction 'Union'. La note finale est calculée à partir d'une requête conditionnelle à partir des notes des 2 descripteurs de base (nature des matériaux, occupation des sols). La note finale obtenue, comprise entre 0 et 5, est renseignée dans un nouveau champ.

L'ajustement se fait ensuite par itération en comparant les résultats du découpage aux données réelles. Une première version peut être réalisée sans l'appui des données d'analyses géochimiques, permettant donc une première cartographie des villes a priori. La notation n'a pas été fondée sur une analyse exploratoire des données d'analyses géochimiques, afin de démontrer son applicabilité sur les agglomérations disposant de peu d'analyses chimiques de leurs sols. La phase itérative d'adaptation des notes peut être réalisée à mesure que de nouvelles données d'analyses chimiques sont collectées.

L'ensemble de cette démarche empirique est récapitulé à la Figure 43.

III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques

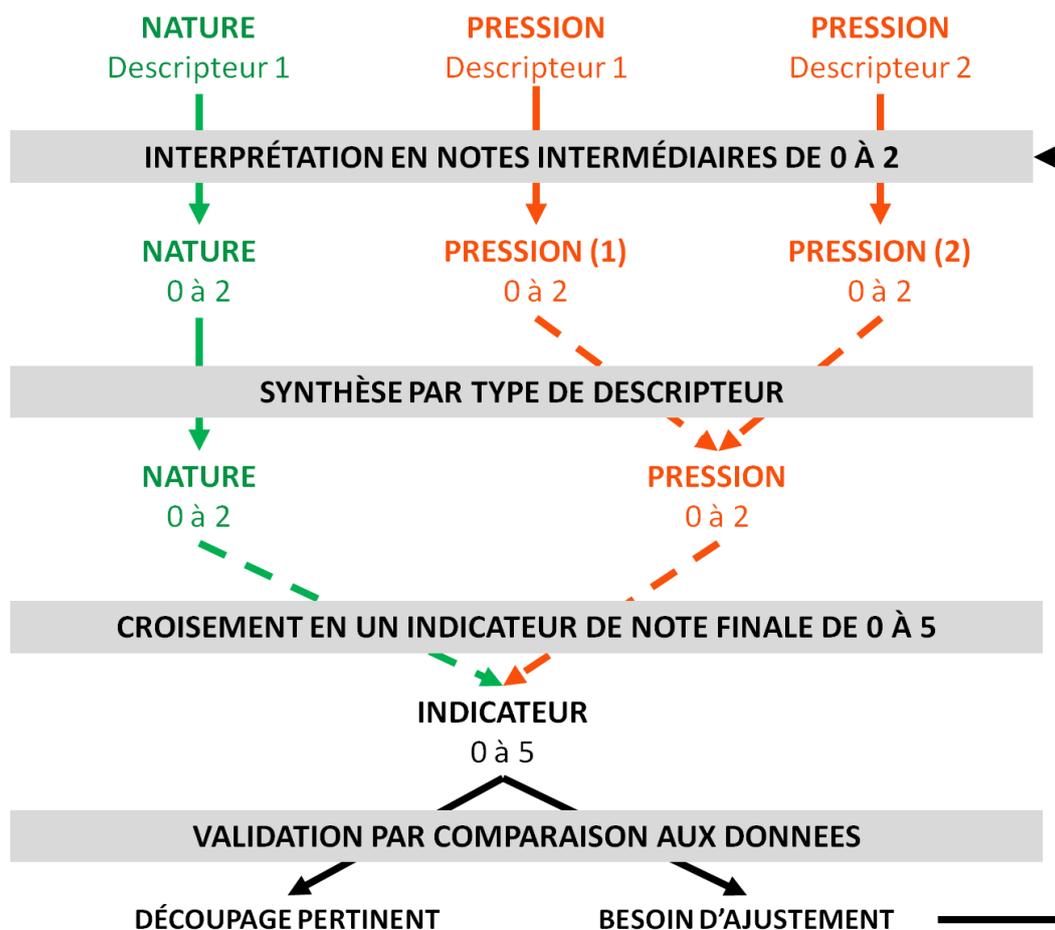


Figure 43: Schéma global de l'analyse multi-critères empirique, dans le cas d'un seul descripteur de la nature des sols et de deux descripteurs de la pression anthropique.

III.C.2. Application à la ville et au quartier (Nantes)

L'ensemble des descripteurs utilisés pour tester le croisement des entités est rappelé à la figure ci-dessous :

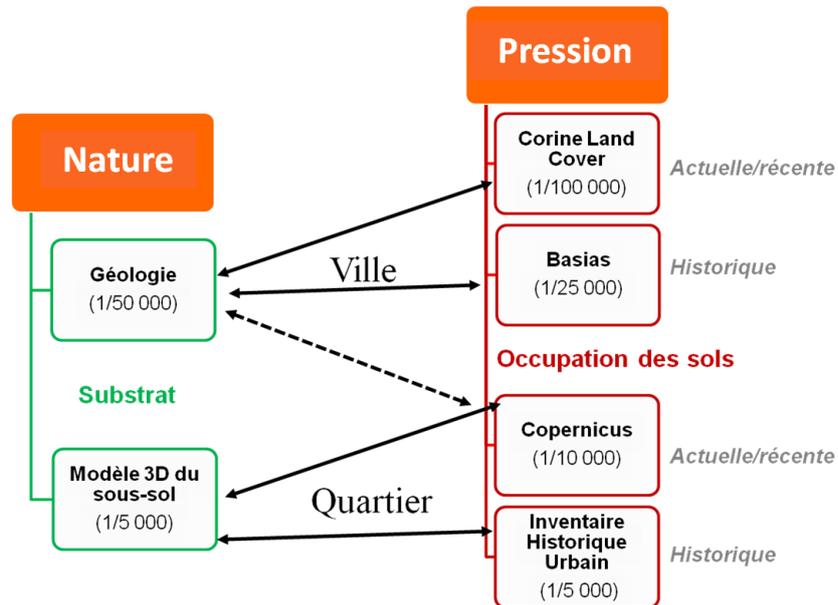


Figure 44: Croisement des descripteurs retenus suivant leur pertinence vis-à-vis de l'objectif de cartographie de potentiel d'anomalies géochimiques et de l'échelle.

Dans le cas de la carte géologique, les notes sont attribuées suivant le retour d'expérience du projet METOTRASS (Le Guern et al., 2013). Elles sont disponibles en annexe 6. Ce projet mettant en relation la lithologie du département avec la géochimie des sols, permet un traitement relativement fiable de ce descripteur. La carte interprétée de ce descripteur est également disponible en annexe 5.

Dans le cas du modèle géologique en trois dimensions, les notes attribuées se basent sur un retour d'expérience lié aux travaux antérieurs sur la typologie des remblais (Le Guern et al., 2016) :

- Remblais suspects : 2
- Remblais divers : 1
- Remblais assimilables naturels : 0
- Alluvions : 0

Les notes attribuées aux codes de Corine Land Cover et de Copernicus (Tableau 7) proviennent d'un retour d'expérience ou d'échanges sur différents projets portant sur la géochimie urbaine (dont GeoBaPa et le Groupe de travail national « Fond pédo-géochimique anthropisé ») et de recherches bibliographiques (Baize, 2000 ; Birke et Rauch, 2000, 1997 ; Kelly et al., 1996 ; Lu et Bai, 2010 ; Wang et al., 2012 ; Xia et al., 2011). Le détail est fourni en annexe 6. Les notes obtenues pour le cuivre, le plomb et le zinc (regroupés comme métaux fortement anthropisés) sont présentées ci-dessous et les figures 45 et 46 procurent un exemple de cartographie :

III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques

CODE	LIBELLE	Pression	Code	Libelle	Pression
111	Tissu urbain continu	2	11100	Tissu urbain continu	2
112	Tissu urbain discontinu	1	11210	Tissu urbain discontinu dense (50-80%)	2
121	Zone industrielle et commerciale	2	11220	Tissu urbain discontinu moyen (30-50%)	1
122	Réseaux routier et ferroviaire et espaces associés	2	11230	Tissu urbain discontinu faible (10-30%)	1
131	Extraction de matériaux	2	11240	Tissu urbain discontinu très faible (<10%)	1
133	Chantiers	1	11300	Structures isolées	0
141	Espaces verts urbains	0	12100	Zone industrielle, commerciale, publique, militaire et privé	2
142	Equipement sportif et de loisirs	0	12210	Routes de transit rapide et terrains associés	2
211	Terres arables hors périmètres d'irrigation	0	12220	Autres routes et terrains associés	2
222	Vergers et petits fruits	1	12230	Chemins de fer et terrains associés	2
231	Prairies	0	12300	Ports	2
242	Systèmes culturaux et parcellaires complexes	0	13100	Extraction de matériaux	2
243	Surfaces essentiellement agricoles	0	13300	Sites de construction	2
311	Forêts de feuillus	0	13400	Terres sans usages actuel	0
312	Forêts de conifères	0	14100	Espaces verts urbains	0
322	Landes et broussailles	0	14200	Equipement sportif et de loisirs	0
324	Forêts et végétation arbustive en mutation	0	21000	Terres arables (cultures annuelles)	0
411	Marais intérieurs	0	22000	Cultures permanentes (vignes, arbres fruitiers, oliveraies)	1
511	Cours et voies d'eau	999	23000	Paturages	0
			31000	Forêts	0
			40000	Zone humide	0
			50000	Eau	999

Tableau 7: Notes de potentiel d'anomalies géochimique attribuées aux codes de Corine Land Cover et de Copernicus dans le cas de métaux liés aux pressions anthropiques (cuivre, plomb et zinc). Les notes sont fournies dans la colonne « Pression ». La note 999 correspond aux surfaces immergées non prises en compte dans le modèle

Les notes de 2 (potentiel fort d'anomalies) attribuées aux tissus urbains continus, aux zones industrielles et aux réseaux de transports, sont appuyées par les résultats des différentes études compilées (Birke et Rauch, 2000 ; Lu et Bai, 2010 ; Wang et al., 2012 ; Kelly et al., 1996), indiquant une corrélation entre la présence de ces types d'occupation des sols et des teneurs élevées en métaux dans les sols de surface.

L'impact du tissu urbain sur les teneurs en cuivre, plomb et zinc étant lié à la densité d'occupation (Birke et Rauch, 2000) nous avons choisi d'attribuer la note de 1 au tissu urbain discontinu. De plus, les classes des « vergers et petits fruits » et des « cultures permanentes » comprenant les vignes, nous leur avons attribué la note 1 afin de tenir compte des pollutions en cuivre liée à la bouillie bordelaise (Baize, 2000).

Enfin, les espaces verts et les zones naturelles ne présentant pas de contaminations liées à l'occupation des sols (dans la bibliographie collectée) la note 0, pour une absence d'anomalie potentielle, leur a été attribué.

Toutes ces notes sont également appuyées par les retours d'expérience des différents projets du BRGM et de l'Ademe (quartiers nantais, GéoBaPa, GT « valeurs de fond »).

Les notes de l'Inventaire Historique Urbain (IHU) se basent quant à elles sur les matrices activité/polluants (présentée en annexe 7 avec un résumé de la démarche et la carte

III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques

interprétée pour ce descripteur) déjà utilisées au sein de cet IHU permettant de lier les emprises identifiées à des polluants potentiels. La nouvelle version de cette matrice est parue en septembre 2018 (<http://ssp-infoterre.brgm.fr/bd-activipoll>), et n'a pas pu être utilisée. Le travail de notation a donc consisté à traduire les résultats dans notre système de 0 à 2. La même méthode a été appliquée aux sites Basias.

Enfin, l'étape finale d'attribution des notes consiste à croiser les descripteurs de pression anthropique et de nature des sols et sous-sols. Pour cela une matrice de croisement est réalisée pour chaque élément considéré.

Lors de la première itération, la note finale correspondait à la somme des deux notes (nature des sols et pression anthropique). Néanmoins, après les premières comparaisons aux données d'analyses chimiques, nous avons adapté la notation à ce cas d'étude (phase d'itération). Nous avons donc choisi de donner plus d'influence à la note de pression anthropique, qu'à celle de nature des sols. Pour cela nous ajoutons 1 aux notes du descripteur des pressions anthropiques pour les notes supérieures à 0. Cet ajout est visible en rouge au Tableau 8 présentant l'exemple des combinaisons de notes attribuées pour le Pb, Cu et Zn, dont la présence dans les sols urbains est fortement liée à la pression anthropique (Alloway, 1990 ; Birke, 2000 ; Birke et Rauch, 1997). Dans le cas de l'Ile de Nantes, la différence d'influence est marquée à deux mètres de profondeur (Le Guern et al., 2016).

Nature	Pression anthropique	Note générale
0	0	0
1	0	1
2	0	2
0	1+1	
0	2+1	3
1	1+1	
1	2+1	4
2	1+1	
2	2+1	5

Tableau 8 : Notes de potentiel d'anomalie géochimique suivant les combinaisons possibles des notes intermédiaires (de 0 pour un potentiel nul à 2 pour un potentiel fort) des descripteurs de pression anthropique et de nature des sols et sous-sols dans le cas de métaux marqués par une origine anthropique (Cu, Pb, Zn) dans les deux premiers mètres de profondeur.

Dans le cas de certains métaux d'origine purement géogénique, seule la nature des sols et proches sous-sols est prise en compte dans la note finale de l'indicateur. Dans le quartier testé, on peut citer le baryum, le chrome ou le nickel.

III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques

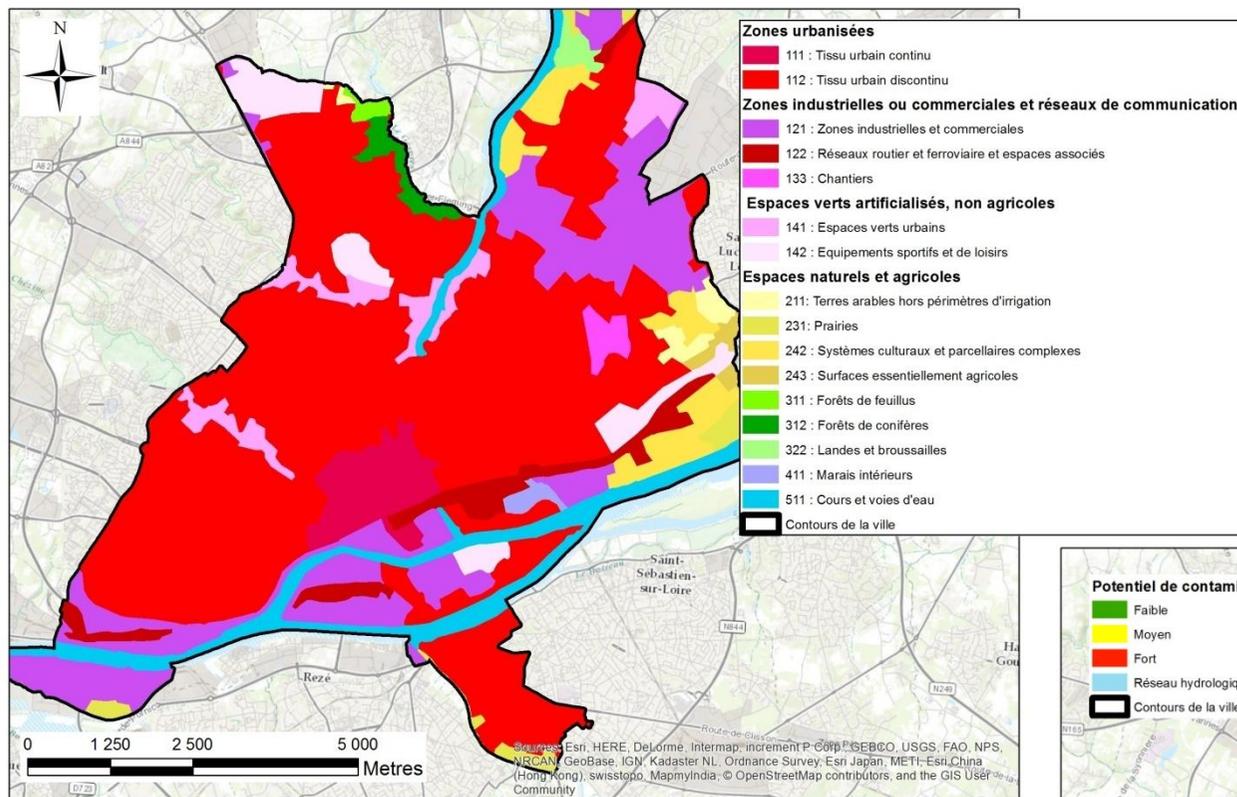
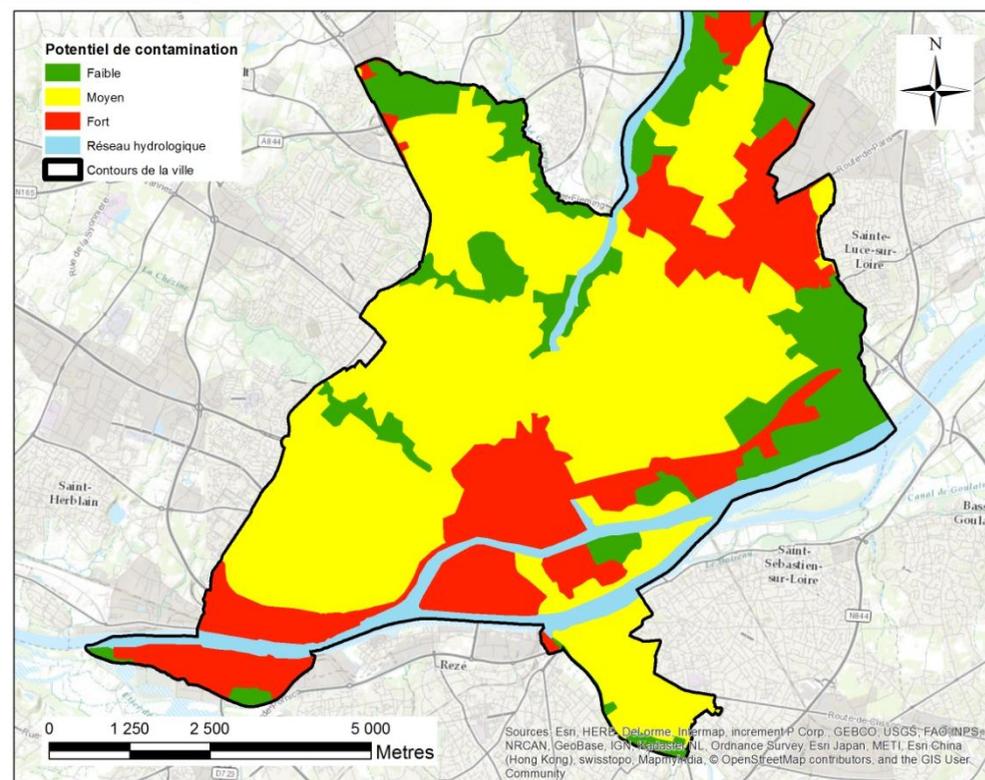


Figure 45 : Occupation des sols fournie par Corine Land Cover (2016) sur la ville de Nantes.

Figure 46 : Potentiel de contamination en plomb, cuivre et zinc, suivant les notes intermédiaires liées à l'occupation des sols fournie par Corine Land Cover sur la ville de Nantes



III.C.3. Résultats cartographiques du croisement des descripteurs

Les notes attribuées empiriquement à chaque descripteur ont ainsi permis de cartographier des entités géographiques suivant le potentiel croissant de présence d'anomalie noté de 0 à 5.

Les résultats obtenus pour les métaux les plus influencés par les activités anthropiques (cuivre, plomb et zinc) sont présentés dans cette partie, à l'échelle de la ville et du quartier.

A l'échelle de la ville, le croisement entre Corine Land Cover et la géologie à 1/50 000 (Figure 47) fait ressortir le centre-ville avec un fort potentiel d'anomalie (en violet ou rouge). Cela reflète la présence conjuguée de remblais anthropiques (Annexe 6) et d'activités industrielles ou de tissu urbain continu (Figure 45 et 46). On notera également la présence d'une forme linéaire (rouge) de potentiel relativement élevé lié à la géologie (formations de micaschistes et gneiss de Landreau) et au tissu urbain discontinu (note 1). Enfin, le reste de la ville étant couvert en majorité par du tissu urbain discontinu et une géologie sans anomalie connue en ces métaux, le potentiel de présence d'anomalie y est relativement faible.

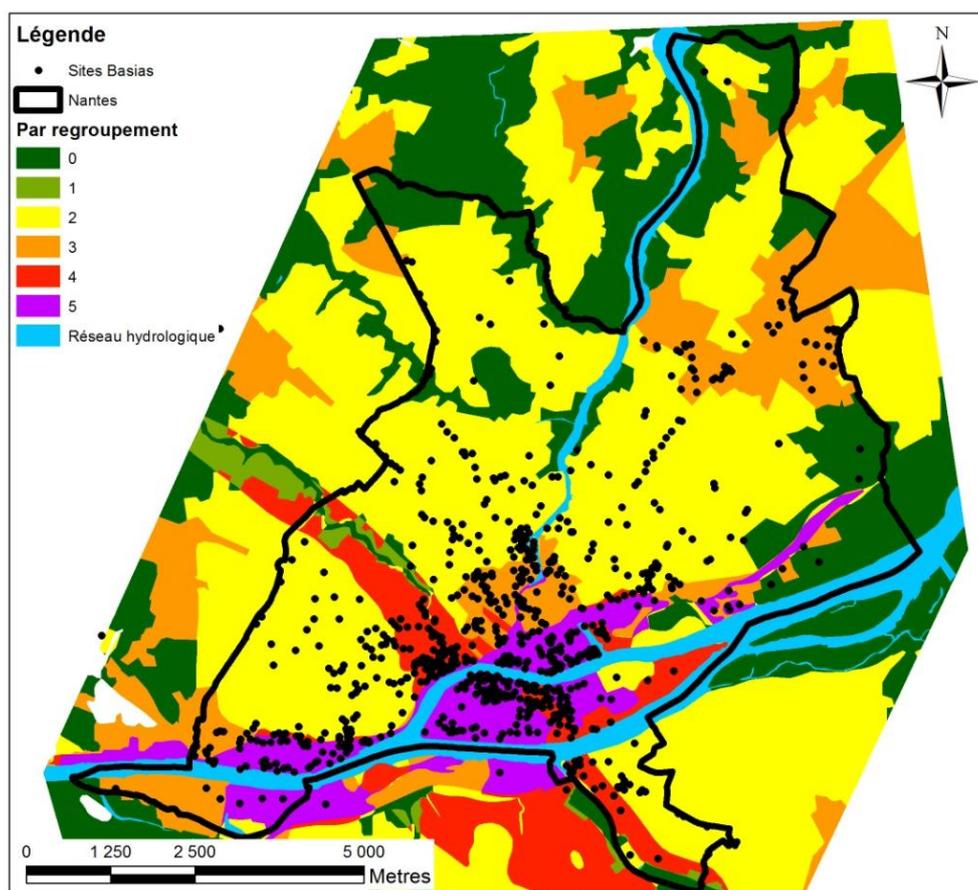


Figure 47 : Carte du potentiel de présence d'anomalie en cuivre, plomb et zinc à l'échelle de la ville, obtenue par le croisement de Corine Land Cover et de la carte géologique au 1/50 000^{ème}.

III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques

Le croisement entre Copernicus (1/10 000) et la carte géologique à 1/50 000 (Figure 48) montre également un fort potentiel d'anomalie géochimique en centre-ville. Le polygone en rouge correspond à une formation géologique (micaschistes et gneiss de Landreau) (Annexe 6) présentant des anomalies en cuivre. Les deux différences principales vis-à-vis de la carte précédente sont :

- la meilleure précision du découpage des entités ;
- les notes 2 de la première carte passent à 3 dans la seconde. Cela s'explique par la différenciation du tissu urbain discontinu dans Copernicus (cf. Figure 48) et l'attribution d'une note différente selon le taux d'urbanisation. Ce code étant très présent dans l'aire urbaine test, une grande surface est concernée par ce changement de note et donc de couleur associée.

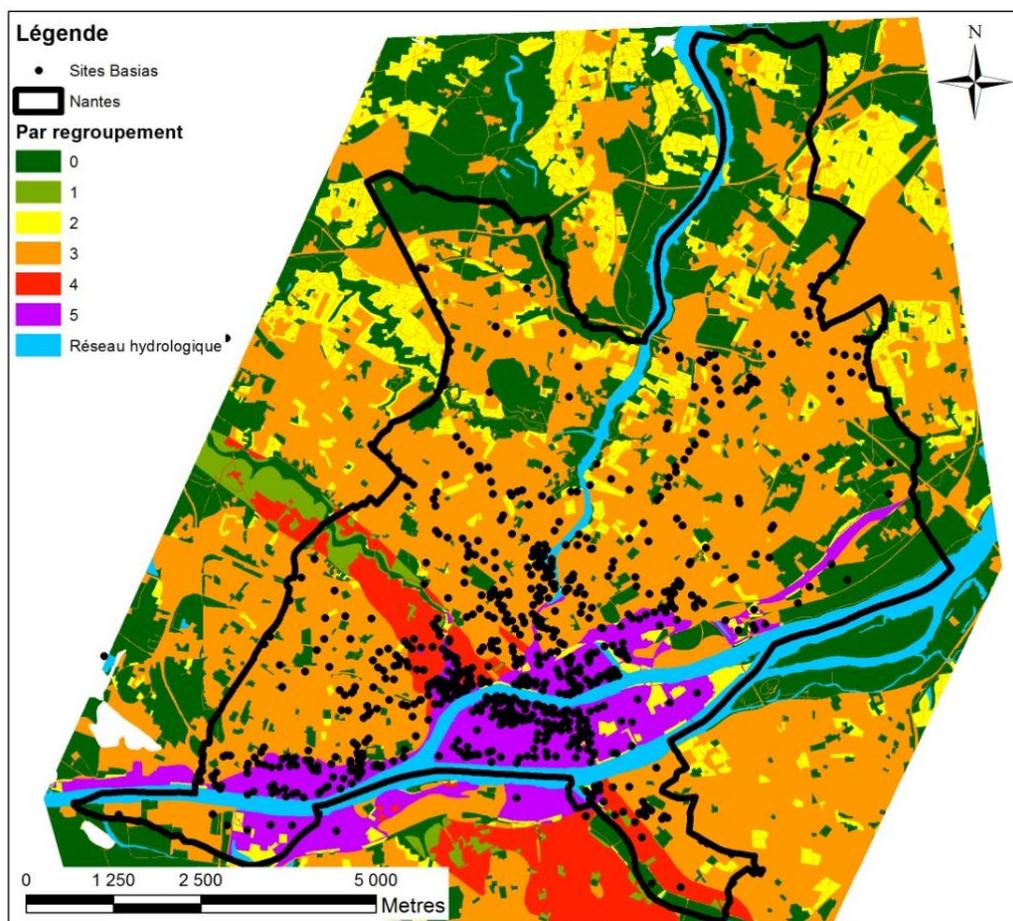


Figure 48 : Carte du potentiel de présence d'anomalie en cuivre, plomb et zinc à l'échelle de la ville obtenue par le croisement de Copernicus et de la carte géologique au 1/50 000^{ème}.

Il est possible de prendre en compte les sites Basias, très concentrés dans le centre de la ville. Ceux-ci étant représentés par des données ponctuelles, ils ne peuvent cependant être directement intégrés au découpage. Une première méthode consiste à affecter à chaque entité une note complémentaire correspondant au nombre de sites Basias par kilomètre carré dans l'entité. Toutefois, la position parfois imprécise de certains sites Basias peut entraîner des erreurs d'affectation du site à une entité (dans les quelques cas où le site est proche d'une

III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques

limite entre deux entités différentes). De plus, pour les entités les plus grandes, la répartition des sites n'est pas prise en compte et peut s'avérer hétérogène sur l'entité. Cette méthode n'a donc pas été conservée.

Une autre méthode consiste à attribuer une zone d'influence aux sites Basias, l'emprise ou la superficie réelle du site n'étant pas indiqué dans Basias, une zone d'influence empirique est attribuée. Dans l'exemple de la Figure 49, cette zone mesure 150 mètres de rayon. La superposition d'au moins deux zones d'influence (symbolisée en gris sur la figure) est considérée comme une pression forte. Cette couche intervient alors comme information complémentaire dans la cartographie. Le rayon de la zone d'influence est, ici, le même pour tous les sites ; il est néanmoins possible d'intégrer des rayons différents suivant le type d'activité. Les emprises précises des sites seraient néanmoins plus fiables, montrant l'intérêt des inventaires historiques urbains à l'échelle des villes.

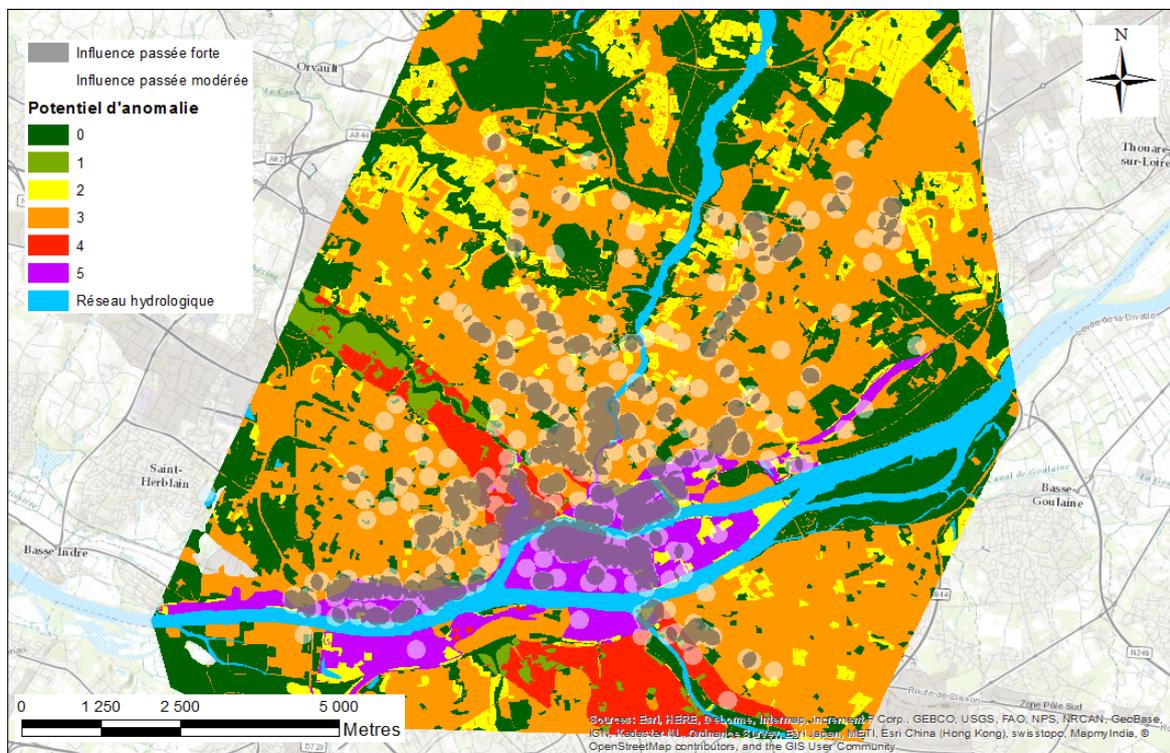


Figure 49: Superposition des zones d'influence des sites Basias sur la carte du potentiel de présence d'anomalie en cuivre, plomb et zinc à l'échelle de la ville obtenue par le croisement de Copernicus et de la carte géologique au 1/50 000^{ème}.

A l'échelle du quartier (Figure 50), les entités obtenues par croisement entre le modèle géologique, Copernicus et l'Inventaire Historique Urbain offrent une précision nettement supérieure (échelle de validité de l'ordre du 1/10 000). On distingue clairement une zone à fort potentiel d'anomalie à l'ouest du quartier (en violet) marquée par l'ancienne zone industrielle et la présence de remblais suspects. A l'inverse, la partie la plus à l'est indique un potentiel d'anomalie faible (en vert). Elle correspond à des parcs situés sur des alluvions. Dans cette zone, la voie ferrée visible en rouge, fait exception. Le reste du quartier, occupé

III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques

par la classe 4 (rouge) et la classe 2 (jaune), est principalement affecté par le tissu urbain dense et la présence de remblais divers ou assimilables naturels.

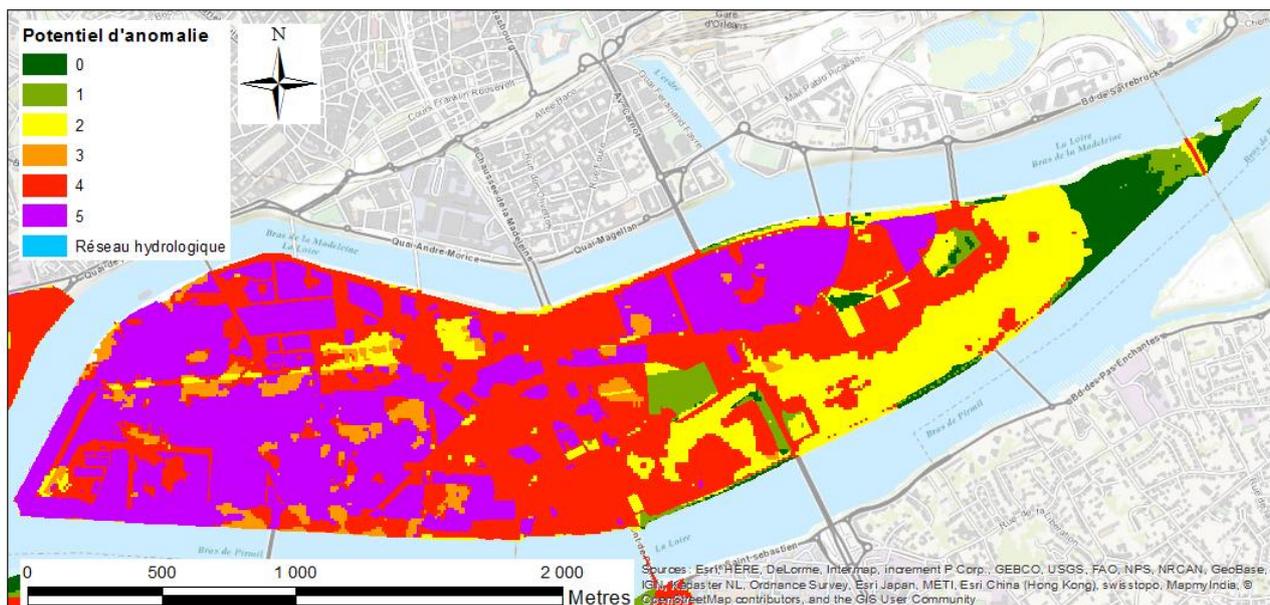


Figure 50: Carte du potentiel de présence d'anomalie en cuivre, plomb et zinc à l'échelle d'un quartier obtenue par le croisement de l'Inventaire Historique Urbain, de Copernicus et du modèle géologique 3D du sous-sol.

Dans tous ces exemples de découpage par Analyse Multi-Critères, les entités sont relativement dispersées et forment généralement des ensembles de polygones. Toutefois, cela reste compatible avec une application opérationnelle à l'échelle du quartier puisque ces polygones correspondent à des critères physiques identifiables sur le terrain. En effet, l'occupation des sols actuelle ou leur nature peuvent être évaluées sur le terrain et peuvent donc être associées à l'une des entités du découpage si l'on en connaît les matrices de pondération. Cela est toutefois plus complexe à l'échelle de la ville, du fait des limites liées à l'échelle de la carte géologique.

III.C.4. Comparaison aux données réelles

Nous disposons donc d'une cartographie a priori du potentiel de présence d'anomalies géochimiques. L'étape suivante consiste à vérifier la pertinence de ce découpage vis-à-vis de la géochimie réelle. Pour cela nous avons recours aux résultats d'analyses collectés et bancarisés par le BRGM. Les données présentées par la suite correspondent aux échantillons compris entre 0 et 2 m de profondeur.

Dans un premier temps nous avons comparé les analyses en plomb et zinc mesurées sur les échantillons aux entités du découpage à l'échelle de la ville via Corine Land Cover (Figure 51) (les résultats du cuivre sont disponibles en annexe 8). Les échantillons de la classe violette (potentiel d'anomalies très fort) présentent les teneurs les plus élevées en plomb, mais comportent toutefois un grand nombre de teneurs faibles. En effet, un potentiel d'anomalie ne

III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques

signifie pas l'absence de valeurs faibles dans les classes à fort potentiel d'anomalies. A l'opposé, la classe en vert (potentiel d'anomalies très faible) présente des teneurs et une moyenne plus faible que les autres classes. Les classes intermédiaires ne montrent pas de tendances claires.

Dans le cas du zinc, les résultats sont moins concluants. En effet un certain nombre de teneurs (entourées en rouge) ne suivent pas les tendances attendues. Cela s'explique par le degré d'approximation lié à l'échelle des descripteurs (Corine Land Cover notamment) et à une plus forte mobilité du zinc (Vaněk et al., 2005). Le découpage tenant compte des sources, les migrations latérales de contaminants ne sont prises en compte qu'au sein des entités et non en dehors de leurs limites.

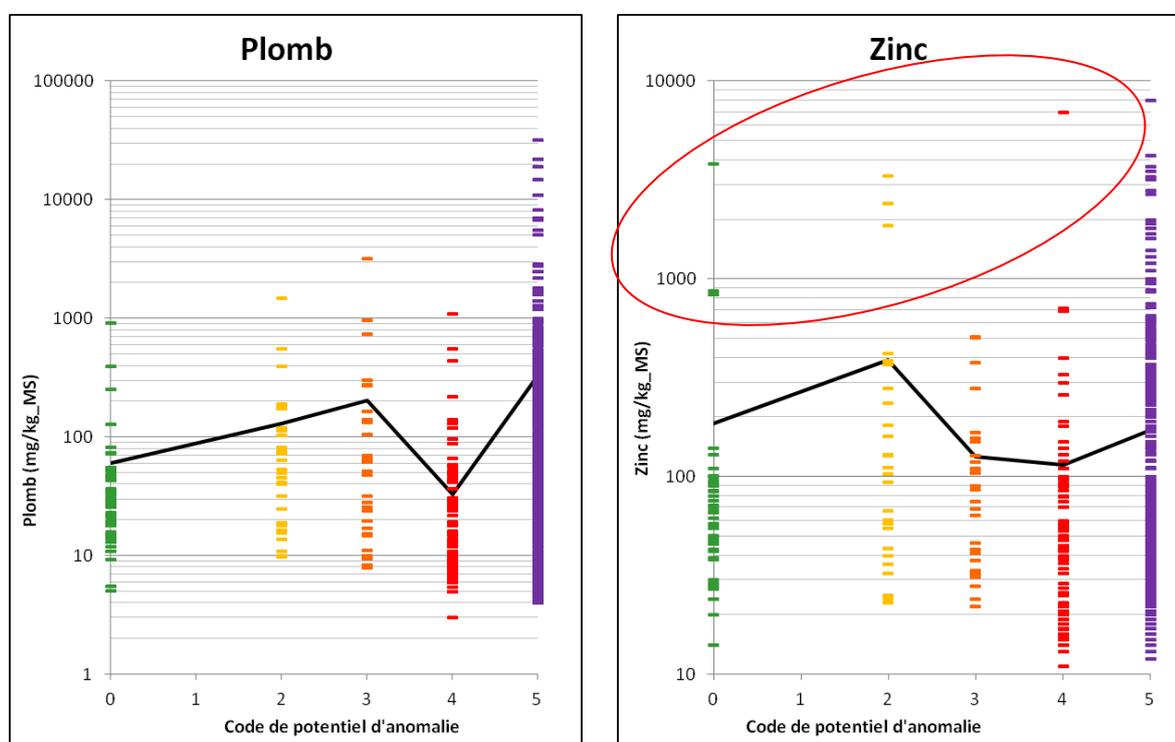


Figure 51 : Teneurs en plomb et zinc des échantillons collectés dans la ville. Les couleurs représentent le potentiel de présence d'anomalie suivant la carte réalisée à partir de Corine Land Cover (figure 45). Les moyennes par classe sont représentées par le trait noir.

Nous comparons ensuite les résultats d'analyses au découpage de la ville obtenu avec Copernicus (Figure 52), la classe 4 (en rouge) n'est plus représentée à cause de la différenciation de la densité urbaine par rapport à Corine Land Cover (cf § III.1.3). Pour ce découpage, les tendances observées pour le plomb sont plus marquées que pour le découpage avec Corine Land Cover. Ainsi les teneurs les plus élevées considérées comme anormales sont plus présentes et plus fortes à mesure que la note augmente. Dans le cas du zinc, les tendances sont aussi mieux marquées avec des anomalies également plus présentes dans les classes à fort potentiel d'anomalies, avec encore quelques teneurs élevées dans les classes à faible potentiel (entourées en rouge). Ces teneurs sont toutefois moins nombreuses que

III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques

précédemment, ce qui illustre bien l'utilité de descripteurs plus fins pour une meilleure caractérisation du potentiel de présence d'anomalie.

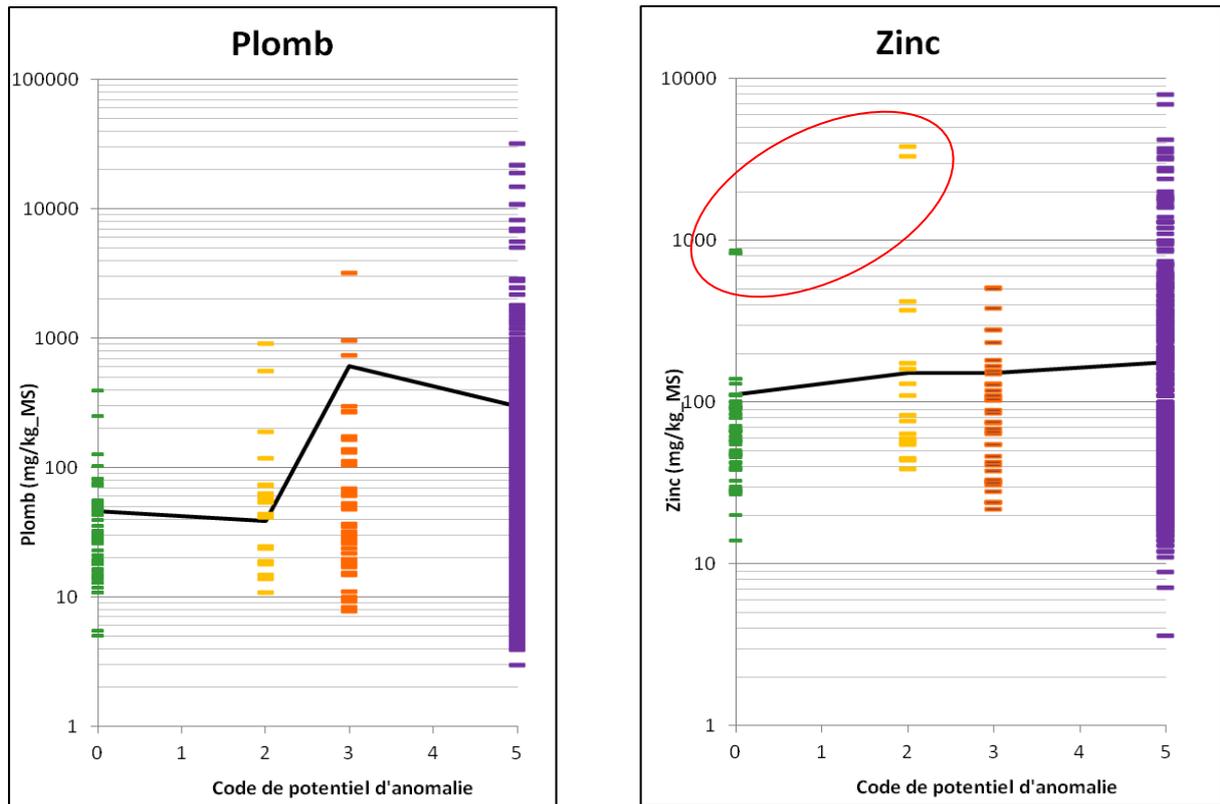


Figure 52 : Teneurs en plomb et zinc des échantillons collectés dans la ville. Les couleurs représentent le potentiel de présence d'anomalie suivant la carte réalisée à partir de Copernicus (figure 46). Les moyennes par classe sont représentées par le trait noir.

Dans le cas du quartier, où le découpage fait intervenir le modèle géologique 3D et l'inventaire Historique Urbain, quatre classes comportent plus de 10 échantillons (classes 2, 3, 4 et 5) la classe 5 étant la plus nombreuse (Figure 53).

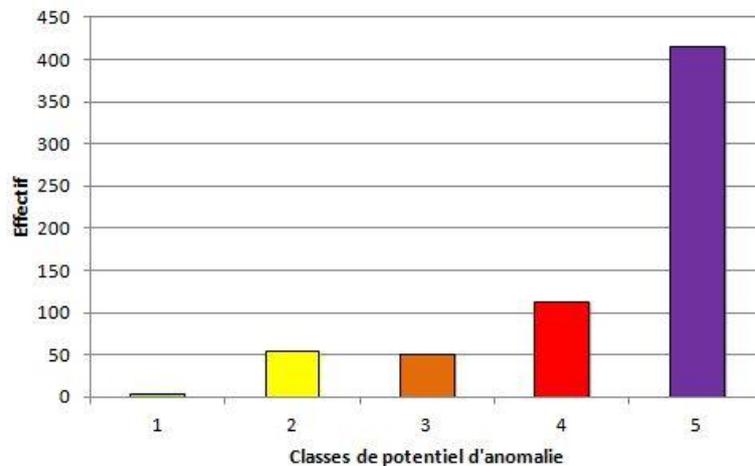


Figure 53: Effectifs des échantillons collectés dans les 2 premiers mètres de profondeur suivant les classes d'anomalies de l'analyse multi-critère sur le quartier.

III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques

Les histogrammes des teneurs du plomb et du zinc (Figure 54), révèlent des tendances similaires entre les deux métaux. Les classes 2 et 3, au potentiel d'anomalies intermédiaire, présentent des teneurs globalement faibles avec une ou deux teneurs plus fortes. Les classes 4 et 5, avec des potentiels d'anomalies fort à très fort, présentent des teneurs plus élevées et en plus grand nombre. Cela est d'autant plus marqué pour la classe 5. Ceci illustre l'importance des descripteurs de bonne qualité pour une meilleure caractérisation de l'environnement urbain. Ainsi un investissement plus conséquent (Inventaire Historique Urbain, Modèle géologique 3D du sous-sol) pour la constitution de ces descripteurs peut être pertinent dans des quartiers en réaménagement, afin de fournir un outil d'anticipation de la qualité des terres excavées suivant un potentiel de présence d'anomalies géochimiques.

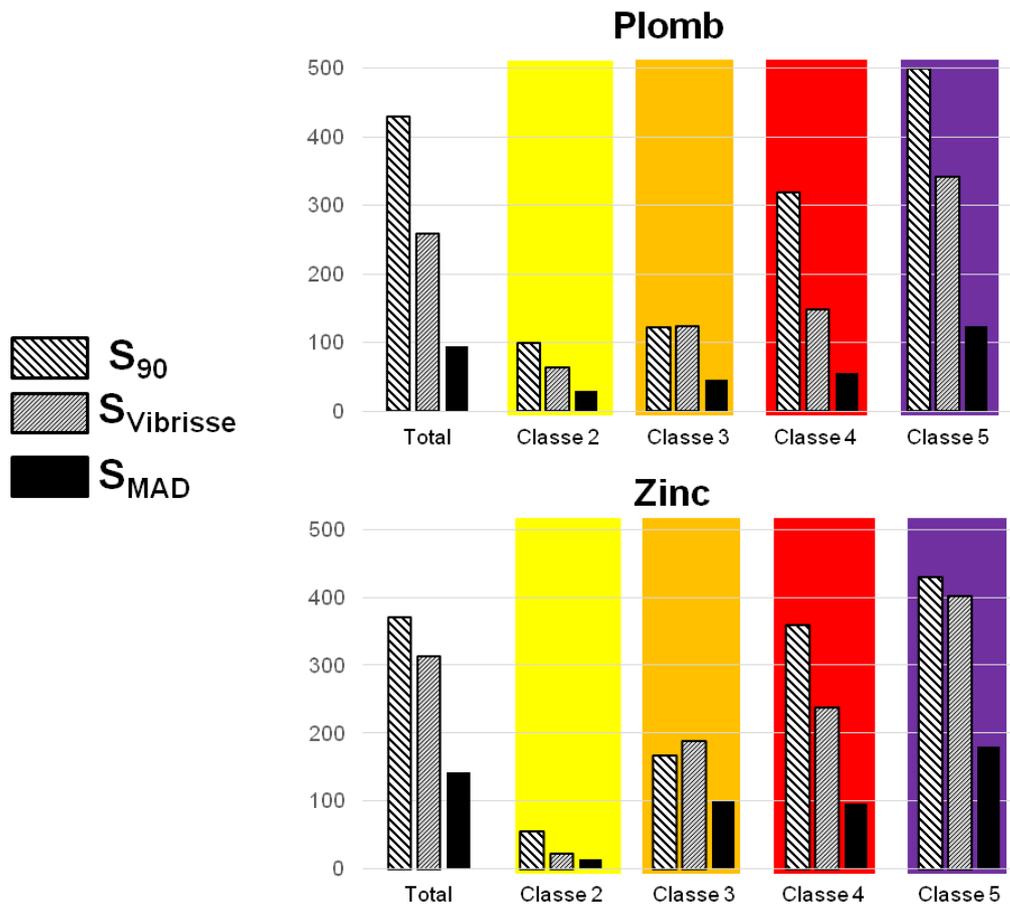


Figure 54: Histogramme des teneurs en plomb et zinc des échantillons collectés sur un quartier de la ville. Les couleurs représentent le potentiel de pollution de la figure 50.

L'Analyse Multi-Critères permet ainsi de cartographier, a priori, le potentiel de présence d'anomalies géochimiques dans les sols urbains, sur la base de descripteurs. L'échelle de validité dépend des données de départ. Des descripteurs plus fins peuvent s'avérer utiles dans des zones à forts enjeux.

Un indice de qualité potentielle des sols urbains peut être construit à partir de descripteurs (données géographiques descriptives, pour certaines très simples d'accès). Si les

III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques

notes attribuées ici visent un indicateur de qualité géochimique, cette méthode de découpage peut être appliquée à d'autres domaines (stockage du carbone, qualité agronomique, perméabilité des sols, ...), en adaptant les descripteurs et les critères de notation.

III.C.5. Calculs statistiques de valeurs seuil

La pertinence du découpage en entités géographiques suivant le potentiel d'anomalie géochimique étant vérifiée, il reste à tester les calculs statistiques usuels de valeurs seuil du fond pédo-géochimique sur ce découpage. La grande majorité des données étant limitées aux quartiers en réaménagement, nous avons calculé ces seuils suivant le découpage le plus précis disponible à l'échelle du quartier, faisant intervenir le modèle géologique 3D et l'Inventaire Historique Urbain. Seules les données des classes 2, 3, 4 et 5 sont utilisées car disposant d'effectifs suffisants (la classe 1 ne comporte que 8 échantillons). Les résultats sont présentés au Tableau 9.

Comme dans l'étude de sensibilité des seuils statistiques, les seuils déduits du quantile 95 sont très élevés et correspondent probablement à des anomalies. Cela est cohérent avec le fait que deux des classes testées (4 et 5) expriment un potentiel fort de présence d'anomalie et donc une proportion supérieure à 5% d'anomalies. De plus, pour les deux autres classes (2 et 3) présentant un potentiel intermédiaire de présence d'anomalies géochimiques, rien n'indique que la proportion d'anomalie soit égale à 5 % (ou 10%).

Les deux autres seuils présentent un intérêt lié aux objectifs visés. Les seuils MAD, plus faibles, sont plus à même de garantir la protection du milieu, alors que les seuils de la vibrisse, plus élevés, permettent de valoriser des volumes plus importants de terres excavées. Rappelons que, la réutilisation des terres excavées doit aussi répondre à des enjeux environnementaux (impact sur les eaux souterraines) et sanitaires (seuils sanitaires selon les usages envisagés, Coussy et al, 2017).

La comparaison des résultats obtenus pour les différentes classes, montre la croissance de la valeur des seuils avec la classe de potentiel d'anomalie, ce qui est cohérent avec l'hypothèse d'une cartographie de la qualité géochimique des sols.

III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques

PLOMB

	TOTAL	CLASSE 2	CLASSE 3	CLASSE 4	CLASSE 5
Effectif	631	54	51	112	414
Minimum	4	14	5	5	4
Moyenne	276	46	57	430	291
Médiane	38	20	21	25	48
Ecart-type	1303	63	85	2271	1077
S_{Q90}	430	100	123	320	500
S_{Q95}	773	114	230	1008	943
$S_{Vibrisse}$	260	64	124	149	343
S_{MAD}	94	29	46	55	124
Maximum	19000	420	450	19000	11000

ZINC

	TOTAL	CLASSE 2	CLASSE 3	CLASSE 4	CLASSE 5
Effectif	631	54	51	112	414
Minimum	4	5	15	16	4
Moyenne	199	34	110	228	223
Médiane	62	9	47	46	75
Ecart-type	587	84	287	637	634
S_{Q90}	370	55	166	358	430
S_{Q95}	619	123	223	838	695
$S_{Vibrisse}$	313	22	189	238	402
S_{MAD}	143	14	100	96	180
Maximum	8000	560	2000	4200	8000

Tableau 9 : Résultats des calculs de seuils statistiques du fond pédo-géochimique anthropisé en plomb et zinc (en mg/kg_MS) par classe de potentiel de présence d'anomalie géochimique d'un quartier urbain sur la tranche 0-2 m.

Nous pouvons comparer les seuils obtenus pour les classes de l'analyse multi-critères (AMC), à ceux obtenus pour les types de remblais (Le Guern, 2018) (Tableau 10). Les données utilisées pour définir les seuils des remblais sont les mêmes que celles utilisées ici, mais un prétraitement avait été effectué afin d'ôter les anomalies (supérieures à dix fois le fond géochimique naturel). Ce prétraitement n'a pas été effectué pour les seuils des classes de l'AMC, car les valeurs de fond pédo-géochimique naturel n'étant pas disponibles pour toutes la ville ou pour tous les quartiers, nous avons considéré que nous n'en disposions pas pour ce traitement.

Dans le cas du plomb, les seuils calculés pour les remblais suspects sont intermédiaires à ceux calculés pour les classes 4 et 5 de l'analyse multi-critères. Cela est également valable pour le S_{Q90} et le $S_{Vibrisse}$ des remblais divers, intermédiaires par rapport aux classes 3 et 4. En revanche, le seuil MAD des remblais divers, ainsi que tous les seuils des remblais assimilables naturels sont inférieurs à ceux calculés pour la classe 2 de l'AMC. Pour le zinc, les seuils calculés sur les remblais suspects sont supérieurs à ceux de toutes les classes de l'AMC. Les seuils des remblais divers correspondent globalement à ceux de la classe 3 de l'AMC et les seuils des remblais assimilables naturels sont du même ordre que ceux de la classe 2 de l'analyse multi-critères.

Ces deux découpages du territoire diffèrent par la prise en compte d'un plus grand nombre de paramètres dans l'analyse multi-critères que dans la typologie des remblais.

Comme pour les remblais suspects (Le Guern et al., 2016), il convient de ne pas tenir compte des valeurs de seuils de fond pédo-géochimique pour la classe 5 de l'AMC. Celle-ci regroupant les anomalies les plus fortes et en plus grand nombre, les seuils statistiques ne sont pas adaptés pour fournir des valeurs de fond pédo-géochimique.

III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques

		Pb	Zn
Remblais suspects	S_{Q90}	336	468
	S_{Vibrisse}	222	546
	S_{MAD}	101	181
Remblais divers	S_{Q90}	172	172
	S_{Vibrisse}	131	183
	S_{MAD}	27	76
Remblais assimilables naturels	S_{Q90}	53	98
	S_{Vibrisse}	50	107
	S_{MAD}	16	39

Tableau 10 : Résultats des calculs de seuils par type de remblais du quartier de l'Ile de Nantes (recalculé pour les échantillons de moins de 2 m de profondeur à partir du jeu de données de Le Guern et al., 2016)

En conclusion, les résultats démontrent l'importance de la prise en compte de données descriptives des sols pour définir une typologie adaptée au fond géochimique, fournissant différents niveaux de fond permettant de valoriser des volumes plus importants de terres excavées tout en garantissant le maintien de la qualité des sols.

De plus, il est possible de fournir des valeurs seuils du fond pédo-géochimique anthropisé pour des entités géographiques définies suivant un potentiel de présence d'anomalie (Figure 55). Toutefois, afin de fournir des valeurs pour chaque niveau de potentiel, des effectifs suffisants par classe sont indispensables. Cela peut nécessiter des échantillonnages complémentaires.

Enfin, la profondeur de validité du modèle dépendra de la nature et de l'occupation des sols (profondeur maximale d'influence des activités anthropiques). Cette profondeur de validité reste toutefois complexe à estimer. Nous avons retenu ici une profondeur de coupure fixe (2 m) sur la base de l'analyse exploratoire des données lors de travaux précédents. Il serait intéressant de développer une méthodologie permettant d'évaluer cette profondeur sur l'ensemble d'un quartier ou d'une ville en fonction des types, et des durées, d'occupation. Des informations liant la profondeur d'impact à l'occupation des sols seront alors nécessaires.

Si un découpage a priori du territoire peut donc être réalisé, une autre approche utilisant directement les échantillons de sols peut aussi être mise en place.

III.C. Découpage d'entités selon le potentiel de présence d'anomalies géochimiques

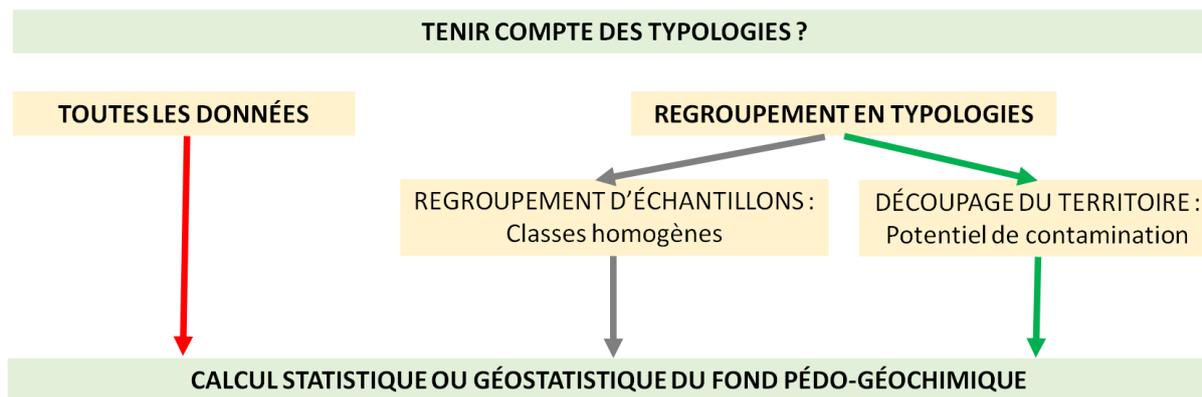


Figure 55 : Schéma conceptuel de la constitution de valeurs de fond pédo-géochimique présentant l'importance de la prise en compte d'une typologie et d'un découpage du territoire par Analyse Multi-Critères.

III.D. Regroupement des échantillons en classes homogènes

III.D.1. Objectif et principe général

Dans la méthode précédente, des entités géographiques ont été construites a priori à partir de descripteurs urbains, pour être ensuite comparées aux données géochimiques.

Dans cette seconde méthode, nous testons la méthodologie inverse en constituant des classes homogènes d'échantillons pour ensuite découper le territoire. Cette approche a pour intérêt de s'appuyer directement sur la géochimie mesurée pour découper le territoire plutôt que sur des retours d'expériences ou des connaissances a priori (Figure 56).

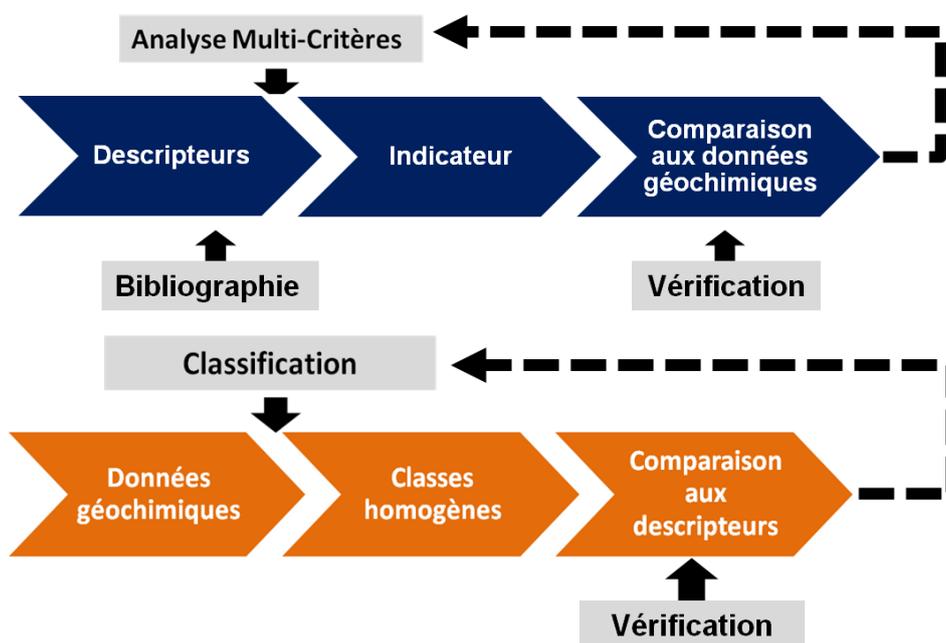


Figure 56: Schéma conceptuel des différentes étapes des deux approches pour définir une typologie des sols et proches sous-sols urbains adaptée au fond pédo-géochimique.

Les échantillons étant majoritairement concentrés sur des quartiers en réaménagement, nous ne pouvons traiter la ville dans son entièreté. Nous avons donc choisi un quartier pour lequel de nombreuses données sont disponibles (plus de 2000 échantillons collectés).

Pour réaliser des classes d'échantillons homogènes nous avons besoin d'un outil permettant de mesurer une « distance » entre les échantillons suivant plusieurs paramètres géochimiques et d'associer le plus ressemblant. Nous avons donc choisi de travailler avec la classification ascendante hiérarchique qui permet de définir le nombre de classes retenues suivant les résultats de la classification et non avant comme dans les classifications non hiérarchiques.

III.D.2. La classification ascendante hiérarchique

La classification est un outil statistique (Cheng, 2009 ; Ducimetiere, 1970 ; Gonzalez, 2008 ; Johnson, 1967 ; Oliver et Webster, 1989 ; Romary, 2012, 2013 ; Steinley et Brusco, 2007) couramment utilisé afin d'affecter des individus à des groupes spécifiques. Elle permet de prendre en compte plusieurs paramètres à la fois et d'éviter de se fonder uniquement sur l'appréciation d'un opérateur pour définir les classes. Dans sa version usuelle, la classification ne fait pas intervenir la position des individus dans l'espace. Afin d'en tenir compte, une variante (Romary, 2012) est utilisée.

a) Principe

La classification ascendante hiérarchique se fonde premièrement sur un calcul de distance (ou dissimilarité) entre des individus (échantillons). Cette distance est faible pour des individus se ressemblant et grande pour des individus très différents.

Cette distance est ensuite utilisée afin d'associer deux individus en minimisant un critère d'agrégation. Cette étape est alors répétée afin de fournir des classes où les individus d'une même classe se ressemblent et où les classes sont les plus différentes possibles. On minimise alors la distance intra-classe et on maximise la distance inter-classe.

A noter qu'ici le choix du nombre de classes retenues est laissé à l'opérateur. En effet les classifications ascendantes hiérarchiques définissent des classes emboîtées de moins en moins nombreuses à mesure que l'algorithme progresse. Il existe également des classifications non hiérarchiques où le nombre de classes recherchées doit être fixé à l'avance. Ces classifications sont plutôt adaptées au traitement de volumes très importants de données (Ducimetiere, 1970 ; Gonzalez, 2008 ; Johnson, 1967).

b) Distance

Plusieurs choix de distance sont possibles.

La distance la plus commune est la distance euclidienne qui consiste à représenter les individus dans un espace euclidien où les dimensions sont définies par les paramètres comparés. La distance entre les individus correspond alors à la distance les séparant dans cet espace (Figure 57). Elle est utilisable ici puisque les données ont toutes la même unité. Si les variations sont toutefois très différentes, il est conseillé de normer les données. Elle reste toutefois arbitraire car sa valeur augmente avec le nombre de paramètres et elle ne tient pas compte des liaisons entre variables.

III.D. Regroupement des échantillons en classes homogènes

Dans cet exemple, la distance entre deux échantillons i et j analysés sur deux paramètres X et Y , serait donc :

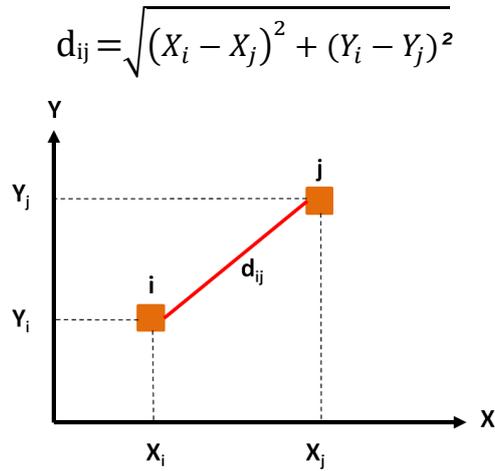


Figure 57: Distance euclidienne dans un repère formé par les paramètres X et Y , pour deux individus i et j .

Dans le cas des classifications hiérarchiques, il est également nécessaire de calculer une distance entre les classes (inter-classe). Celle-ci peut être définie par la plus petite distance entre les individus des deux classes ou par la plus grande. La distance entre le centre de gravité des classes. Il existe enfin un paramètre de Ward qui équivaut à la perte d'inertie inter-classe résultant de leur regroupement (Murtagh et Legendre, 2014).

c) Dendrogramme

Les classes identifiées sont emboîtées et peuvent être représentées selon un arbre de classification appelé dendrogramme (Figure 58). La hauteur des racines correspond au critère d'agrégation. Plus la racine est haute et plus les classes sont différentes.

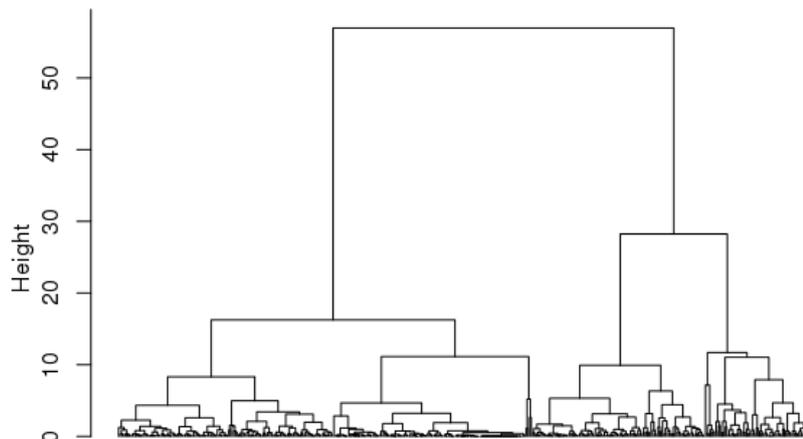


Figure 58: Exemple de dendrogramme.

d) Classes

Le résultat de la classification hiérarchique est donc un dendrogramme présentant les regroupements successifs des individus. Pour définir les classes, il est nécessaire de choisir une coupure dans cet arbre. Cette coupure peut être basée sur une appréciation graphique (augmentation forte de la hauteur avant le prochain regroupement) ou sur des critères numériques (rapport entre la distance inter et intra classe).

Une fois cette coupure choisie, un identifiant de classe est attribué à chaque individu et ceux-ci peuvent être représentés suivant cet identifiant.

e) Spatialisation

Comme évoqué précédemment, l'algorithme classique de la classification ne fait pas intervenir la position des données. Il est néanmoins possible d'intégrer un critère spatial afin que les classes identifiées ne soient pas trop dispersées spatialement.

Pour cela une triangulation des données (Romary, 2012) est utilisée pour que des individus ne puissent être ajoutés à une classe que s'ils sont « voisins » d'un autre point de cette classe. Cela garantit une meilleure continuité spatiale des classes. La Figure 59 présente les résultats de trois classifications classiques (sans prise en compte de la position des points) (a, b et c) et d'une classification spatialisée (d), sur les mêmes données. On note la meilleure connexité de la classe violette (d) vis-à-vis des classes rouges identifiées par les trois autres méthodes de classification.

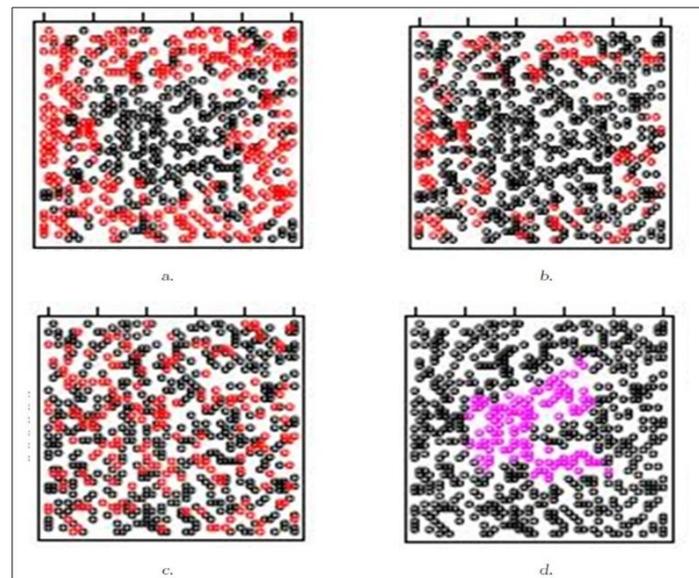


Figure 59: Résultats sur un même jeu de données de a : K-moyennes, b : classification hiérarchique, c : méthode Oliver et Webster et d : classification spatialisée (Romary, 2012)

III.D.3. Application aux données

Pour l'application à des données réelles, nous travaillons comme précédemment avec les résultats d'analyses géochimiques en cuivre, plomb et zinc, d'échantillons de diagnostics de pollution. Différents paramètres sont testés afin d'étudier leur influence sur le résultat de la classification.

Afin d'étudier **l'influence de la profondeur et du type de matériau**, nous testons deux sélections de données en prenant uniquement le premier échantillon de remblais de chaque sondage ou tous les échantillons à moins de deux mètres de profondeur. Pour la comparaison, les calculs ont aussi été effectués en tenant compte de toutes les données, quel que soit le type de matériau ou la profondeur.

En plus des tests sur les données brutes (tous les paramètres ayant la même unité), deux **transformations** sont examinées. La première consiste à réduire les données. Après transformation, elles ont une espérance nulle et un écart-type de 1.

Avec μ l'espérance et σ l'écart-type :

$$Z_i = (X_i - \mu) / \sigma$$

La seconde transformation est une log-translation des données normalisées par la médiane :

$$Z_i = \log (1 + (X_i / \text{Médiane}(X)))$$

Comme décrit précédemment, différentes **distances** sont utilisables pour la classification. Nous en comparons deux, la distance euclidienne et une distance basée sur la matrice de corrélation des données.

Enfin, l'objectif étant de cartographier des ensembles d'échantillons géochimiquement homogènes, des classes spatialement connexes faciliteraient la cartographie. Nous comparons donc les résultats des **classifications classiques** à ceux d'une **classification spatialisée** avec une triangulation de Delaunay.

Nous obtenons un nombre de tests relativement important (108) puisque nous considérons 3 familles de paramètres géochimiques, et que 36 tests sont réalisés par famille. Seule une sélection des résultats est présentée ici.

III.D.4. Résultats et discussion

a) Le premier paramètre testé est la **sélection des échantillons**, Nous comparons ici la sélection des échantillons à **moins de deux mètres de profondeur** (Figure 60) à celle de **l'échantillon de remblais le plus en surface** de chaque sondage (Figure 61). Dans les deux cas nous travaillons avec les données brutes et une distance euclidienne. Nous ne tenons pas compte de la spatialisation des données.

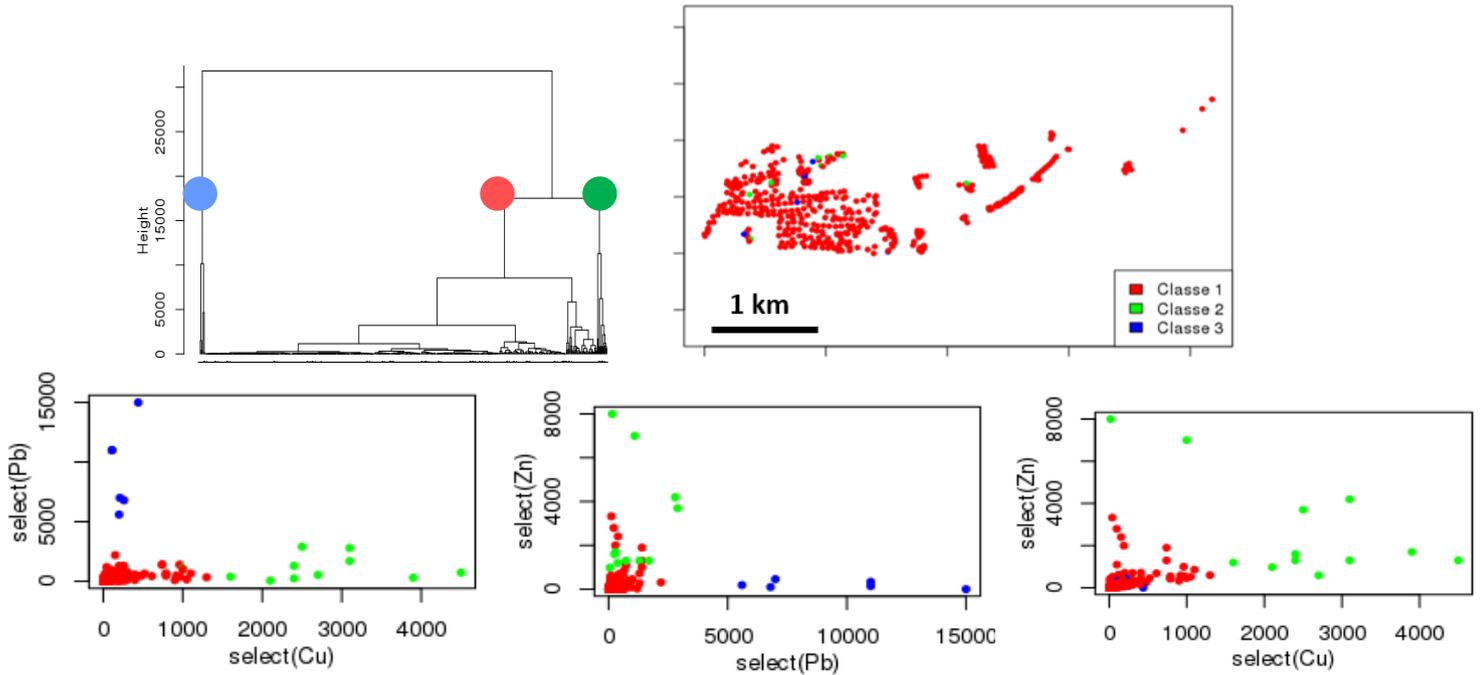


Figure 60 : Résultats de classification sur les données brutes des échantillons de moins de 2m de profondeur. Distance euclidienne, sans spatialisation.

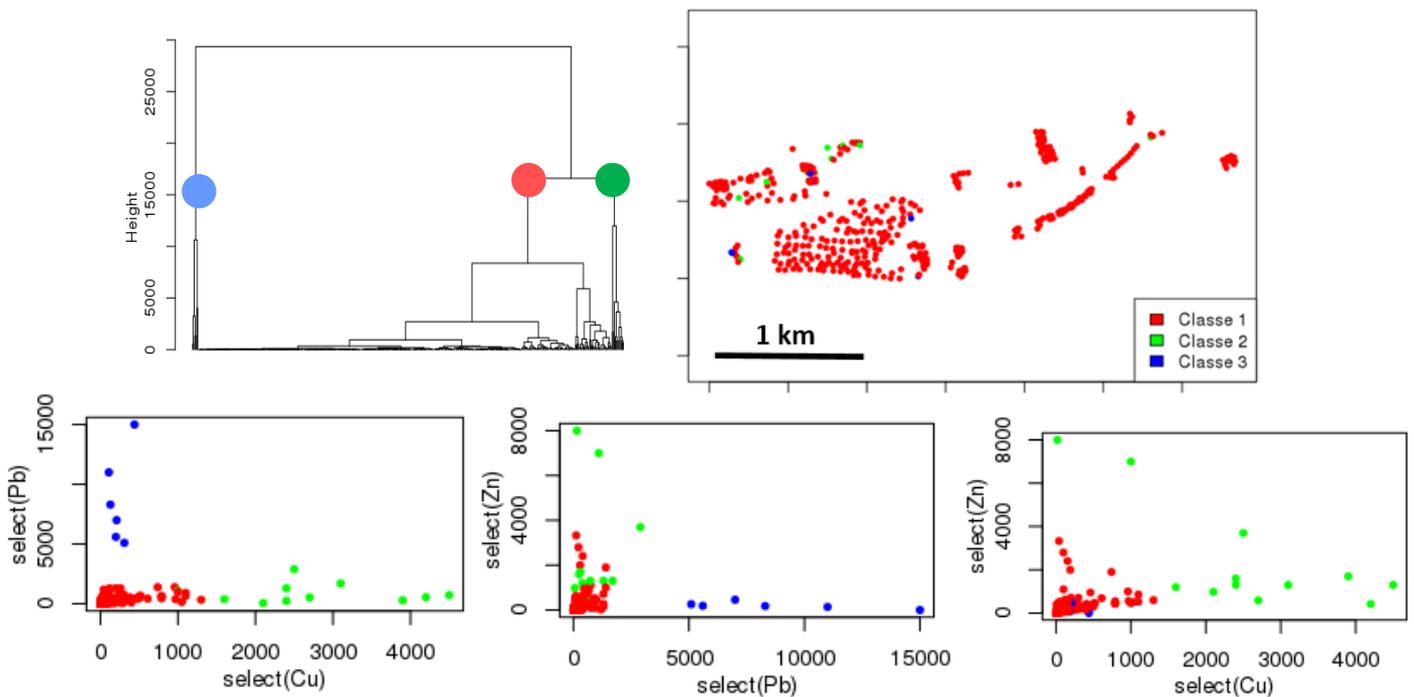


Figure 61 : Résultats de classification sur les données brutes des échantillons de remblais les moins profond de chaque sondage. Distance euclidienne, sans spatialisation.

III.D. Regroupement des échantillons en classes homogènes

Les résultats des deux sélections sont très semblables, tant sur la carte que sur les nuages de corrélation entre les métaux. En effet, la majorité du quartier étant remblayée, les deux sélections correspondent en **grande majorité aux mêmes échantillons**. On note la présence de deux classes bien distinctes. La première (en bleu) correspond à des échantillons riches en plomb mais pauvres en zinc et cuivre. A l'inverse la seconde (en vert) comprend des échantillons riches en cuivre et zinc mais plus pauvres en plomb. La troisième classe (en rouge) comprend le reste des échantillons.

Les premiers échantillons de chaque sondage ayant des profondeurs variables il est complexe de déterminer quelle tranche de sols est réellement concernée par la classification. Nous choisissons de travailler avec la sélection des **échantillons de moins de deux mètres de profondeur** afin d'isoler les teneurs plus impactées par l'activité anthropique et de définir clairement la tranche de sols traitée. Cette sélection conserve 640 échantillons correspondant à 75% des données initiales.

b) Nous comparons ensuite les résultats obtenus avec les **données brutes** (Figure 60) avec ceux obtenus avec les données **centrées-réduites** (Figure 62) et **log-translatées** (Figure 63).

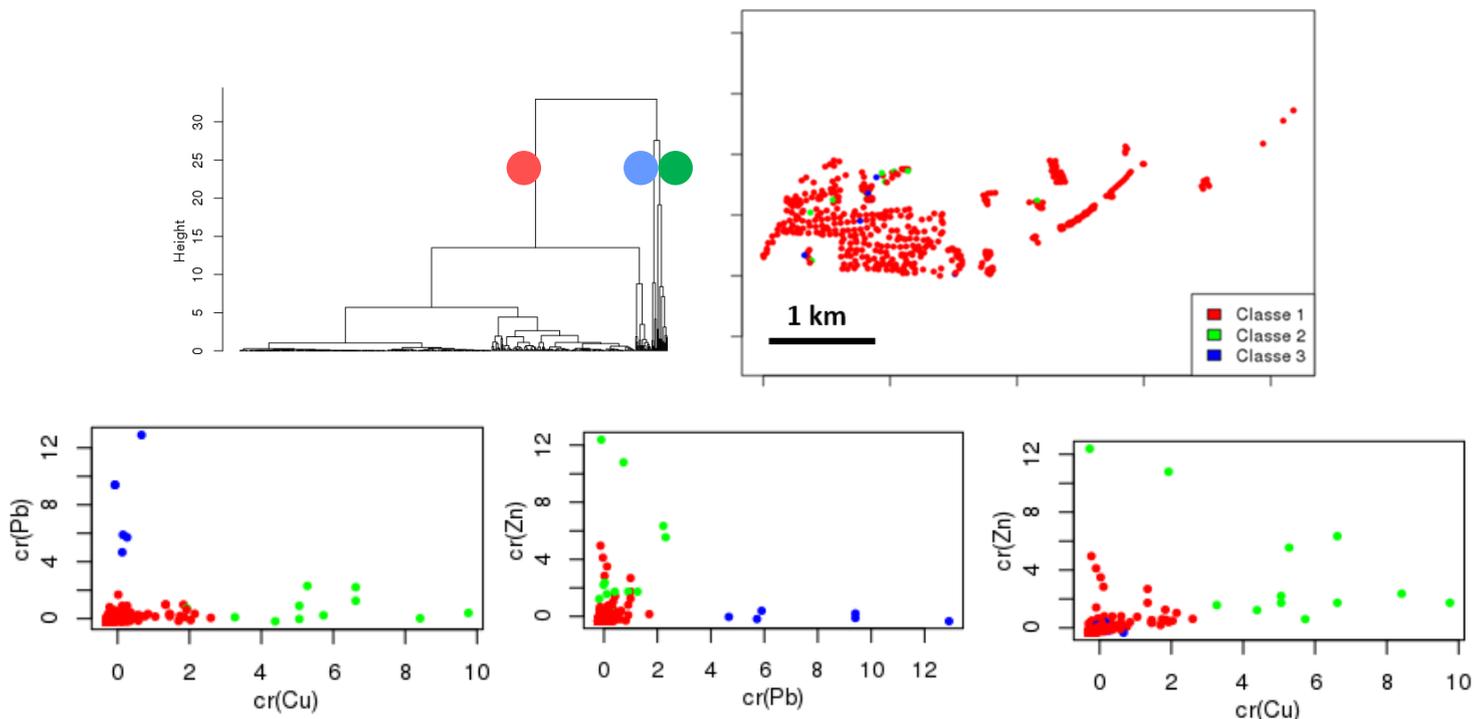


Figure 62: Résultats de classification sur les données centrées-réduites des échantillons de moins de 2 m de profondeur. Distance euclidienne, sans spatialisation.

III.D. Regroupement des échantillons en classes homogènes

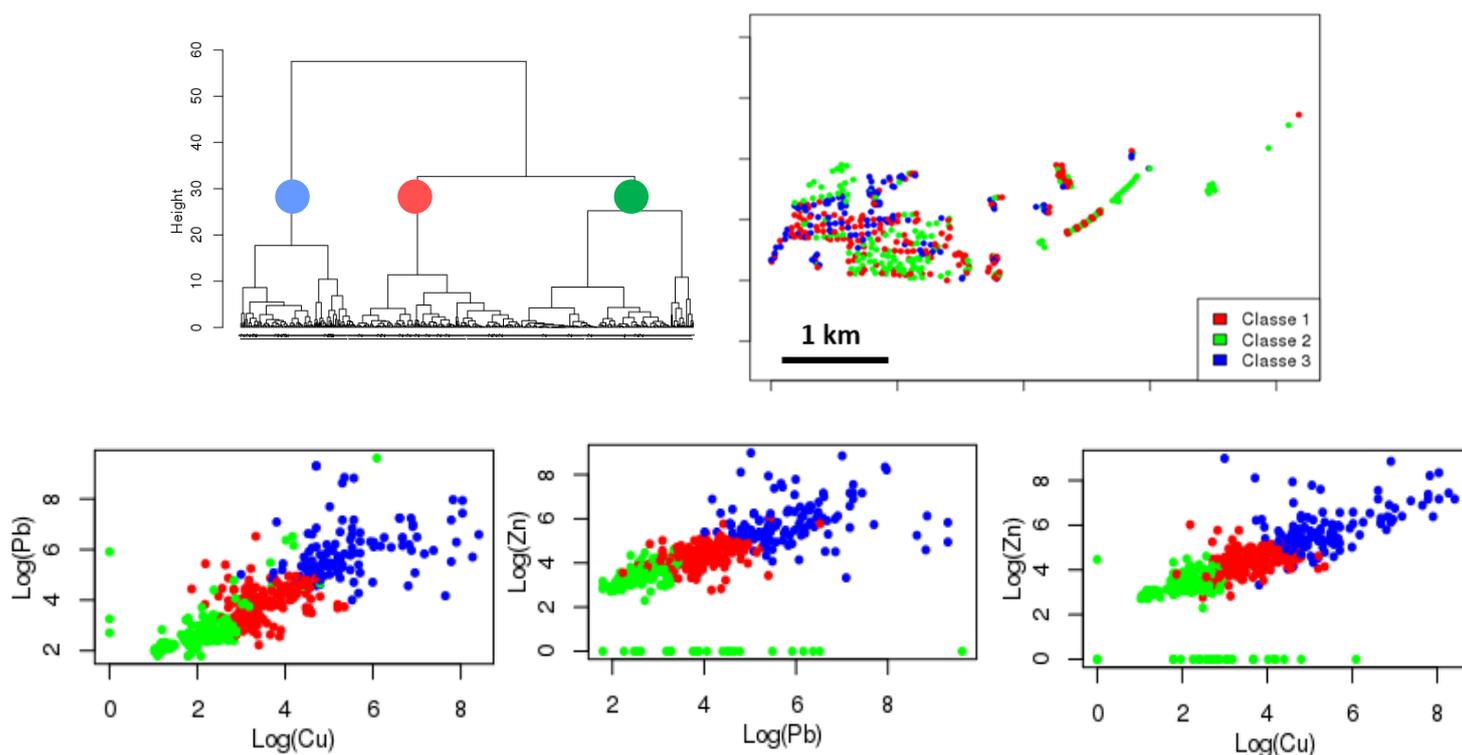


Figure 63 : Résultats de classification sur les données log-translatées des échantillons de moins de 2 m de profondeur. Distance euclidienne, sans spatialisation.

Les résultats indiquent une forte similarité entre les données brutes (Figure 60) et les données réduites (Figure 62), avec les trois mêmes classes identifiées. Les données brutes ayant en effet les mêmes unités et variant dans la même gamme de valeurs, la **réduction a, dans ce cas, peu d'influence sur la classification** pour ces paramètres.

En revanche, les données log-translatées (Figure 63) fournissent des résultats différents, avec une gradation de la contamination, toujours en trois classes. En effet, la classe deux en vert comprend plutôt des teneurs faibles en plomb, cuivre et zinc, alors que la classe trois en bleu comprend les teneurs les plus fortes. La classe une en rouge correspond à des teneurs intermédiaires. Nous obtenons dans ce cas **des classes de teneurs homogènes** présentant trois niveaux de contamination, ce qui correspond davantage à ce que nous recherchons. Cette transformation a pour effet de mieux différencier les valeurs faibles (très présentes dans les données) tout en donnant moins d'influence aux valeurs fortes. Nous utiliserons donc les **données log-translatées** pour la suite des tests.

c) Nous comparons ensuite deux **distances** entre les échantillons, la **distance euclidienne** utilisée précédemment (Figure 63), et la distance basée sur la **matrice de corrélation** des données (Figure 64).

III.D. Regroupement des échantillons en classes homogènes

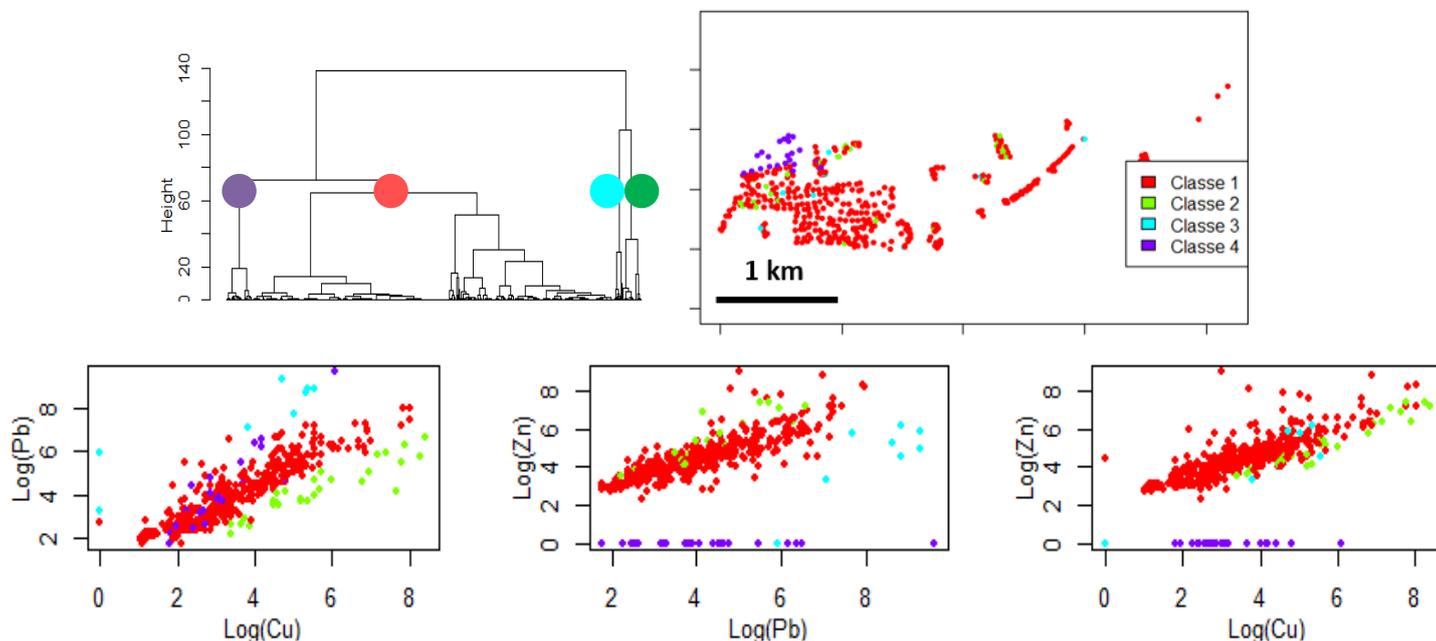


Figure 64 : Résultats de classification sur les données log-translatées des échantillons de moins de 2 m de profondeur. Distance par la matrice de corrélation, sans spatialisation

Avec la distance par la matrice de corrélation (Figure 64) les classes sont construites suivant la **relation entre les paramètres** (métaux). On retrouve une classe majoritaire (rouge) regroupant les données aux teneurs en cuivre, plomb et zinc très bien corrélées, deux autres classes, en cyan et vert, sont différenciés par la corrélation avec le plomb. La classe cyan présente des données aux teneurs en plomb plus fortes qu'en cuivre et zinc, et inversement pour la classe verte. Enfin, une nouvelle classe (violet) apparaît, regroupant des échantillons très pauvres en zinc mais riches en plomb et cuivre. Ces échantillons sont regroupés dans la partie nord-ouest du quartier et correspondent à d'anciens chantiers navals. Les autres classes identifiées via ce calcul de distance ne présentent pas d'application opérationnelle claire.

Pour la suite nous conservons donc la **distance euclidienne** fournissant des résultats plus **en accord avec l'objectif de cartographie de classes homogènes suivant différents niveaux de contamination**. Il est toutefois intéressant d'utiliser d'autres type de distances afin, comme ici, d'identifier de nouvelles classes (violet) potentiellement utiles.

d) Enfin, nous comparons les résultats obtenus **sans spatialisation** (Figure 63) à ceux obtenus par la méthode de *Romary, 2012* avec la **triangulation de Delaunay** (Figure 65).

III.D. Regroupement des échantillons en classes homogènes

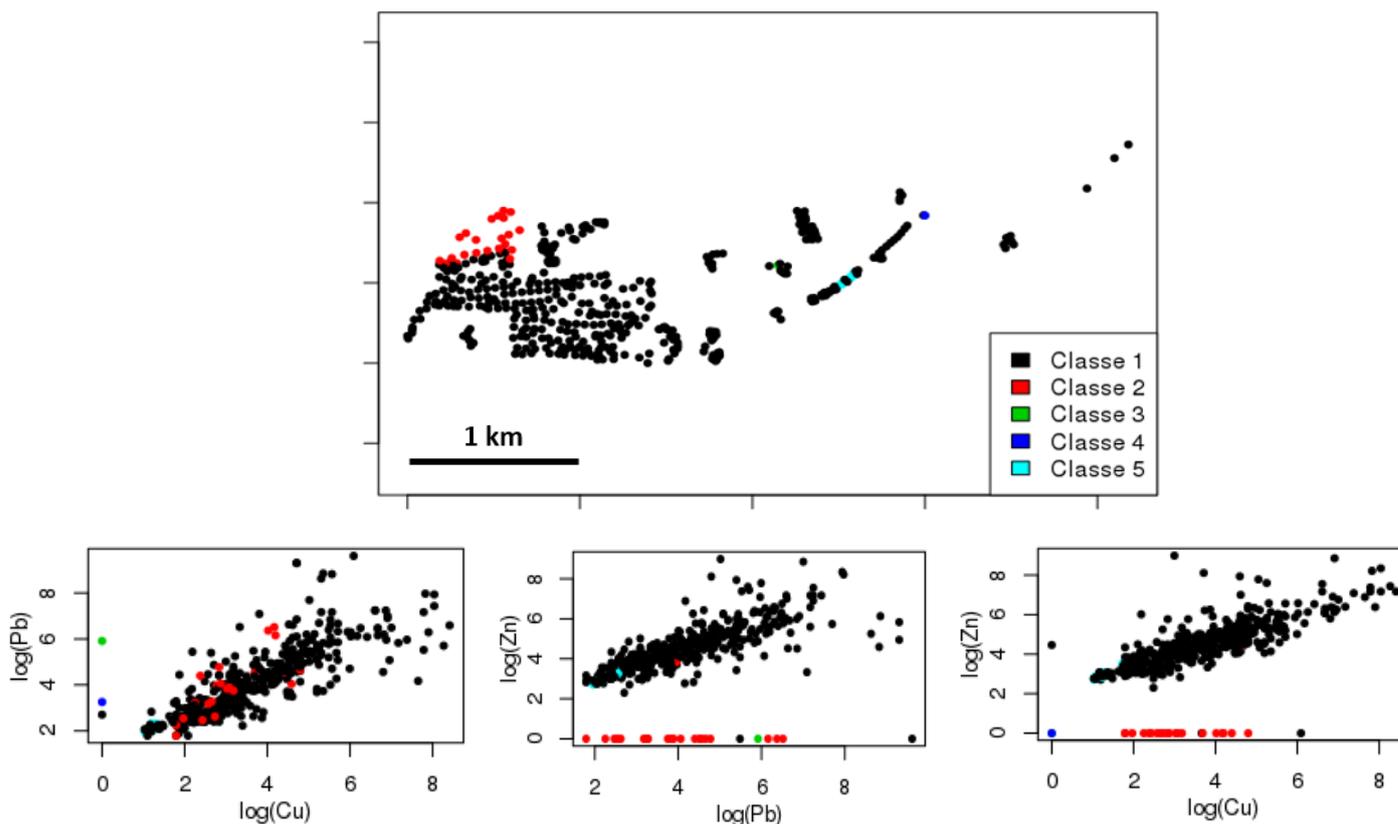


Figure 65 : Résultats de classification sur les données log-translatées des échantillons de moins de 2 m de profondeur. Distance euclidienne, spatialisation par triangulation de Delaunay.

Les résultats avec spatialisation montrent une classe (noire) regroupant la **quasi-totalité des échantillons**. Une seconde classe (rouge) est néanmoins visible correspondant à la classe violette de la Figure 60. Les trois autres classes ne regroupent ensuite qu'un seul échantillon chacune. Le regroupement des échantillons dans une classe très majoritaire s'explique ici par **l'influence du critère de spatialisation**. Le fait que des échantillons doivent être liés par la triangulation pour être associés dans une classe représente une contrainte importante **diminuant l'influence des teneurs** elles-mêmes.

En introduisant une condition de voisinage sur la classification, la version spatialisée entraîne **l'appariement de données plus dissemblables** que la classification classique. En effet, deux données ayant une distance très faible mais éloignées spatialement sont séparées par la classification spatialisée alors qu'une classification sans spatialisation les appairerait. Le gain de connexité des classes est alors compensé par des classes moins homogènes. Ainsi, la méthode **sans spatialisation** permet de constituer des classes plus homogènes.

En conclusion, la **log-translation dans une classification sans spatialisation** semble à privilégier pour une meilleure comparaison des données, puisqu'elle permet, ici, d'obtenir des **classes homogènes présentant des niveaux croissants de contamination**, répondant à l'objectif de regroupement d'échantillons en **classes homogènes et géochimiquement cohérentes**.

III.D.5. Test sur un autre quartier

A des fins de comparaison, nous avons appliqué ces méthodes de classification à un autre quartier de la ville de Nantes. Ce quartier péri-urbain était historiquement occupé par des terrains maraîchers, aujourd'hui en friche. Comme précédemment, les données sont des teneurs en cuivre, plomb et zinc, toutes exprimées en mg/kg de MS. Les dendrogrammes suivants (Figure 66) présentent les résultats des classifications sur les **données brutes, log-translatées et centrée-réduites** (par distance euclidienne). Les nuages de corrélation des teneurs log-translatées sont représentées à la Figure 67 avec le même code couleur que les classes de la Figure 66 qui représente les classes identifiées après étude des dendrogrammes. Les nuages de corrélation des autres classifications (très similaires) ne sont pas représentés.

Les trois classifications isolent toutes le même échantillon (bleu) présentant des teneurs très fortes en cuivre, plomb et zinc. La classe verte présente (à quelques échantillons près) les mêmes échantillons aux teneurs faibles en cuivre, plomb et zinc dans les trois classifications. Enfin, la classe rouge présente les échantillons aux teneurs intermédiaires.

Pour ce quartier, avec un effectif réduit d'échantillons, les résultats des deux méthodes de classifications sont très similaires et présentent une gradation des contaminations (des plus faibles aux plus élevées).

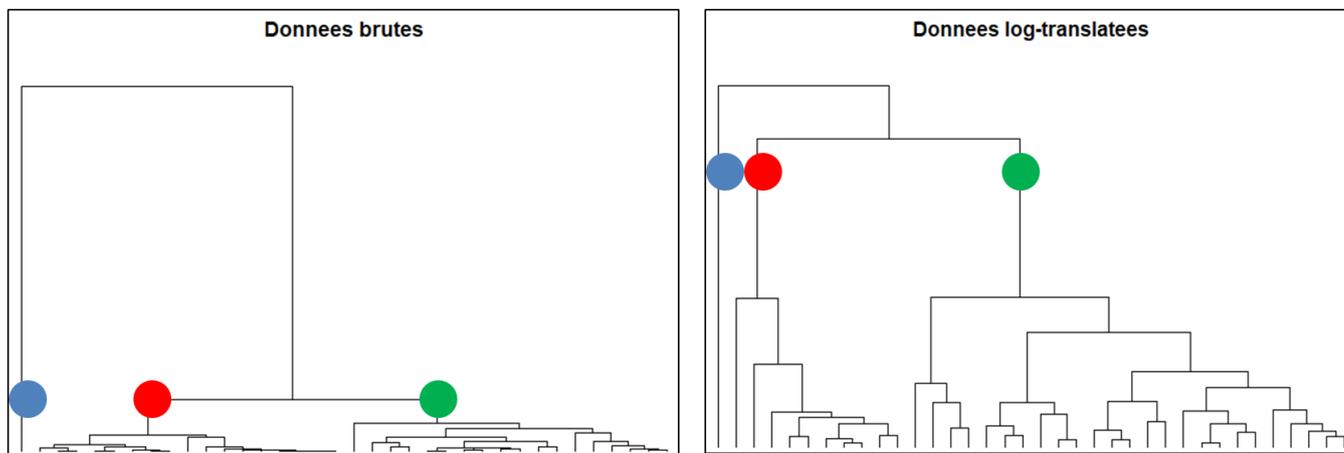


Figure 66 : Dendrogrammes des classifications sur les teneurs en cuivre, plomb et zinc brutes et log-translatées du quartier des Gohards. Les points de couleurs représentent les trois classes distinguées.

III.D. Regroupement des échantillons en classes homogènes

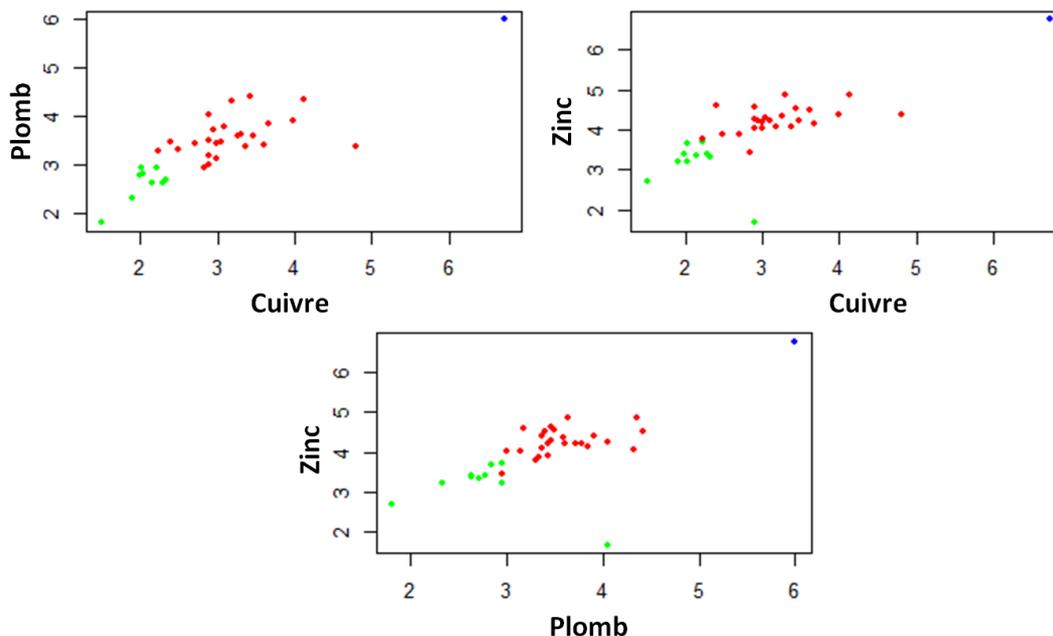


Figure 67: Nuages de corrélation entre les teneurs en cuivre, plomb et zinc suivant les trois classes identifiées par la classification sur données log-translatées du quartier des Gohards.

En revanche, les résultats obtenus via la classification par un calcul de distance avec la matrice de corrélation sont toutefois différents. On remarque, en effet la présence d'une classe fortement majoritaire mais relativement homogène (en rouge dans la figure 68). Les deux autres classes (en bleu et vert sur la figure 68) ne comprennent respectivement qu'un et deux échantillons. La classe bleue isole un échantillon aux teneurs très faibles en zinc mais pas en cuivre, s'éloignant de la corrélation habituellement observée entre les deux substances. La classe verte, quant-à-elle isole des teneurs relativement fortes dans les deux substances. Les résultats présentent donc bien trois classes bien distinctes aux comportements différents, contrairement au cas précédent où la distinction est plutôt faite sur une gradation de la contamination. Le nuage de corrélation entre les teneurs en cuivre et en zinc est fourni en Figure 68.

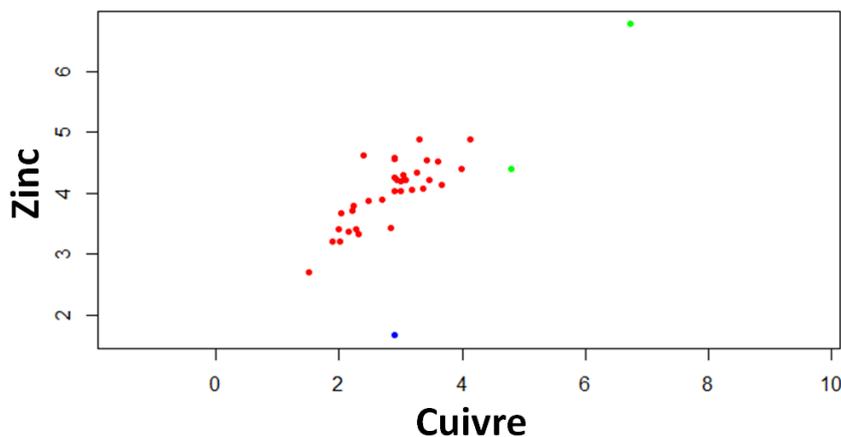


Figure 68: Nuage de corrélation des teneurs en cuivre et zinc suivant les classes identifiées par la classification sur les teneurs log-translatées du quartier des Gohards, avec une distance calculée via la matrice de corrélation.

III.D. Regroupement des échantillons en classes homogènes

On remarque toutefois que du fait de la forte hétérogénéité de la géochimie urbaine, il est complexe d'obtenir des classes homogènes d'échantillons qui seraient spatialement continues. Cela complique notamment le travail de cartographie. En effet, s'il est **possible d'interpoler les résultats**, notamment par des polygones d'influence, les **limites seraient alors définies par l'échantillonnage**. Le découpage ne correspondrait donc **pas à des limites ou des critères physiques** et serait donc difficilement identifiable et exploitable sur le terrain.

III.E. Conclusions sur la prise en compte de l'hétérogénéité du milieu et les indicateurs possibles

Les données descriptives sont une source d'informations importante pour la connaissance du sol urbain, il est donc judicieux d'en tenir compte. Cela est d'autant plus vrai pour le fond pédo-géochimique anthropisé, pour lequel les algorithmes statistiques fournissant les valeurs seuils ne tiennent pas compte de la composante spatiale des données.

Un moyen de tenir compte de cette spatialisation dans ces algorithmes est donc de les appliquer à des données triées selon une typologie cartographiable.

Nous avons comparé deux définitions différentes de la typologie, en partant soit des descripteurs, soit des analyses géochimiques.

L'analyse multi-critère à partir des descripteurs déterminant la géochimie des sols urbains permet de cartographier un indice de potentiel de présence d'anomalie géochimique dont la précision dépend fortement de la résolution des données cartographiques de base. En outre, l'indicateur ainsi cartographié peut, en lui-même, représenter un outil d'aide à la décision pour les aménageurs en permettant de tenir compte des zones potentiellement contaminées dans les plans d'aménagements. Néanmoins, le traitement à l'échelle d'un quartier nécessite, pour une bonne précision, de recourir à des données complexes et coûteuses à obtenir (IHU, modèle 3D). Un équilibre est donc à trouver entre l'apport de ces outils et leur coût par rapport à l'acquisition de données selon un maillage suffisamment serré, la recherche des origines des anomalies éventuelles et le risque de moins bien anticiper les anomalies. De plus, la profondeur de validité du modèle établi est liée à la profondeur d'influence des activités anthropiques. Si une estimation de cette profondeur limite ne peut pas être fournie, une profondeur limite fixe peut être choisie.

La classification a permis de mettre en évidence des classes géochimiquement homogènes d'échantillons. Si l'application des algorithmes statistiques sur ces classes peut permettre d'obtenir différents niveaux de fond pédo-géochimique anthropisés, dans ce cas il n'est pas possible de fournir une cartographie claire de l'emprise géographique concernée par ces valeurs qui aurait un sens physique. Une solution pourrait être d'introduire des covariables descriptives en utilisant les facteurs résultant d'une analyse en composantes principales (ACP).

En conclusion, pour l'utilisation des algorithmes statistiques de calcul d'une valeur seuil du fond pédo-géochimique, il est conseillé de recourir à un découpage du territoire par analyse multi-critère. Ceci afin de tenir compte de l'hétérogénéité du milieu et de fournir une emprise géographique claire aux résultats.

Néanmoins il est nécessaire de tenir compte de certaines limites de ce découpage. La première est la taille des polygones. En effet, avec l'augmentation de la précision des données

III.E. Conclusions sur la prise en compte de l'hétérogénéité du milieu et les indicateurs possibles

descriptives de base, la taille des polygones diminue. Si cela permet une plus grande précision, ce n'est pas toujours optimal pour l'utilisateur. Un regroupement des polygones est toutefois possible en cherchant un équilibre entre simplicité de lecture et précision.

De plus, certaines entités ne présentent qu'un nombre limité d'analyses. Des données complémentaires seraient donc nécessaires afin de vérifier plus précisément la cohérence de ces découpages.

A l'échelle de la ville, afin de tenir compte des données ponctuelles Basias, il est possible de leur attribuer une zone d'influence. Cette information n'ayant pas la même précision spatiale que les autres données descriptives (une donnée ponctuelle), elle ne peut être utilisée que comme une information complémentaire superposée à la carte de potentiel d'anomalie.

La prise en compte de co-variables comme l'occupation du sol permet donc de cartographier la qualité géochimique des sols urbains à partir des calculs statistiques de seuils. Toutefois, les seuils statistiques ne peuvent pas garantir la séparation du « fond » et des « anomalies » dans tous les cas de figure

Nous allons donc examiner une autre démarche, de calcul d'un fond pédogéochimique anthropisé tenant compte de l'implantation des données et de la variabilité spatiale des teneurs, et permettant de séparer le fond des anomalies suivant un critère d'échelle de variabilité spatiale. Cette démarche est fondée sur la géostatistique.

Partie IV : SEPARER FOND ET ANOMALIES PAR FILTRAGE GEOSTATISTIQUE

Compte-tenu de la difficulté à disposer d'un nombre suffisant de données dans les entités géographiques géochimiquement cohérentes pour déterminer le fond géochimique à partir d'une approche statistique, nous la contribution de la géostatistique à l'estimation du fond pédo-géochimique anthropisé.

A la différence de la statistique usuelle, dans laquelle les données sont considérées comme des tirages aléatoires indépendants d'une même Variable Aléatoire, et qui ne tient donc pas compte de leur localisation, la géostatistique prend en compte cette localisation. La géostatistique traite des variables dites régionalisées (Emery, 2001), c'est-à-dire présentant une structure spatiale. Intuitivement, les teneurs de deux points voisins se ressemblent plus que celles de deux points éloignés. Un exemple de variable régionalisée est l'altitude ; connaissant l'altitude en un point, il est plus précis de prévoir celle en un autre point distant d'un mètre qu'en un point distant d'un kilomètre.

La géostatistique (cf. §1.A.2.3) est l'application de la modélisation probabiliste à l'étude des variables régionalisées et à leur estimation (Matheron, 1962 ; 1970). La géostatistique linéaire nécessite des modèles à l'ordre deux, c'est-à-dire posant des hypothèses sur l'espérance, la variance et covariance de la variable ou de ses incréments. Le variogramme expérimental, qui représente (au facteur $\frac{1}{2}$ près), l'écart-quadratique moyen en fonction de la distance, permet de caractériser la variabilité spatiale (régularité de la variable, stationnarité ou non stationnarité, anisotropies, etc.). Le variogramme expérimental, qui ne vérifie pas les propriétés mathématiques voulues, est ajusté par un modèle variographique, utilisé pour l'estimation. Le variogramme croisé est la covariance des accroissements de deux variables. Parmi les modèles multivariés, le modèle linéaire de corégionalisation consiste à ajuster l'ensemble des variogrammes simples et croisés des variables étudiées par des combinaisons linéaires de variogrammes élémentaires, vérifiant certaines contraintes (Wackernagel, 2003).

Le krigeage et sa version multivariée, le cokrigeage est l'estimateur linéaire optimal, au sens de l'absence de biais de l'estimateur, et de la meilleure précision (par minimisation de la variance de l'erreur d'estimation). Ces estimations permettent de tracer des cartes de la variable estimée, à partir des données acquises aux points expérimentaux. Le krigeage (et le co-krigeage) fournissent également la précision de l'estimation, par la carte de l'écart-type d'erreur d'estimation.

Les estimateurs non-linéaires (espérance conditionnelle, krigeage disjonctif ; Matheron, 1976 ; Hani et Pazira, 2011) permettent une estimation cohérente de la variable et de (presque) toute fonction de cette variable. En particulier, on en déduit des cartes de probabilité de dépassement d'un seuil de qualité. Le krigeage disjonctif nécessite de spécifier la loi bi-

variable, tandis que l'ensemble de la loi spatiale est requise pour l'estimation par espérance conditionnelle et pour les simulations.

D'autres modèles ont été développés récemment, comme le modèle d'écrêtage (Rivoirard et al., 2013). Dans ce modèle, les teneurs supérieures au seuil d'écrêtage sont supposées délocalisées dans le lieu des teneurs fortes, et sont remplacées, pour l'estimation, par la moyenne des teneurs dépassant ce seuil. Cependant, il n'y a pas de raison pour que le seuil d'écrêtage corresponde au fond pédo-géochimique recherché.

L'analyse krigeante, qui est une variante du krigeage (cas monovarié) ou du cokrigeage (cas multivarié), consiste à décomposer une variable en composantes spatiales associées aux différentes échelles de variabilité, mises en évidence par la variographie. Ceci permet de définir comme « anomalies » les composantes présentant une variabilité à petite distance (courte portée), et comme « fond » les composantes de plus grande portée, éventuellement intrinsèques strictes (par exemple, admettant un variogramme linéaire, donc sans portée finie). La séparation du fond et des anomalies peut s'effectuer alors suivant un critère d'échelle de variabilité spatiale. C'est ce que s'attache à faire l'analyse krigeante (Bourennane et al, 2003 ; Goovaerts, 1992 ; Saby et al., 2009; Wackernagel, 1993).

Revenant à la définition d'un « fond » et « d'anomalies », nous examinons dans la suite si leur séparation par analyse krigeante (Arnaud et al., 2001 ; Matheron, 1982 ; Wackernagel, 1993) fournit des résultats satisfaisants. Le « filtrage » désigne l'estimation de la composante de « grande portée », calculée par l'analyse krigeante. Deux échelles de modélisation sont examinées : l'échelle du site, et celle d'un quartier urbain.

Dans sa forme actuelle cette démarche ne vise plus à fournir directement une unique valeur seuil mais une cartographie du fond pédo-géochimique anthropisé, à partir de laquelle une valeur seuil pourra être déterminée (Figure 69).

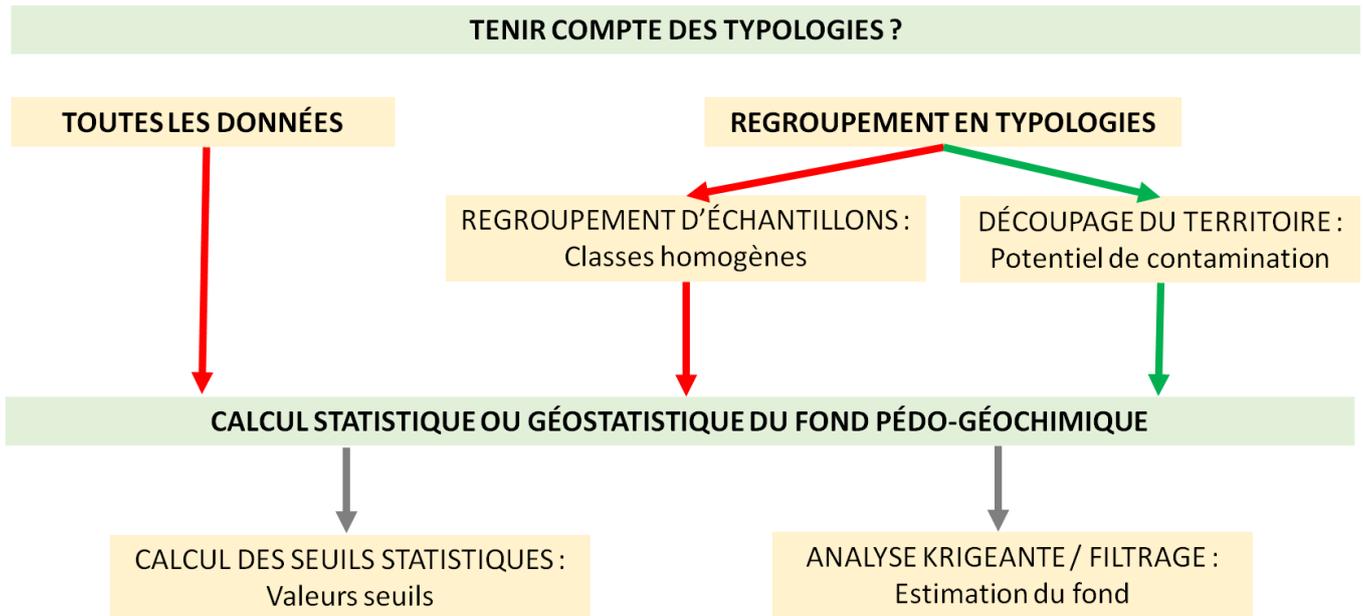


Figure 69: Schéma conceptuel de l'évolution de la méthodologie de détermination du fond pédogéochimique anthropisé intégrant la géostatistique pour l'estimation d'une gamme de valeurs de fond.

IV.A. Le filtrage par analyse krigéante

IV.A.1. Une autre définition d'une anomalie

Selon Reimann (2005), « le fond pédo-géochimique (anthropisé) est une gamme de valeurs au-delà de laquelle les concentrations ne peuvent plus être considérées comme reflétant l'environnement général ». Nous cherchons donc comment séparer le fond pédo-géochimique des concentrations élevées dans le cas où les données utilisées n'ont pas été acquises de façon systématique.

Comme on l'a vu précédemment, les algorithmes statistiques de calcul de valeurs seuil du fond pédo-géochimique sont mal adaptés au traitement des données non spécifiques au fond pédo-géochimique, même après pondération selon la densité locale des données. Ces seuils peuvent éventuellement permettre de séparer le fond des anomalies géochimiques présentes dans le jeu de données.

Lorsque les anomalies sont définies comme ne reflétant pas le milieu, elles ne correspondent pas nécessairement aux valeurs les plus fortes. Le « milieu » considéré dépend évidemment de l'extension de la zone d'étude. L'anomalie varie donc selon l'échelle de travail.

En cas d'échantillonnage non spécifique au fond géochimique, les anomalies peuvent représenter une part importante des données. Les seuils issus du calcul statistique peuvent conduire à intégrer une part des anomalies dans le fond.

La question suivante se pose donc : comment définir l'anomalie à supprimer pour en déduire le fond ? Examinons le schéma suivant (Figure 70) :

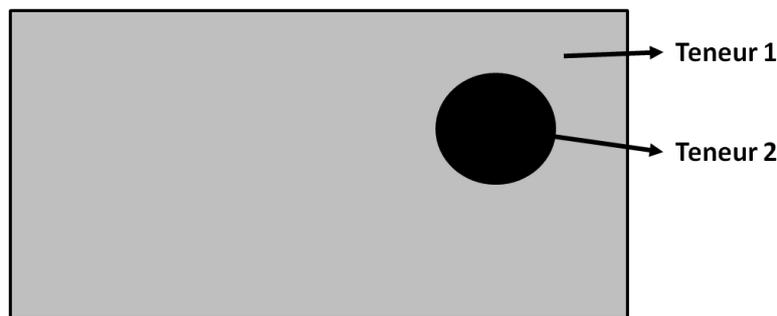


Figure 70: Schéma conceptuel de l'anomalie (en noir) et du fond (en gris).

Dans ce schéma, la teneur 2 de l'anomalie (en noir) est bien différente de la teneur 1 présente sur le reste de la zone. C'est donc une anomalie dans le sens où elle ne reflète pas la zone d'étude. Néanmoins, il n'est pas précisé que la teneur 2 de l'anomalie soit nécessairement supérieure à la teneur 1 du fond. C'est le fait qu'elle soit différente et ne concerne qu'une zone restreinte qui la définit comme anomalie.

IV.A. Le filtrage par analyse krigéante

Deux aspects sont donc à préciser :

- la notion de « différence » entre les teneurs : à partir de quel écart peut-on les séparer ?
- la notion de zone représentative : quelle part de la zone l'anomalie doit-elle couvrir pour être considérée comme le fond ?

Dans la réalité, les différences ne sont pas aussi tranchées entre teneurs du fond et anomalies.

Dans l'hypothèse où nous serions à même de connaître la géochimie exhaustivement, comment différencier l'anomalie ?

Dans le cas de la géochimie, il semble qu'un phénomène diffus puisse être associé aux variations à grande portée de la teneur tandis qu'une anomalie, liée à un phénomène ponctuel ou très localisé, correspondrait aux variations à courte portée de la teneur. Ce principe va permettre de différencier les anomalies suivant l'échelle de variabilité spatiale (Figure 71).

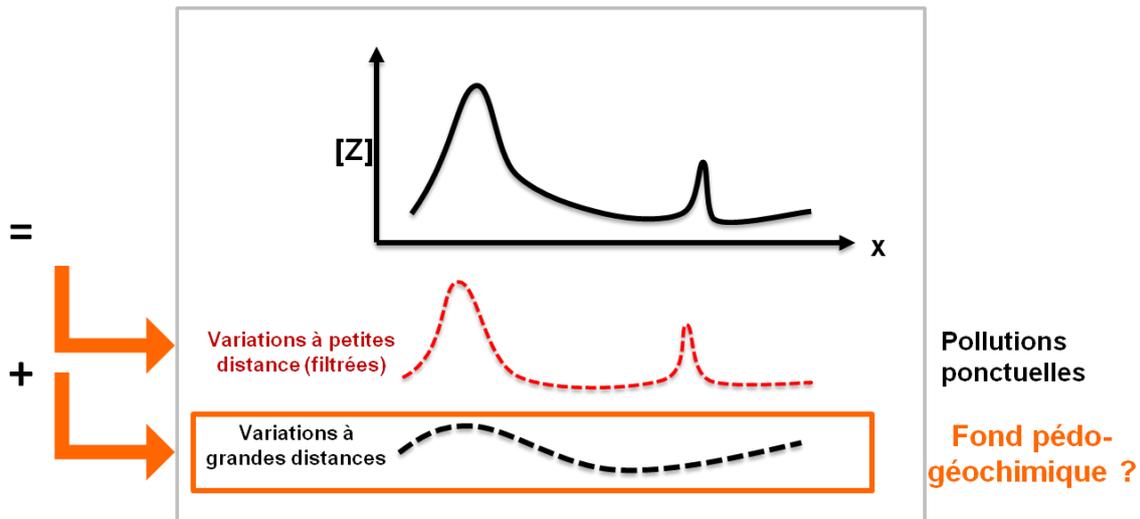


Figure 71: Séparation des anomalies (en rouge) du fond (pointillés noirs) d'une variable Z le long d'un linéaire x , selon l'échelle de variabilité spatiale.

Dans cette figure la variable initiale Z est composée de deux termes associés aux deux échelles de variabilité spatiale, bien distinctes. Les variations à petites distances (variations rapides) peuvent être filtrées afin de ne garder que les variations à grandes distances (lentes). Cette décomposition s'effectue par le filtrage monovarié ou multivarié fondé sur le modèle linéaire de corégionalisation (Bourennane et al, 2003 ; Goovaerts, 1992 ; Saby et al., 2009 ; Wackernagel, 1993), appelé aussi analyse krigéante.

IV.A.2. Principe de l'analyse krigéante

L'analyse krigéante est d'abord présentée dans le cas stationnaire. Puis le modèle linéaire de corégionalisation et l'analyse krigéante sont introduits dans le cas bivariable.

a) Cas monovarié

On se place d'abord dans le cas stationnaire.

Considérons une teneur d'espérance (de moyenne probabiliste) μ et dont le variogramme γ comporte deux structures γ^1 et γ^2 :

$$\gamma(h) = b^1\gamma^1(h) + b^2\gamma^2(h)$$

Cette teneur peut être modélisée par une Fonction aléatoire $Z(x)$, combinaison linéaire de deux composantes d'espérance nulle $Z^1(x)$ et $Z^2(x)$, spatialement non corrélées (et donc de variogramme croisé nul), et de variogrammes respectifs $\gamma^1(h)$ et $\gamma^2(h)$, à laquelle se rajoute l'espérance μ :

$$Z(x) = Z^1(x) + Z^2(x) + \mu$$

Lorsque les portées des deux composantes sont clairement différentes, il est logique d'associer l'espérance à la composante de plus grande portée, ce qui s'écrit :

$$Z(x) = Z^1(x) + Y^2(x)$$

Où Y^2 admet la même espérance μ que l'estimation de la composante $Y^2(x) = Z^2(x) + \mu$ à partir des données de teneurs s'effectue par krigéage. Y^2 ayant même espérance que Z , la condition de non biais impose classiquement que la somme des poids de cokrigéage soit égale à l'unité. Classiquement, la matrice de cokrigéage fait intervenir le variogramme γ entre les données, tandis que le second membre du système fait intervenir le variogramme croisé entre données et valeur à estimer, c'est-à-dire $b^1\gamma^1(h)$.

Pour l'estimation de la composante $Z^1(x)$, supposée d'espérance nulle, on montre que la condition de non biais impose que la somme des poids soit nulle. Le second membre du système de cokrigéage fait intervenir le variogramme croisé entre données et valeur à estimer, dans ce cas $b^2\gamma^2(h)$.

Ces résultats se généralisent au cas intrinsèque, lorsque l'une des composantes admet un variogramme linéaire par exemple. Les conditions de non-biais deviennent des conditions d'autorisation.

b) Cas multivariable

Dans la suite nous utiliserons les propriétés du modèle linéaire de corégionalisation (Rivoirard, 2003).

Dans ce modèle, les variogrammes simples et croisés de toutes les variables sont des combinaisons linéaires des mêmes composantes structurales, ce qui s'écrit

$$(6) \quad \gamma_{ij}(h) = \sum_k b_{ij}^k \gamma^k(h)$$

où i et j sont deux variables, k l'indice des différentes composantes du modèle linéaire, et h la distance entre les points. Sous forme matricielle :

$$(7) \quad [\gamma_{ij}(h)] = \sum_k [b_{ij}^k] \gamma^k(h)$$

où les matrices des paliers $[b_{ij}^k]$ doivent être de type positif.

Les variables elles-mêmes se décomposent en combinaisons linéaires de composantes spatiales correspondant aux différentes échelles de variabilité :

$$Z_i(x) = \mu_i + \sum_k Z_i^k(x)$$

où les composantes spatiales sont de moyenne nulle.

Comme précédemment, pour chaque variable, la « moyenne » μ_i peut être ajoutée à la composante Z_i^k de plus grande portée.

Afin de faire le lien avec notre problématique, considérons le cas de deux variables fictives (Rivoirard, 2003) (indicées 1 et 2) dont les variogrammes simples et croisés s'écrivent :

$$\gamma_1(h) = b_{11}^1 \gamma^1(h) + b_{11}^2 \gamma^2(h)$$

$$\gamma_2(h) = b_{22}^1 \gamma^1(h) + b_{22}^2 \gamma^2(h)$$

$$\gamma_{12}(h) = b_{12}^1 \gamma^1(h) + b_{12}^2 \gamma^2(h)$$

Ce modèle à deux structures gigognes (1 et 2), comporte par exemple un effet de pépite et une structure sphérique de portée 100m, tels que :

$$\gamma_1(h) = 10 \text{Pep}(h) + 50 \text{Sph}(h/100)$$

IV.A. Le filtrage par analyse krigéante

$$\gamma_2(h) = 20 \text{Pep}(h) + 60 \text{Sph}(h/100)$$

$$\gamma_{12}(h) = 30 \text{Sph}(h/100)$$

Les matrices de paliers $(b^1) = (b_{ij}^1) = \begin{pmatrix} 10 & 0 \\ 0 & 20 \end{pmatrix}$ et $(b^2) = (b_{ij}^2) = \begin{pmatrix} 50 & 30 \\ 30 & 60 \end{pmatrix}$, correspondent aux matrices de variances-covariances pour chaque structure spatiale du modèle.

Dans cet exemple, l'effet de pépite est absent du variogramme croisé. En effet, une composante structurale présente dans une structure croisée doit nécessairement apparaître dans les structures simples des deux variables. Mais une composante structurale présente dans les deux structures simples, n'est pas nécessairement présente dans la structure croisée.

Chaque variable peut s'interpréter comme la somme de deux composantes $Z_i^1(x)$ (effet de pépite) et $Z_i^2(x)$ (sphérique), sans corrélation spatiale mutuelle et d'espérance nulle :

$$(8) \quad Z_i(x) = Z_i^1(x) + Z_i^2(x) + \mu_i$$

où μ_i est la moyenne de la variable.

Nous cherchons à estimer la composante associée au fond géochimique. Nous l'avons vu précédemment, les structures à filtrer seraient donc celles à courte portée vis-à-vis de la zone d'étude. Dans notre exemple nous considérons que la composante pépitique reflète les contaminations ponctuelles et la composante sphérique, le fond pédo-géochimique.

Pour préciser ce choix, il est également possible de s'intéresser à la matrice de corrélation déduite des matrices de variances-covariances calculable globalement ou par structure selon la formule :

$$(9) \quad r = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

Ce qui donne $\begin{pmatrix} 1 & 0.55 \\ 0.55 & 1 \end{pmatrix}$ c'est-à-dire un coefficient de corrélation de 0.55 pour la deuxième structure, et $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ c'est-à-dire un coefficient de corrélation nul pour la première structure, puisque l'effet de pépite est absent du variogramme croisé.

La matrice de corrélation par structure spatiale traduit des liaisons différentes entre les variables selon l'échelle de variabilité spatiale. Ce résultat est particulièrement intéressant entre la teneur étudiée et les covariables descriptives, permettant de sélectionner la ou les composantes à garder dans l'estimation du fond pédo-géochimique.

IV.A. Le filtrage par analyse krigéante

D'après la formule (8), sélectionnant uniquement la composante sphérique de la teneur Z_1 pour l'estimation du fond, nous posons par exemple :

$$(10) \quad Z_{Fond}(x) = Z_1^2(x) + \mu$$

Cette composante peut être estimée par cokrigéage à partir des données de Z_1 , et d'une covariable Z_2 , éventuellement connue de façon plus dense, et dont la composante sphérique est suffisamment corrélée à celle de la teneur. Le filtrage se généralise au cas de plusieurs covariables.

Remarques :

- Les composantes $Z^i(x)$ identifiées dans la formule (8) sont chacune d'espérance nulle sauf l'une d'entre elles à laquelle on ajoute la moyenne. Cela implique que les composantes filtrées sont d'espérance nulle, et comportent donc des valeurs négatives. Cette méthode filtre donc les anomalies positives comme les contaminations ponctuelles, mais aussi des teneurs localement basses (anomalies négatives). Contrairement aux méthodes statistiques, cette méthode de filtrage met en évidence des anomalies qui ne sont pas exclusivement des contaminations.

La carte du fond pédo-géochimique (estimé) ainsi obtenue est localement supérieure à celle de la teneur (estimée). Ce résultat est intéressant pour la gestion des terres excavées.

- Les teneurs estimées seront influencées par la moyenne des données. Cela peut avoir comme conséquence un biais dans l'estimation si la moyenne des données utilisées n'est pas représentative de la zone, par exemple si les teneurs basses ne sont pas échantillonnées (ce qui peut être le cas d'un échantillonnage exclusivement issu de la reconnaissance de sites pollués, par exemple).

- Le filtrage a été présenté dans sa version de référence ; des variantes peuvent être étudiées.

IV.B. Application à un site

IV.B.1. Sélection du site et des données

Pour le filtrage nous utilisons comme précédemment les données de diagnostics de pollution. Provenant de différents bureaux d'études, ces données correspondent à des protocoles d'échantillonnage (plan et mode d'échantillonnage) et d'analyse différents. Afin de tester la pertinence de la méthode sans être influencé par les différents protocoles, nous avons choisi de traiter, dans un premier temps, un seul site du quartier.

Le site sélectionné est un projet de réaménagement de grande ampleur du quartier, sur lequel 208 échantillons ont été collectés par un même bureau d'études au cours de deux opérations d'échantillonnage (Figure 72). A noter que le faible nombre d'échantillons par sondage empêche la modélisation d'un variogramme vertical.

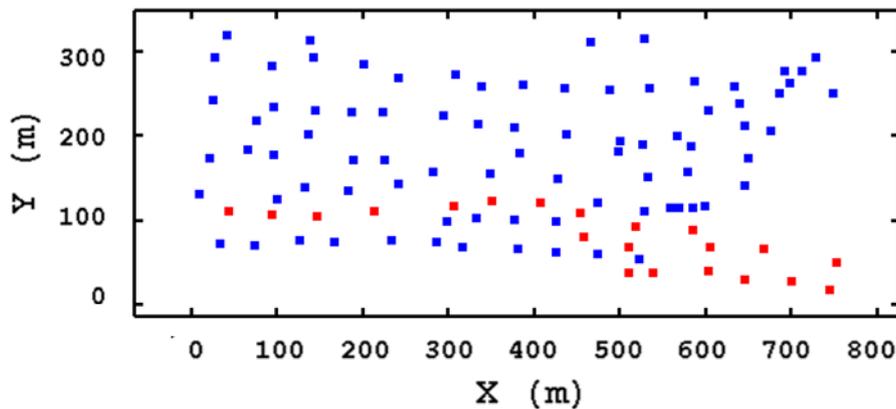


Figure 72: Plan d'implantation des sondages sur lesquels ont été prélevés les 208 échantillons. Les deux opérations sont indiquées en bleu et rouge.

Le protocole d'échantillonnage mis en place par le bureau d'étude, consiste à prélever des échantillons en séparant les différentes couches rencontrées (§ I.B.2). Une représentation de ces échantillons est fournie en annexe 9.

Une première étude exploratoire a été menée, afin de comparer les échantillons aux typologies de remblais (Le Guern et al, 2016) (Figure 73). Nous n'utiliserons pas, ici, l'occupation du sol actuelle (Corine Land Cover ou Copernicus) identique sur l'ensemble du site. La Figure 73 montre que les trois types de remblais sont présents depuis la surface jusqu'à 5 m de profondeur, mais sont plus nombreux en surface. En revanche, les alluvions (en vert) sont limitées aux échantillons les plus profonds (plus de 3,5 m de profondeur).

IV.B. Application à un site

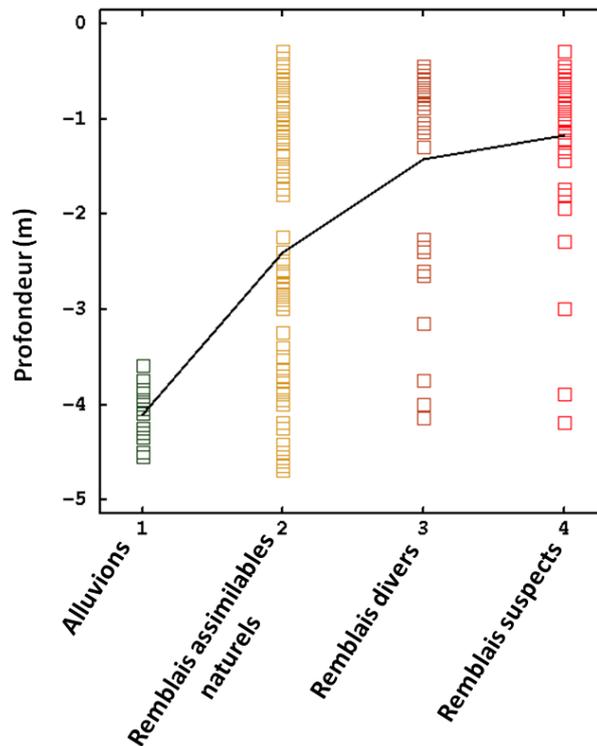


Figure 73: Nuage de corrélation entre les profondeurs des échantillons du site et les typologies de matériaux définies par (Le Guern et al. 2016). Le trait plein relie les moyennes des profondeurs par classe.

En représentant les teneurs d'une substance de chacune des trois familles identifiées dans la partie 1 (métaux anthropogéniques, géogéniques et hydrocarbures) en fonction de la profondeur des échantillons et du type de remblais (Figure 74), on observe deux tendances. Les métaux/métalloïdes anthropogéniques (Cu) et les hydrocarbures (C10-C40) sont plus présents en surface et dans les remblais, avec des teneurs plus fortes dans les premiers mètres de profondeur. A l'inverse, les métaux/métalloïdes géogéniques (Cr) sont majoritairement présents en profondeur dans les alluvions. On peut donc parler ici d'un fort impact anthropique dans les deux premiers mètres de profondeur qui s'atténue au-delà.

IV.B. Application à un site

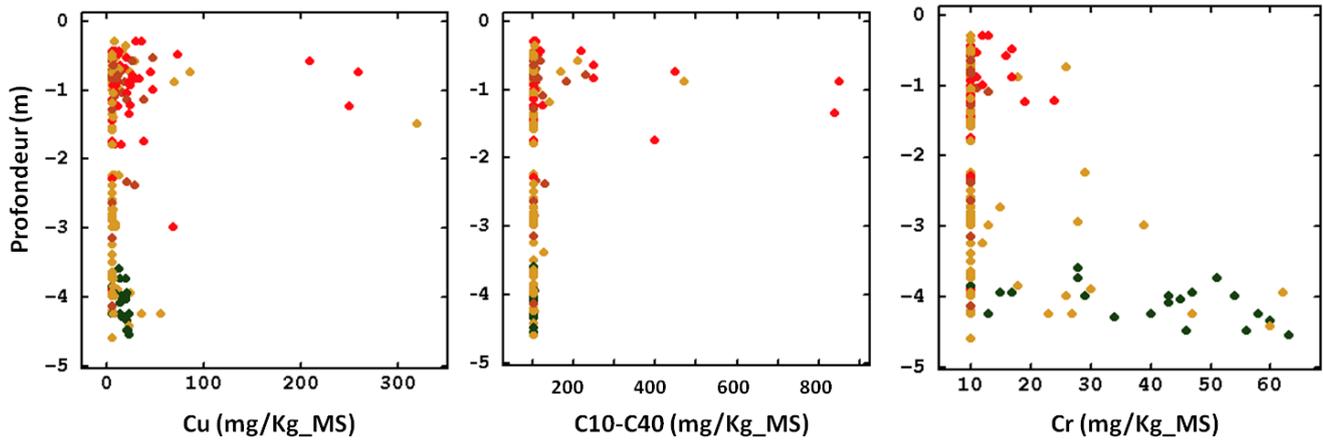


Figure 74: Nuages de corrélation des concentrations en cuivre, hydrocarbures et chrome dans les échantillons du site en fonction de la profondeur et du type de remblais (rouge : remblais suspects, marron : remblais divers, orange : remblais assimilés naturels et vert foncé : alluvions).

Comme le protocole mis en place pour l'échantillonnage sépare les différentes couches rencontrées, la longueur des échantillons est variable.

Pour s'affranchir de l'influence du support, l'analyse exploratoire est d'abord menée sur des échantillons de même longueur ou de longueur voisine.

La Figure 75 présente l'histogramme des longueurs des 208 échantillons, qui varient entre 0,1 et 2,3 m. Pour obtenir un groupe d'échantillons de longueur (support) comparable nous avons sélectionné des échantillons de longueur comprise entre 50 et 70 cm (inclus), en gris sur la Figure 75, correspondant au mode principal, avec 86 échantillons.

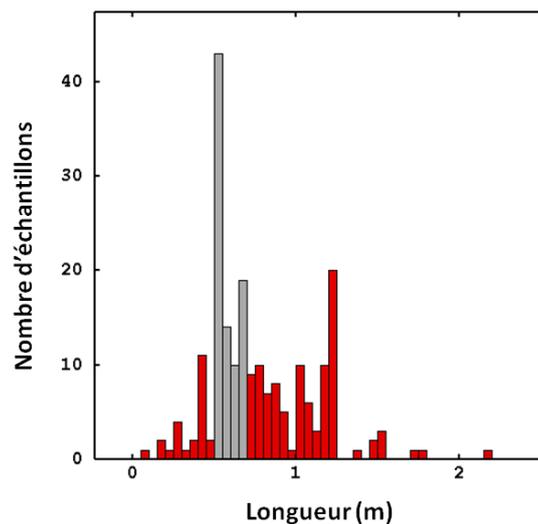


Figure 75: Longueur des 208 échantillons du site. En gris les 86 échantillons de longueur comprise entre 50 et 70 cm, sélectionnés pour la suite de l'analyse exploratoire.

L'histogramme de la profondeur moyenne des 86 échantillons (Figure 76) présente deux modes. Le premier regroupe les échantillons des deux premiers mètres de profondeur et

IV.B. Application à un site

le second les plus profonds, autour de quatre mètres. Or, les résultats de l'étude exploratoire (Figure 73 et Figure 74) indiquent que les deux premiers mètres présentent principalement des remblais et les couches plus profondes des alluvions naturels, l'impact anthropique étant particulièrement marqué dans ces deux premiers mètres. Afin de ne pas mélanger ces deux environnements différents dans le traitement nous avons d'abord sélectionné les 53 échantillons des deux premiers mètres.

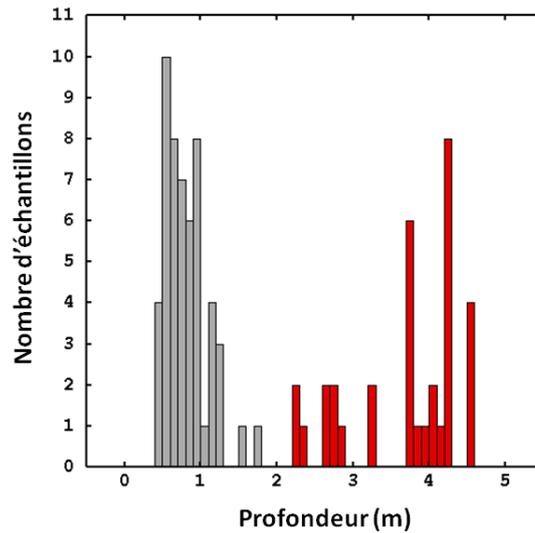


Figure 76 : Profondeur moyenne des 86 échantillons sélectionnés précédemment. En gris les 53 échantillons à moins de 2 m de profondeur.

Ainsi, nous avons fortement restreint le nombre de données (de 208 à 53) mais celles-ci présentent maintenant des supports comparables et concernent des matériaux et des pressions anthropiques semblables. La carte d'implantation de ces échantillons (Figure 77) indique une répartition homogène dans la zone. Néanmoins, les échantillons sélectionnés sont en moyenne espacés de 60 m, ce qui détermine notamment le pas de calcul du variogramme horizontal.

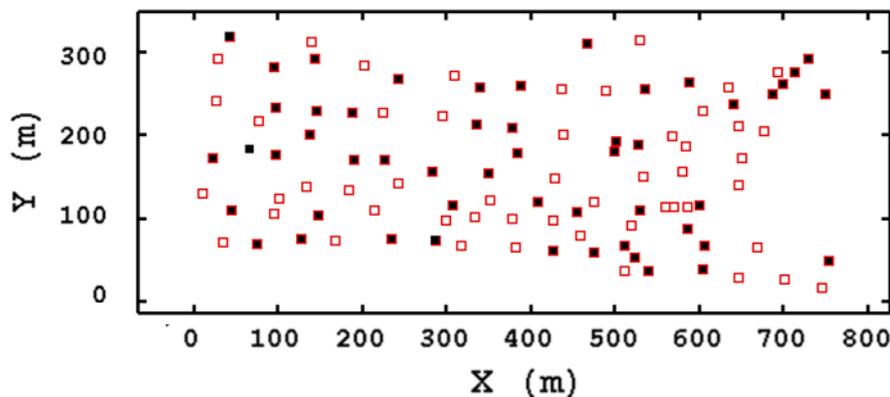


Figure 77: Implantation des 53 échantillons sélectionnés (en noir) par rapport aux 208 d'origine (en rouge).

IV.B. Application à un site

Les résultats sont illustrés sur la présence d'anomalies pour la famille des métaux anthropisés (cuivre, plomb et zinc). Les données inférieures à la limite de quantification ont été ramenées à cette limite.

Nous présentons les résultats obtenus sur les données initiales. En effet, les estimations sur les données log-translatées, ne peuvent pas être directement retransformées pour revenir à l'unité de départ (mg/kg_MS). L'hypothèse de normalité des transformées n'étant pas vérifiée, les calculs du cas log-normale ne s'appliquent pas.

Les variogrammes expérimentaux simples et croisés du cuivre, du plomb et du zinc sont tracés et croisés avec ceux des indicatrices des types de remblais (Figure 79 et Figure 80). Chacune de ces indicatrices renseigne si l'échantillon appartient à un des trois types de remblais, en prenant la valeur 1 lorsque l'échantillon est dans le type de remblais visé et 0 sinon, cela définit une nouvelle variable sur laquelle le variogramme peut-être calculé. Les remblais assimilables naturels ne présentant pas de corrélation avec les teneurs en cuivre, plomb et zinc, l'indicateur n'a pas été conservée pour les traitements. Néanmoins, les deux premiers mètres d'épaisseurs du sol du site étant composés des trois types de remblais, la présence des remblais assimilables naturels est prise en compte par la valeur 0 des deux autres indicatrices.

Par la suite, un modèle est ajusté de façon semi-automatique (en trait plein sur la Figure 78, Figure 79 et Figure 80). Les structures sont choisies au préalable et les paliers sont ajustés automatiquement. Le modèle comporte trois composantes : un effet de pépite qui correspond aux variations à des distances inférieures à 60 m, une composante sphérique intermédiaire de portée 70 m et une composante linéaire. Comme nous l'avons présenté précédemment, le modèle linéaire de corégionalisation permet non seulement d'identifier les différentes composantes mais également d'étudier les coefficients de corrélation des variables par structure, dont il sera tenu compte dans la sélection des composantes à filtrer.

Les lignes pointillées des figures 79 et 80 représentent les enveloppes de corrélation maximales entre les variogrammes croisés. Plus les variables sont corrélées et plus le variogramme sera proche de la ligne pointillée positive.

IV.B. Application à un site

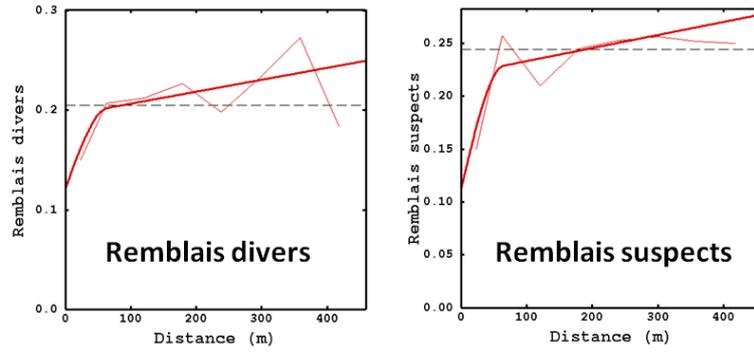


Figure 78 : Variogrammes expérimentaux simples des indicatrices de remblais. Le modèle ajusté est représenté en trait plein. Les données sont les 53 échantillons sélectionnés pour leur longueur comprise entre 50 et 70cm et leur profondeur inférieure à 2 m.

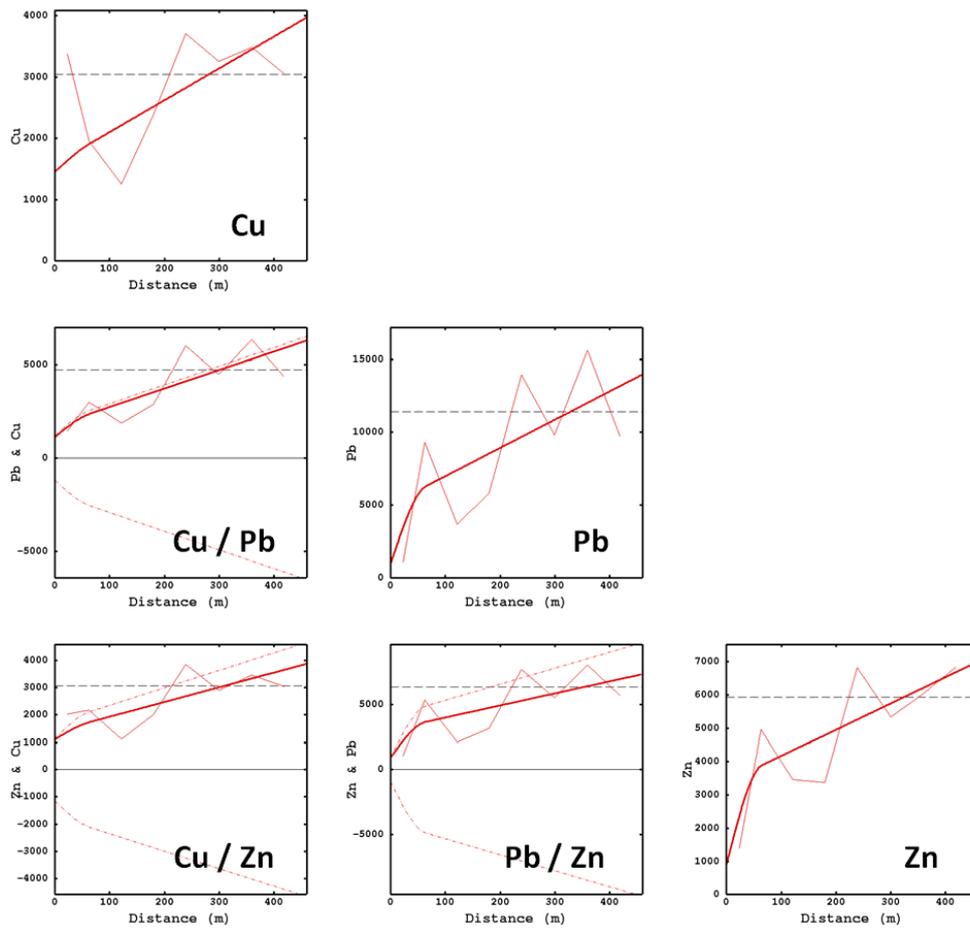


Figure 79 : Variogrammes expérimentaux simples et croisés des teneurs en cuivre, plomb et zinc. Le modèle ajusté est représenté en trait plein. Les données sont les 53 échantillons sélectionnés avec une longueur comprise entre 50 et 70cm et une profondeur inférieure à 2m.

IV.B. Application à un site

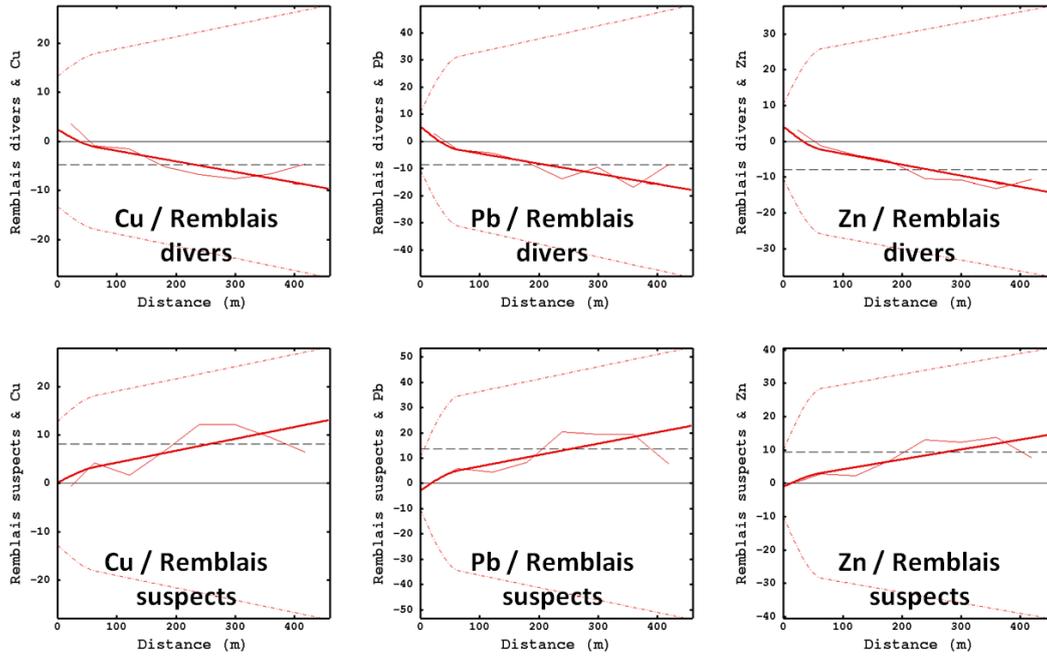


Figure 80 : Variogrammes expérimentaux croisés des teneurs en cuivre, plomb et zinc avec les indicatrices de remblais divers et suspects. Le modèle ajusté est représenté en trait plein.

Les composantes péritique et sphérique de portée 60 m, supposées d'espérance nulle, vont donc être filtrées de façon à estimer la composante linéaire (Z_3).

Z_3 étant une composante linéaire, donc non-stationnaire, la moyenne disparaît par passage aux accroissements, et n'est donc pas explicitement modélisée.

L'estimation par filtrage consiste à cokriquer les teneurs par les indicatrices de remblais et les autres paramètres de la même famille, en filtrant les composantes de plus petites portées pour ne garder que la composante non stationnaire (ou de plus grande portée dans un cas stationnaire). Pour comparaison, le cokripage du même paramètre est effectué en considérant toutes les composantes.

IV.B.2. Résultats et discussion

Les coefficients de corrélations régionalisés associés aux composantes du modèle linéaire de corégionalisation obtenus pour l'effet de pépite, présentés au Tableau 11, indiquent une très bonne corrélation entre le cuivre, le plomb et le zinc. Il n'y a pas de corrélation marquée avec les types de remblais. Les trois métaux étant souvent présents dans les mêmes types de contamination, il est cohérent de les trouver bien corrélés dans cette composante. L'absence de corrélation avec les remblais indiquerait une influence extérieure sur les teneurs plutôt qu'un effet du matériau composant le sol pour cette structure spatiale.

Effet de pépite			
	Cu	Pb	Zn
Cu	1,00		
Pb	0,94	1,00	
Zn	0,97	0,98	1,00
R_divers	0,19	0,49	0,40
R_suspects	0,01	-0,28	-0,10

Tableau 11 : Coefficients de corrélation régionalisés du cuivre, du plomb, du zinc et des remblais divers et suspect pour l'effet de pépite du modèle linéaire de corégionalisation.

Pour la composante de portée intermédiaire (sphérique de portée 70 m) (Tableau 12), les coefficients de corrélation entre le cuivre, le plomb et le zinc diminuent. De plus, les coefficients des remblais restent relativement faibles, à l'exception d'une anti-corrélation (marquée par un coefficient de corrélation négatif) entre le cuivre et les remblais divers. Le changement de signe entre les remblais divers et suspects indique que les teneurs les plus fortes des métaux sont globalement plus présentes dans les remblais suspects que divers, ce qui correspond bien à la description des remblais suspects comme présentant des contaminations. La faible corrélation (ou anti-corrélation) entre les teneurs des métaux et les indicatrices de remblais indiquent, pour cette composante également, une prédominance de l'apport extérieur sur la présence des teneurs fortes plutôt qu'un lien avec la nature du matériau. De plus la diminution de la corrélation entre les métaux indiquerait des origines de contaminations différentes selon les substances (une variation des teneurs en zinc n'étant pas nécessairement liée aux mêmes variations en plomb ou cuivre, à cette distance).

Les variations relativement rapides (portée de 70 m) et la plus faible corrélation entre les paramètres appuient le choix du filtrage de cette composante pour l'estimation d'un fond relativement homogène.

IV.B. Application à un site

Sphérique : 70m			
	Cu	Pb	Zn
Cu	1,00		
Pb	0,85	1,00	
Zn	0,51	0,68	1,00
R_divers	-0,72	-0,36	-0,34
R_suspects	0,46	0,25	0,14

Tableau 12 : Coefficients de corrélation régionalisés du cuivre, du plomb, du zinc et des remblais divers et suspect pour la composante sphérique du modèle linéaire de corrégalisation.

Enfin, les coefficients de corrélation pour la composante linéaire du modèle (Tableau 13), indiquent une bonne corrélation entre les métaux. De plus, on remarque une forte anti-corrélation avec les remblais divers et, à l'inverse, une très bonne corrélation avec les remblais suspects. Ceci reflèterait l'influence des types de remblais sur les teneurs en plomb du site, les teneurs les plus fortes étant liées aux remblais suspects, et les plus faibles aux remblais divers. Nous associons cette composante au fond pédo-géochimique, et cela est appuyé par la cohérence entre les teneurs des 3 métaux, mais aussi par leur relation avec les types de remblais.

Linéaire			
	Cu	Pb	Zn
Cu	1,00		
Pb	0,99	1,00	
Zn	0,84	0,75	1,00
R_divers	-0,86	-0,78	-1,00
R_suspects	0,97	0,92	0,95

Tableau 13 : Coefficients de corrélation régionalisés du cuivre, du plomb, du zinc et des remblais divers et suspect pour la composante linéaire du modèle linéaire de corrégalisation.

Les estimations obtenues par cokrigage et par filtrage dans le cas du zinc (Figure 81), présentent un lissage des teneurs, plus marqué par filtrage. On remarque la disparition dans le filtrage des zones de fortes et faibles teneurs très localisées. Ces zones sont liées à des variations à courte portée (<70 m) et appartiennent à des contaminations locales ou à des zones de teneurs anormalement faibles pour le site. Des zones de teneurs élevées dans les parties nord-est et nord-ouest du site, ainsi qu'une zone de faibles teneurs au sud sont tout de même présentes dans l'estimation par filtrage. Ces zones présentent des variations lentes et globalement faibles, cohérentes avec la notion de contaminations diffuses du fond pédo-géochimique.

IV.B. Application à un site

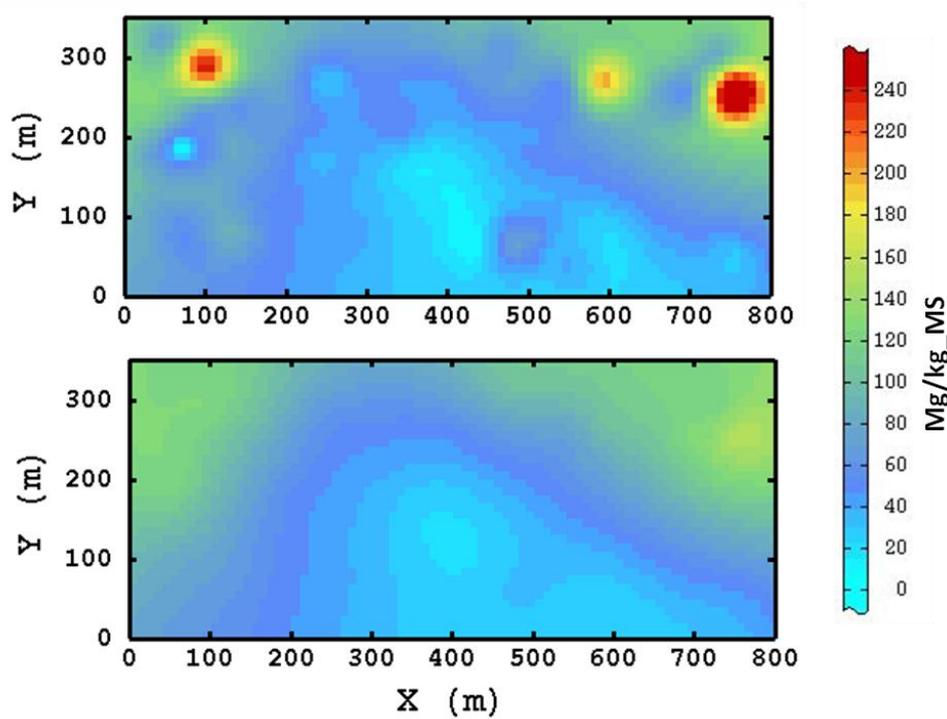


Figure 81 : Estimation des teneurs en zinc par cokrigage (en haut) et estimation du fond pédogéochimique de la même substance par filtrage (en bas).

Les résultats sont analogues pour les estimations des teneurs en plomb (Figure 82), avec la disparition des anomalies positives ou négatives par le filtrage.

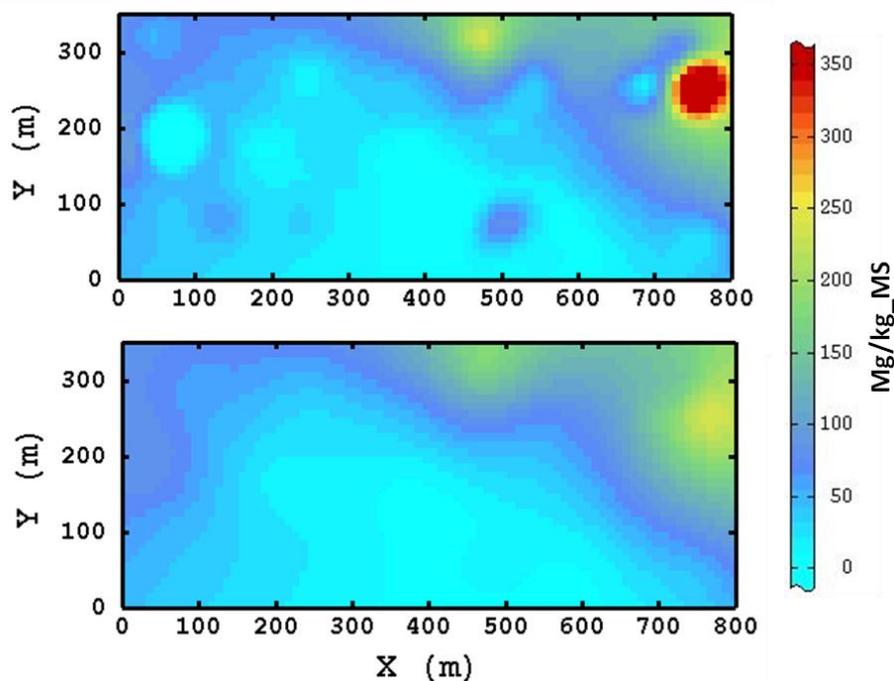


Figure 82 : Estimation des teneurs en plomb par cokrigage (en haut) et estimation du fond pédogéochimique de la même substance par filtrage (en bas).

IV.B. Application à un site

Le filtrage fournit donc une cartographie des teneurs du site exempte de variations à petite distance des teneurs interprétées comme des anomalies locales. Les teneurs estimées sont alors plus homogènes et donc plus cohérentes avec la définition du fond pédogéochimique.

L'histogramme des teneurs en zinc dans les échantillons (Figure 83 haut), indique la présence de quelques teneurs très fortes mais la majorité reste inférieure à 100 mg/kg_MS. Bien que la distribution des valeurs estimées diffère nécessairement de celle des données, on retrouve une tendance similaire dans les estimations (Figure 83 bas), où la majorité des résultats est inférieure à 150 mg/kg_MS. On note toutefois la présence de teneurs supérieures. Pour le cokrigage la fréquence diminue progressivement à mesure que la teneur augmente mais nous observons des teneurs jusqu'à 430 mg/kg_MS (maximum des échantillons). Pour plus de lisibilité ces dernières ne sont pas représentées sur la Figure 83. Les données obtenues par filtrage montrent le maximum le plus faible à 160 mg/kg_MS du fait de l'absence d'anomalies positives. On y observe toutefois nettement un second mode dans la distribution autour de 120 mg/kg_MS.

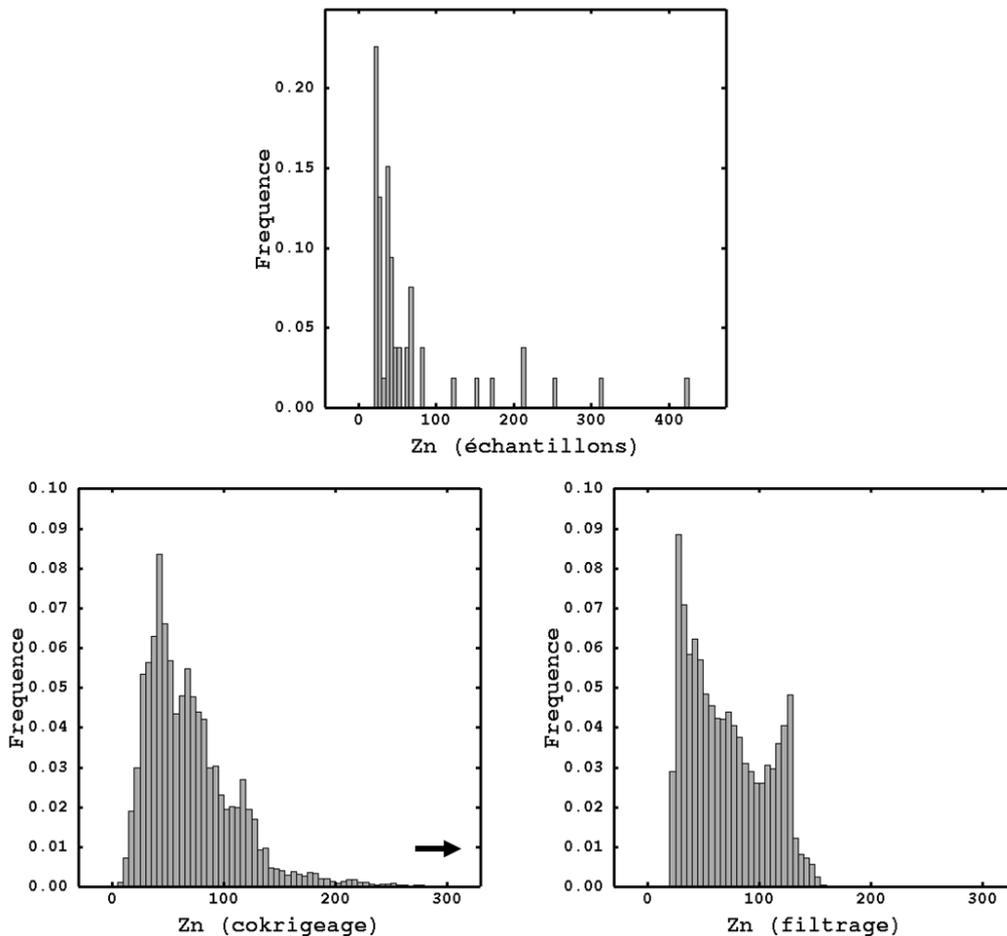


Figure 83: Teneurs en zinc (mg/kg_MS) dans les échantillons du site (en haut) et dans les estimations (gauche : cokrigage, droite : filtrage). L'histogramme du cokrigage est tronqué et ne présente pas les teneurs supérieures à 300 mg/kg_MS.

IV.B. Application à un site

Ces changements dans les distributions sont également visibles via des diagrammes de dispersion (Figure 84), qui montrent clairement l'absence des anomalies positives (outliers) dans les résultats du filtrage alors que ceux-ci sont présents dans les résultats du cokrigage. Les estimations par filtrage sont moins dispersées que les résultats du cokrigage. Cela est confirmé par les résultats statistiques du Tableau 14, avec la forte diminution du maximum dans les résultats de filtrage et des variations dans la distribution à travers la valeur des quantiles : la médiane et le quantile à 75% sont plus élevés sur les estimations du fond.

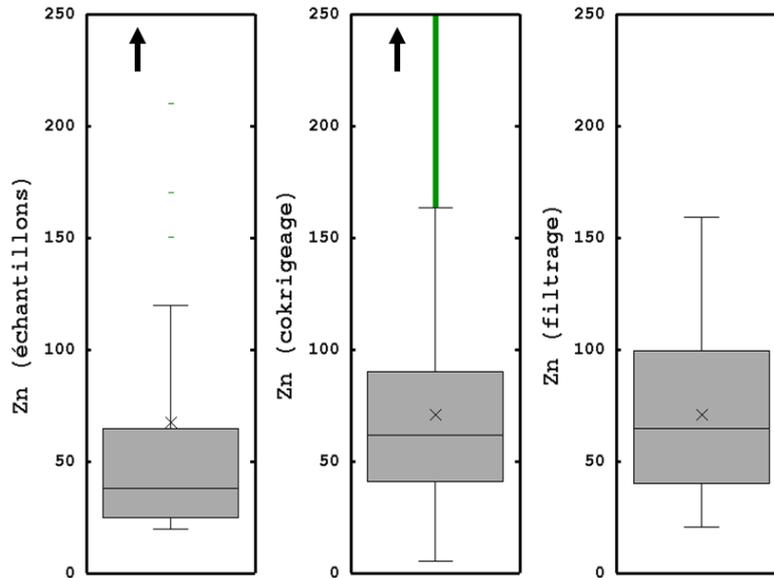


Figure 84: Diagrammes de dispersion des échantillons de départ et des résultats de cokrigage et de filtrage du zinc en mg/kg_MS. La moyenne est représentée par le x. Les outliers supérieurs à 250 mg/kg_MS dans les diagrammes de dispersion des échantillons et du cokrigage ne sont pas représentés.

	Zinc		
	Echantillons	Cokrigage	Filtrage
Minimum	20	5	21
Q25	25	41	40
Q50	38	62	65
Q75	65	90	100
Q90	170	122	123
Q95	250	144	128
Maximum	430	419	160
Vibrisse	125	/	/
MAD	50	/	/

Tableau 14: Quantiles calculés sur les échantillons et les estimations par cokrigage et filtrage des teneurs en zinc du site.

Pour le plomb (Figure 85, Tableau 15), les teneurs les plus fortes, là encore, n'apparaissent pas dans les résultats de filtrage. Le second mode visible dans les histogrammes des estimations du zinc l'est également dans les résultats d'estimation du plomb et correspond à la même zone du site ; cela est expliqué par la forte corrélation entre

IV.B. Application à un site

les deux métaux. Le faible nombre d'échantillons entraîne de grandes variations dans les valeurs des quantiles 90 et 95 calculés qui sont inférieurs à ceux des estimations. Là encore, la médiane et la quantile d'ordre 75% sont supérieurs après filtrage.

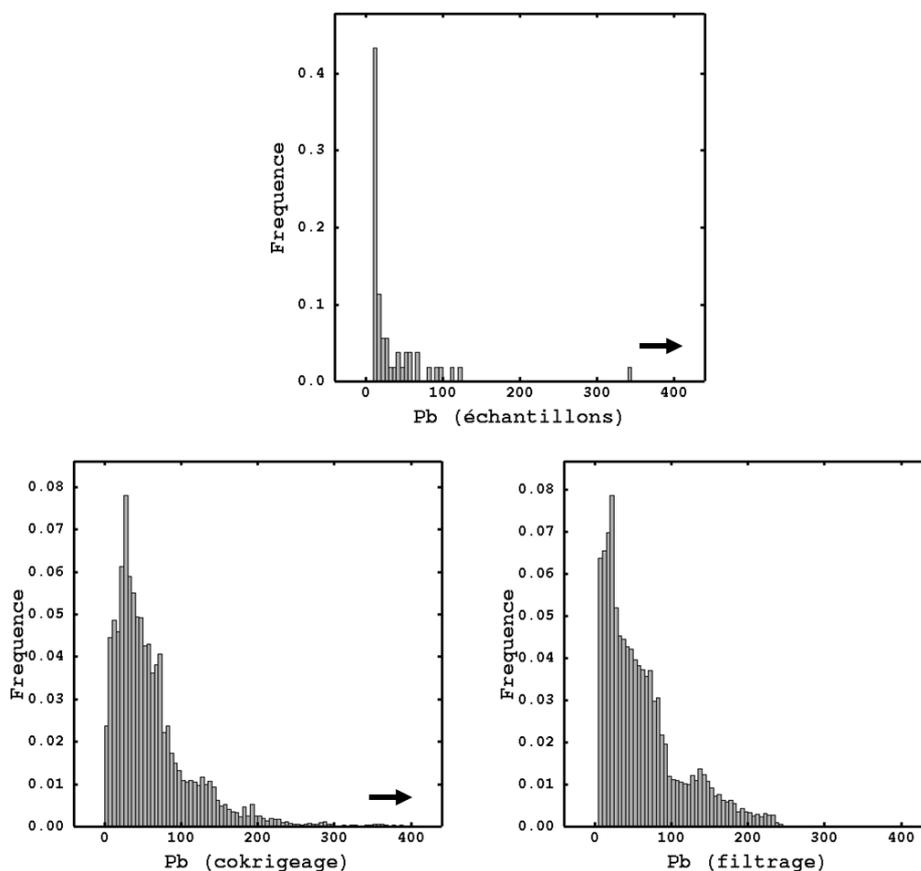


Figure 85: Teneurs en plomb (mg/kg_{MS}) dans les échantillons du site (en haut) et dans les estimations (gauche : cokrigage, droite : filtrage). L'histogramme des échantillons et du cokrigage est tronqué et ne présente pas les teneurs supérieures à 400 mg/kg_{Ms}.

	Plomb		
	Echantillons	Cokrigage	Filtrage
Minimum	10	0	5
Q25	10	25	23
Q50	17	47	49
Q75	46	78	85
Q90	93	135	142
Q95	120	179	170
Maximum	720	694	252
Vibrise	100	/	/
MAD	25	/	/

Tableau 15: Quantiles calculés sur les échantillons et les estimations par cokrigage et filtrage des teneurs en plomb du site.

IV.B. Application à un site

En conclusion, sur ce site, le filtrage via un modèle linéaire de corégionalisation fournit bien une estimation exempte des anomalies locales. Les anomalies filtrées regroupent des contaminations mais aussi des zones de teneurs anormalement faibles, contrairement aux seuils statistiques qui ne visaient à séparer que les anomalies positives. Il convient alors de déterminer si les anomalies négatives mises en évidence font, ou non, partie du fond pédogéochimique. Considérant une approche spatiale des anomalies où celles-ci sont marquées par des variations à courte portée des teneurs, nous choisissons de ne pas les inclure dans le fond. Il paraît, en effet, cohérent que le fond estimé soit supérieur aux teneurs les plus faibles. Néanmoins, des modifications du protocole de traitement peuvent être mises en place afin d'inclure ces zones de teneurs anormalement faibles dans l'estimation du fond.

Les estimations peuvent toutefois être fortement influencées par quelques valeurs fortes dans les données, conduisant à une surestimation globale des teneurs. Pour limiter cela, plusieurs approches sont possibles. La première serait de réaliser un premier tri des données avant les traitements. Néanmoins, la sélection des échantillons suivant leur longueur et leur profondeur ayant déjà diminué fortement le nombre de données, cela reste à éviter. L'autre approche consisterait à réaliser des sondages complémentaires répondant aux critères de sélection sur le site pour limiter l'influence des quelques teneurs fortes relativement isolées et pour assurer une densité minimale de données sur l'ensemble du site.

Enfin, la sélection suivant la profondeur des échantillons a un sens physique puisqu'elle tient compte de l'influence de l'occupation des sols. En revanche, celle sur la longueur des échantillons est un parti pris et peut être corrigée en incluant des échantillons de longueurs différentes dans le krigeage. En effet, pour une grandeur additive, la prise en compte de données de supports différents ne pose pas de difficulté théorique. Il suffit de discrétiser le support des données ou de la « cible », en se ramenant à un modèle « ponctuel » de variogramme.

IV.C. Application à un quartier

IV.C.1. Différences et limites par rapport à l'application à un site

L'application du filtrage à un quartier entier nécessite de tenir compte d'un découpage de ce territoire afin de fournir différents niveaux du fond suivant la nature des sols et les pressions anthropiques.

Les différents types de remblais présentent des gammes de teneurs très différentes, et leur répartition sur le quartier est relativement complexe. De plus les transitions de teneurs entre les différents types de remblais sont brutales. A cause de cela, un cokrigeage global tenant compte des indicatrices de remblais, comme pour le site, n'est pas adapté. Nous choisissons d'effectuer des estimations par types de remblais séparément. Ces estimations sont premièrement réalisées pour l'ensemble du quartier et sont ensuite découpées suivant le modèle géologique en trois dimensions. Cela permet d'obtenir différentes gammes de valeurs (niveaux de fond pédo-géochimiques anthropisés) suivant les types de remblais. Le découpage fourni par l'analyse multi-critères n'est pas utilisé du fait de l'incertitude sur la profondeur d'influence des activités anthropiques, néanmoins si cette profondeur pouvait être connue ou estimée sur l'ensemble du quartier, ce découpage fournirait une covariable descriptive importante pouvant remplacer ou venir compléter l'utilisation de la typologie des remblais.

Nous l'avons vu précédemment, certains échantillons recoupent et mélangent différentes typologies de matériaux. Nous avons choisi de ne traiter que les échantillons ne recoupant qu'un seul type de matériau. Nous cherchons en effet à fournir une estimation du fond selon la nature des sols et les échantillons mélangeant plusieurs types de matériaux nécessitant un traitement géostatistique spécifique. Les échantillons correspondant à plusieurs types de remblais, ou à des remblais et des alluvions, ont donc été enlevés. Nous passons alors de 4 063 échantillons à 3 238 pour les traitements. A noter que parmi ces échantillons, seuls 745 disposent d'analyses en cuivre, plomb et zinc sur solide.

Nous comparons la longueur des échantillons aux teneurs mesurées suivant le type de matériau, afin d'identifier une relation potentielle entre la longueur de l'échantillon et les teneurs en cuivre, plomb et zinc (Figure 86).

IV.C. Application à un quartier

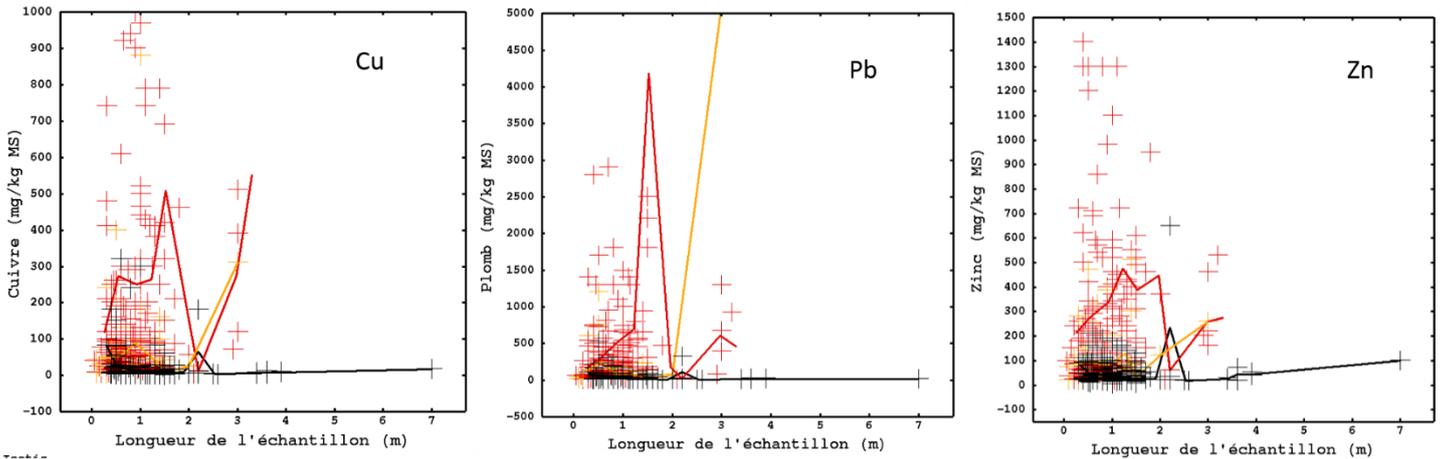


Figure 86: Nuage de corrélation entre les teneurs en cuivre, plomb et zinc et la longueur de l'échantillon et leur moyenne par classe (de 20 échantillons) (trait plein).

Rouge : remblais suspects, Orange : Remblais divers, Noir : Remblais assimilables naturels.

La moyenne par classe des remblais suspects évolue avec la longueur de l'échantillon. En revanche, les moyennes pour les remblais assimilables naturels sont relativement stables et semblent donc montrer que la longueur de l'échantillon aurait peu d'impact sur la teneur dans ces remblais. Cela est probablement expliqué par leur homogénéité. Les remblais divers montrent des variations intermédiaires dans leurs moyennes par classe, vis-à-vis des deux autres remblais. Ils semblent donc plus homogènes que les remblais douteux, mais moins que les remblais assimilables naturels.

Ensuite, certains échantillons ne recoupent pas la totalité de l'épaisseur des couches qu'ils échantillonnent, nous définissons une nouvelle variable, le « pourcentage représentatif » correspondant à :

$$\text{Pourcentage représentatif} = \frac{\text{Longueur de l'échantillon}}{\text{Épaisseur de la couche}}$$

Par type de remblais on obtient les histogrammes suivants :

IV.C. Application à un quartier

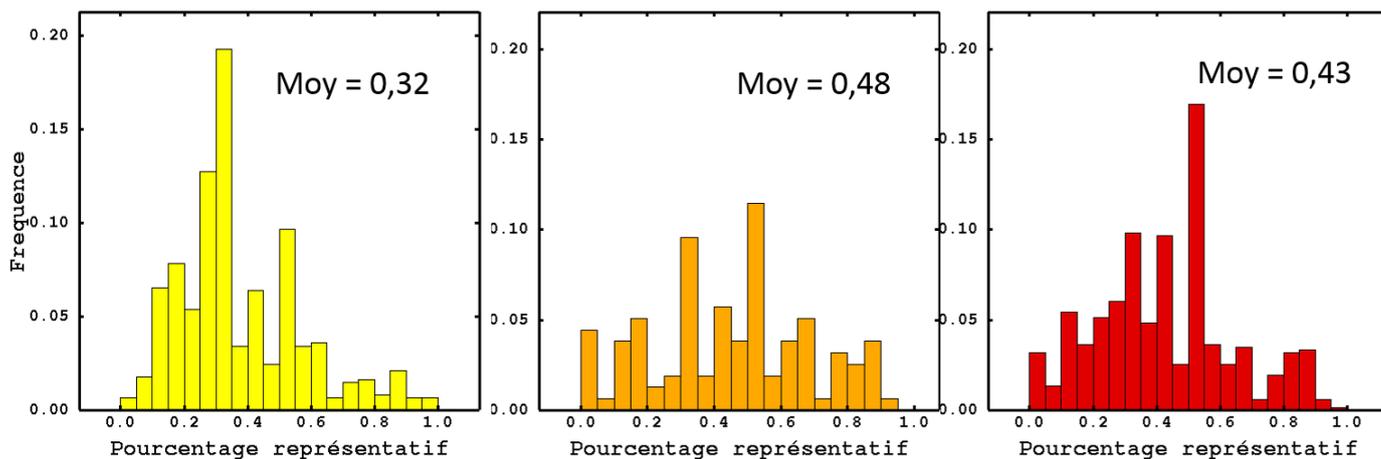


Figure 87: Fréquence des pourcentages représentatifs suivant le type de remblais.

Rouge : remblais suspects, Orange : Remblais divers, Jaune : Remblais assimilables naturels.

Très peu d'échantillons sont réellement représentatifs de l'ensemble de la couche. On note un mode autour de 50% pour les remblais suspects et divers et plutôt autour de 30% pour les remblais assimilables naturels. Cela s'explique par l'épaisseur parfois importante des remblais, et plus particulièrement des remblais assimilables naturels.

Nous pouvons ensuite comparer ce « pourcentage représentatif » aux teneurs mesurées (Figure 88) afin d'identifier des relations potentielles suivant le type de matériau.

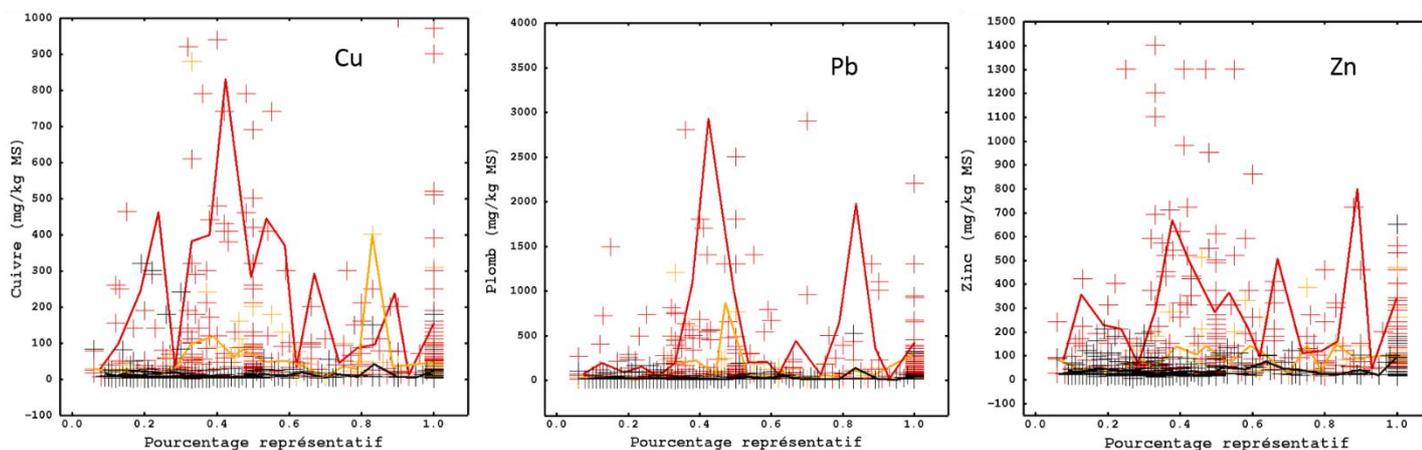


Figure 88: Nuage de corrélation entre les teneurs en cuivre, plomb et zinc et le pourcentage représentatif et leur moyenne par classe (de 20 échantillons) (trait plein).

Rouge : remblais suspects, Orange : Remblais divers, Noir : Remblais assimilables naturels.

Comme pour la longueur des échantillons, les moyennes par classes calculées sur les échantillons des remblais suspects sont très variables, celles calculées sur les remblais divers et assimilés naturels étant moins variables selon le pourcentage.

Nous choisissons de travailler avec les échantillons des remblais assimilables naturels et divers (603 pour les remblais assimilables naturels et 157 pour les remblais divers). Comme

IV.C. Application à un quartier

précédemment une sélection à été effectuée suivant la longueur des échantillons de façon à ne pas mélanger des supports trop différents influençant la variance et de façon à correspondre au mode principal de la distribution des données. Dans le cas des remblais assimilables naturels, les échantillons sélectionnés mesurent entre 0,75 et 1,25 m de long et représentent environ 70% des données d'origine, alors que dans le cas des remblais divers la sélection se fait entre 0,5 et 0,8 m avec environ 65% des données sélectionnées.

Nous ne traiterons pas les données des remblais suspects, à cause de leur hétérogénéité et de leur mauvaise qualité géochimique compromettant leur valorisation (nous préférons donc ne pas fournir d'estimation afin que celles-ci ne soient pas mal utilisées).

Contrairement au cas précédent, aucune distinction ne sera, ici, faite selon la profondeur, en raison de l'homogénéité apparente des couches de remblais traitées (divers et assimilables naturels).

Pour le filtrage nous choisissons de fournir une cartographie séparée des variations latérales des teneurs du fond d'une part dans les remblais divers et d'autre part dans les remblais assimilables naturels. Ces couches se situant plus en profondeur ou dans des zones d'occupation des sols relativement homogènes, seule la nature des sols sera prise en compte.

Afin de réaliser ces traitements sur un quartier, il est donc nécessaire d'étudier au préalable :

- La variabilité des teneurs suivant la nature ou l'occupation des sols afin de déterminer si des estimations doivent être réalisées séparément.
- L'influence de la longueur des échantillons et leur « représentativité » afin de déterminer si une sélection est nécessaire.

Nous illustrerons les traitements et les résultats sur les remblais assimilables naturels.

Les variogrammes simples et croisés obtenus pour les remblais assimilables naturels sont illustrés à la Figure 89. Le pas de calcul est fixé à 90m. Pour 15 pas de calcul, le modèle considéré est isotrope pour les variations latérales et calculé pour des tranches de 1m d'épaisseur. Le modèle ajusté comprend trois composantes :

- un effet de pépite
- une composante sphérique de portée 200 m
- une composante sphérique de portée 1000 m

IV.C. Application à un quartier

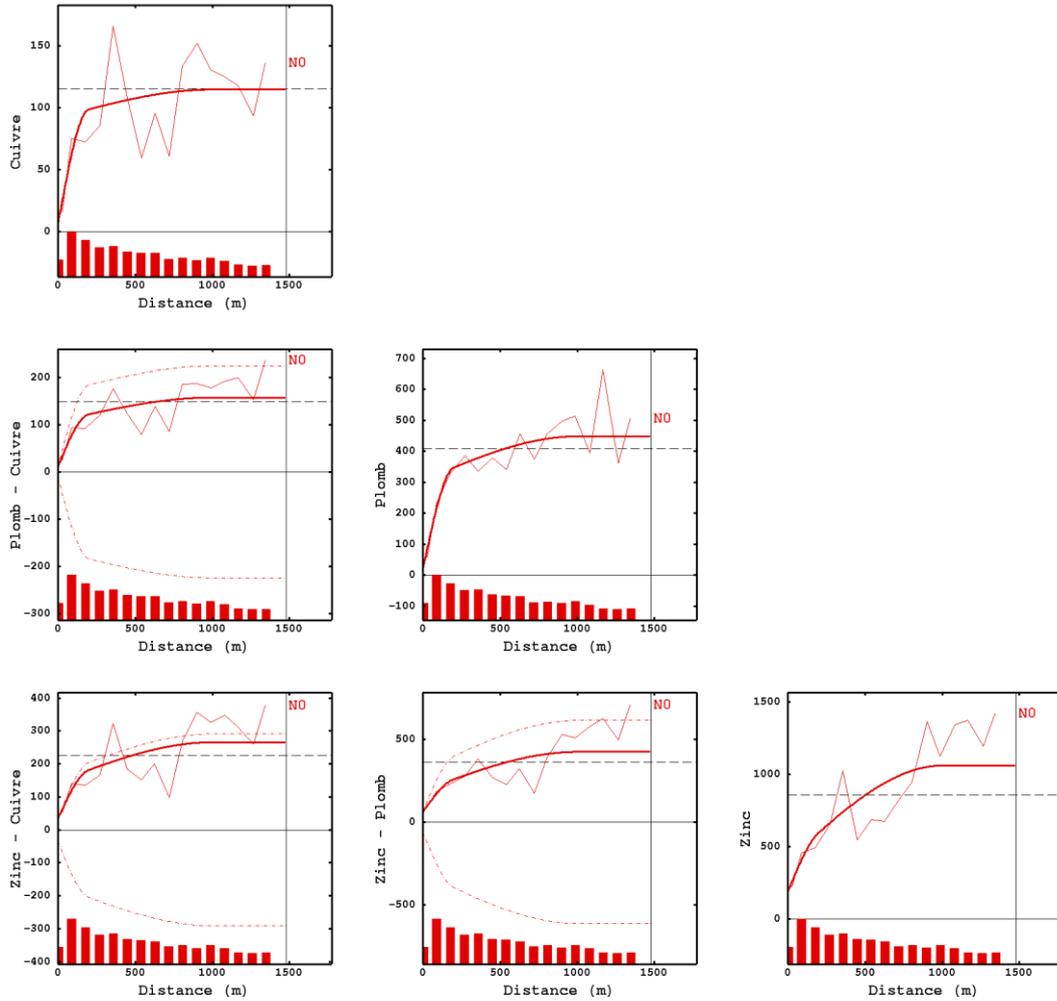


Figure 89 : Variogrammes expérimentaux simples et croisés des teneurs en cuivre, plomb et zinc. Le modèle ajusté est représenté en trait plein. L'histogramme des paires de calcul est reporté en rouge.

IV.C.2. Résultats et discussion

Examinons d'abord les coefficients de corrélations des différentes composantes du modèle.

La composante pépitique (Tableau 16) est marquée par une très bonne corrélation entre les trois variables. Comme précédemment, les contaminations ponctuelles sont donc semblables pour le cuivre, le plomb et le zinc. Cette composante est filtrée car elle correspond à des variations à petite distance.

Effet de pépité			
	Cu	Pb	Zn
Cu	1,00		
Pb	1,00	1,00	
Zn	1,00	1,00	1,00

Tableau 16 : Coefficients de corrélation régionalisés du cuivre, du plomb et du zinc pour l'effet de pépité du modèle linéaire de corégionalisation, à l'échelle du quartier.

La composante sphérique intermédiaire (portée de 200 m) (Tableau 17) présente des coefficients de corrélation régionalisés plus faibles. Les variations prises en compte par cette composante sont parfois différentes d'une variable à l'autre. Cette composante est également filtrée, de façon à ne garder dans le fond que des variations bien corrélées entre les variables.

Sphérique : 200m			
	Cu	Pb	Zn
Cu	1,00		
Pb	0,61	1,00	
Zn	0,83	0,51	1,00

Tableau 17 : Coefficients de corrélation régionalisés du cuivre, du plomb et du zinc pour la composante sphérique de portée 200 m du modèle linéaire de corégionalisation, à l'échelle du quartier.

Enfin, la troisième composante (Tableau 18), sphérique, de portée 1000 m, montre une meilleure corrélation entre le cuivre, le plomb et le zinc que la précédente. On note une corrélation plus faible entre le plomb et le zinc, majoritairement expliquée par quelques valeurs fortes en zinc sur des échantillons dont les teneurs en plomb sont plus faibles. On conserve toutefois cette composante pour l'estimation du fond puisqu'elle représente les variations à longues portées et reste globalement bien corrélée entre les trois métaux.

IV.C. Application à un quartier

Sphérique : 1000m			
	Cu	Pb	Zn
Cu	1,00		
Pb	0,89	1,00	
Zn	0,98	0,77	1,00

Tableau 18 : Coefficients de corrélation régionalisés du cuivre, du plomb et du zinc pour la composante sphérique de portée 1000 m du modèle linéaire de corégionalisation, à l'échelle du quartier.

Les cartes des estimations de teneurs en cuivre, plomb et zinc (Figure 90, Figure 91, Figure 92) obtenues par cokrigage de toutes les composantes du modèle sont présentées ainsi que celles obtenues par filtrage, dans les remblais assimilables naturels.

Les cartes du cuivre (Figure 90), montrent la présence de nombreux points rouges et bleu clairs dans la partie ouest de la carte du cokrigage. Ces points représentent les contaminations ponctuelles de la zone et des sites anormalement pauvres en cuivre vis-à-vis du quartier. Cela reste cohérent avec le passé industriel très marqué dans la partie ouest du quartier, alors que la partie est, longtemps laissée naturelle, est bien plus homogène. Ces différents points sont absents de l'estimation par filtrage (en bas). On observe une représentation bien plus homogène de la géochimie du quartier après élimination des anomalies.

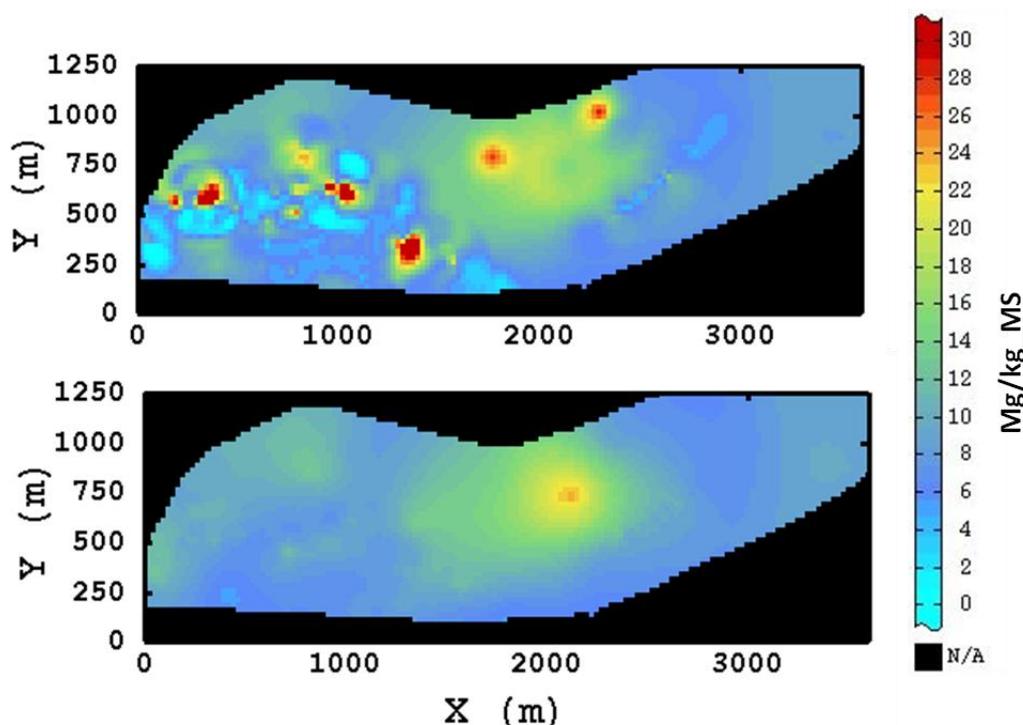


Figure 90 : Estimation des teneurs en cuivre par cokrigage (en haut) et du fond pédo-géochimique de la même substance par filtrage (en bas).

IV.C. Application à un quartier

Les résultats sont similaires pour le plomb (Figure 91), avec une partie ouest du quartier relativement hétérogène dans les résultats de cokrigage (en haut) et des résultats bien plus homogènes pour le filtrage. Comme pour le cuivre, il apparaît une zone de plus fortes teneurs au centre du quartier.

Enfin, les résultats du zinc (Figure 92), bien que moins marqués que les précédents montrent bien aussi un lissage des teneurs via le filtrage. On remarque encore la zone de plus fortes teneurs au centre du quartier, plus marquée que pour le cuivre et le plomb.

Les résultats cartographiques paraissent donc cohérents avec la définition du fond en filtrant les anomalies positives ou négatives pour ne garder dans l'estimation qu'un fond plus homogène. Se pose toutefois la question de la conservation de la zone de plus fortes teneurs au centre du quartier. Celle-ci restant cohérente avec des variations lentes du fond, nous avons choisi, ici, de la conserver, ce choix dépend toutefois de l'échelle de travail.

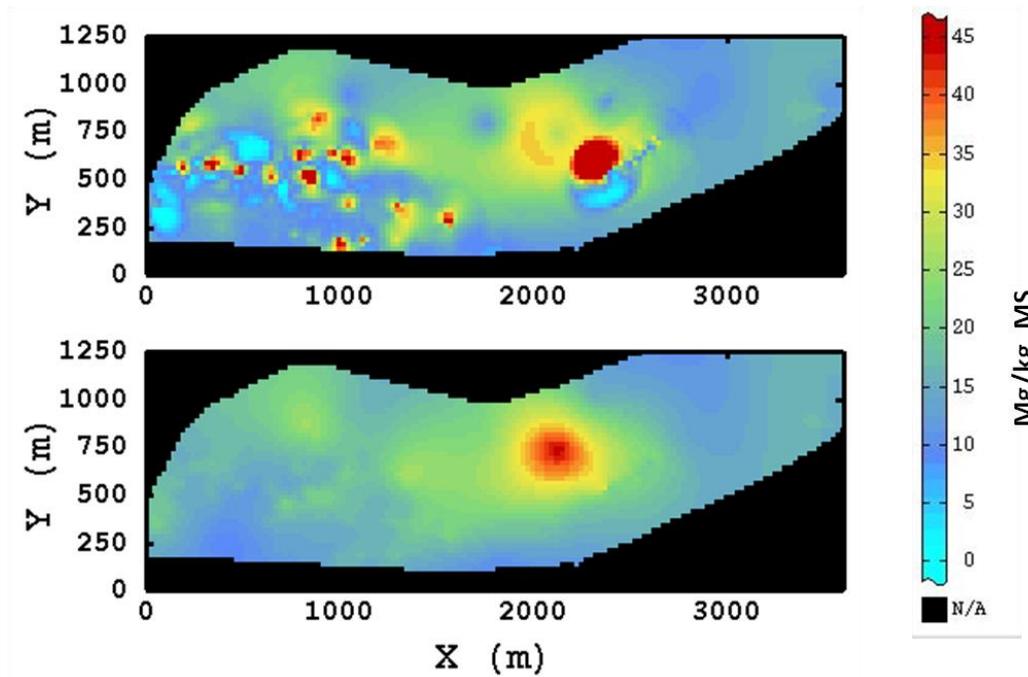


Figure 91: Estimation des teneurs en plomb par cokrigage (en haut) et du fond pédo-géochimique de la même substance par filtrage (en bas).

IV.C. Application à un quartier

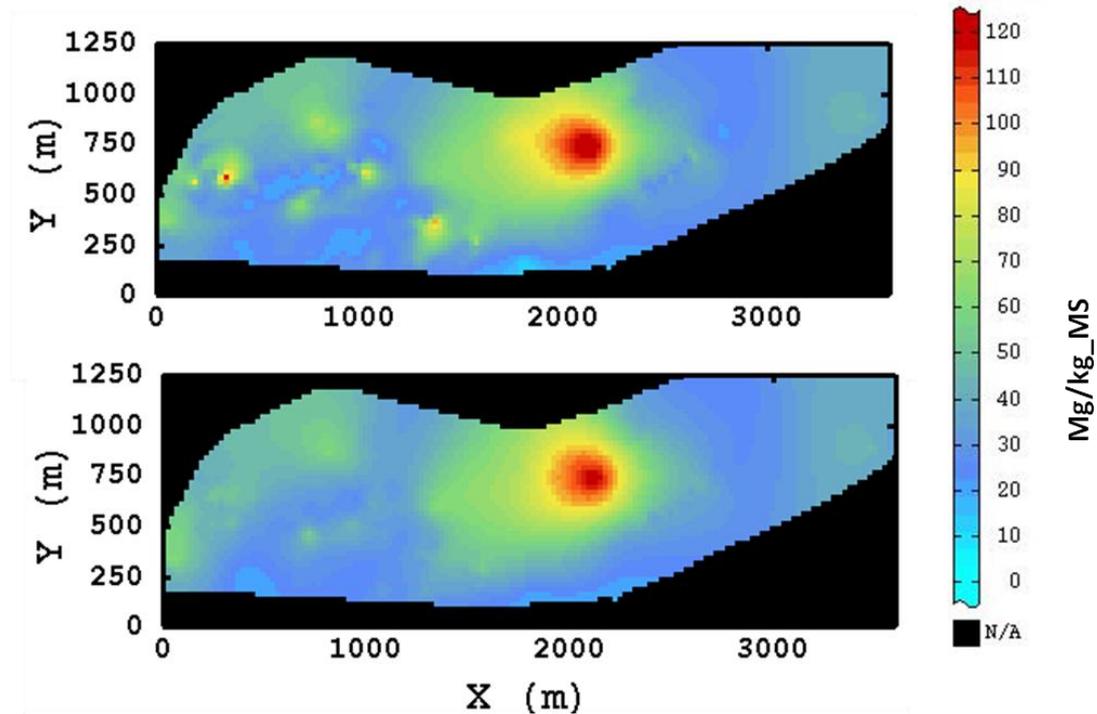


Figure 92 : Estimation des teneurs en zinc par cokrigage (en haut) et du fond pédo-géochimique de la même substance par filtrage (en bas).

Après avoir étudié les résultats cartographiques, il est important de s'intéresser aux histogrammes des estimations. Les histogrammes des échantillons sont également représentés.

Dans le cas du cuivre (Figure 93), les teneurs des estimations par filtrage sont plus resserrées que celle du cokrigage. En effet, le fond pédo-géochimique estimé ne comprenant que des variations lentes, les teneurs sont globalement semblables à l'échelle du quartier. Le maximum diminue avec le filtrage, ce qui reste cohérent avec la disparition des anomalies positives ; et le minimum augmente en lien avec le filtrage des zones de teneurs anormalement faibles.

IV.C. Application à un quartier

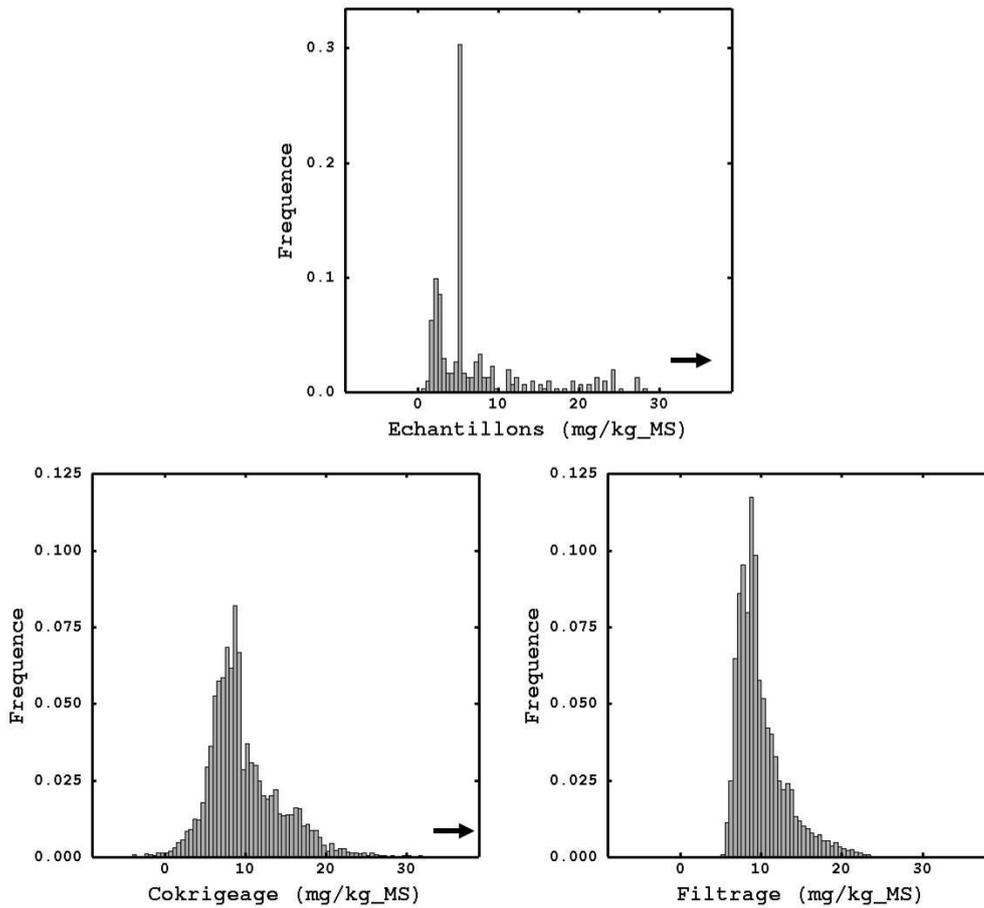


Figure 93: Teneurs en cuivre (mg/kg_MS) dans les échantillons des remblais assimilables naturels du quartier (en haut) et dans les estimations (gauche : cokrigeage, droite : filtrage). Les flèches noires indiquent des données supérieures qui ne sont pas représentées.

Les résultats sont similaires pour le plomb et le zinc (Figure 94 et Figure 95), avec une diminution du maximum et une augmentation du minimum et des teneurs logiquement plus resserrées pour le filtrage que pour le cokrigeage. Les résultats sont toutefois moins visibles pour le zinc, notamment à cause de la présence de moins d'anomalies.

IV.C. Application à un quartier

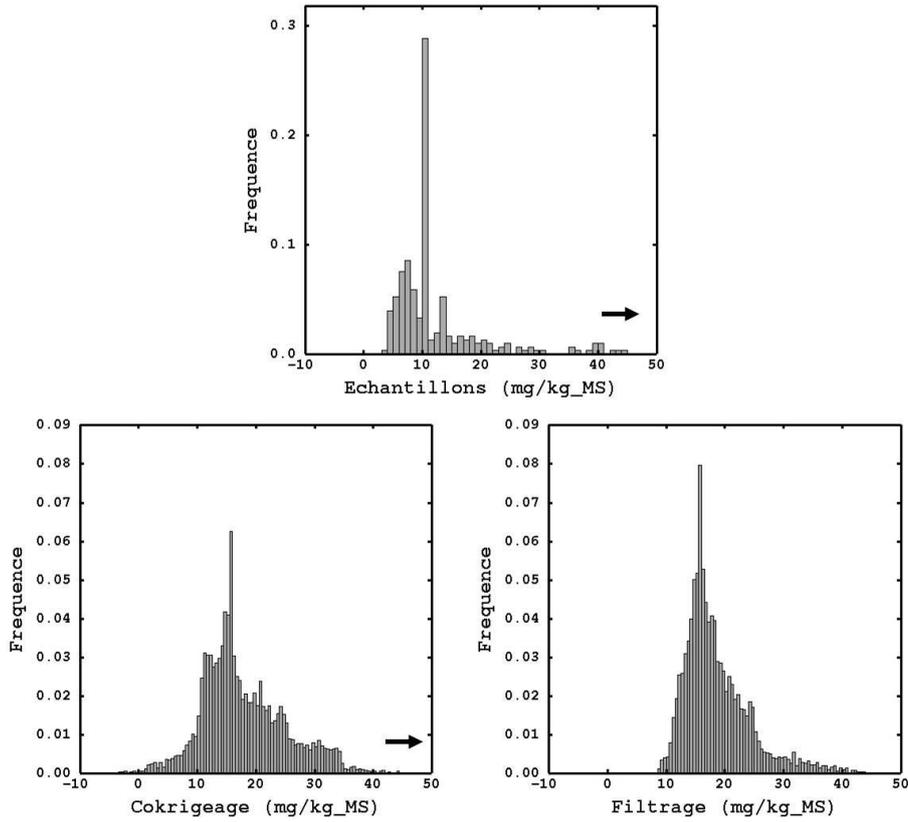


Figure 94: Teneurs en plomb (mg/kg_{MS}) dans les échantillons des remblais assimilables naturels du quartier (en haut) et dans les estimations (gauche : cokrigeage, droite : filtrage). Les flèches noires indiquent des données supérieures qui ne sont pas représentées.

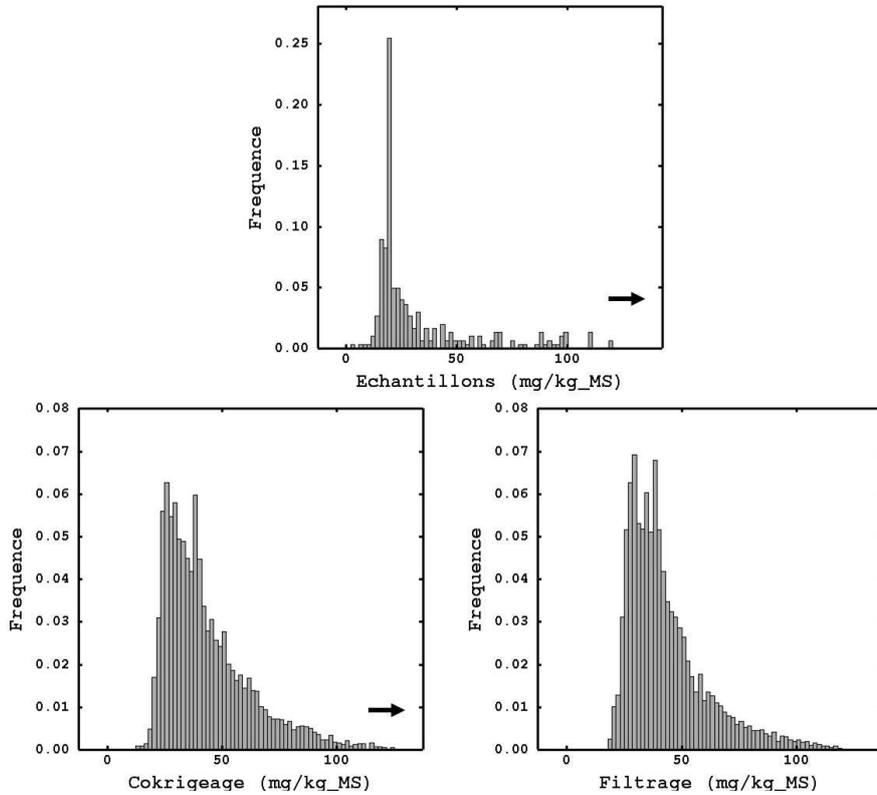


Figure 95: Teneurs en zinc (mg/kg_{MS}) dans les échantillons des remblais assimilables naturels du quartier (en haut) et dans les estimations (gauche : cokrigeage, droite : filtrage). Les flèches noires indiquent des données supérieures qui ne sont pas représentées.

IV.C. Application à un quartier

Enfin, comme pour le cas du site, les seuils statistiques calculés sur les échantillons sont comparés aux résultats d'estimation (Tableau 19, Tableau 20, Tableau 21).

Dans le cas du cuivre (Tableau 19), on retrouve une diminution du maximum dans les estimations et surtout dans les résultats du filtrage et une augmentation du minimum et de la médiane. Comme évoqué précédemment cela s'explique par la disparition des anomalies positives et négatives. De la même façon les quantiles 90 et 95 diminuent également. De plus, le maximum des estimations par filtrage correspond ici au quantile 95 calculé sur les échantillons.

Si on représente les seuils calculés (Q95, Vibrisse supérieure et seuil MAD) pour les échantillons sur l'histogramme correspondant aux estimations par filtrage, le seuil du quantile 95 marque le maximum des estimations par filtrage ; le seuil MAD est proche de la médiane et correspond au mode des estimations du fond. Enfin, la vibrisse supérieure est comprise entre les quantiles 90 et 95 des estimations.

	Cuivre		
	Echantillons	Cokrigeage	Filtrage
Minimum	0,8	-3,7	5,4
Q25	2,9	6,9	7,8
Q50	5	8,8	9,1
Q75	7,9	12	11,2
Q90	19	16,5	14,1
Q95	24	18,6	16,5
Maximum	83	67,9	24,1
Vibrisse	15,4	/	/
MAD	9,6	/	/

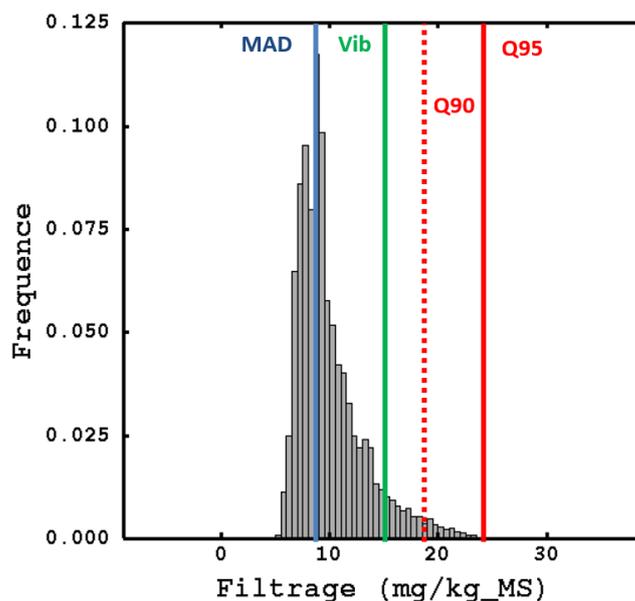


Tableau 19: Quantiles calculés sur les échantillons et les estimations par cokrigeage et filtrage des teneurs en cuivre des remblais assimilables naturels du quartier, et leur représentation sur l'histogramme des teneurs estimées par filtrage (Q95 : rouge, Vibrisse supérieure : vert, seuil MAD : bleu).

Les mêmes tendances s'observent pour le plomb et le zinc (Tableau 20 et Tableau 21), avec un seuil MAD relativement faible par rapport aux résultats d'estimation du fond, un quantile 95 plutôt élevé (supérieur au maximum pour le plomb et légèrement inférieur au maximum pour le zinc) et la vibrisse supérieure intermédiaire mais toujours proche du quantile 90 des estimations du fond. Dans les trois cas, le seuil MAD correspond au mode principal (ou inférieur) des estimations filtrées.

IV.C. Application à un quartier

	Plomb		
	Echantillons	Cokrigage	Filtrage
Minimum	3	-4,3	8,6
Q25	7,9	13,1	14,9
Q50	10	16,2	17,1
Q75	16	22,4	21,2
Q90	40	29,5	25,7
Q95	56	33	30,8
Maximum	190	130	44,9
Vibrisse	28,1	/	/
MAD	16	/	/

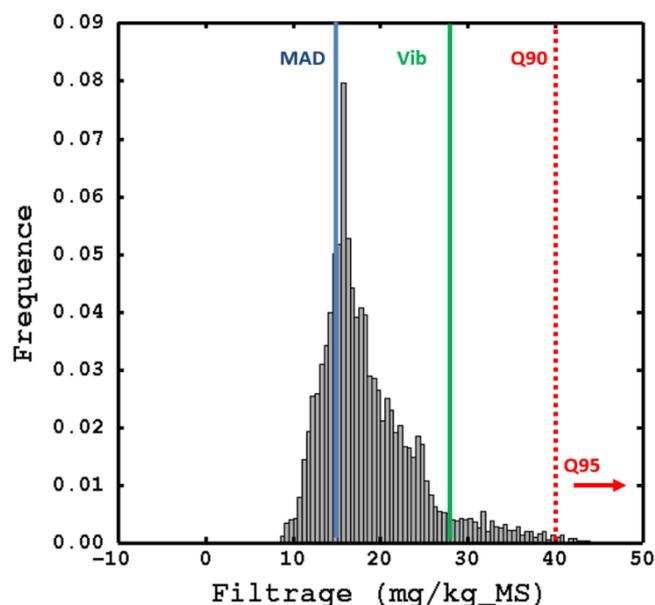


Tableau 20: Quantiles calculés sur les échantillons et les estimations par cokrigage et filtrage des teneurs en plomb des remblais assimilables naturels du quartier, et leur représentation sur l'histogramme des teneurs estimées par filtrage (Q95 : rouge, Vibrisse supérieure : vert, seuil MAD : bleu).

	Zinc		
	Echantillons	Cokrigage	Filtrage
Minimum	3,6	13,4	18,4
Q25	20	29,1	30,6
Q50	21	38,5	38,7
Q75	38	52	50,3
Q90	79	69,9	67,8
Q95	100	83,7	80,9
Maximum	180	132,2	122,7
Vibrisse	65	/	/
MAD	31	/	/

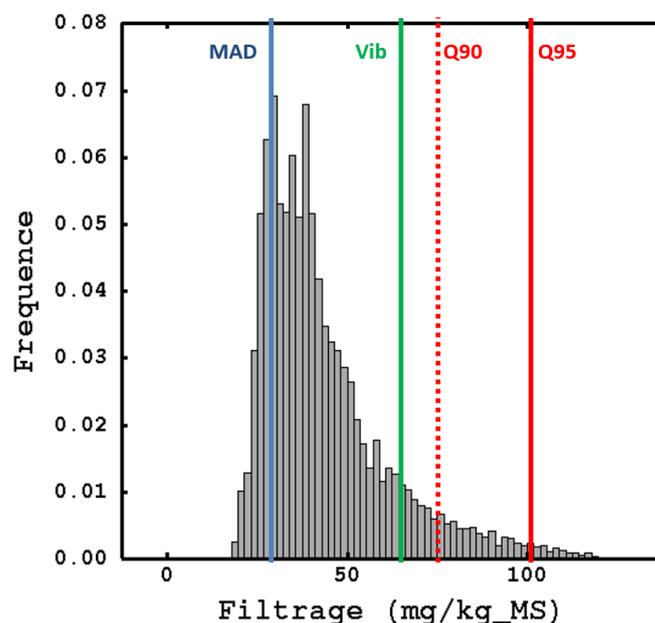


Tableau 21: Quantiles calculés sur les échantillons et les estimations par cokrigage et filtrage des teneurs en zinc des remblais assimilables naturels du quartier, et leur représentation sur l'histogramme des teneurs estimées par filtrage (Q95 : rouge, Vibrisse supérieure : vert, seuil MAD : bleu).

IV.C. Application à un quartier

Ainsi, une cartographie filtrant les anomalies des teneurs dans des sols relativement homogènes et de même nature est possible. Cette cartographie peut être associée à l'estimation de l'épaisseur de la couche de remblais concernée. Dans les cas où il aurait été réalisé, le modèle géologique en trois dimensions du sous-sol est utile.

Pour l'exploitation des résultats, des cartes par tranche de profondeur sont recommandées. Il est alors possible de réaliser des cartes d'affleurement des remblais (Figure 96) à partir de l'estimation de l'épaisseur des remblais ou du modèle géologique en trois dimensions. Les estimations du fond pédo-géochimique pour chaque type de remblais sont ensuite attribuées à ces cartes d'affleurement pour les tranches correspondantes. Une cartographie des teneurs du fond pédo-géochimique anthropisé, suivant la profondeur et le type de remblais peut alors être obtenue (Figure 97).

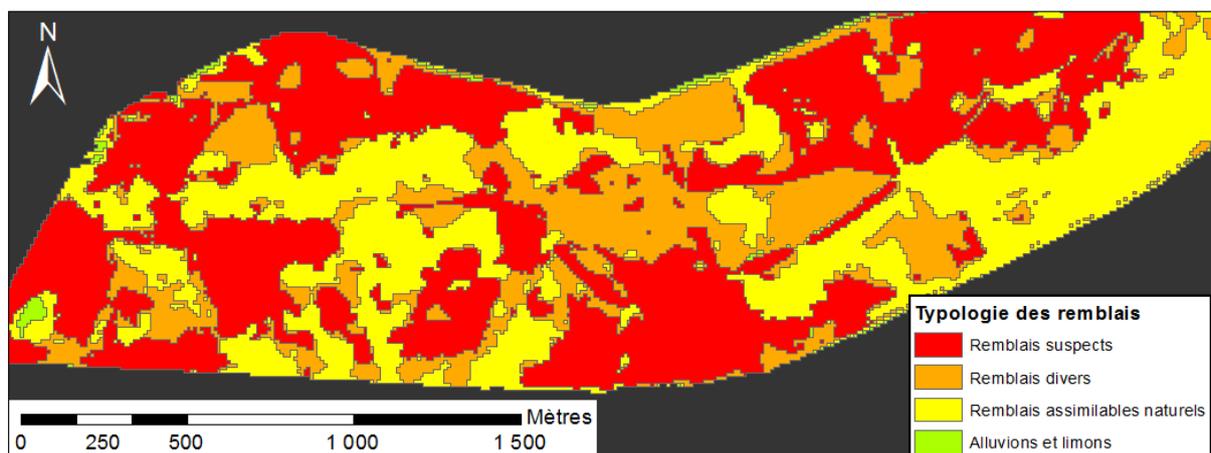


Figure 96: Carte d'affleurement des remblais à 1m de profondeur, selon le modèle géologique 3D réalisé à partir des sondages et de données historiques.

IV.C. Application à un quartier

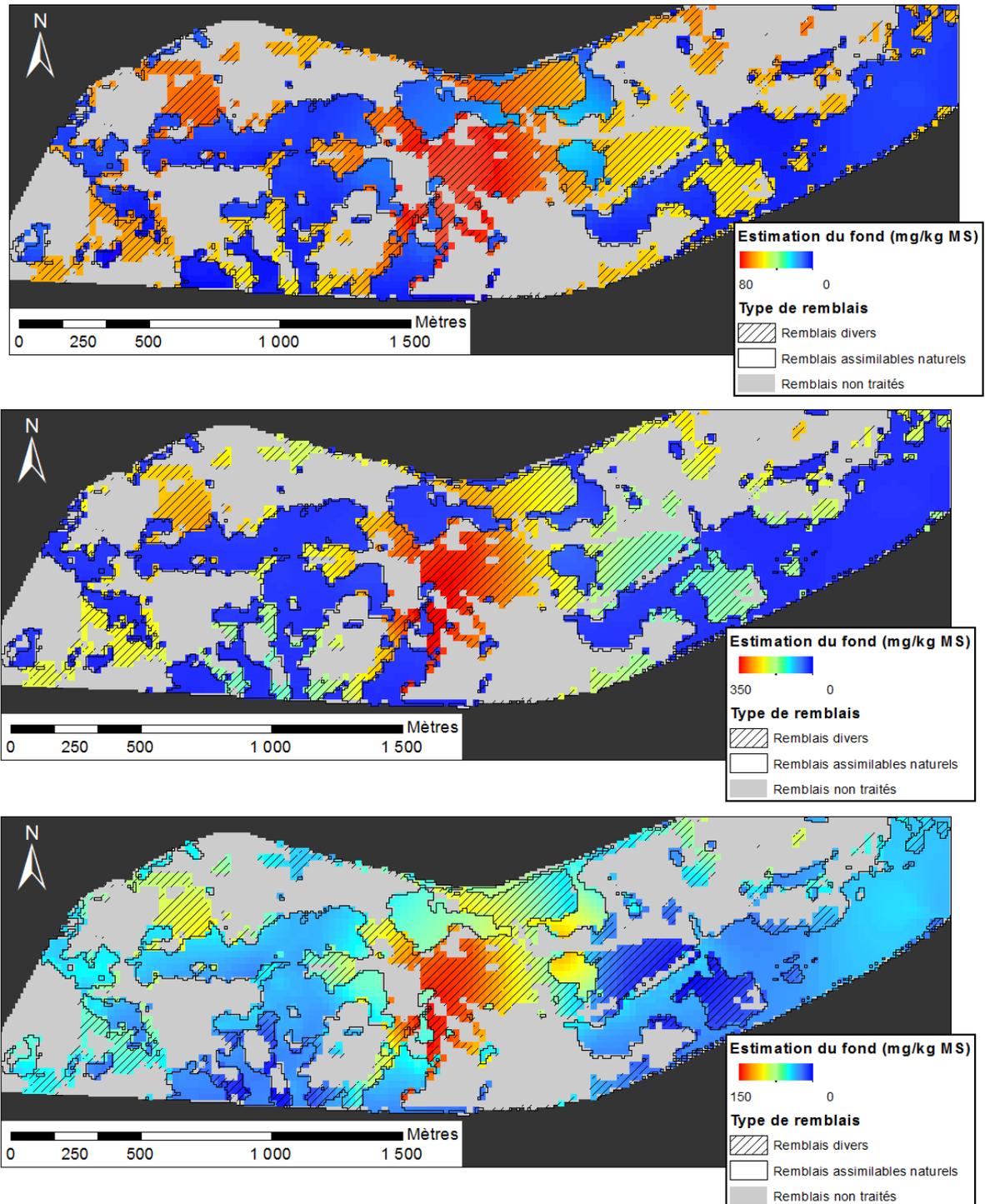


Figure 97: Estimation du fond pédo-géochimique anthropisé dans les remblais divers et assimilables naturels à une profondeur de 1m.

Haut : Cuivre, Milieu : Plomb, Bas : Zinc

IV.D. Prise en compte d'échantillons de supports variables

Les échantillons utilisés pour l'estimation ont des longueurs variables. Précédemment, nous avons sélectionné des échantillons de longueur (support) similaire. Il est toutefois préférable de conserver les échantillons de supports variables dans le krigeage. Pour cela nous utilisons une discrétisation des échantillons et adaptons le modèle variographique à la dimension du pas de discrétisation. Le logiciel R 3.5.1 avec le package RGeostats 11.2.9, a été utilisé pour cette estimation.

Le but de ce travail est de fournir une estimation du fond pédo-géochimique par analyse krigeante (filtrage) en utilisant tous les échantillons quel que soit leur support.

a) Principe

La méthode consiste tout d'abord, à réaliser une analyse variographique sur une sélection d'échantillons de longueur 'L' (ici entre 50 et 70 cm, nous utiliserons $L = 60\text{cm}$). Comme nous ne disposons généralement pas de plusieurs échantillons le long des sondages, seul le variogramme horizontal est calculé. On supposera ce variogramme (pépite, sphérique 200 m et sphérique 1000 m) isotrope.

On souhaite ensuite déduire de ce variogramme au support 'L', un variogramme tridimensionnel anisotrope au support 'd', plus petit que 'L' suivant lequel seront discrétisés les données. Ici on choisit $d = 10\text{ cm}$ comme longueur de discrétisation verticale. Cette modification du variogramme défini pour 'L' influence l'effet de pépite, qui augmente puisque 'd' est inférieur à 'L', et les portées qui sont alors diminuées de la valeur $L-d$ dans la direction de régularisation, ici verticale (les portées étant de 200 et 1000 m, leur correction est ici négligeable). Les paliers du variogramme déduits pour le support 'd' doivent ensuite être recalculés.

Le variogramme déduit peut ensuite être utilisé pour l'estimation en discrétisant les supports au pas 'd' dans le premier et le deuxième membre du système de krigeage.

b) Application

Nous appliquons cette démarche à l'ensemble des échantillons de remblais assimilables naturels du quartier. Les données sélectionnées suivant leur longueur d'environ 60 cm sont utilisées dans le krigeage en trois dimensions comme des données ponctuelles situées au milieu de l'échantillon.

Dans RGeostats, la longueur de l'échantillon est renseignée dans la base de données par la fonction « `db.locate` ». Le paramètre « `dbl` » spécifie la variable décrivant la longueur du support dans l'une des directions de l'espace (le support ne varie que dans la direction

IV.D. Prise en compte d'échantillons de supports variables

verticale, la valeur 0 étant attribuée aux deux autres directions de l'espace). La fonction de krigeage utilisée de façon classique renvoie alors un message indiquant qu'une taille de support est spécifiée dans la base de données mais n'est pas traitée par la fonction de krigeage et que les échantillons sont considérés comme « ponctuels ». Cela permet de choisir ou non d'intégrer la taille du support au calcul.

Une des limites de l'outil actuellement disponible dans R est que le nombre de points de discrétisation des supports est fixe et déclaré par le paramètre « `ndisc` », et ce n'est donc pas la longueur du pas de discrétisation qui est fixée. Cela ne correspond pas réellement à notre objectif de discrétisation du support suivant une longueur 'd' fixe, puisque c'est ici le nombre de point de discrétisation qui est fixé (la longueur 'd' est alors variable). Les calculs actuels constituent donc une approximation, et sont effectués à titre indicatif, afin de montrer que la prise en compte d'échantillons de longueur variable est possible. Nous choisissons de fixer `ndisc=6`.

Le support de chaque donnée sera alors discrétisé en « `ndisc` » points dans chacune des dimensions de l'espace. Dans notre cas, 2 des 3 dimensions ayant une longueur de support nulle, seule la direction verticale sera discrétisée. La valeur de « `ndisc` » est transmise au krigeage par l'intermédiaire de la fonction « `keypair` » en utilisant le mot clef « `DataDiscretization` ». Le krigeage est ensuite effectué en tenant compte de la discrétisation et donc de la taille du support.

Ces traitements ont nécessité l'ajout de nouvelles fonctionnalités à la fonction de krigeage du package RGeostats 11.2.9, les ajouts ont été effectués par Didier Renard.

c) Résultats

Pour illustrer cette prise en compte du support, les estimations « ponctuelles » aux nœuds d'une grille de maille 25x25x0,5 m (un pas de 50 cm est utilisé pour la direction verticale pour comparaison aux cartes d'affleurement des remblais si nécessaire) et leurs écarts-types d'estimation obtenus par filtrage sans prise en compte du support (sélection d'échantillons de support 'L'), sont comparées à celles obtenues en discrétisant le support. Les résultats sont présentés sous forme de nuages de corrélation.

La Figure 98, présente les estimations par filtrage du cuivre, du plomb et du zinc après sélection d'échantillons de longueur 'L' d'environ 60 cm (ordonnée) et sans sélection des données suivant la longueur (abscisse). Les résultats montrent des valeurs globalement plus fortes dans le cas des estimations prenant en compte le support (discrétisation), cela étant plus marqué à mesure que les valeurs augmentent. Cela s'explique par l'augmentation du nombre de données disponibles (nous faisons intervenir toutes les données via la discrétisation du support et non une sélection), mais aussi par leur meilleure répartition dans l'espace. Les données écartées par la sélection comportant des valeurs plus fortes que celles sélectionnées, leur intégration par discrétisation a alors tendance à augmenter les valeurs estimées. Les

IV.D. Prise en compte d'échantillons de supports variables

maxima pour les données sélectionnées sont alors : cuivre (110 mg/kg_MS), plomb (140 mg/kg_MS) et zinc (310 mg/kg_MS). Pour les données sans sélection ils sont de : cuivre (130 mg/kg_MS), plomb (240 mg/kg_MS) et zinc (370 mg/kg_MS).

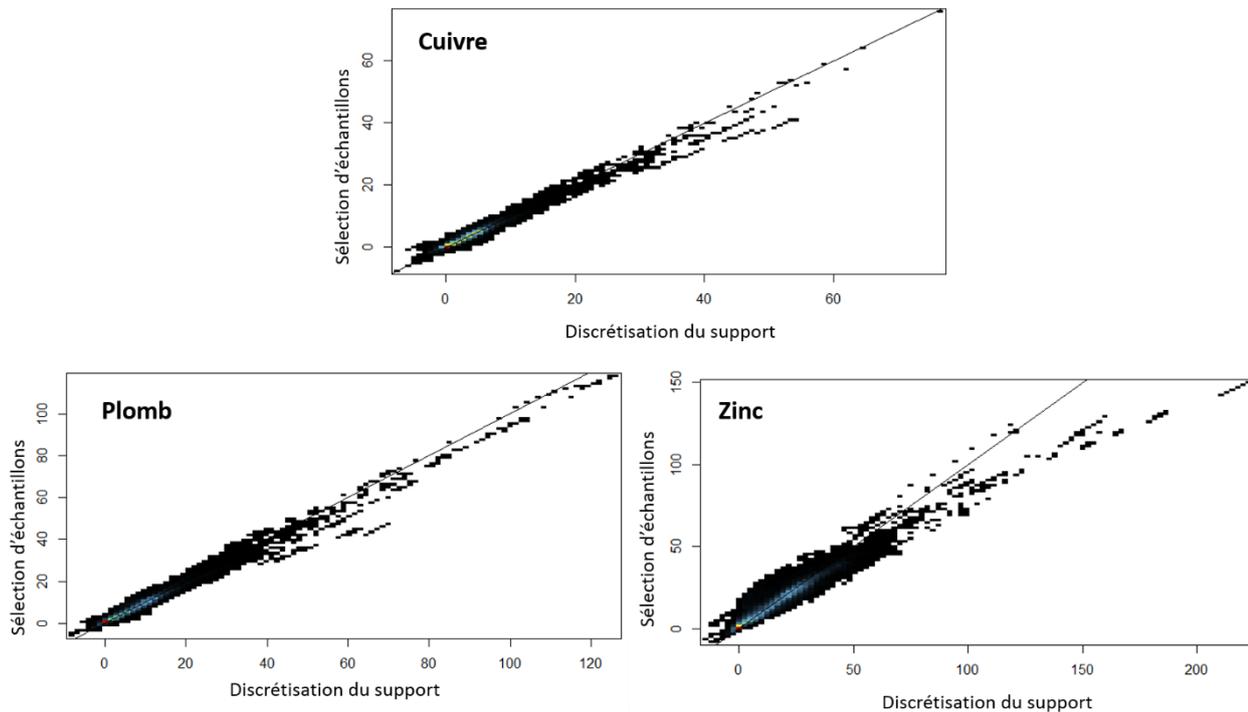


Figure 98 : Nuages de corrélation des estimations en cuivre, plomb et zinc obtenues par filtrage avec sélection de données (ordonnée) et en discrétisant le support en n_{disc} points (abscisse).

L'écart-type d'estimation (Figure 99) est systématiquement plus élevé dans les estimations après la sélection d'échantillons que par discrétisation en l'absence de sélection, cette différence étant plus marquée pour les valeurs faibles d'écart-types. Cela s'explique par une meilleure répartition des données dans l'espace. En effet, les données sont alors représentées par « n_{disc} » points répartis dans un certain volume de l'espace (la taille réelle du support) plutôt que par un point unique dans le cas ponctuel. Cela est accentué par la présence d'un plus grand nombre de données du fait de l'absence de sélection (138 échantillons pour la sélection à 342 sans sélection)

IV.D. Prise en compte d'échantillons de supports variables

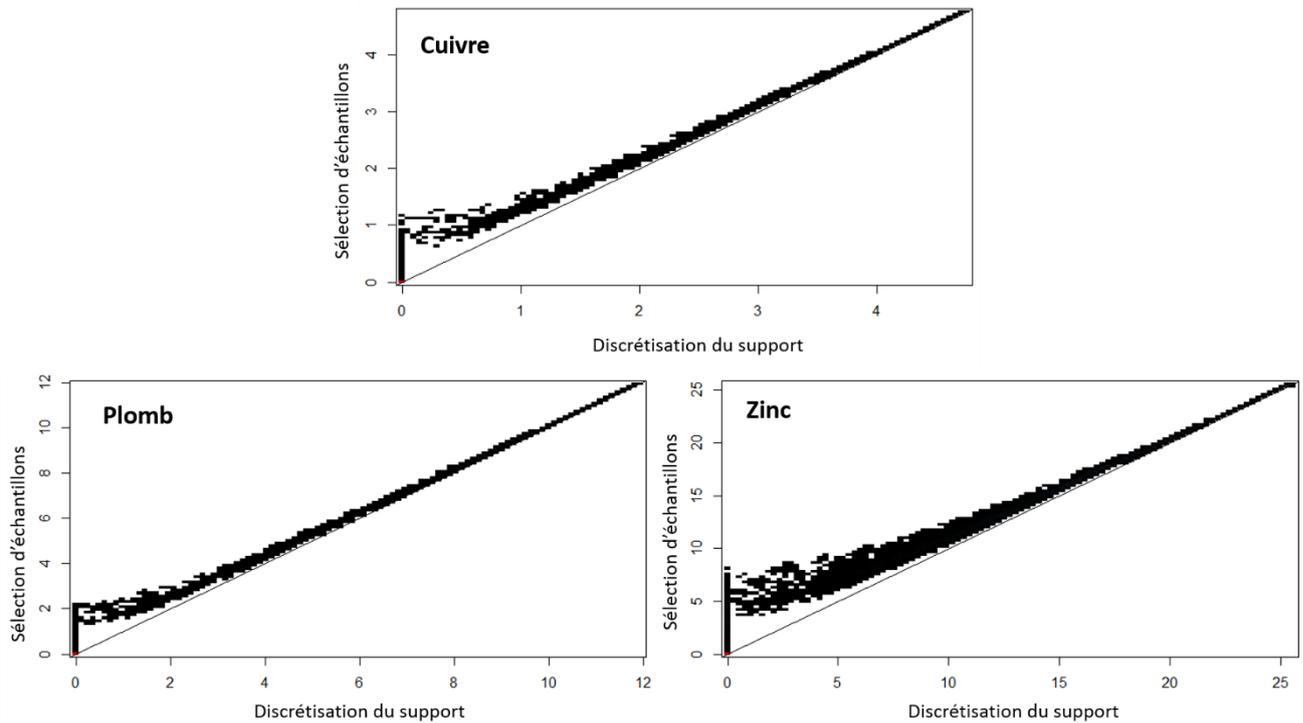


Figure 99 : Nuages de corrélation des écarts-type d'estimation du cuivre, plomb et zinc obtenues par filtrage avec sélection de données (ordonnée) et en discrétisant le support en n disc points (abscisse).

Il est donc possible de traiter conjointement des échantillons de longueurs variables comme c'est généralement le cas suite à une collecte de données préexistantes d'origines diverses. La prise en compte de l'ensemble des supports permet de diminuer l'écart-type d'estimation et donc d'apporter de la robustesse à l'estimation réalisée, sous réserve que le modèle variographique utilisé soit correct.

Cette prise en compte du support est préconisée dans les cas de données au support très variable afin d'éviter de réduire le nombre de données. Des développements informatiques sont néanmoins encore nécessaires afin de passer d'un système au nombre de points de discrétisation fixe à un système fixant le pas de discrétisation, afin d'améliorer la modélisation et l'estimation.

IV.E. Conclusion et perspectives

Il est possible cartographier le fond pédo-géochimique anthropisé, à l'échelle d'un site ou d'un quartier en s'appuyant sur des méthodes géostatistiques et en particulier en ayant recours au filtrage via un modèle linéaire de corégionalisation utilisant des co-variables.

Le filtrage fournit une image lissée de la géochimie en supprimant les variations à courte portée des teneurs. Cela prend en compte les anomalies positives (contaminations locales), mais aussi négatives (zone de teneurs faibles). Dans le cas d'une utilisation du fond pédo-géochimique pour la valorisation des terres excavées ces anomalies négatives sont intéressantes. Ainsi, le filtrage ou non de ces zones de faibles teneurs reste donc une question à débattre, notamment au sein de futurs groupes de travaux nationaux sur le fond pédo-géochimique anthropisé.

Si les teneurs estimées pour le fond sont plus homogènes que celles estimées par un cokrigeage standard, elles restent néanmoins influencées par la moyenne des teneurs dans les échantillons. Dans les exemples traités, les quantiles centraux (Q25, Q50, Q75) sont plus élevés pour le fond pédo-géochimique estimé. Les anomalies étant filtrées, les quantiles d'ordre plus faibles ou plus élevés ont toutefois des valeurs plus centrées autour de la médiane. De plus, la gestion de la moyenne dans le traitement peut également être à débattre. En effet, la moyenne (même pondérée) des échantillons collectés reflète-t-elle bien la moyenne spatiale du type de matériau considéré ?

Ensuite, dans nos tests, la longueur variable des échantillons est un obstacle important puisque lié à la variance. Celle-ci a alors nécessité une sélection d'échantillons limitant grandement l'effectif, ou une hypothèse d'homogénéité du matériau qui n'est pas toujours vérifiable. Le recours à des échantillons de volume constant n'est toutefois pas compatible avec les méthodologies actuelles d'échantillonnage visant à échantillonner au maximum l'ensemble de la hauteur de la couche concernée.

Une autre approche envisagée a été un modèle en puissance – accumulation (testée mais non concluante avec le jeu de données exploité, elle n'est pas présentée). Celle-ci nécessite de connaître l'épaisseur de la couche (puissance), la teneur moyenne le long d'un échantillon vertical dans cette couche, mais également la masse volumique. On peut alors travailler en accumulation correspondant au produit de la teneur par la puissance. Cette approche est toutefois compliquée par le manque de connaissance sur la variabilité de la masse volumique des matériaux en place. De plus, l'avantage principal de tel modèle repose sur le recours de modèle à résidus où l'estimation de l'accumulation dépend en majeure partie de l'estimation de la puissance à laquelle s'ajoute un résidu lié aux variations latérales des teneurs. Cela ne s'est toutefois pas vérifié dans notre cas d'étude sur le quartier nantais. Enfin, la présentation de résultats sous la forme d'une accumulation reste compliquée dans le cadre d'une utilisation dans un référentiel de qualité géochimique pour la compréhension d'un large panel d'acteurs de l'aménagement.

IV.E. Conclusion et perspectives

Enfin, si l'hypothèse d'homogénéité verticale du matériau ciblé ne peut être vérifiée, la prise en compte d'échantillons de supports différents est un problème pratique sans difficulté théorique par la discrétisation des échantillons. Celle-ci renforce l'estimation en permettant l'utilisation de toutes les données, mais nécessite encore des développements pour être mise en place. L'étude exploratoire préalable des données est donc une étape clé de la méthodologie.

La représentation des résultats par des cartes interpolées, bien qu'utile pour visualiser les variations, n'est toutefois pas adaptée pour fournir des valeurs de gestion pour les terres excavées. Un traitement complémentaire est donc nécessaire afin de fournir des résultats exploitables.

Dans une première approche, une gamme de valeurs correspondant aux estimations par filtrage sur l'étendue d'un polygone (exemple : affleurement des remblais à une certaine profondeur) pourrait être fournie. La question se pose toutefois de la définition de la gamme de valeurs. Doit-on considérer toutes les données ou sélectionner un quantile pour le minimum et un second pour le maximum (à la manière des seuils statistiques) ?

Dans une approche plus précise, des cartes interpolées en considérant la teneur estimée par le filtrage en un point comme la teneur limite du fond, pourraient être réalisées. Ainsi, les analyses d'un lot de terres pourraient être comparées aux estimations par filtrage et fournir en retour une cartographie de l'ensemble des zones où les teneurs analysées sont compatibles pour une réutilisation (en croisant les résultats par substance) (Figure 100).

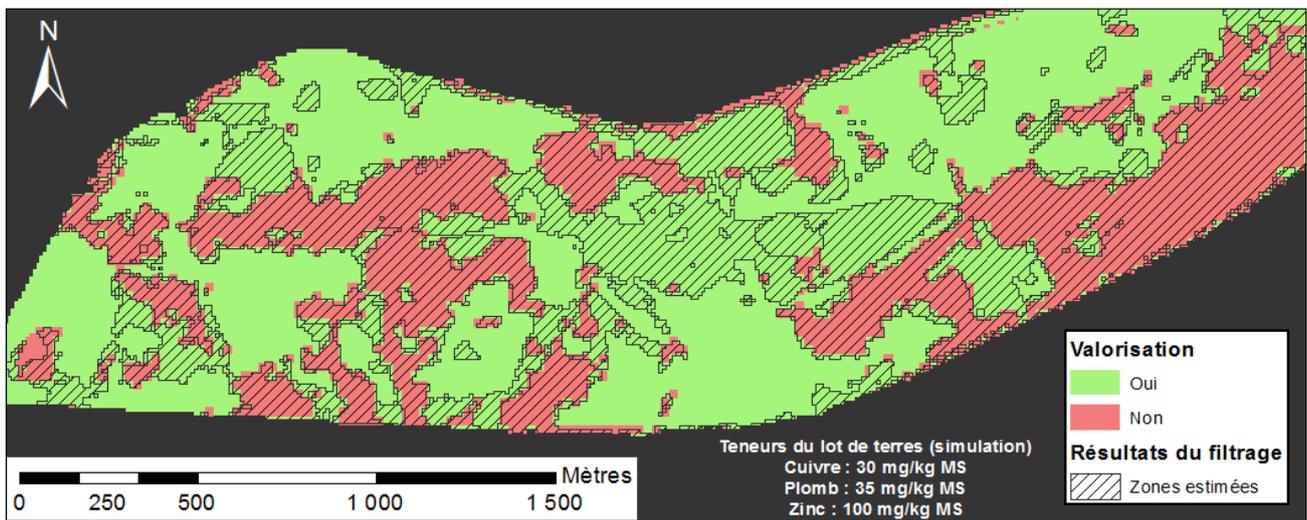


Figure 100: Résultats de cartographie des zones de valorisation d'un lot de terres excavées dans la tranche 0-1m de profondeur. Les teneurs utilisées pour simuler le lot sont :

Cuivre = 30 mg/kg MS, Plomb = 35 mg/kg MS, Zinc = 100 mg/kg MS

Les résultats du filtrage géostatistique sont complétés (remblais suspects et alluvions) d'après les seuils statistiques de (Le Guern et al., 2016).

Recommandations :

Le filtrage géostatistique par un modèle linéaire de corégionalisation, nécessite un grand nombre de données. Son application à des aires urbaines relativement grandes (un quartier par exemple), requiert une bonne connaissance des différents types de matériaux et d'occupation de sols. Ces données descriptives pouvant être associées comme covariables pour les estimations.

Son utilisation est donc plutôt conseillée sur les quartiers aux enjeux importants et présentant un grand nombre de données. A l'échelle de la ville une collecte et une étude exploratoire importante des données est nécessaire avant de pouvoir procéder au filtrage.

IV.E. Conclusion et perspectives

CONCLUSION

Ce travail de thèse s'est donné pour but d'évaluer les possibilités de réutilisation de données géochimiques disponibles pour la détermination de valeurs de référence de la qualité géochimique des sols urbains.

L'application des méthodes statistiques usuelles avec des données non dédiées au fond pédo-géochimique anthropisé ont été examinées.

Les critères statistiques seuls, ne peuvent pas toujours garantir un résultat cohérent avec ce type de données non dédiées au fond pédo-géochimique et ce type d'environnement fortement hétérogène. Si l'étude de la distribution, de la limite de quantification et de l'effectif sont, parmi d'autres, des paramètres à prendre en compte, c'est principalement la part des anomalies dans le jeu de données qui déterminera la cohérence du résultat. En effet, afin que les seuils calculés puissent représenter le fond pédo-géochimique, ils doivent séparer le fond des anomalies. Les anomalies doivent représenter une part limitée des données. Nous avons montré que les seuils calculés par la vibrisse supérieure sont plus adaptés à une valorisation de volumes plus importants de terres excavées, alors que l'utilisation du seuil MAD, plus conservatif, convenait mieux pour un objectif de protection de l'environnement.

Dans cette première démarche statistique, une seule valeur de fond est déterminée pour l'ensemble des données traitées. Toutefois, la géochimie des sols des aires urbaines étant liée à l'occupation de sols et à la nature de matériaux, différents niveaux de fond pédo-géochimique peuvent-être nécessaires.

Nous avons donc ensuite cherché à cartographier les sols urbains suivant leur potentiel de présence d'anomalies géochimiques ou leur qualité générale, afin de fournir plusieurs valeurs de fond suivant différents niveaux de qualité des sols liés à l'activité anthropique actuelle ou passée et/ou à la nature même de ces sols.

La première méthode testée fut de cartographier le potentiel de présence d'anomalie a priori, à partir d'une analyses multi-critères sur des données descriptives de la nature et de l'occupation des sols. Les résultats ont montré que l'application à l'échelle de la ville ne pouvait tenir compte de l'ensemble des variations de la géochimie du fait de l'échelle relativement large des descripteurs cartographiques. Si cette échelle de travail reste utile pour un aperçu général des grandes tendances du potentiel d'anomalies, elle n'est toutefois pas optimale pour une étude approfondie sur un secteur précis. La méthode a également été testée sur des données cartographiques plus précises disponibles pour certains quartiers de la ville. Bien que toutes les classes cartographiées ne soient pas uniformément échantillonnées, les résultats sont toutefois plus cohérents qu'à l'échelle de la ville avec des teneurs moyennes plus élevées dans la classe au potentiel d'anomalie le plus fort, et un plus grand nombre de teneurs fortes associées aux anomalies géochimiques.

CONCLUSION

L'analyse multi-critère se fonde sur la notation a priori des descripteurs suivant des recherches bibliographiques et des retours d'expérience. Néanmoins une phase d'adaptation des notes à la ville ou au quartier étudié est nécessaire. Si dans notre cas d'étude cette phase d'adaptation a été relativement simple du fait des connaissances acquises via différentes études antérieures, elle pourrait être plus longue sur d'autres villes dont la géochimie a été moins étudiée. L'application à d'autres villes serait donc nécessaire pour tester la précision des notes et le degré d'adaptation nécessaire.

Par la suite, l'application des critères statistiques peut se faire suivant ce découpage du territoire permettant l'identification de différents niveaux de fond.

La seconde méthode de découpage du territoire reposait sur une démarche inverse, à savoir une classification des échantillons suivant les teneurs mesurées, puis une comparaison de ces classes aux données cartographiques descriptives du découpage. Si des classes homogènes présentant différents niveaux de contamination ont pu être identifiées, aucune correspondance n'a toutefois été trouvée avec les données cartographiques descriptives. Si un découpage est néanmoins possible, par exemple par polygones d'influence, celui-ci ne reflète pas de limites physiques et est donc difficilement utilisable pour un outil de gestion des terres excavées.

Enfin, s'il est possible de découper le territoire pour plus de cohérence vis-à-vis des seuils calculés par les critères statistiques, ceux-ci restent néanmoins développés pour des données spécifiques au fond pédo-géochimique et restent influencés par la proportion d'anomalie dans les données. Ainsi, même avec le découpage du territoire, la séparation des anomalies géochimiques par ces critères statistiques n'est pas garantie. Un travail préliminaire de tri des données est donc nécessaire afin d'éliminer les anomalies, mais le tri reste arbitraire.

Nous avons donc étudié les apports de la géostatistique à la séparation des anomalies dans les données, pour l'estimation du fond pédo-géochimique.

Se basant sur une approche spatiale où l'anomalie géochimique est définie comme une variation à courte portée des teneurs sur une zone restreinte, le filtrage via un modèle linéaire de corégionalisation a été testé sur un site et à l'échelle d'un quartier.

Une sélection des échantillons suivant leur longueur et leur profondeur a été nécessaire pour l'étude variographique et a fortement réduit l'effectif disponible. Néanmoins, les résultats obtenus sur le cuivre, le plomb et le zinc montrent bien une différence marquée dans les estimations obtenues par cokrigage et par filtrage, ces résultats étant utilisables pour définir le fond pédo-géochimique anthropisé.

L'application au quartier a ensuite nécessité la séparation des échantillons suivant les types de remblais. Cela permet de séparer des matériaux de qualité géochimique parfois très différente dans l'estimation. Comme pour le site, les différences entre les estimations par cokrigage et filtrage sont nettes et le fond alors estimé apparaît bien plus homogène et aux

CONCLUSION

variations plus lentes que la simple estimation de la teneur. Enfin, nous avons également montré qu'il était possible de transformer ces résultats interpolés en des cartes de sites potentiels pour la valorisation en couplant les résultats d'estimation du fond par filtrage à des cartes d'affleurement des types de remblais obtenus via le modèle 3D du sous-sol.

Cette approche par filtrage comporte néanmoins des limites, dont la question du caractère préférentiel de l'échantillonnage pouvant induire un biais. Une des solutions est de recourir à des échantillonnages complémentaires.

Malgré ces limites, le filtrage géostatistique apparaît comme une méthode fiable pour séparer le fond et les anomalies.

Il est donc possible d'utiliser des données non spécifiques au fond pédo-géochimique pour fournir des valeurs de référence de la qualité géochimique des sols urbains, mais en les complétant.

Si les critères statistiques seuls ne sont pas adaptés pour traiter l'ensemble d'une aire urbaine, le recours à un découpage du territoire fournit différents niveaux de seuils pour la ville. Ce genre d'approche à l'échelle de la ville peut alors être réalisée comme étude préliminaire avec peu de données mais réparties sur l'ensemble de la ville afin de fournir un premier aperçu des grandes tendances de la géochimie dans l'aire urbaine. La marche à suivre est alors synthétisée dans la Figure 101. Toutefois, dans les cas d'effectifs réduits et du fait du recours aux critères statistiques nous conseillons de ne valoriser que les seuils les plus conservatifs afin de garantir la préservation de la qualité géochimique des sols.

Des résultats plus précis sont cependant possibles dans les cas où des effectifs plus importants de données sont disponibles. En raison des coûts de cette approche nous la conseillons pour des quartiers à fort enjeu économique pour la valorisation des terres excavées (importants travaux générant des volumes conséquents de terres excavées). Cette approche repose alors sur une collecte et la conception de descripteurs urbains plus précis tels qu'un modèle géologique 3D et si possible un Inventaire Historique Urbain. Ces données permettant à la fois un découpage plus précis du territoire via l'analyse multi-critères, mais aussi de disposer de données précises comme co-variables pour l'étude géostatistique. Par la suite, le recours à la géostatistique à travers le filtrage est conseillé pour éliminer les anomalies des échantillons collectés. Si l'approche géostatistique n'est pas possible, le recours à la vibrissse supérieure et au seuil MAD est envisageable. Contrairement à l'échelle de la ville, il est alors possible de proposer des seuils moins conservatifs du fait de la précision plus importante des résultats. Cette approche est résumée à la Figure 102.

Les travaux de cette thèse ont apporté des éléments de réflexion au groupe de travail national sur le fond pédo-géochimique anthropisé, sur la pertinence des critères statistiques et sur les principaux paramètres les influençant. Ils ont notamment mis en évidence l'influence de la limite de quantification et de sa valeur de remplacement en montrant que le recours au 0

CONCLUSION

comme valeur de substitution n'implique pas nécessairement des seuils plus faibles que pour une substitution par la valeur de la LQ.

Ces travaux ont également montré l'intérêt du découpage du territoire en entités géographiques cohérentes, en fournissant une méthodologie différente de celle déjà mise en œuvre dans GeoBaPa. De plus, si le découpage du territoire est ici orienté vers le potentiel d'anomalies géochimiques, il peut aussi s'appliquer à la cartographie de la qualité agronomique potentielle des sols, du potentiel de stockage de carbone ou de leur qualité écologique. Ces travaux ouvrent des perspectives de recherches sur l'intégration des données ponctuelles telles que Basias dans les méthodes de segmentation du territoire à l'échelle des villes. L'intégration de ces données permettant l'utilisation d'autres sources d'informations. La suite des recherches peut également concerner les traitements topologiques de simplification du découpage facilitant l'utilisation des résultats dans des outils d'aide à la décision.

Enfin, ces résultats présentent un exemple d'apport de la géostatistique au référentiel de qualité géochimique des sols, ont permis de sensibiliser les acteurs du milieu à l'intérêt de tels outils, et ouvrent la voie vers de nouveaux questionnements tels que l'influence d'un échantillonnage orienté dans ces méthodes de traitement. La prise en compte d'échantillons de longueurs variables est également un point d'intérêt pour les recherches futures de façon à pouvoir traiter la majorité des données disponibles en géochimie urbaine. Enfin, la définition du fond pédo-géochimique à partir des résultats de filtrage (estimation d'une valeur seuil, d'un majorant ou d'une gamme de valeur) reste à statuer en collaboration avec les différents groupes de travail sur la valorisation des terres excavées et de valeurs de fond pédo-géochimique.

Ces travaux n'ayant pu être appliqués qu'à une ville pour des questions de temps, leur application à d'autres aires urbaines permettra d'en étudier la robustesse et si besoin d'adapter les outils.

Une étape de transition entre ces résultats de recherche et un outil opérationnel et compréhensible pour des opérateurs divers est nécessaire. Cela doit également passer par le traitement des résultats de façon à les rendre compréhensibles et utilisables par les aménageurs et les différents acteurs de la valorisation des terres excavées. Le projet « Fond Pédo-géochimique Urbain – FGU » conduit par l'ADEME et le BRGM, et qui conduit au développement de la BDSolU, pourrait être propice à de tels développements.

Ces travaux permettent de développer des outils d'aide à la décision facilitant la mise en place de plateformes de valorisation des terres excavées, dans le but d'organiser une meilleure gestion économique et écologique de ces terres, en limitant notamment le comblement des sites de stockage et l'apport de terres extérieures.

Enfin, ces travaux sont également l'occasion de mieux comprendre et cartographier un aspect des sols urbains, la géochimie.

CONCLUSION

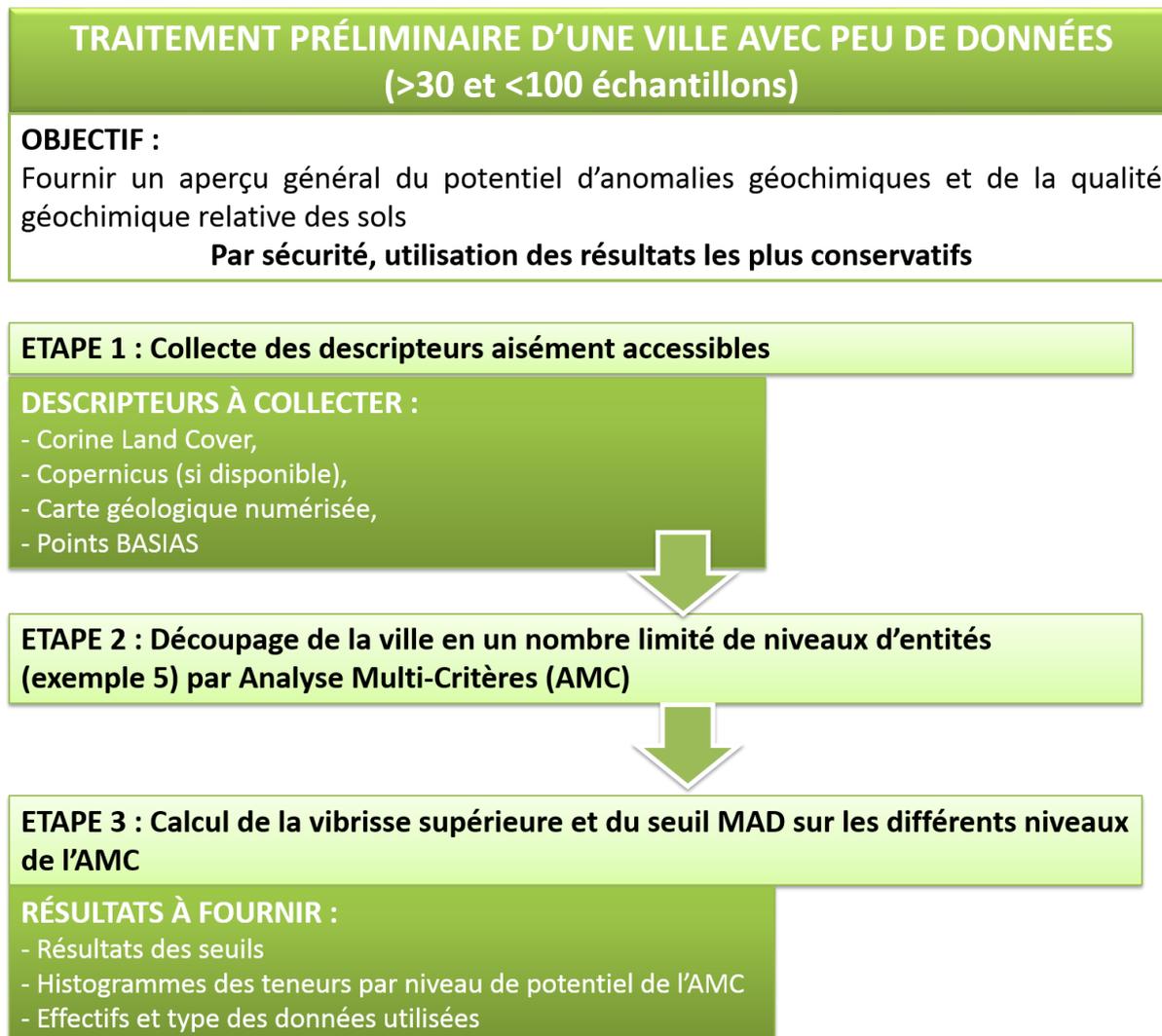


Figure 101: Schéma conceptuel des traitements à appliquer pour la constitution de valeurs de référence de la qualité géochimique des sols urbains, à l'échelle de la ville en cas d'un effectif réduit de données disponibles.

CONCLUSION

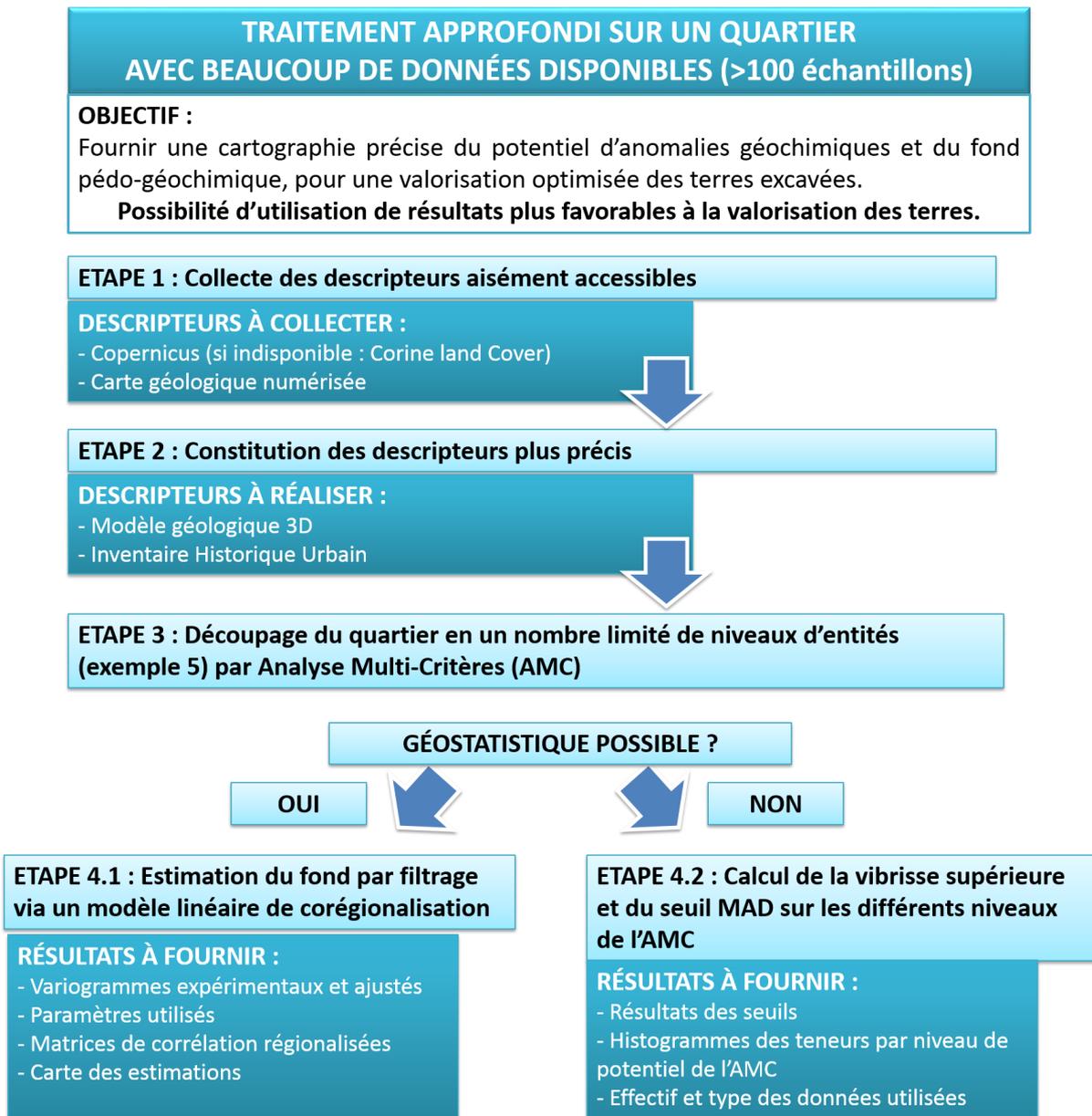


Figure 102: Schéma conceptuel des traitements à appliquer pour la constitution de valeurs de référence de la qualité géochimique des sols urbains, à l'échelle d'un quartier en cas d'un effectif important de données disponibles.

REFERENCES

- Acosta J. A., Faz A., Martínez-Martínez S., & Arocena J. M.** (2011). Enrichment of metals in soils subjected to different land uses in a typical Mediterranean environment (Murcia City, southeast Spain). *Applied Geochemistry*, 26(3), 405-414.
<https://doi.org/10.1016/j.apgeochem.2011.01.023>
- ADEME.** (2018). Méthodologie de détermination des valeurs de fonds dans les sols : Echelle territoriale. Groupe de travail sur les valeurs de fonds. 112 p.
- Ahmed I., Nazzal Y., Zaidi, F. K., Al-Arifi N. S., Ghrefat H., & Naeem M.** (2015). Hydrogeological vulnerability and pollution risk mapping of the Saq and overlying aquifers using the DRASTIC model and GIS techniques, NW Saudi Arabia. *Environmental Earth Sciences*, 74(2), 1303-1318.
<https://doi.org/10.1007/s12665-015-4120-5>
- Albanese S., De Vivo B., Lima A., & Cicchella D.** (2007). Geochemical background and baseline values of toxic elements in stream sediments of Campania region (Italy). *Journal of Geochemical Exploration*, 93(1), 21-34.
<https://doi.org/10.1016/j.gexplo.2006.07.006>
- Allen M. A., Cave M. R., Chenery S., Gowing C. J., & Reeder S.** (2011), Sample preparation and inorganic analysis for urban geochemical survey soil and sediment samples. *Mapping the chemical environment of urban areas*, 28-46.
<https://doi.org/10.1002/9780470670071.ch3>
- Alloway B. J.** (1990). Soil processes and the behaviour of metals. *Heavy metals in soils.*, 7-28.
- Ander E. L., Johnson C. C., Cave M. R., Palumbo-Roe B., Nathanail C. P., & Lark R. M.** (2013). Methodology for the determination of normal background concentrations of contaminants in English soil. *Science of the Total Environment*, 454, 604-618.
<https://doi.org/10.1016/j.scitotenv.2013.03.005>
- Appleton J. D., & Adlam K. A. M.** (2012). Geogenic control on soil chemistry in urban areas: a novel method for urban geochemical mapping using parent material classified data. *Applied Geochemistry*, 27(1), 161-170.
- Appleton J. D., Johnson C. C., Ander E. L., & Flight D. M. A.** (2013). Geogenic signatures detectable in topsoils of urban and rural domains in the London region, UK, using parent material classified data. *Applied Geochemistry*, 39, 169-180.

REFERENCES

- Argyrazi A., & Kelepertzis E.** (2014). Urban soil geochemistry in Athens, Greece: the importance of local geology in controlling the distribution of potentially harmful trace elements. *Science of the Total Environment*, 482, 366-377. <https://doi.org/10.1016/j.scitotenv.2014.02.133>
- Arnaud M., Emery X., De Fouquet C., Brouwers M., & Fortier M.** (2001). L'analyse krigeante pour le classement d'observations spatiales et multivariées. *Revue de statistique appliquée*, 49(2), 45-67.
- Artmann M.** (2015). Managing urban soil sealing in Munich and Leipzig (Germany)—From a wicked problem to clumsy solutions. *Land Use Policy*, 46, 21-37.
- Baize D.**(2008), Eléments traces dans les sols : ne plus parler de “Bruit de fond”, *Environnement et Technique* N°281, 5, 25-30.
- Baize, D.** (2000). Teneurs totales en « métaux lourds» dans les sols français: résultats généraux du programme ASPITET. *Le Courrier de l'Environnement de l'INRA*, (40), 39-54. hal-01203415, version 1
- Ban Y., Jacob A., & Gamba P.** (2015), Spaceborne SAR data for global urban mapping at 30m resolution using a robust urban extractor. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, 28-37.
- Bastian O.** (2000). Landscape classification in Saxony (Germany)—a tool for holistic regional planning. *Landscape and Urban Planning*, 50(1), 145-155.
- Bavec Š., Biester H., & Gosar M.** (2014). Urban sediment contamination in a former Hg mining district, Idrija, Slovenia. *Environmental geochemistry and health*, 36(3), 427-439. <https://doi.org/10.1007/s10653-013-9571-6>
- Béchenec F.** (2007) *Carte géologique harmonisée du département de Loire-Atlantique*. BRGM/RP-55703-FR, 369 p., 2 fig., 2 tab., 3 pl. h.-t.
- Béchet, B., Le Bissonais, Y., Ruas, A., Aguilera, A., André, M., Andrieu, H., & Belton-Chevallier, L.** (2017). *Sols artificialisés et processus d'artificialisation des sols: déterminants, impacts et leviers d'action* (INRA).
- Béchet B., Carré F., Florentin L., Leyval C., Montanarella L., Morel J. L., & Schwartz C.** (2009). Caractéristiques et fonctionnement des sols urbains. *Cheverry et Gascuel (éd) Sous les pavés la terre, Omniscience, Montreuil*, 45-74.
- Bernard-Michel C., de Fouquet C.** (2005). Geostatistical indicators of nutrients concentrations in streams. In :Cheng Q. & Bonham-Carter G. (eds), *Proceedings IAMG2005 Toronto* , York University, Toronto Canada &GPMR, Wuhan, China. 716-721 <https://doi.org/10.1016/j.gexplo.2014.10.011>
- Bernard-Michel, C.** (2006). *Indicateurs géostatistiques de la pollution des cours d'eau* (Thèse de doctorat, École Nationale Supérieure des Mines de Paris), 195p.

REFERENCES

- Birke, M., & Rauch, U.** (2000). Urban geochemistry: investigations in the Berlin metropolitan area. *Environmental Geochemistry and Health*, 22(3), 233-248. <https://doi.org/10.1023/A:1026554308673>
- Birke M., & Rauch U.** (1997). *Geochemical investigations in the Berlin metropolitan area*. *Z. Angew. Geol.* 43(1), 58–65, Hannover.
- Blanc C., Lefevre F., Boissard G., Scamps M. & B. Hazebrouck B.** (2012). Guide deréutilisation hors site des terres excavées en technique routière et dans des projets d'aménagement. BRGM/RP-60013-FR, 53p.
- Blanchart A.**, (2018) *Vers une prise en compte des potentialités des sols dans la planification territoriale et l'urbanisme opérationnel* (Thèse de doctorat, Université de Lorraine, Nancy), 381 p
- Bourennane H., Salvador-Blanes S., Cornu S., & King D.** (2003). Scale of spatial dependence between chemical properties of topsoil and subsoil over a geologically contrasted area (Massif central, France). *Geoderma*, 112(3-4), 235-251. [https://doi.org/10.1016/S0016-7061\(02\)00309-9](https://doi.org/10.1016/S0016-7061(02)00309-9)
- Branchu P., Joimel S., Douay F., Lefebvre G., Rémy E., Béchet B., & Cambier P.** (2016, Octobre). Risques environnementaux et sanitaires dans les jardins. In *AFPP-4e Conférence sur l'Entretien des Jardins Végétalisés et Infrastructures*.
- Cambou A.**, (2018) *Évaluation du stock et de la stabilité du carbone organique dans les sols urbains* (Thèse de doctorat, AGROCAMPUS OUEST, Angers),
- Cave M. R., Johnson C. C., Ander E. L., & Palumbo-Roe B.** (2012). Methodology for the determination of normal background contaminant concentrations in English soils. Nottingham, UK, British Geological Survey, 32p.
- Cheng W., Zhang X., Wang K., & Dai X.** (2009). Integrating classification and regression tree (CART) with GIS for assessment of heavy metals pollution. *Environmental monitoring and assessment*, 158(1-4), 419-431.
- Cicchella D., De Vivo B., Lima A., Albanese S., & Fedele L.** (2008). Urban geochemical mapping in the Campania region (Italy). *Geochemistry: Exploration, Environment, Analysis*, 8(1), 19-29. <https://doi.org/10.1144/1467-7873/07-147>
- Clanché F., & Rascol O.** (2011). Le découpage en unités urbaines de 2010. L'espace urbain augmente de 19% en une décennie, département de démographie. *Insee Première*, (1364).
- Commissariat Général au Développement Durable**, (2012), *Le point sur la dépense de réhabilitation des sites et sols pollués en France*, Observation et statistiques environnement n°142, 4 p.

REFERENCES

- Council of Europe, 6th European Ministerial Conference on the Environment, Brussels** (1990) (MEN6 (90) 2), Feasibility study on possible National and/or European actions in the field of soils protection.
- Coussy S., Hulot C., Billard A.,** (2017). Guide de valorisation hors site des terres excavées issues de sites et solspotentiellement pollués dans des projets d'aménagement. Ministère de la transition écologique et solidaire, 62p.
- Daniau C., Dor F., Denys S., Floch-Barneaud A., & Dab W.** (2009), Problèmes posés par la définition de l'état de référence des sols en santé environnementale. In *Annales des Mines-Responsabilité et environnement* (No. 2, pp. 70-77). ESKA.
<https://doi.org/10.3917/re.054.0070>
- DE, P.** (2001). Union européenne. *Ordonnance n 2010-1579 du 17 décembre 2010 portant diverses dispositions d'adaptation au droit de l'Union européenne dans le domaine des déchets*, 22301.
- de Fouquet.** (2012). C. Environmental Statistics revisited: is the Mean reliable? *Environ. Sci. Technol.*,46(4), 1964-1970. DOI : 10.1021/es2024143
- de Fouquet C.** (2011). From exploratory data analysis to geostatistical estimation: examples from the analysis of soils pollutants. *European Journal of Soil Science, special issue: Pedometrics*. 62(3) 454-466.
- de Fouquet C.** (2006). La modélisation géostatistique des milieux anthropisés. Habilitation à diriger des recherches. *Mémoire des Sciences de la Terre* n° 2006-13. Université Pierre-et-Marie Curie, Académie de Paris
- DÉCHET,** (2003) Article L541-1-1 modifié par la Loi n° 2003-591 du 2 juillet 2003 – art. 31 (V) JORF 3 juillet 2003
- Demetriades, A and Birke, M.** (2015). Urban Topsoil Geochemical Mapping Manual (URGE II).EuroGeoSurveys, Brussels
- Demolon A., & Leroux D.** (1952). *Guide pour l'étude expérimentale du sol* (Vol. 74, No. 5, p. 405).
- Demougeot-Renard H., Haouche-Belkessam L., & Denys S.** (2008). Reconnaissance assistée de sites pollués par l'utilisation conjointe de mesures rapides sur site et de traitements géostatistiques. *Conception et validation d'une démarche itérative de reconnaissance. Rapport final REPERAGE. Rapport FSSADEME2007002.*
- Desse-engrand F.** (2017) Cartographie des sources potentielles de pollution sur les anciens sites industriels et activités de services à Nantes Métropole (Rapport de stage, BRGM, Nantes), 89p

REFERENCES

- Donati M., de Fouquet C.** (2018). How to describe organic contamination in soils: A model comparison for chlorinated solvent concentrations at industrial site scale. *Science of The Total Environment*. 633 (2018) 1480–495.
<https://doi.org/10.1016/j.scitotenv.2018.03.257>
- Douay F., Pruvot C., Mazzuca M., Howsam M., Fourrier H., & De Saint Mahieu A. L.** (2005). Cadmium, lead and zinc concentrations in soil and vegetables from kitchen gardens in urban and highly-contaminated areas of northern France: evaluation of the risk of population exposure, ConSoil 2005. In *Proceedings of the Ninth International FZK/TNO Conference on Soil–Water Systems, Bordeaux, France* (pp. 667-676). DOI [10.1007/s11270-007-9541-7](https://doi.org/10.1007/s11270-007-9541-7)
- Ducimetiere P.** (1970). Les méthodes de la classification numérique. *Revue de Statistique appliquée*, 18(4), 5-25.
- Effland W. R., & Pouyat R. V.** (1997). The genesis, classification, and mapping of soils in urban areas. *Urban Ecosystems*, 1(4), 217-228.
- Emery X.** (2001). Géostatistique linéaire, Ecole des Mines de Paris.
- Erhardt W., Höpker K. A., & Fischer I.** (1996), Bewertungsverfahren Verfahren zur Bewertung von immissionsbedingten Stoffanreicherungen in standardisierten Graskulturen. *Umweltwissenschaften und Schadstoff-Forschung*, 8(4), 237-240.
<https://doi.org/10.1007/BF02940467>
- Ersoy H., & Bulut F.** (2009). Spatial and multi-criteria decision analysis-based methodology for landfill site selection in growing urban regions. *Waste Management & Research*, 27(5), 489-500. <https://doi.org/10.1177/0734242X08098430>
- Fersmann A.** (1935). *The Scientific Study of Soviet Mineral Resources: By AE Fersman. Ed. By CP Dutt.* Indra Co-operative Pub. Soc..
- Fetzer K., Enricht E., Grenzius R., Kubiniok J., Schwartz C., & Morel J. L.** (1998, August). Garden soils in south-western Germany (Saarland) and north-eastern France (Lorraine). In *16ème Congrès Mondial de Science du Sol* (p. 7).
- Fondrillon, M.** (2007). *La formation du sol urbain: étude archéologique des terres noires à Tours (4e-12e siècle)* (Thèse de doctorat, Université François Rabelais-Tours).
- Ford J., Kessler H., Cooper A. H., Price S. J., & Humpage A. J.** (2010). *An enhanced classification of artificial ground*. Nottingham, UK, British Geological Survey, 34pp. (OR/10/036) (Unpublished)

REFERENCES

- Fordyce F. M., Brown S. E., Ander E. L., Rawlins B. G., O'Donnell K. E., Lister T. R., & Johnson C. C.** (2005). *GSUE: urban geochemical mapping in Great Britain. Geochemistry: Exploration, Environment, Analysis*, 5(4), 325-336.<https://doi.org/10.1144/1467-7873/05-069>
- Franck-Néel C., Borst W., Diome C., & Branchu P.** (2015). Mapping the land use history for protection of soils in urban planning: what reliable scales in time and space?. *Journal of soils and sediments*, 15(8), 1687-1704.<https://doi.org/10.1007/s11368-014-1017-y>
- Garrett R. G., Reimann C., Smith D. B., & Xie X.** (2008). From geochemical prospecting to international geochemical mapping: a historical overview. *Geochemistry: Exploration, Environment, Analysis*, 8(3-4), 205-217.
- Guagliardi I., Cicchella D., & De Rosa R.** (2012). A geostatistical approach to assess concentration and spatial distribution of heavy metals in urban soils. *Water, Air, & Soil Pollution*, 223(9), 5983-5998.<https://doi.org/10.1007/s11270-012-1333-z>
- Gavrilenko E. G., Ananyeva N. D., & Makarov O. A.** (2013). Assessment of soil quality in different ecosystems (with soils of Podolsk and Serpukhov districts of Moscow oblast as examples). *Eurasian soil science*, 46(12), 1241-1252.
<https://doi.org/10.1134/S1064229313120041>
- Gasperi J.** (2006). *Introduction et transfert des hydrocarbures à différentes échelles spatiales dans le réseau d'assainissement parisien* (Thèse de doctorat : Marne-la-vallée, ENPC).
- Girard, M. C., & Baize, D.** (2009). *Référentiel pédologique 2008*.
- Göbel P., Dierkes C., & Coldewey W. G.** (2007). Storm water runoff concentration matrix for urban areas. *Journal of contaminant hydrology*, 91(1-2), 26-42
- Gong M., Wu L., Bi X. Y., Ren L. M., Wang L., Ma Z. D., & Li Z. G.** (2010). Assessing heavy-metal contamination and sources by GIS-based approach and multivariate analysis of urban–rural topsoils in Wuhan, central China. *Environmental geochemistry and health*, 32(1), 59-72.<https://doi.org/10.1007/s10653-009-9265-2>
- Gonzalez P. L.** (2008). Méthodes de classification.
- Goovaerts P.** (1992). Factorial kriging analysis: a useful tool for exploring the structure of multivariate spatial soil information. *Journal of soil science*, 43(4), 597-619.<https://doi.org/10.1111/j.1365-2389.1992.tb00163.x>

REFERENCES

- Greinert A.** (2015). The heterogeneity of urban soils in the light of their properties. *Journal of Soils and Sediments*, 15(8), 1725-1737. <https://doi.org/10.1007/s11368-014-1054-6>
- Guillén M. T., Delgado J., Albanese S., Nieto J. M., Lima A., & De Vivo B.** (2011). Environmental geochemical mapping of Huelva municipality soils (SW Spain) as a tool to determine background and baseline values. *Journal of Geochemical Exploration*, 109(1-3), 59-69. <https://doi.org/10.1016/j.gexplo.2011.03.003>
- Ha H., Olson J. R., Bian L., & Rogerson P. A.** (2014). Analysis of heavy metal sources in soil using kriging interpolation on principal components. *Environmental science & technology*, 48(9), 4999-5007.
- Hani A., & Pazira E.** (2011). Heavy metals assessment and identification of their sources in agricultural soils of Southern Tehran, Iran. *Environmental monitoring and assessment*, 176(1-4), 677-691.
- Hawkes H. E., & Webb J. S.** (1963). Geochemistry in mineral exploration. *Soil Science*, 95(4), 283.
- Huang I. B., Keisler J., & Linkov I.** (2011). Multi-criteria decision analysis in environmental sciences: Ten years of applications and trends. *Science of the total environment*, 409(19), 3578-3594. <https://doi.org/10.1016/j.scitotenv.2011.06.022>
- International Union of Soil Sciences Working Group.** (2015). World reference base for soil resources 2014. International soil classification system for naming soils and creating legends for soil maps. *Update*.
- ISO 11074.** (2005). Qualité du sol -Vocabulaire.
- ISO19258.** (2005) « Qualité du sol - Guides pour la détermination des valeurs de bruit de fond », ISO 19258 :2005
- Ivezić V., Lončarić Z., Engler M., Kerovec D., & Singh B. R.** (2013), Comparison of different extraction methods representing available and total concentrations of Cd, Cu, Fe, Mn and Zn in soil. *PoljoPrivreda*, 19(1), 53-58.
- Jarva J., Ottesen R. T., & Tarvainen T.** (2014). Geochemical studies on urban soil from two sampling depths in Tampere Central Region, Finland. *Environmental earth sciences*, 71(11), 4783-4799. <https://doi.org/10.1007/s12665-013-2869-y>
- Jarva J., Tarvainen T., Reinikainen J., & Eklund M.** (2010). TAPIR—Finnish national geochemical baseline database. *Science of the Total Environment*, 408(20), 4385-4395. <https://doi.org/10.1016/j.scitotenv.2010.06.050>

REFERENCES

- Jean-Soro L., Le Guern C., Bechet B., Lebeau T., & Ringard M. F.** (2014). Qualité des sols urbains et risques sanitaires-Étude des anomalies en métaux traces dans un jardin familial de la ville de Nantes. *Techniques Sciences Méthodes*, (4), 72-79.
<https://doi.org/10.1051/tsm/201404072>
- Jeannée N., & de Fouquet C.** (2003). Apport d'informations qualitatives pour l'estimation des teneurs en milieux hétérogènes: cas d'une pollution de sols par des hydrocarbures aromatiques polycycliques (HAP). *Comptes rendus Geoscience*, 335(5), 441-449.
- Jeannot E., Carreau M., Le Guern C., Baudouin V., Bâlon P., Blanc C., Conil P.** (2014) La gestion des terres excavées sur les zones d'aménagement de l'île de Nantes, Journées *Techniques Nationales « Reconversion des friches urbaines polluées »*, 25-26 mars 2014, Paris
- Jeannot R., Lemièrre B., Chiron S., Augustin F., & Darmendrail D.** (2001). *Guide méthodologique pour l'analyse des sols pollués* (Vol. 298). BRGM. 85p
- Johnson C. C., & Ander E. L.** (2008), Urban geochemical mapping studies: how and why we do them. *Environmental Geochemistry and Health*, 30(6), 511-530.
<https://doi.org/10.1007/s10653-008-9189-2>
- Johnson S. C.** (1967). Hierarchical clustering schemes. *Psychometrika*, 32(3), 241-254.
<https://doi.org/10.1007/BF02289588>
- Karnatak H. C., Saran S., Bhatia K., & Roy P. S.** (2007). Multicriteria spatial decision analysis in web GIS environment. *Geoinformatica*, 11(4), 407-429.
<https://doi.org/10.1007/s10707-006-0014-8>
- Karim Z., Qureshi B. A., & Mumtaz M.** (2015). Geochemical baseline determination and pollution assessment of heavy metals in urban soils of Karachi, Pakistan. *Ecological Indicators*, 48, 358-364. <https://doi.org/10.1016/j.ecolind.2014.08.032>
- Kelly J., Thornton I., & Simpson P. R.** (1996). Urban geochemistry: a study of the influence of anthropogenic activity on the heavy metal content of soils in traditionally industrial and non-industrial areas of Britain. *Applied geochemistry*, 11(1-2), 363-370.
[https://doi.org/10.1016/0883-2927\(95\)00084-4](https://doi.org/10.1016/0883-2927(95)00084-4)
- Kheir R. B., Shomar B., Greve M. B., & Greve M. H.** (2014). On the quantitative relationships between environmental parameters and heavy metals pollution in Mediterranean soils using GIS regression-trees: The case study of Lebanon. *Journal of Geochemical Exploration*, 147, 250-259.
- Krishna R., Iqbal J., Gorai A. K., Pathak G., Tuluri F., & Tchounwou P. B.** (2015). Groundwater vulnerability to pollution mapping of Ranchi district using GIS. *Applied water science*, 5(4), 345-358. <https://doi.org/10.1007/s13201-014-0198-2>

REFERENCES

- Lark R. M., & Scheib C.** (2013). Land use and lead content in the soils of London. *Geoderma*, 209, 65-74. <https://doi.org/10.1016/j.geoderma.2013.06.004>
- Laurent, A.** (2007). *Evaluation du potentiel archéologique du sol en milieu urbain* (Thèse de doctorat, Université François Rabelais-Tours).
- Laurent A., & Fondrillon M.** (2010), Mesurer la ville par l'évaluation et la caractérisation du sol urbain: l'exemple de Tours. *Revue archéologique du Centre de la France*, (Tome 49). URL : <http://journals.openedition.org/racf/1485>
- LCSQA-INERIS** (2015), Conception, implantation et suivi des stations françaises de surveillance de la qualité de l'air. Rapport LCSQA, <https://www.actu-environnement.com/media/pdf/news-24464-guide-lcsqa.pdf>
- Le Guern C.** (2017) Urban Geochemistry: from 2D to 3D, *Procedia Engineering* 209:26-33, DOI10.1016/j.proeng.2017.11.126
- Le Guern C. & Sauvaget B. with contributions from Campbell S.D.G. & Pfeleiderer S.,** (2017). A review of good practice and techniques in sub-urban geochemistry; to ensure optimal information use in urban planning, COST TU1206 Working Group 2. Work package 2.6 report; TU1206-WG2-008. (available from www.sub-urban.eu)
- Le Guern C., Baudouin V., Bridier E., Cottineau C., Delayre M., Desse-Engrand F., Grellier M., Milano E., Mouny A., Pollett S., Sauvaget B., Ménoury A., Bâlon P. avec la collaboration de Conil P. & Rouvreau L.** (2016) - *Développement d'une méthodologie de gestion des terres excavées issues de l'aménagement de l'Ile de Nantes - Phase 1 : Caractérisation des sols et recensement des sources de pollution potentielles*. Rapport BRGM/RP-66013-FR, 122 p., 27. Ill., 7 Annexes.
- Le Guern C., Baudouin V., Sauvaget B., Delayre M., & Conil P.** (2016). A typology of anthropogenic deposits as a tool for modeling urban subsoil geochemistry: example of the Ile de Nantes (France). *Journal of Soils and Sediments*, 18(2), 373-379. <https://doi.org/10.1007/s11368-016-1594-z>
- Le Guern C., Baudouin V., Conil P., Courtois B., Houel M. avec la collaboration de Coffinet A., Latourte C. & Gourmelen F.** (2013) *METOTRASS : méthodologie optimisée pour l'évaluation des teneurs en éléments traces (As, Pb, Cu, Zn) dans les sols en domaine de socle – test sur le département de la Loire-Atlantique*, Rapport BRGM/RP-63998-FR, 112p.
- Le Guern C., Baudouin V., Barbaud F., Journaud C. et Roy E.**(2011) –*Inventaire historique urbain des anciens sites industriels et activité de service sur les communes de Saint-Nazaire et Trignac*(année 2011), RR 99–04.105p.
- Le Guern C., Fauconnier D., Baudouin V. avec la collaboration de Conil P.** (2001) *Sites industriels anciens de l'Agglomération Nantaise : Méthodologie de définition des priorités d'intervention – Application sur les Rives de Loire – Cas de l'Ile de Nantes*, Rapport Confidentiel BRGM/RP-51411-FR.

REFERENCES

- Le Guern C., & al.** (en préparation) Anticipating soil pollution linked to former industrial activities: key issues for urban planning and redevelopment
- Lepeltier C.** (1969), A simplified statistical treatment of geochemical data by graphical representation. *Economic geology*, 64(5), 538-550.
- Leys C., Ley C., Klein O., Bernard P., & Licata L.** (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764-766. <https://doi.org/10.1016/j.jesp.2013.03.013>
- Li X., & Feng L.** (2012). Multivariate and geostatistical analyzes of metals in urban soil of Weinan industrial areas, Northwest of China. *Atmospheric Environment*, 47, 58-65.
- Li J., Pu L., Liao Q., Zhu M., Dai X., Xu Y., & Jin Y.** (2015). How anthropogenic activities affect soil heavy metal concentration on a broad scale: a geochemistry survey in Yangtze River Delta, Eastern China. *Environmental Earth Sciences*, 73(4), 1823-1835. <https://doi.org/10.1007/s12665-014-3536-7>
- Li X., Lee S. L., Wong S. C., Shi, W., & Thornton I.** (2004). The study of metal contamination in urban soils of Hong Kong using a GIS-based approach. *Environmental Pollution*, 129(1), 113-124.
- Liebens J., Mohrherr C. J., & Rao K. R.** (2012). Trace metal assessment in soils in a small city and its rural surroundings, Pensacola, FL, USA. *Environmental Earth Sciences*, 65(6), 1781-1793. <https://doi.org/10.1007/s12665-011-1158-x>
- Lovering T. S., Huff L. C., & Almond H.** (1950), Dispersion of copper from the San Manuel copper deposit, Pinal County, Arizona. *Economic Geology*, 45(6), 493-514. <https://doi.org/10.2113/gsecongeo.45.6.493>
- Lu S. G., & Bai S. Q.** (2010). Contamination and potential mobility assessment of heavy metals in urban soils of Hangzhou, China: relationship with different land uses. *Environmental Earth Sciences*, 60(7), 1481-1490. <https://doi.org/10.1007/s12665-009-0283-2>
- Lv J., Liu Y., Zhang Z., Zhou R., & Zhu Y.** (2015). Distinguishing anthropogenic and natural sources of trace elements in soils undergoing recent 10-year rapid urbanization: a case of Donggang, Eastern China. *Environmental Science and Pollution Research*, 22(14), 10539-10550. <https://doi.org/10.1007/s11356-015-4213-4>

REFERENCES

- Madrid L, Diaz-Barrientos E, Ruiz-Cortés E, Reinoso R, Biasioli M, Davidson CM, Duarte AC, Grcman H, Hossack I, Hursthouse A, Kralj T, Ljung K, Otabbong E, Rodrigues S, Urquhart GJ, Ajmone-Marsan F** (2006) Potentially toxic metals in urban soils of six European cities: a pilot study on selected parks. *J Environ Monitor* 8:1158–1165
- Malczewski J.** (2006). GIS-based multicriteria decision analysis: a survey of the literature. *International journal of geographical information science*, 20(7), 703-726. <https://doi.org/10.1080/13658810600661508>
- Malczewski, J.** (2010). Multiple criteria decision analysis and geographic information systems. In *Trends in multiple criteria decision analysis* (pp. 369-395). Springer, Boston, MA.
- Malik R. N., Jadoon W. A., & Husain S. Z.** (2010). Metal contamination of surface soils of industrial city Sialkot, Pakistan: a multivariate and GIS approach. *Environmental geochemistry and health*, 32(3), 179-191. <https://doi.org/10.1007/s10653-009-9274-1>
- Manta D. S., Angelone M., Bellanca A., Neri R., & Sprovieri M.** (2002). Heavy metals in urban soils: a case study from the city of Palermo (Sicily), Italy. *Science of the Total Environment*, 300(1), 229-243. [https://doi.org/10.1016/S0048-9697\(02\)00273-5](https://doi.org/10.1016/S0048-9697(02)00273-5)
- Matheron G.** (1982). Pour une analyse krigéante des données régionalisées. *Centre de Géostatistique, Rapport N-732, Fontainebleau.*
- Matheron G.** (1970). La théorie des variables régionalisées, et ses applications. *Les Cahiers du CMM, fascicule 5*, Ecole des Mines de Paris.
- Matheron G.** (1965). Les variables régionalisées et leur estimation. Masson et C^{ie}, Paris.
- Matheron, G.** (1962). *Traité de Géostatistique appliquée*. Tome 1, Editions [Technip](#), Paris, 334 p.
- Matschullat J., Ottenstein R., & Reimann C.** (2000). Geochemical background—can we calculate it? *Environmental Geology*, 39(9), 990-1000.
- McIlwaine R., Cox S. F., Doherty R., Palmer S., Ofterdinger U., & McKinley J. M.** (2014). Comparison of methods used to calculate typical threshold values for potentially toxic elements in soil. *Environmental geochemistry and health*, 36(5), 953-971. <https://doi.org/10.1007/s10653-014-9611-x>
- Ministère de l'écologie** (1993) Circulaire du 03/12/93 relative à la politique de réhabilitation et de traitement des sites et sols pollués.

REFERENCES

- Ministère de l'écologie** (2007) Circulaire du 08/02/07 relative aux sites et sols pollués - Modalités de gestion et de réaménagement des sites pollués
- Murtagh F., & Legendre P.** (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?. *Journal of classification*, 31(3), 274-295. <https://doi.org/10.1007/s00357-014-9161-z>
- Nachtergaele F.**, (2005). The "soils" to be classified in the World Reference Base for Soil Resources. *Euras. Soil Sci.*, 38(Suppl. 1): 13–19.
- Négre P., Ladenberger A., Reimann C., Birke M., & Sadeghi M.** (2018). Distribution of Rb, Ga and Cs in agricultural land soils at European continental scale (GEMAS): Implications for weathering conditions and provenance. *Chemical Geology*, 479, 188-203.
- Oliver M. A., & Webster R.** (1989). A geostatistical basis for spatial weighting in multivariate classification. *Mathematical Geology*, 21(1), 15-35. <https://doi.org/10.1007/BF00897238>
- Paustenbach D. J.** (2000), The practice of exposure assessment: a state-of-the-art review. *Journal of Toxicology and Environmental Health Part B: Critical Reviews*, 3(3), 179-291. <https://doi.org/10.1080/10937400050045264>
- Pickett S. T., & Cadenasso M. L.** (2009). Altered resources, disturbance, and heterogeneity: a framework for comparing urban and non-urban soils. *Urban Ecosystems*, 12(1), 23-44. <https://doi.org/10.1007/s11252-008-0047-x>
- Poggio L., & Vrščaj B.** (2009). A GIS-based human health risk assessment for urban green space planning—An example from Grugliasco (Italy). *Science of the total environment*, 407(23), 5961-5970. <https://doi.org/10.1016/j.scitotenv.2009.08.026>
- Pouyat, R. V., Yesilonis, I. D., Russell-Anelli, J., & Neerchal, N. K.** (2007). Soil chemical and physical properties that differentiate urban land-use and cover types. *Soil Science Society of America Journal*, 71(3), 1010-1019. DOI : 10.2136/sssaj2006.0164
- Puskás I., & Farsang A.** (2009). Diagnostic indicators for characterizing urban soils of Szeged, Hungary. *Geoderma*, 148(3), 267-281. <https://doi.org/10.1016/j.geoderma.2008.10.014>
- Quevauviller Ph., Lachica M., Barahona E., Rauret G., Ure A., Gomez A., Muntau H.** (1996), Interlaboratory comparison of EDTA and DTPA procedures prior to certification of extractable trace elements in calcareous soil. *Science of Total Environment* 178, 127–132. [https://doi.org/10.1016/0048-9697\(95\)04804-9](https://doi.org/10.1016/0048-9697(95)04804-9)

REFERENCES

- Ramos-Miras J. J., Roca-Perez L., Guzmán-Palomino M., Boluda R., & Gil C.** (2011). Background levels and baseline values of available heavy metals in Mediterranean greenhouse soils (Spain). *Journal of Geochemical Exploration*, 110(2), 186-192. <https://doi.org/10.1016/j.gexplo.2011.05.009>
- Recatalá L., & Sacristán D.** (2014). A minimum indicator set for assessing resources quality and environmental impacts at planning level in a representative area of the European Mediterranean Region. *Ecological Indicators*, 45, 160-170.
- RECORD** (2010), Réalisation d'un avis d'experts : "Bruit de fond" des polluants dans l'environnement, analyse des données existantes et identification des lacunes, 163p, n°09-0812/1A
- Reimann C., Filzmoser P., & Garrett R. G.** (2005). Background and threshold: critical comparison of methods of determination. *Science of the Total Environment*, 346(1), 1-16.
- Reimann C., & de Caritat P.** (2017). Establishing geochemical background variation and threshold values for 59 elements in Australian surface soil. *Science of The Total Environment*, 578, 633-648. <https://doi.org/10.1016/j.scitotenv.2016.11.010>
- Reimann C., Fabian K., Birke M., Filzmoser P., Demetriades A., Négrel P., & Anderson M.** (2018). GEMAS: Establishing geochemical background and threshold for 53 chemical elements in European agricultural soil. *Applied geochemistry*, 88, 302-318.
- Reimann C., Fabian K., Flem B., & Englmaier P.** (2019). The large-scale distribution of Cu and Zn in sub-and topsoil: Separating topsoil bioaccumulation and natural matrix effects from diffuse and regional contamination. *Science of The Total Environment*, 655, 730-740.
- Rivoirard J.** (2003). Cours de géostatistique multivariable. *Cours C-172, CG, Ecole des Mines de Paris*.
- Rivoirard J., Demange C., Freulon X., Lécureuil A., & Bellot N.** (2013), A top-cut model for deposits with heavy-tailed grade distribution. *Mathematical Geosciences*, 45 (8), 967-982. <https://doi.org/10.1007/s11004-012-9401-x>
- Rodrigues S., Urquhart G., Hossack I., Pereira M. E., Duarte A. C., Davidson C., & Roberston D.** (2009). The influence of anthropogenic and natural geochemical factors on urban soil quality variability: a comparison between Glasgow, UK and Aveiro, Portugal. *Environmental chemistry letters*, 7(2), 141-148. <https://doi.org/10.1007/s10311-008-0149-y>

REFERENCES

- Romano G., Dal Sasso P., Liuzzi G. T., & Gentile F.** (2015). Multi-criteria decision analysis for land suitability mapping in a rural area of Southern Italy. *Land Use Policy*, 48, 131-143. <https://doi.org/10.1016/j.landusepol.2015.05.013>
- Romary T.** (2013). Deux algorithmes pour la classification non supervisée de données géostatistiques. In *45e Journées de Statistique*. hal-00842826
- Romary T., Rivoirard J., Deraisme J., Quinones C., & Freulon X.** (2012). Domaining by clustering multivariate geostatistical data. In *Geostatistics Oslo 2012* (pp. 455-466). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-4153-9_37
- Rothwell K. A., & Cooke M. P.** (2015). A comparison of methods used to calculate normal background concentrations of potentially toxic elements for urban soil. *Science of the Total Environment*, 532, 625-634. <https://doi.org/10.1016/j.scitotenv.2015.06.083>
- Saby N., Arrouays D., Boulonne L., Jolivet C., & Pochot A.** (2006). Geostatistical assessment of Pb in soil around Paris, France. *Science of the total environment*, 367(1), 212-221. <https://doi.org/10.1016/j.scitotenv.2005.11.028>
- Saby N. P. A., Thioulouse J., Jolivet C. C., Ratié C., Boulonne L., Bispo A., & Arrouays D.** (2009). Multivariate analysis of the spatial patterns of 8 trace elements using the French soil monitoring network data. *Science of the Total Environment*, 407(21), 5644-5652. <https://doi.org/10.1016/j.scitotenv.2009.07.002>
- Šajn R., Gosar M., Bidovec M., Pirc S., & Alijagić J.** (2011). Geochemical mapping of Ljubljana urban and suburban area, Slovenia. In : Johnson C.C., Demetriades A., Locutura O., & Ottosen R. T. (eds) , *Mapping the Chemical Environment of Urban Areas*, John Wiley, Oxford, pp. 375-392. <https://doi.org/10.1002/9780470670071.ch22>
- Sancho L.** (2016). Protocole d'analyse statistique pour la construction d'un Fond Pédogéochimique Anthropisés Sols Urbains. *Rapport final. BRGM/RP-66501-FR*, 111p.
- Schaefer K., Einax J. W., Simeonov V., & Tsakovski S.** (2010). Geostatistical and multivariate statistical analysis of heavily and manifoldly contaminated soil samples. *Analytical and bioanalytical chemistry*, 396(7), 2675-2683.
- Schindelbeck R. R., van Es H. M., Abawi G. S., Wolfe D. W., Whitlow T. L., Gugino B. K., & Moebius-Clune B. N.** (2008). Comprehensive assessment of soil quality for landscape and urban management. *Landscape and Urban Planning*, 88(2), 73-80.
- Shaheen D. G.** (1975). *Contributions of urban roadway usage to water pollution* (Vol. 1). Office of Research and Development, US Environmental Protection Agency.

REFERENCES

- Smith D. B., Garrett R. G., Closs L. G., Ellefsen K. J., Kilburn J. E., Horton J. D., & Smith S. M.** (2011), Geochemical mapping of the Denver, Colorado (USA) urban area: a comparison of studies in 1972 and 2005. *Mapping the Chemical Environment of Urban Areas*. Wiley-Blackwell, Oxford, 521-546.
- Sobocká J., & Fulajtár E.** (2009). ENVIRONMENTAL HAZARDS RESULTING FROM URBAN SOILS IN BRATISLAVA CITY. *Local land & soil news*, (28/29).
- Sokolov I.A.**, (1997). Soil Formation and Exogenesis. Moscow. 241pp. [in Russian].
- SPAQuE : Pereira, B., Sonnet, P., & Capette, L.** (2010). Guide pour la définition des concentrations de fond en polluants dans les sols de Wallonie. Cahier de Bonnes Pratiques n° 10.
- Steinley D., & Brusco, M. J.** (2007). Initializing K-means batch clustering: A critical evaluation of several techniques. *Journal of Classification*, 24(1), 99-121. <https://doi.org/10.1007/s00357-007-0003-0>
- Tarvainen T., & Jarva J.** (2011). Using geochemical baselines in the assessment of soil contamination in Finland. In : Johnson C.C., Demetriades A., Locutura O., & Ottosen R. T. (eds) , *Mapping the Chemical Environment of Urban Areas*, John Wiley, Oxford, pp. 223-231. <https://doi.org/10.1002/9780470670071.ch15>
- Teng Y., Wu J., Lu S., Wang Y., Jiao X., & Song L.** (2014). Soil and soil environmental quality monitoring in China: a review. *Environment international*, 69, 177-199.
- Thackway R., & Specht A.** (2015). Reprint of: Synthesising the effects of land use on natural and managed landscapes. *Science of The Total Environment*, 534, 14-30.
- Thornton I., & Webb J. S.** (1979), Geochemistry and health in the United Kingdom. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 288(1026), 151-168. <https://doi.org/10.1098/rstb.1979.0097>
- Tukey J. W.** (1977), Exploratory data analysis (Vol. 2)
- URGE II.** EuroGeoSurveys, Brussels, 52 pp.
- Vaněk A., Borůvka L., Drábek O., Mihaljevič M., & Komárek M.** (2005). Mobility of lead, zinc and cadmium in alluvial soils heavily polluted by smelting industry. *Plant, Soil and Environment*, 51(7), 316-321.
- Velasquez M., & Hester P. T.** (2013). An analysis of multi-criteria decision making methods. *International Journal of Operations Research*, 10(2), 56-66.
- Wackernagel, H.** (2003). *Multivariate Geostatistics*, Third edition, Springer-Verlag, Berlin, 387 p

REFERENCES

- Wackernagel H. & Sanguinetti H.** (1993) Gold prospecting with factorial cokriging in the Limousin, France. In: Davis, J.C. and Herzfeld, U.C. (Ed.) *Computers in Geology: 25 years of progress. Studies in Mathematical Geology*, Vol. 5, 33–43, Oxford University Press, Oxford.
- Wang M., Markert B., Chen W., Peng C., & Ouyang Z.** (2012). Identification of heavy metal pollutants using multivariate analysis and effects of land uses on their accumulation in urban soils in Beijing, China. *Environmental monitoring and assessment*, 184(10), 5889-5897. <https://doi.org/10.1007/s10661-011-2388-9>
- Webb J. S., Thornton I., Howarth R. J., & Thompson M.** (1979). The Wolfson geochemical atlas of England and Wales. *Chemistry in Britain*, 15, 500-505.
- Wragg J., & Cave M. R.** (2003), *In-vitro methods for the measurement of the oral bioaccessibility of selected metals and metalloids in soils: a critical review*. Environment Agency.
- WRB, I. W. G.** (2015). World reference base for soil resources 2014, update 2015. *International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports*, 106.
- Xia X., Chen X., Liu R., & Liu H.** (2011). Heavy metals in urban soils with various types of land use in Beijing, China. *Journal of Hazardous Materials*, 186(2-3), 2043-2050. <https://doi.org/10.1016/j.jhazmat.2010.12.104>
- Yuan G. L., Sun T. H., Han P., & Li J.** (2013). Environmental geochemical mapping and multivariate geostatistical analysis of heavy metals in topsoils of a closed steel smelter: Capital Iron & Steel Factory, Beijing, China. *Journal of Geochemical Exploration*, 130, 15-21.
- Zeilhofer P., & Topanotti V. P.** (2008). GIS and ordination techniques for evaluation of environmental impacts in informal settlements: A case study from Cuiaba, central Brazil. *Applied Geography*, 28(1), 1-15.
- Zhang X. Y., Lin F. F., Wong M. T., Feng X. L., & Wang K.** (2009). Identification of soil heavy metal sources from anthropogenic activities and pollution assessment of Fuyang County, China. *Environmental Monitoring and Assessment*, 154(1-4), 439-449. <https://doi.org/10.1007/s10661-008-0410-7>
- Zheng Y. M., Chen T. B., & He J. Z.** (2008). Multivariate geostatistical analysis of heavy metals in topsoils from Beijing, China. *Journal of Soils and Sediments*, 8(1), 51-58. <https://doi.org/10.1065/jss2007.08.245>

La référence “décret sols belge, 2008” renvoie au décret du Parlement wallon du 5 décembre 2008 relatif à la gestion des sols, publié au Moniteur belge du 18 février 2009.

REFERENCES

Sites internet

BRGM, ADEME, *BDSolU* (2019). Available at : <http://www.bdsolu.fr/>

British Geological Survey, *G-BASE*(2019). Available at :
<https://www.bgs.ac.uk/gbase/home.html>

Copernicus (2019). Available at : <https://www.copernicus.eu/en/access-data>

CORINE LAND COVER (2019). Available at : <https://www.statistiques.developpement-durable.gouv.fr/corine-land-cover-0>

Données de la Banque Mondiale(2019). Available at :
<https://donnees.banquemondiale.org/indicateur/sp.urb.totl.in.zs>

FOREGS (2019). Available at : <http://weppi.gtk.fi/publ/foregsatlas/index.php>

GEMAS (2019). Available at : <http://www.eurogeosurveys.org/projects/gemas/>

GisSol, *RMQS* (2019). Available at :<https://www.gissol.fr/donnees>

Georisques, *BASIAS* (2019). Available at :
<http://www.georisques.gouv.fr/dossiers/inventaire-historique-des-sites-industriels-et-activites-de-service-basias>

Matrice activités polluantes (2019). Available at:
<http://ssp-infoterre.brgm.fr/matrice-activites-polluants>

SAMOA (2019) Ile de Nantes-Make the city differently. Available at:
<http://www.iledenantes.com>

US Geological Survey (2019). Available at:
<https://mrdata.usgs.gov/soilgeochemistry/#/periodictable>

TABLE DES ABREVIATIONS

ACP : Analyse en composante principale
AMC : Analyse multi-critères
ASPITET : Apports d'une Stratification Pédologique à l'Interprétation des Teneurs en Eléments Traces
BASIAS : Base de Données des Anciens Sites Industriels et Activités de Service
BDAT : Base de Données d'Analyses des Terres
BDETM : Base de Données des Eléments Trace Métalliques
BDSolU : Base de Données des analyses de Sols Urbains
BGS : British Geological Survey
DLI : Dépôt de liquide inflammable
BRGM : Bureau de Recherches Géologiques et Minières
BTEX : Benzène, Toluène, Éthylbenzène et Xylènes
CLC : Corine Land Cover
COP : Copernicus
FPG : Fond pédo-géochimique
FPGA : Fond pédo-géochimique anthropisé
G-Base : Geochemical Baselines
GS MS : Chromatographie en phase gazeuse couplée à la spectrométrie de masse
HAP : Hydrocarbures aromatiques polycycliques
ICP MS : Spectrométrie de masse à plasma à couplage inductif
IGCS : Inventaire, gestion et Conservation des Sols
IHR : Inventaire historique régional
IHU : Inventaire historique urbain
IMN : Inventaire minier national
MAD : Median absolute deviation
NMA : Nantes Métropole Aménagement
PCB : Polychlorobiphényles
pXRF : Spectromètre à fluorescence X portable
RMQS : Réseau de Mesure de la Qualité des Sols
SAMOA : Société d'Aménagement de la Métropole Ouest Atlantique
SIG : Système d'information géographique
URGE : Urban Geochemistry
USGS : United States Geological Survey
WRB : World Reference Base

Annexe 1 : Présentation du variogramme

Le variogramme (Figure 103) (ou la covariance) est utilisé en krigeage et notamment en cartographie géochimique (Lark, 2013 ; Lv, 2015 ; McIlwaine, 2014 ; Poggio, 2009 ; Zhang, 2009 ; Zheng, 2008). Le variogramme expérimental (calculé à partir de données) se calcule par l'équation ci-dessous :

$$\gamma_e(h) = \frac{1}{2 N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2$$

Avec :

- h = Distance et orientation entre deux points d'une même paire
- $N(h)$ = nombre de paires de points séparés d'une distance h
- $Z(x_i)$ = valeur au premier point de la paire située en x_i
- $Z(x_i+h)$ = valeur au deuxième point de la paire à une distance h de x_i

Toutefois le calcul du variogramme se basant sur un ensemble de paires de points de mesures, si l'un de ces points présente une erreur conséquente d'analyse celle-ci se propagera alors de façon importante aux résultats finaux. La phase de vérification des données et d'évaluation des erreurs d'analyse et d'échantillonnage est donc primordiale dans ce type de traitement.

Le variogramme expérimental, est ajusté par un modèle ayant les propriétés mathématiques adéquates. Le modèle comporte souvent plusieurs structures emboîtées.

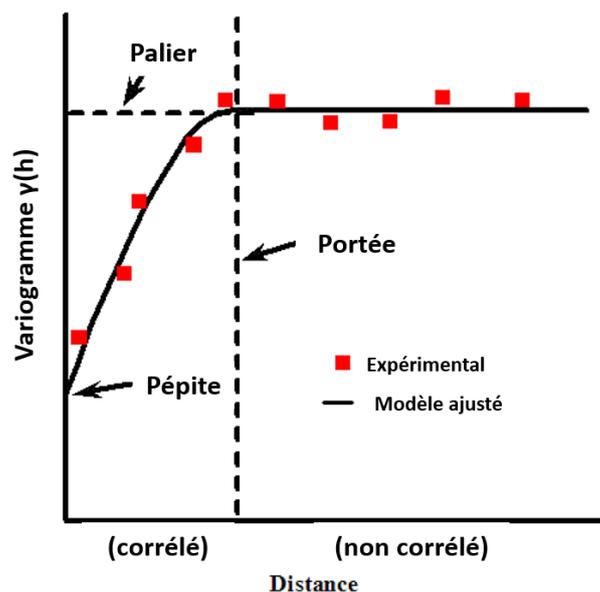


Figure 103: Exemple de variogramme illustrant le palier, la portée et l'effet de pépite.

Portée : Distance à partir de laquelle deux observations ne sont plus corrélées (ou avec une corrélation négligeable).

Palier : Les variogrammes de portée finie présentent, un palier. Celui-ci correspond alors à la variance de la variable dans l'hypothèse d'un champ infini.

Effet de pépite : Il représente les variations à très courte distance (inférieure à la distance entre les deux points les plus proches) dues aux erreurs d'analyse ou d'échantillonnage ou à l'hétérogénéité intrinsèque du milieu. Il s'exprime par une ordonnée à l'origine différente de 0.

Annexe 2 : Exemple de sondage et échantillonnage d'un diagnostic de pollution

COUPE DE SONDAGE

Site :					
Opérateur :		Date : 12/04/10	Heure : 11 H 15		
Outils de sondage :		Tarière mécanique	N° sondage : T2		
Profondeur d'arrêt :		4,00 m			
Coupe géologique				Echantillon	
Grad. (m)	Prof. (m)	Coupe	Description	N°	Observations (aspect, odeur, couleur)
0		[Symbole géologique]	Remblais limoneux		Couleur marron, présence de briques, pas d'odeur
1					
	1,5	[Symbole géologique]	Remblais limoneux	T2-1	Couleur marron bordeaux, présence de briques et de verres, pas d'odeur
2					
	3	[Symbole géologique]	Remblais limoneux	T2-2	Couleur marron bordeaux, présence de briques et de verres, pas d'odeur
4					
5					
Remarques :					
Arrêt sondage : 4,00 m					

Figure 104: Exemple de résultats de sondage extrait d'un diagnostic de pollution. Les échantillons sont indiqués comme « T2-1 » et « T2-2 ». Figure extrait d'un rapport fourni par la SAMOA.

Annexe 3 : Liste des laboratoires et principaux protocoles d'analyse recensés dans la base de données locale

Modalités d'analyse chimique

Les méthodes d'analyse varient selon les protocoles qui évoluent au fil des années et d'un laboratoire à un autre. La démarche reste cependant généralement assez similaire. Les échantillons de sol préalablement séchés sont tout d'abord passés au tamis à maille de 2 mm afin d'éliminer les plus gros éléments. Ils sont ensuite broyés finement puis mis en solution par des attaques à l'acide.

Les solutions sont ensuite analysées par différents appareils de mesure. La méthode de mesure la plus couramment utilisée ces dernières années est l'ICP-MS (Spectrométrie de masse à plasma à couplage inductif) pour la mesure des teneurs en éléments inorganiques tels que les métaux. Des paramètres physico-chimiques sont également mesurés, comme le pH ou le pourcentage de matière sèche. La Figure 105 présente un exemple de protocole d'analyse mis en place pour un échantillon dans le cadre de l'étude géochimique des sols urbains.

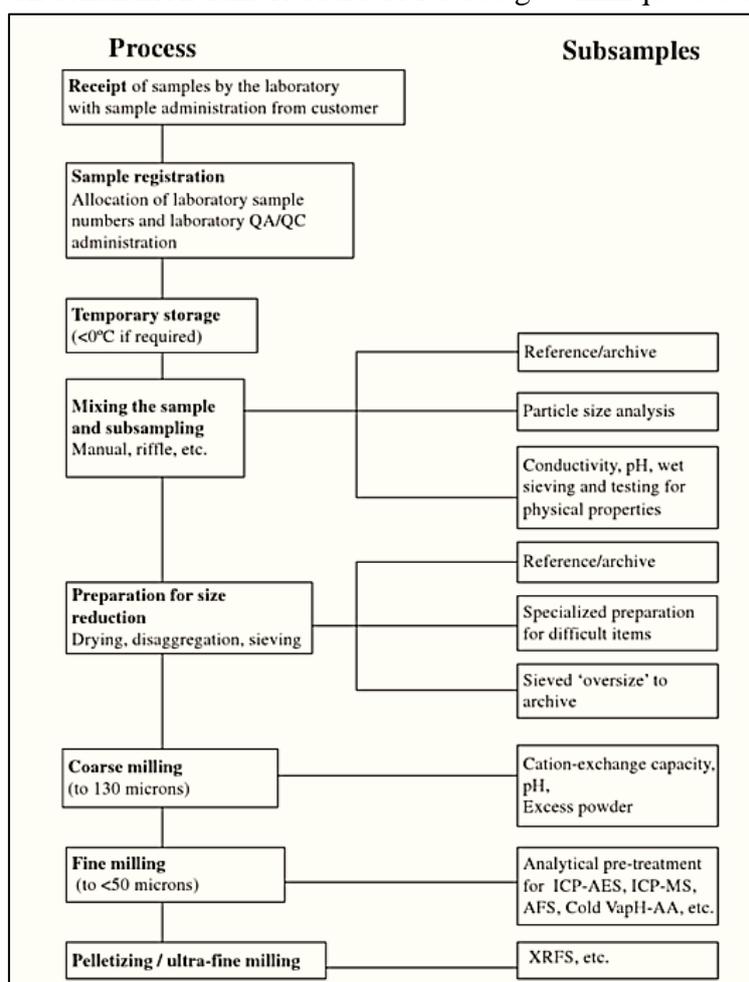


Figure 105 : Exemple de diagramme des étapes de la préparation des analyses de sols urbains. (Allen et Cave, 2011)

Annexe 3 : Liste des laboratoires et principaux protocoles d'analyse recensés dans la base de données locale

Dans la majorité des cas, l'extraction sera pseudo-totale et consistera à mesurer la teneur des métaux sans tenir compte des spéciations (Manta, 2002). Cette extraction se fait le plus souvent avec de l'eau régale (aqua regia) capable de mettre en solution la plupart des métaux. On parle ici d'extraction pseudo-totale, l'eau régale n'étant pas capable de détruire la totalité des liaisons silicatées et donc de mettre en solution la totalité des ions métalliques. Par de l'acide fluorhydrique est aussi possible mais elle reste coûteuse et rarement mise en place à grande échelle. Bien que ces méthodes d'extraction portent sur des approches différentes du problème des sols pollués, des comparaisons entre elles sont parfois effectuées (Ivezić, 2014). Les conclusions sont que dans le cadre d'études de la géochimie sur de larges territoires, l'extraction pseudo-totale est à privilégier. En effet, cela permet un plus grand nombre d'analyses. De plus, les métaux non extraits par l'eau régale ne sont de toute façon pas bioaccessibles et présentent donc moins d'intérêt vis-à-vis des enjeux liés au fond pédogéochimique anthropisé.

Pour une quantification totale d'un grand nombre d'éléments, la spectroscopie à fluorescence aux rayons X est parfois utilisée. L'avantage de cette méthode est qu'elle est non destructive. Dans les études sur la géochimie urbaine elle reste toutefois plutôt utilisée pour les éléments majeurs, la méthode par ICP étant préférée pour les éléments traces métalliques (Manta, 2002).

Dans le cas d'URGE 2, les deux extractions (totale et pseudo-totale) sont recommandées. L'extraction totale ayant l'intérêt de mettre en solution tous les métaux alors que la pseudo-totale à l'eau régale est généralement moins coûteuse et recommandée dans la plupart des normes environnementales.

La majorité des études du fond pédogéochimique anthropisé se concentrent sur des analyses de métaux et métalloïdes, ceux-ci étant, à la fois de bons indicateurs de l'influence de l'activité anthropique sur les sols, et des enjeux sanitaires majeurs du fait de leur toxicité potentielle (plomb, cadmium, arsenic, etc.). Ceux-ci présentent toutefois des formes diverses dans les sols (spéciations) liées, entre autres, aux conditions d'oxydo-réduction (Bavec, 2014). En fonction de leur spéciation, les métaux et métalloïdes peuvent devenir bio-disponibles. Ce terme, tel qu'il est utilisé en pédologie, définit un élément qui, une fois ingéré, a la capacité d'être absorbé par l'organisme et donc de passer dans le sang. Ce terme ne doit pas être confondu avec la bio-accessibilité qui définit un élément soluble dans les conditions d'une digestion (Paustenbach, 2000). Si la bio-disponibilité ne peut être mesurée que par des prélèvements de sang, la bio-accessibilité peut, quant à elle, être simulée. Certaines études vont ainsi se baser sur une analyse partielle des métaux afin de ne tenir compte que des teneurs bio-accessibles en simulant le travail du système digestif (Lu, 2010 ; Quevauviller, 1996 ; Ramos-Miras, 2011). Néanmoins, plusieurs protocoles existent et dans le but de produire des valeurs de référence pour le fond pédogéochimique, des méthodes plus couramment utilisées, comme l'extraction pseudo-totale à l'eau régale, sont préférées.

Des paramètres organiques sont aussi fréquemment analysés. Comme les différents types d'hydrocarbures et notamment les hydrocarbures aromatiques polycycliques (HAP), les

Annexe 3 : Liste des laboratoires et principaux protocoles d'analyse recensés dans la base de données locale

PCB, les organochlorés et les BTEX. Les analyses sont généralement faites par chromatographie en phase gazeuse couplée à la spectrométrie de masse (GS MS), l'extraction étant réalisée par acétone/hexane.

Enfin, d'autres méthodes sont également utilisées pour des analyses in-situ rapides et peu coûteuses. On peut ainsi citer la spectroscopie à fluorescence X portable (pXRF) pour les éléments métalliques et métalloïdes. Dans le cadre de Metotrass (*Le Guern et al, 2013*) et de travaux sur des jardins familiaux (*Jean-Soro et al, 2014 ; Branchu et al, 2016*), les données acquises par pXRF ont été exploitées pour des représentations cartographiques et statistiques simples après vérification de la corrélation avec les résultats d'analyses obtenues par ICP-MS sur une sélection d'échantillons et correction des teneurs obtenues.

Données disponibles :

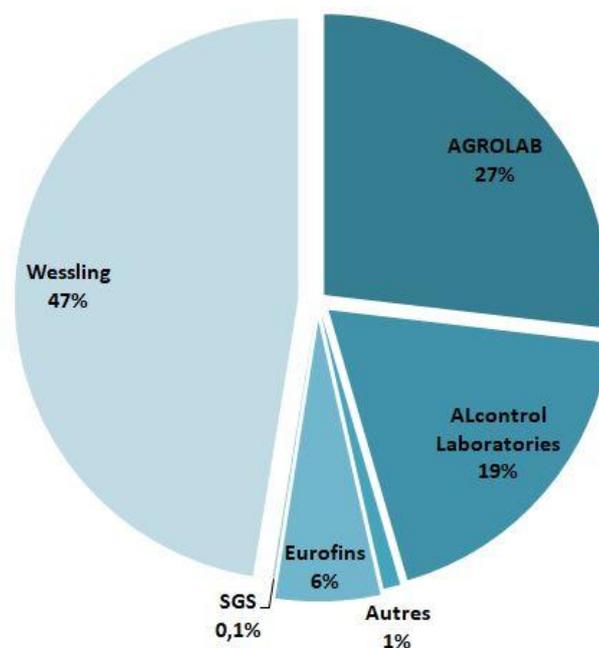


Figure 106: Diagramme de la répartition de toutes les données collectées sur les quartiers suivant le laboratoire d'analyse.

La base de données présente un manque important de renseignements sur les méthodes d'analyse utilisées dans les diagnostics de pollution traités. Cela provient principalement du manque d'information dans les documents d'origine et d'une saisie incomplète ou inégale de certaines données.

Le tableau suivant regroupe le pourcentage de données dont les méthodes d'analyses sont renseignées dans la base ainsi que la plus représentée pour les métaux/métalloïdes, les hydrocarbures, les HAP, les PCB et le carbone organique total.

Annexe 3 : Liste des laboratoires et principaux protocoles d'analyse recensés dans la base de données locale

Fraction	Famille	Pourcentage renseigné	Méthode principale
Solide	COT	35%	Combustion sèche
Solide	PCB	21%	Extraction Hexane/Acétone et dosage par GC/MS
Solide	Métaux-Métalloïdes	29%	Minéralisation à l'eau régale et dosage par ICP/AES
Solide	Hydrocarbures	25%	Extraction Hexane / Acétone et dosage par GC/FID
Solide	HAP	23%	Extraction Hexane/Acétone et dosage par GC/MS
Eluat	Métaux-Métalloïdes	21%	Dosage par ICP/AES

Figure 107: Récapitulatif des méthodes d'analyse principale renseignées dans la base de données suivant la fraction et la famille analysée, ainsi que le pourcentage de données ayant une méthode d'analyse renseignée.

Annexe 4 : Résultats des sondages complémentaires et comparaisons (pXRF / ICP)

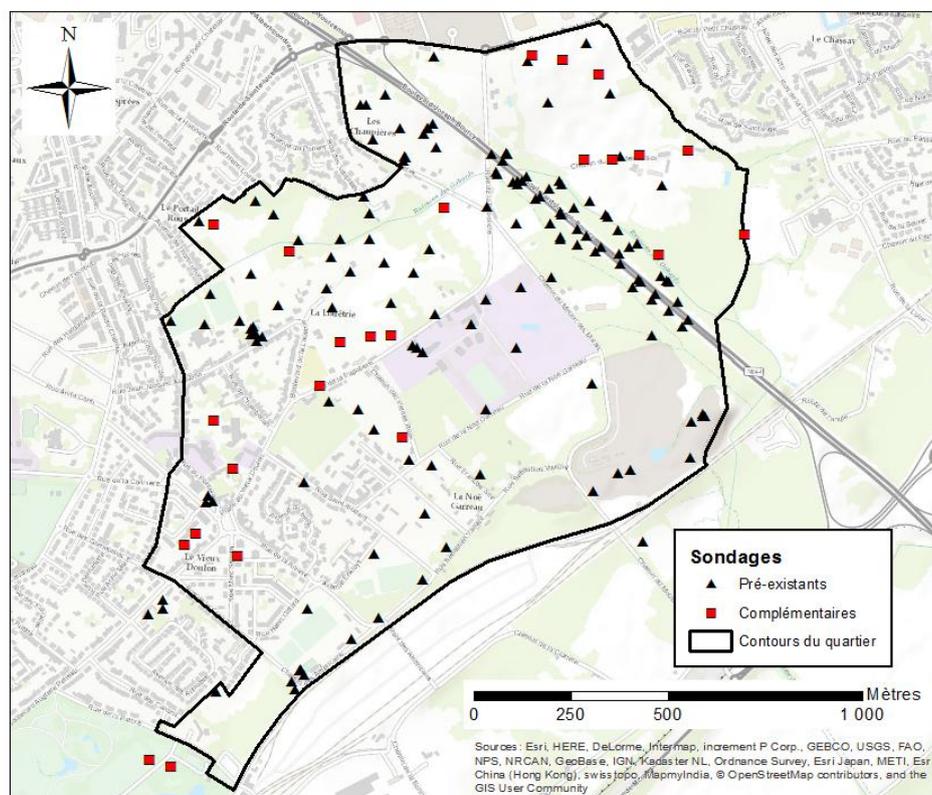


Figure 108 : Position des sondages complémentaires

Sur les 43 échantillons complémentaires analysés par pXRF, 10 ont également été envoyés au laboratoire Eurofins pour des analyses par ICP. Les nuages de corrélation ainsi que les droites de régression obtenues sont illustrés par les figures suivantes.

Les résultats obtenus pour le nickel (Figure 109), montrent une bonne corrélation entre les résultats obtenus par pXRF et ceux obtenus par ICP. La droite de régression obtenue indique alors que les résultats par pXRF sont surestimés vis-à-vis de ceux par ICP.

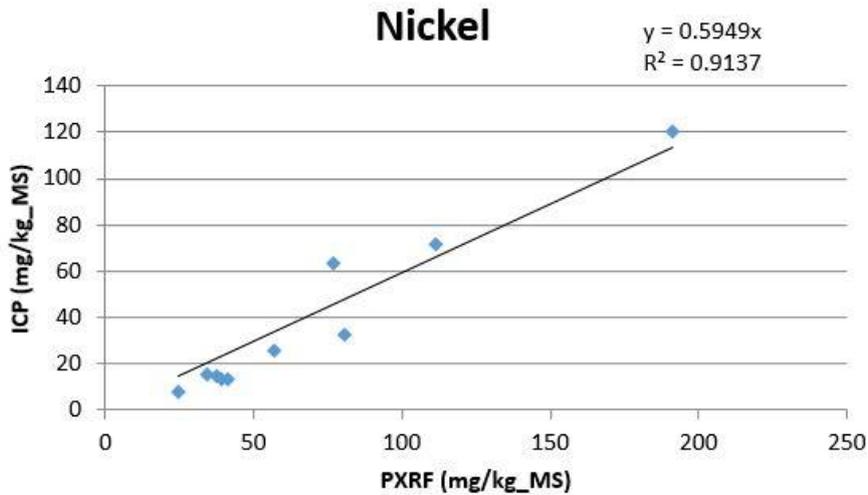


Figure 109: Nuage de corrélation et droite de regression des analyses en nickel par pXRF et ICP.

On remarque toutefois, que la corrélation est parfois moins nette entre les résultats d'ICP et de pXRF. On peut citer le cas du plomb (Figure 110), où les résultats les plus forts sont plus dispersés. A l'inverse du cas précédent, les résultats par pXRF semblent sous-estimés par rapport à ceux par ICP.

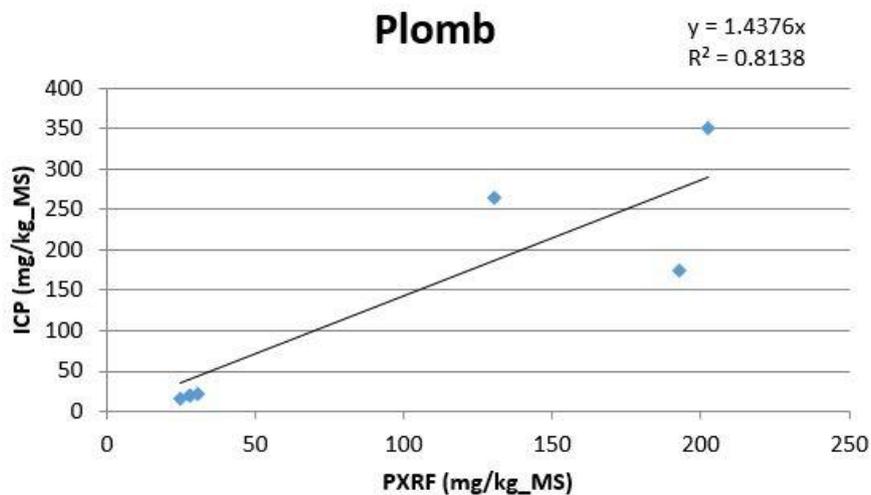


Figure 110 : Nuage de corrélation et droite de regression des analyses en plomb par pXRF et ICP.

D'autres exemples, comme le cuivre (Figure 111), présentent des R^2 très proches de 1. Toutefois, cela est principalement dû à un échantillon aux teneurs plus importantes que le reste des analyses. Le coefficient calculé par la droite de régression est alors très similaire à celui obtenu pour le plomb.

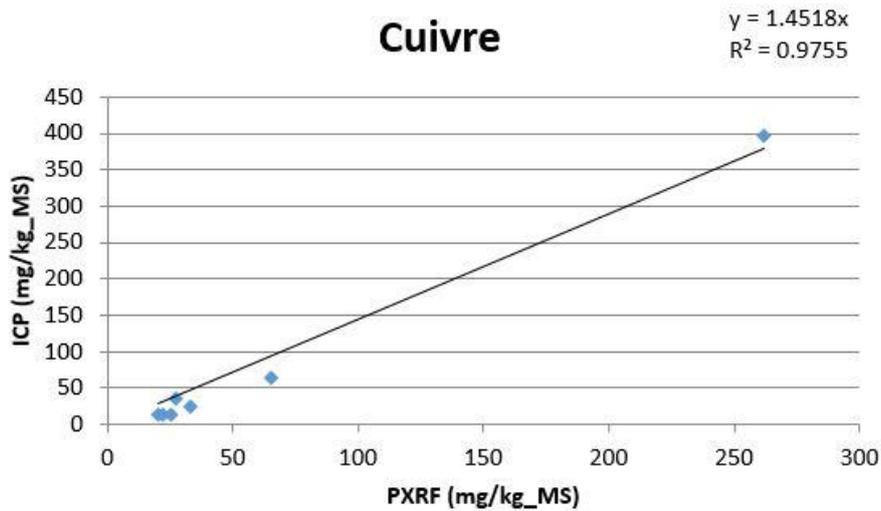


Figure 111: Nuage de corrélation et droite de régression des analyses en cuivre par pXRF et ICP.

Enfin, dans certains cas, comme l'antimoine (Figure 112), aucune corrélation n'est identifiable entre les résultats par pXRF et ceux par ICP. Les données ne peuvent donc pas être corrigées ou comparées. Ceci peut s'expliquer par le fait que pXRF correspond à une mesure du métal total, alors que l'analyse par ICP-MS fait après extraction à l'eau régale ne correspond pas à une analyse du métal total. Il est probable ici qu'une partie de l'antimoine soit lié à des minéraux silicatés, non dissous par l'eau régale. Une meilleure corrélation devrait être obtenue après extraction totale (incluant HF pour dissoudre les silicates) (C. Le Guern, 2019, communication personnelle).

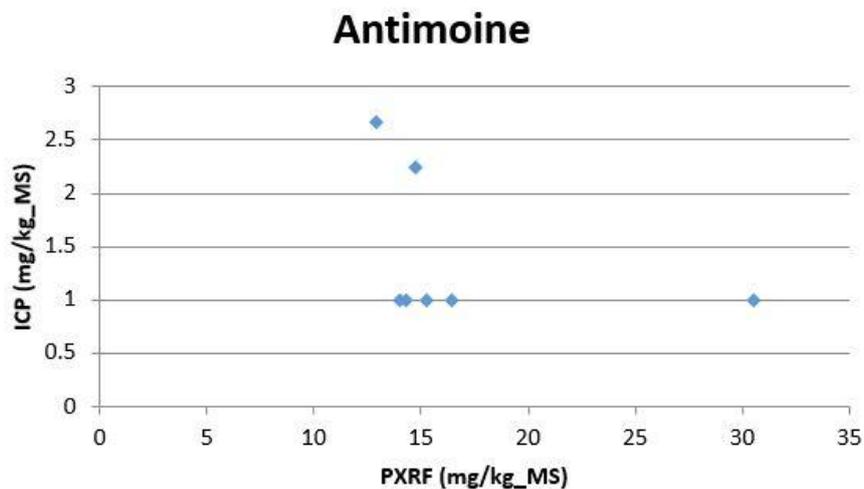


Figure 112: Nuage de corrélation des analyses en antimoine par pXRF et ICP

Ainsi, il semble que pour certains métaux et métalloïdes, des comparaisons puissent être faites entre des données de laboratoire (ICP) et des données obtenues par pXRF plus rapides et moins coûteuses, en considérant toutefois une correction des mesures par pXRF suivant les coefficients des droites de régression. L'ajout de nouvelles données permettrait toutefois d'améliorer la précision de cette correction.

Annexe 4 : Résultats des sondages complémentaires et comparaisons (pXRF / ICP)

Les résultats obtenus pour l'arsenic, le baryum, le chrome et le zinc sont illustrés dans la Figure 113.

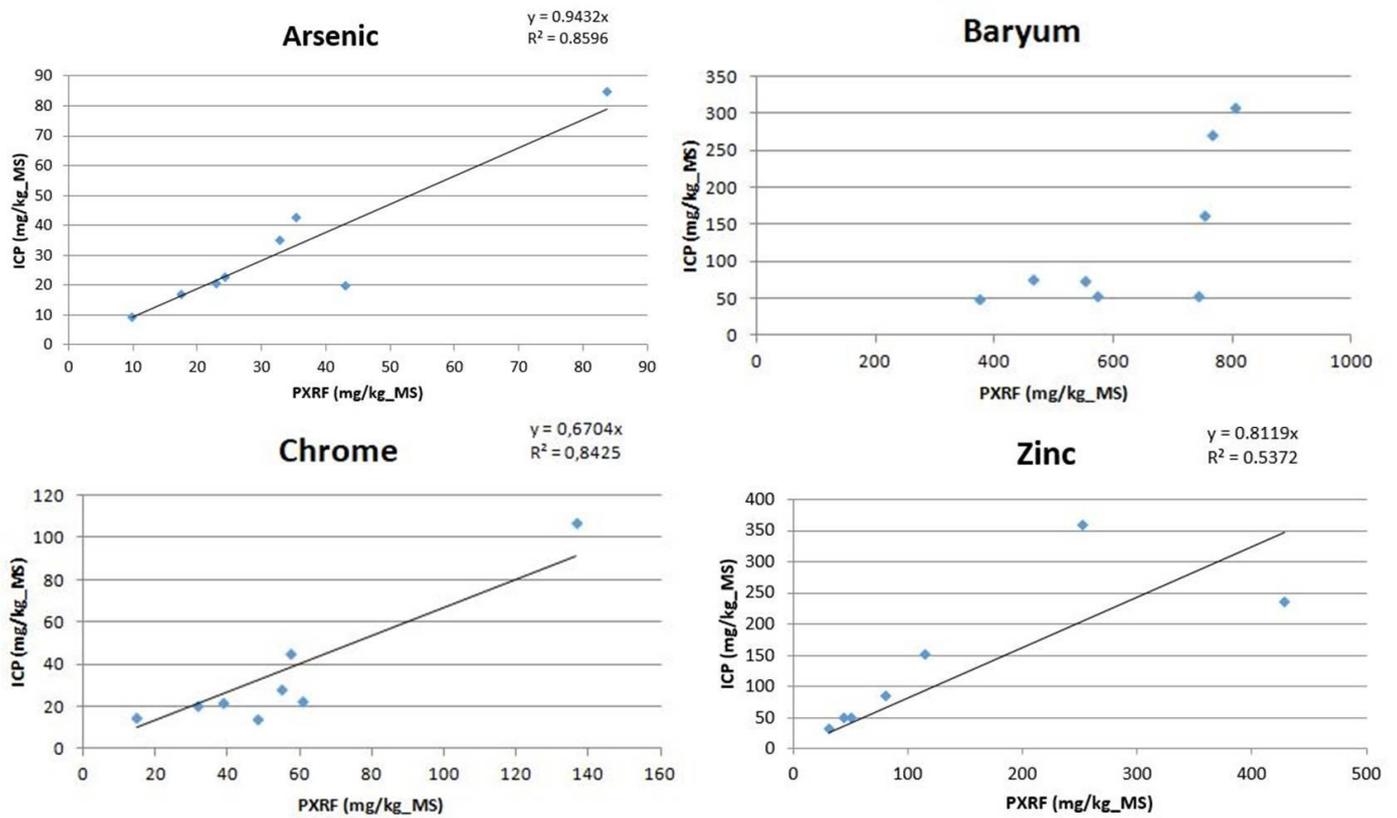


Figure 113: Nuages de corrélation et droite de régression des analyses en arsenic, baryum, chrome et zinc par pXRF et ICP.

Annexe 5 : Résultats de METOTRASS utilisés pour la notation du potentiel d'anomalie géochimique de la ville de Nantes.

Le tableau suivant regroupe les différents faciès géologiques rencontrés sur Nantes et le potentiel d'anomalie géochimique en arsenic, plomb, zinc et cuivre décrit dans METOTRASS (Le Guern et al., 2013). Les colonnes As, Pb, Zn et Cu décrivent le potentiel en anomalies de Metotrass pour chaque formation (note maximale des différents niveaux d'anomalie de la base Metotrass, par paramètre et faciès). L'absence de note est liée au fait que, ce faciès n'était pas présent dans la base de données METOTRASS. Les notes des métaux anthropisés et de l'arsenic (dernières colonnes du tableau) présentent les notes intermédiaires (de 0 à 2) qui ont été attribuées pour l'analyse multi-critère.

- 0 si les notes des METOTRASS sont toutes égales à 0
- 1 si au moins une note de METOTRASS est égale à 1
- 2 si au moins une note est supérieure à 1

Pour les faciès non décrits par METOTRASS et en l'absence de description d'anomalies géochimiques dans la bibliographie, seuls les remblais se voient attribué une note de 2 pour le cuivre, le plomb et le zinc afin de tenir compte de la qualité très hétérogène de ces formations anthropiques et de leur potentiel d'anomalies géochimiques, suite au retours d'expérience du traitement des quartiers urbains nantais.

Description	Unité géologique	As	Pb	Zn	Cu	Note métaux anthropisés	Note Arsenic
Formations anthropiques, remblais						2	0
Colluvions indifférenciées						0	0
Alluvions récentes et actuelles, colluvions de fond de vallon, Holocène						0	0
Alluvions fluvio-marines de l'estuaire de la Loire, vases et sables, Holocène						0	0
Levées sableuses des berges de la Loire, Holocène		0	0	0	0	0	0
Alluvions de la moyenne terrasse, graviers, sables, argiles, Pléistocène moyen-supérieur						0	0
Limons éoliens						0	0
Altérites argileuses indifférenciées, Paléocène supposé						0	0
Pliocène, sables		0	0	0	0	0	0
Pliocène, faciès redoniens, faluns		0	0	0	0	0	0

Annexe 5 : Résultats de METOTRASS utilisés pour la notation du potentiel d'anomalie géochimique de la ville de Nantes.

Massif d'Orvault-Mortagne, granite à biotite et muscovite		2	0	0	0	0	2
Massif d'Orvault-Mortagne, granite à biotite et muscovite, orienté		2	0	0	0	0	2
Massif d'Orvault-Mortagne, granite à biotite et muscovite, mylonitique		2	0	0	0	0	2
Massif d'Orvault-Mortagne, mylonites fluidales, ultramylonites		1	0	0	0	0	1
Formation de Drain, amphibolites	Unité des Mauges	0	0	0	0	0	0
Formation du Landreau, micaschistes et gneiss à biotite et muscovite	Unité des Mauges	2	0	0	1	1	2
Formation de Mauves-sur-Loire, micaschistes albitiques à muscovite et chlorite	Unité de Mauves-sur-Loire	3	0	0	0	0	2
Formation de Mauves-sur-Loire, micaschistes albitiques à biotite et muscovite	Unité de Mauves-sur-Loire	3	0	0	0	0	2
Formation de Mauves-sur-Loire, micaschistes siliceux, quartzites	Unité de Mauves-sur-Loire	3	0	0	0	0	2
Formation du Pellerin-Montaigu, orthogneiss oeilés leucocrates et gneiss feuilletés à muscovite	Unité de Paimboeuf-Montaigu	1	0	0	0	0	1
Formation du Pellerin-Montaigu, orthogneiss oeilés à muscovite et biotite	Unité de Paimboeuf-Montaigu	1	0	0	0	0	1
Formation du Pellerin-Montaigu, leptynites et gneiss anatectiques	Unité de Paimboeuf-Montaigu	1	0	0	0	0	1
Formation du Pellerin-Montaigu, mylonites hétérogènes	Unité de Paimboeuf-Montaigu	0	0	0	0	0	0
Massifs de l'Ortay et du Bignon, granite à biotite et muscovite		0	0	0	0	0	0
Formation de Saint-Paul-en-Pareds, micaschistes à biotite et grenat	Unité de Paimboeuf-Montaigu	0	0	0	0	0	0
Formation de Montaigu, amphibolites massives	Unité de Paimboeuf-Montaigu	0	0	0	0	0	0
Filons de quartz		0	0	0	0	0	0
Filons de pegmatite à béryl		2	0	0	0	0	2
Filons de pegmatite à lépidolite		2	0	0	0	0	2

Tableau 22 : Potentiel d'anomalie géochimique des faciès géologiques présents sur la ville de Nantes d'après la base de données METOTRASS (Le Guern et al, 2013) et interprétés pour l'analyse multi-critères du potentiel de présence d'anomalies géochimiques.

Annexe 5 : Résultats de METOTRASS utilisés pour la notation du potentiel d'anomalie géochimique de la ville de Nantes.

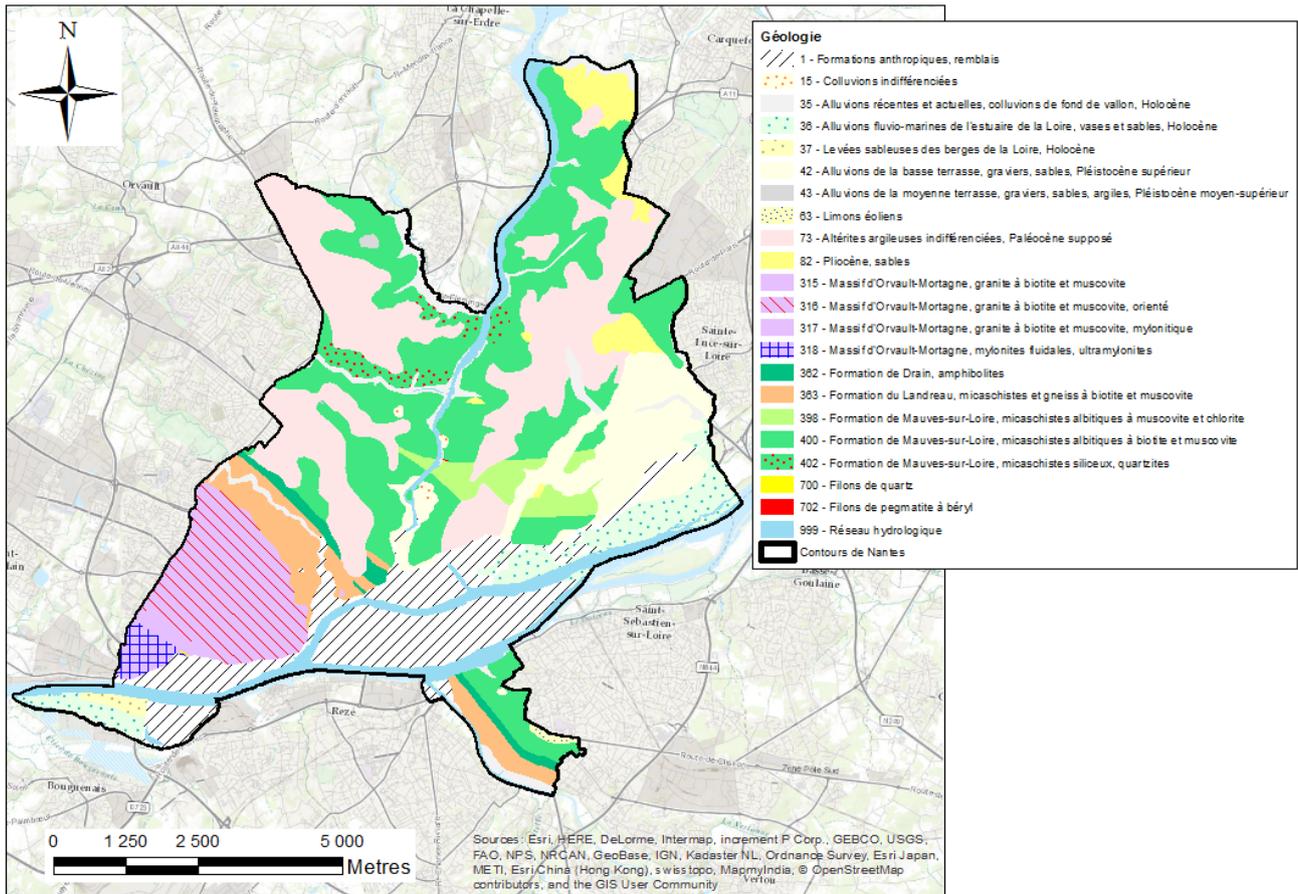


Figure 114 : Carte géologique harmonisée (d'après Bechenec, 2007) à 1/50 000 de la ville de Nantes.

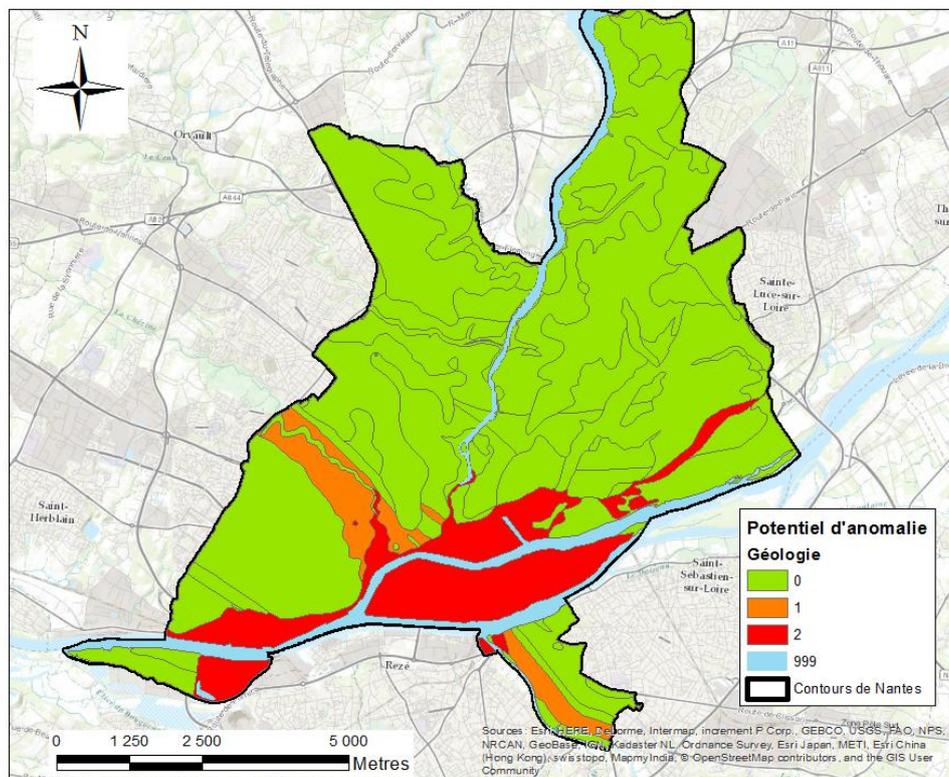


Figure 115 : Potentiel d'anomalie géochimique en cuivre, plomb et zinc suivant la géologie

Annexe 6 : Tableau de synthèse des choix de notes de potentiel d'anomalie géochimique de Corine Land Cover et copernicus

Type d'occupation	Codes CLC	Codes Copernicus	Note attribuée (potentiel d'anomalie)	Sources de contamination	Types de sources	Sources bibliographiques
Tissu urbain continu	111	1110 , 11210	2 (Fort)	Matériaux de construction, circulation automobile, forte densité urbaine	Retours d'expérience (SAMOA, GeoBaPa) et bibliographie	Birke et Rauch, 2000 Lu et Bai, 2010 Wang et al., 2012 Kelly et al., 1996
Tissu urbain discontinu	112	11220, 11230, 11240	1 (Moyen)	Matériaux de construction, Circulation automobile, Densité urbaine moyenne	Retours d'expérience (SAMOA, GeoBaPa) et bibliographie	Birke et Rauch, 2000
Zone industrielle	121	12100	2 (Fort)	Fonderies, chantiers navals, ...	Retours d'expérience (SAMOA, GeoBaPa) et bibliographie	Birke et Rauch, 2000 Lu et Bai, 2010 Wang et al., 2012 Kelly et al., 1996
Réseau de transport	122	12210, 12220, 12230, 12300	2 (Fort)	Circulation automobile, traitement des voies ferrées et gares de triage industrielles	Retours d'expérience (SAMOA, GeoBaPa) et bibliographie	Lu et Bai, 2010 Wang et al., 2012 Kelly et al., 1996
Extraction de matériaux	131	13100	2 (Fort)		Comparaison aux analyses chimiques	
Chantiers	133	13300	1(Moyen)		Comparaison aux analyses chimiques	

Annexe 6 : Tableau de synthèse des choix de notes de potentiel d'anomalie géochimique de Corine Land Cover et copernicus

Vergers et petits fruits	222	22000	1(Moyen)	Cuivre utilisé dans les vignes	Retours d'expérience (Pollusols) et bibliographie	Baize, 2000
Autres cultures	211, 242, 243	21000	0 (Nul)	Pas de contamination connue	Retours d'expérience (SAMOA, GeoBaPa) et bibliographie	Birke et Rauch, 2000 Wang et al., 2012
Espaces verts, zones naturelles	141, 142, 231, 311, 312, 322, 324, 411	13400, 14100, 14200, 23000, 31000, 40000	0 (Nul)	Pas de contamination connue	Retours d'expérience (SAMOA, GeoBaPa) et bibliographie	Birke et Rauch, 2000 Lu et Bai, 2010
Cours d'eau	511	50000	999	Absence de sol	/	/

Tableau 23 : Détail des choix de note de potentiel d'anomalie en cuivre, plomb et zinc attribuées à Corine Land Cover.

Annexe 6 : Tableau de synthèse des choix de notes de potentiel d'anomalie géochimique de Corine Land Cover et copernicus

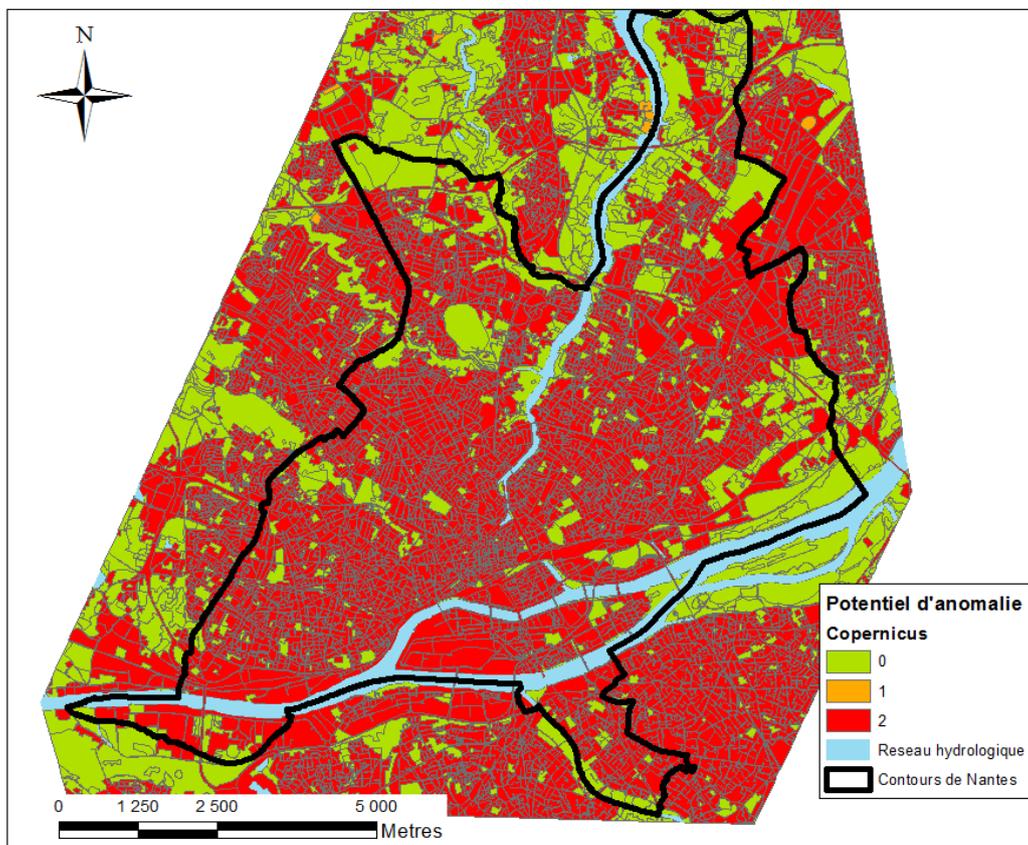


Figure 116 : Potentiel d'anomalie géochimique en cuivre, plomb et zinc suivant l'occupation des sols de la cartographie Copernicus.

Annexe 7 : Exemple des matrices de l'Inventaire Historique Urbain utilisé et principe de fonctionnement.

L'inventaire Historique Urbain réalisé sur plusieurs quartiers de Nantes (Le Guern et al., 2016, 2017) associe une emprise à une présence potentielle de différents polluants s'appuyant sur l'interprétation des activités sur la base de matrices activités polluants et des produits utilisés. Selon le principe suivant :

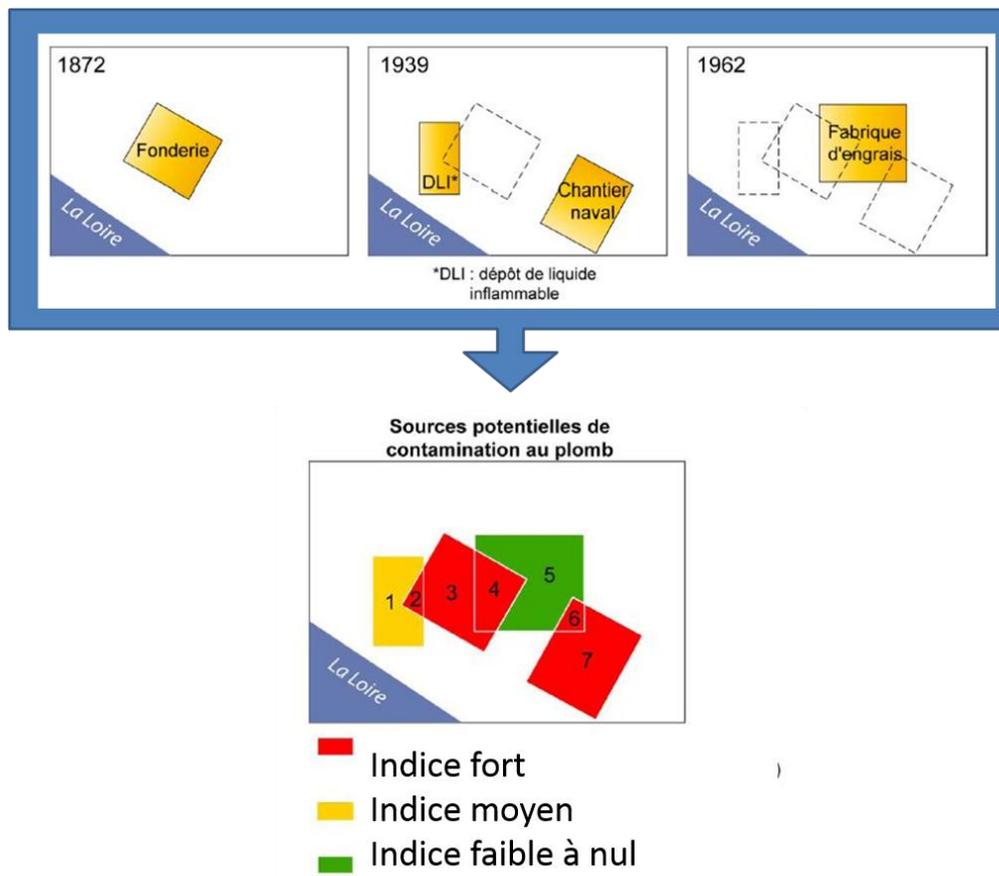


Figure 117: Cartographie de synthèse du potentiel de présence de contamination en plomb à partir des résultats d'IHU (modifiée d'après Desse-Engrand, 2015).

Enfin, de façon à traiter ces résultats dans l'analyse multi-critère, une interprétation en note intermédiaire de ce descripteur est nécessaire. Dans le cas des trois métaux cuivre, plomb, zinc, une interprétation globale est effectuée.

- 0 si les notes de l'IHU des trois métaux sont égales à 0
- 1 si au moins une des trois notes est supérieure ou égale à 1
- 2 si l'une des trois notes est supérieure à 3 ou si au moins deux notes sont égales à 3

Annexe 7 : Exemple des matrices de l'Inventaire Historique Urbain utilisé et principe de fonctionnement.

Un exemple est fourni par le Tableau 24, les tables regroupent alors 269 contours de sites et 1262 intrasites.

OID	Note_Cuivre	Note_Plomb	Note_Zinc	Note_intermédiaire
16	1	1	1	1
20	1	3	1	1
21	1	3	3	2
37	4	3	1	2
783	6	1	1	2
39	5	6	5	2
44	0	0	0	0
101	2	0	0	1

Tableau 24: Extrait de la table d'attribution des notes intermédiaires pour l'analyse multi-critère suivant les résultats des matrices polluant de l'Inventaire Historique Urbain de Nantes.

OID : Identifiant du polygone, Note_Métal : Note attribuée par l'IHU au potentiel de contamination en métal pour le site (de 0 à 6 : 0 : nul à très faible, 1 ou 2 : faible, 3 ou 4 : moyen, 5 ou 6 : fort), Cuivre : Traduction du code, Note_intermédiaire : Note attribuée pour l'analyse multi-critère pour les trois métaux (cuivre, plomb et zinc) (de 0 à 2).

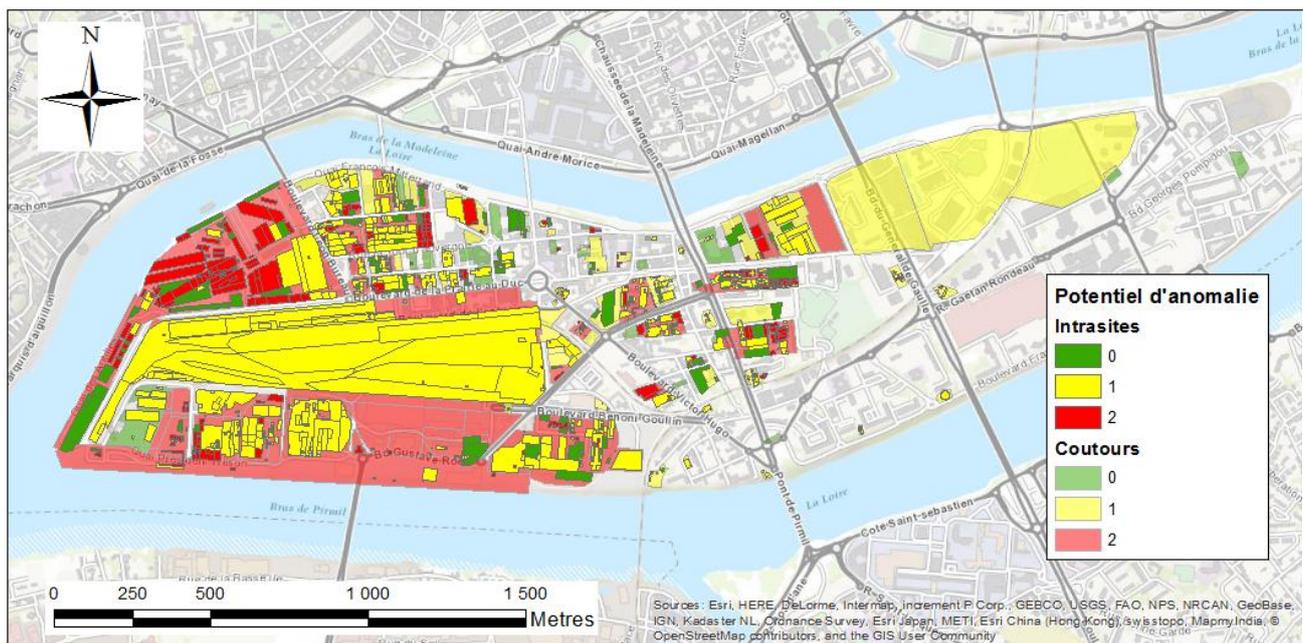


Figure 118 : Potentiel d'anomalie géochimique en cuivre, plomb et zinc lié aux activités industrielles et de service potentiellement polluantes de 1850 à nos jours.

Annexe 8 : Résultats d'analyses en cuivre suivant les découpages du territoire par analyse multi-critères

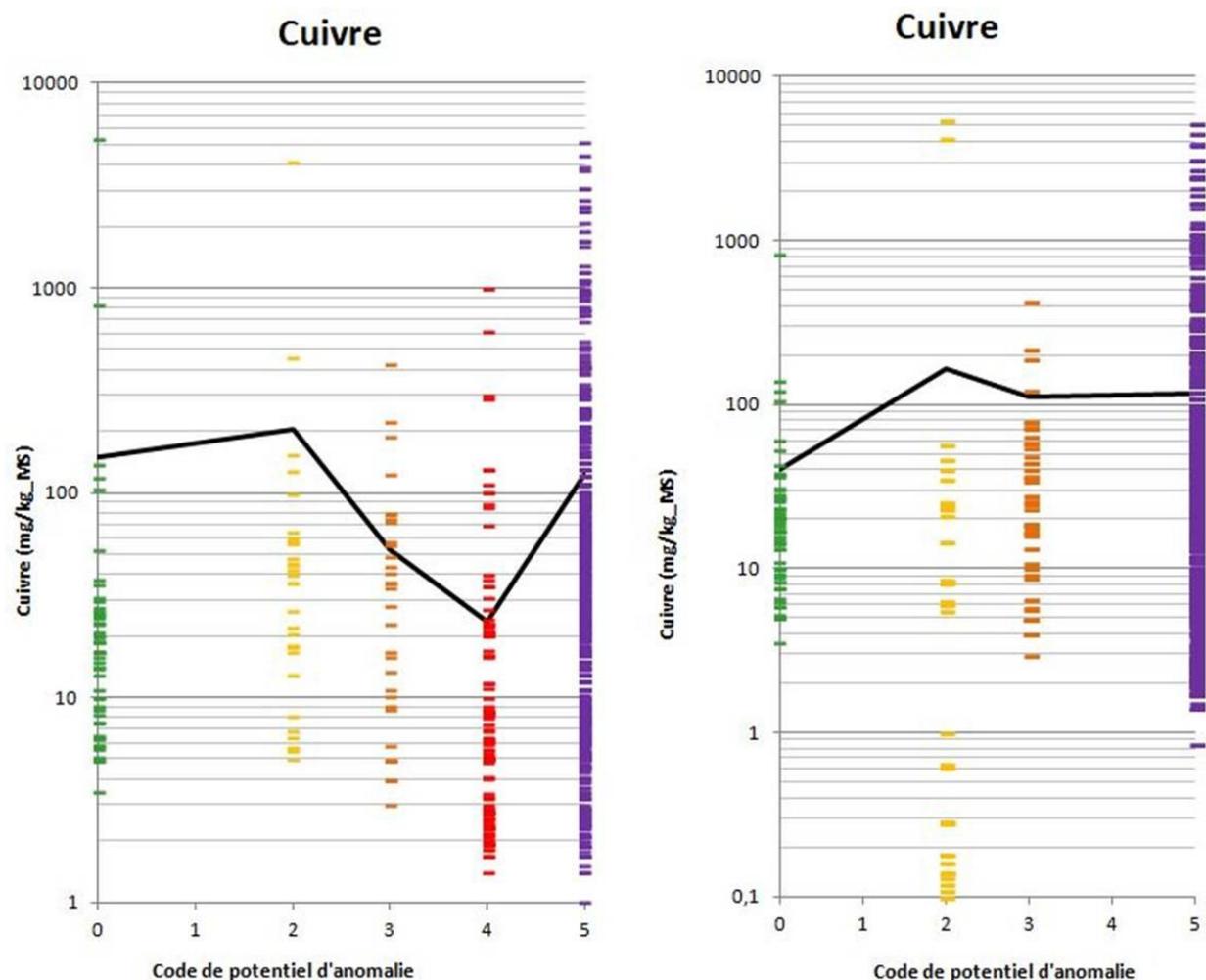


Figure 119: Teneurs en cuivre des échantillons collectés dans la ville. Les couleurs représentent le potentiel de présence d'anomalie suivant la carte réalisé à partir de Corine Land Cover (Gauche) et celle réalisée à partir de Copernicus (Droite). Les moyennes par classe sont représentées par le trait noir.

Annexe 9 : Représentation des échantillons du site sélectionnés pour l'essai de filtrage

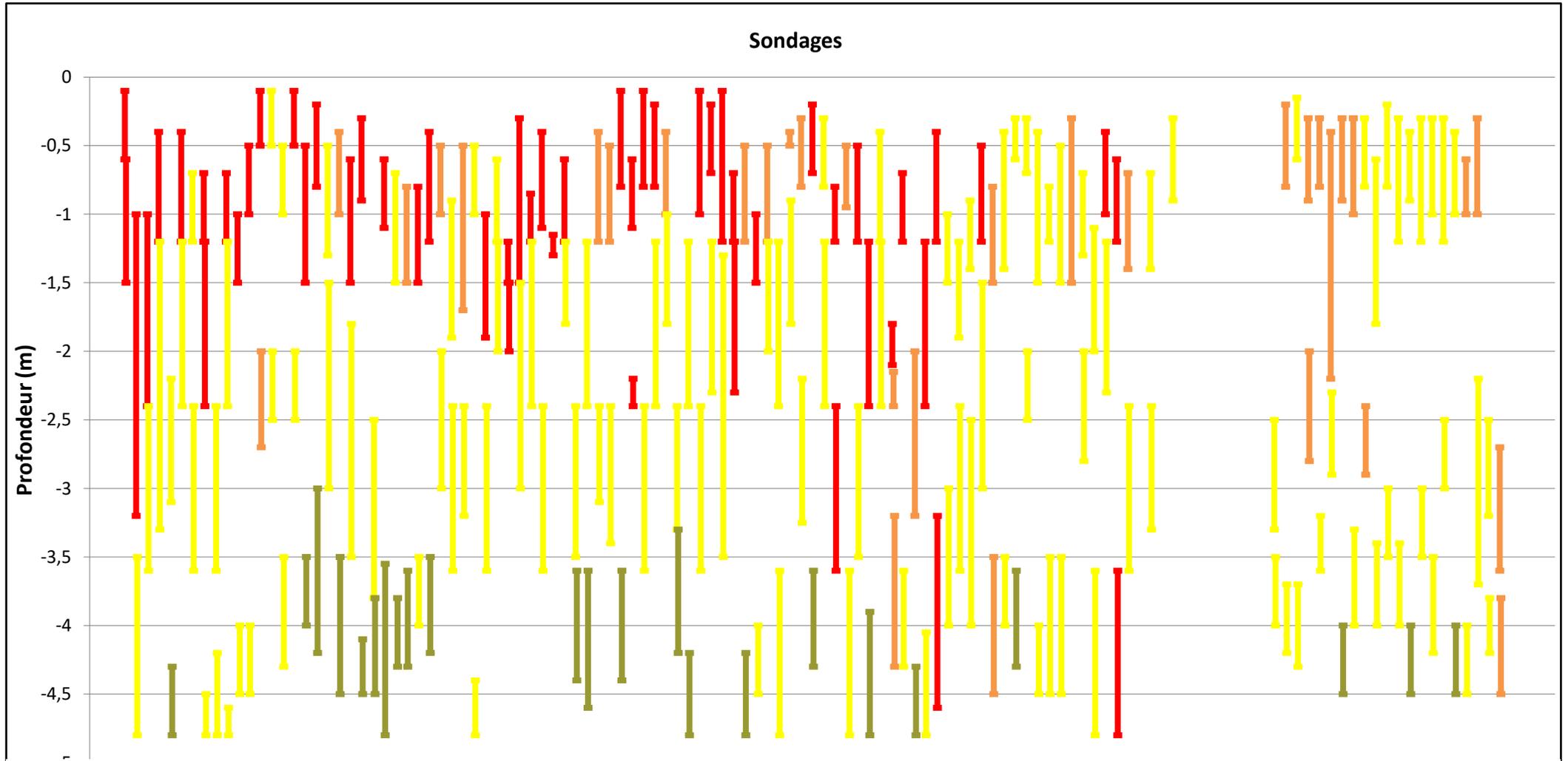


Figure 120 : Echantillons du site sélectionné pour l'essai de filtrage le long de leur sondage en fonction de la profondeur. Le code couleur correspond aux typologies de remblais (Le Guern et al, 2016).

Rouge : remblais suspects, orange : remblais divers, jaune : remblais assimilables naturels, vert : alluvions

RÉSUMÉ

Les grands réaménagements urbains génèrent d'importants volumes de terres excavées dont la valorisation est désormais possible si elles sont compatibles avec le fond pédo-géochimique du site receveur. Afin de faciliter cette valorisation, cette thèse s'intéresse à des méthodologies de traitements de données acquises à d'autres fins que la détermination du fond pédo-géochimique, afin de fournir un référentiel de qualité géochimique des sols urbains à l'échelle d'un quartier ou d'une ville. Les premiers tests montrent les limites des calculs statistiques, usuellement utilisés avec des données dédiées au fond pédo-géochimique du fait de leur sensibilité à la limite de quantification, à la loi de distribution ou encore au pourcentage d'anomalie. Le découpage du territoire en entités géographiques cohérentes à partir de la nature des sols et sous-sols et des pressions anthropiques (actuelles ou passées) permet la constitution de différents niveaux de fond suivant la qualité géochimique des sols, mais une des difficultés concerne le renseignement de l'ensemble des entités. La classification statistique des échantillons, même spatialisée, n'apporte, actuellement pas de résultats exploitables pour la gestion des terres excavées. Enfin, le filtrage géostatistique via un modèle linéaire de corégionalisation présente une alternative solide pour le filtrage des anomalies et l'estimation du fond à l'échelle d'un quartier.

MOTS CLÉS

Fond pédo-géochimique, sols urbains, filtrage géostatistique, découpage du territoire, analyse multi-critère, classification, seuil, statistique

ABSTRACT

Large urban redevelopments generate large volumes of excavated soils whose reusability is now possible if they are compatible with the pedo-geochemical background of the receiving site. In order to facilitate this valorization, this thesis focuses on processing methodologies on data acquired for other purposes than the determination of the pedo-geochemical background, in order to provide a geochemical quality reference system for urban soils at the scale of one neighborhood or city. The first tests show the limits of the statistical thresholds, usually used with data dedicated to the pedo-geochemical background because of their sensitivity to the limit of quantification, to the distribution law or to the percentage of anomaly. The segmentation of the territory into coherent geographical entities based on the nature of the soil and subsoil and the anthropogenic pressures (current or past) allows the constitution of different geochemical background levels according to the geochemical quality of the soils, but one of the difficulties concerns the information of all entities. The statistical classification of the samples, even if spatialized, does not currently provide usable results for the management of the excavated earth. Lastly, geostatistical filtering via a linear model of coregionalisation provides a solid alternative for anomaly filtering and background estimation at a neighborhood scale.

KEYWORDS

Pedo-geochemical background, urban soils, geostatistical filtering, territorial segmentation, multi-criteria analysis, classification, threshold, statistics