

SERVICE PUBLIC DE WALLONIE (SPW)
DIRECTION GENERALE OPERATIONNELLE AGRICULTURE,
RESSOURCES NATURELLES ET ENVIRONNEMENT (DGO3)
Département des Sols et des Déchets (DSD)
Direction de la Protection des Sols (DPS)

CONVENTION DGO3-UCL

Gestion des concentrations de fond élevées à l'échelle
régionale : mesure et diagnostic du risque adapté aux
spécificités du contexte wallon
(acronyme : CAPASOL 6)

Rapport final

Dr. B. PEREIRA, Ir. A. VANDEUREN et Pr. P. DELMELLE
Earth and Life Institute - Environmental Sciences
Université catholique de Louvain

Louvain-la-Neuve, septembre 2018

Table des matières

Introduction	7
1 Acquisition de données pour la cartographie des concentrations de fond	9
1.1 Base de données CAPASOL 6	9
1.1.1 Encodage des nouvelles données	10
1.1.2 Constitution de la base de données CAPASOL 6	10
1.2 Campagne d'échantillonnage et d'analyse effectuée lors de la présente convention	12
1.3 Données disponibles pour les cartes de concentrations de fond	16
2 Correction des biais et constitution de jeux de données pour la cartographie	19
2.1 Méthode établie et appliquée lors de la convention CAPASOL 3	19
2.1.1 Principe général et fondement de la méthode	19
2.1.2 Résumé de la méthode	21
2.2 Méthode établie et appliquée lors de la convention CAPASOL 6	23
2.2.1 Préparation des jeux de données	23
2.2.2 Procédures statistiques appliquées pour comparer les jeux de données .	24
2.2.3 Etape 1 : Traitement des biais temporels intralaboratoire	27
2.2.4 Etape 2 : Constitution de groupes de jeux de données non-biaisés	28
2.2.5 Etape 3 : Correction des biais entre les jeux de données biaisés	29
2.3 Constitution des jeux de données « de référence »	31
2.3.1 Principe général	31
2.3.2 Statistiques descriptives des données disponibles	32

3	Cartographie de l'impact des retombées atmosphériques de polluants	35
3.1	Contexte du travail et méthodologie	35
3.2	Sélection de cheminées et calcul des « auréoles de retombées unitaire »	36
3.3	Calcul de l'enrichissement des sols en polluants par voie atmosphérique	37
3.3.1	Principe	37
3.3.2	Calcul de l'enrichissement en polluant d'origine atmosphérique	38
3.4	Estimation de la contribution des cheminées à l'aide des points de contrôle	40
3.5	Résultats	40
3.5.1	Qualité des modèles prédictifs	40
3.5.2	Carte des retombées atmosphériques de polluants	43
4	Cartographie des concentrations de fond	47
4.1	Introduction	47
4.2	Groupe de sols et niveau d'appartenance	50
4.3	Algorithme de prédiction des concentrations de fond	54
4.3.1	Introduction	54
4.3.2	Description de l'algorithme	54
4.4	Validation des cartes de concentration de fond	58
4.4.1	Cartes du centile 50	58
4.4.2	Carte des centiles 90, 95 et 99	71
4.5	Cartes des concentrations de fond	73
4.5.1	Statistiques descriptives des concentrations de fond en Wallonie	73
4.5.2	Présentation des résultats cartographiques	74
4.6	Cartes d'incertitude sur les concentrations de fond	90
4.7	Comparaison des cartes produites avec des valeurs normatives wallonnes	90
4.7.1	Arrêté réglementant l'utilisation des boues	91
4.7.2	Décret sols	92
5	Etude de sensibilité du modèle S-Risk et applications de l'outil « concentration de fond limite »	107
5.1	Etude de sensibilité du modèle S-Risk et stratégies de gestion à l'échelle régionale des jardins potagers et parcelles agricoles	107
5.1.1	Evaluation des risques pour un jardin potager	109

5.1.2	Cas de Pb et As	110
5.1.3	Impact de l'ingestion de légumes dans le calcul de l'indice de risque . . .	111
5.1.4	Stratégie de gestion des potagers et parcelles agricoles contaminés	113
5.1.5	Comment évaluer la qualité d'une production végétale ?	114
5.1.6	Calcul de nouvelles valeurs limites	115
5.2	Outil « concentration de fond limite » : applications possibles	117
Conclusions et perspectives		121
Annexe 1 : Cartographie de la répartition des erreurs de prédiction		123
Annexe 2 : Cartes de l'incertitude sur les concentrations de fond		131

Introduction

La présente étude se situe dans le cadre général de la recherche de l'origine des polluants dans les sols, de l'évaluation des risques liés à ceux-ci pour l'écosystème, la santé humaine et la production agricole ainsi que sur les méthodes de prédiction les plus performantes pour quantifier leur présence dans les sols. La Wallonie présente un cas d'étude très favorable pour étudier ces thématiques, car une grande proportion de son territoire a été touchée par des contaminations en polluants. L'amélioration de la connaissance et de la quantification des polluants dans les sols constitue une thématique de recherche dont les acquis peuvent se révéler cruciaux pour la gestion des sols de Wallonie, notamment par leur exploitation dans une méthodologie d'évaluation des risques adaptés aux spécificités du contexte wallon.

La convention comporte deux grandes thématiques de recherche.

La première thématique a pour objectif l'établissement d'un diagnostic de risque pour la santé humaine et les écosystèmes des concentrations de fond élevées adapté aux spécificités des sols wallons. Cette première thématique a été en partie réorientée vers les travaux de la convention SANISOL. Cette convention de recherche, débutée en janvier 2018, a pour objectif le développement d'un outil pilote permettant de fournir des recommandations particulières de gestion et d'utilisation à tout producteur de biomasse alimentaire produite sur un sol contaminé en Wallonie. Une série de tâches de CAPASOL 6 liées à la thématique 1 ont été redéfinies pour les besoins de la convention SANISOL. Ces tâches sont en cours de réalisation, elles ne seront donc pas l'objet du présent rapport.

La seconde thématique porte sur la mise à jour de la recherche visant à établir les concentrations de fond en polluants et l'amélioration des cartes des concentrations de fond en Wallonie. Une première version de la carte des concentrations de fond avait été réalisée en 2008. Cette dernière avait été mise à jour en 2012 lors de la convention CAPASOL 3. Au cours de CAPASOL 4, des partages de connaissances sur le thème de l'établissement et la définition des concentrations de fond avec les acteurs wallons et européens a été réalisé. Ces échanges ont permis de comparer l'approche wallonne et d'en dégager ses spécificités. Ce travail de recherche s'inscrit dans la continuité de ces différents travaux. Les nouvelles cartes concentrations de fond réalisées dans le cadre de CAPASOL 6 utilisent un modèle prédictif amélioré qui bénéficie à la fois des nouvelles données géochimiques disponibles en 2018, des méthodes les plus récentes pour la constitution de jeu de données fiable à partir de ces données et des nouvelles cartes de l'impact des retombées atmosphériques de polluants sur les sols de Wallonie.

Le rapport est structuré de la façon suivante :

Le premier Chapitre traite de l'acquisition de données pour la cartographie des concentrations de fond. Il s'agit de la collecte des jeux de données géochimiques disponibles en Wallonie, de l'encodage de nouvelles données dans la base de données CAPASOL et de la campagne d'échantillonnage et d'analyse réalisée dans le cadre de la présente convention.

Le second Chapitre traite de la méthodologie utilisée pour la constitution de jeux de données fiables pour la cartographie des concentrations de fond. Ce Chapitre détaille les étapes de traitement des biais temporels intralaboratoires des données de la BD CAPASOL et des biais interlaboratoire entre l'ensemble des jeux de données disponibles (décrits au Chapitre 1).

Le troisième Chapitre traite de la cartographie de l'impact des retombées atmosphériques de polluants inorganiques sur les sols de Wallonie. Ce Chapitre expose la méthodologie utilisée, inspirée des travaux de POLLUSOL 2 (SPAQuE), et les résultats cartographiques sont présentés. Les cartes produites ont pour objectif d'améliorer le modèle géostatistique utilisé pour prédire les concentrations de fond.

Le quatrième Chapitre traite de la cartographie des concentrations de fond. L'algorithme utilisé pour obtenir les cartes des concentrations de fond est présenté, ainsi que l'ensemble des résultats obtenus. Les cartes produites sont comparées aux valeurs normatives du décret « sols » (M.B. 22.03.2018) et de l'arrêté réglementant l'utilisation des boues (M.B. 12.04.1995).

Le dernier Chapitre reprend certains travaux liés à la thématique 1 de la convention (qui vise l'établissement d'un diagnostic de risque adapté aux spécificités des sols wallons) qui n'ont pas été réorientés vers la convention SANISOL.

Chapitre 1

Acquisition de données pour la cartographie des concentrations de fond

1.1 Base de données CAPASOL 6

La base de données CAPASOL 6 constitue la grande majorité des données disponibles en Wallonie pour réaliser la cartographie des concentrations de fond. Elle reprend les bulletins d'analyse de sol des parcelles qui ont fait l'objet d'une demande d'épandage de boues de station d'épuration à l'Administration de la Région wallonne conformément à l'arrêté du gouvernement wallon du 12 avril 1995 réglementant l'utilisation des boues¹.

Lors de la convention CAPASOL 3, les données parvenues à l'Administration avant 2012 ont rendu possible la constitution d'une base de données de 17 668 analyses de parcelles agricoles localisées par les coordonnées géographiques du centre de la parcelle. Sur chaque parcelle, les teneurs pseudo-totales (eau régale) pour sept ETM (Cd, Cr, Cu, Hg, Ni, Pb, Zn) ont été mesurées, ainsi que le pH. Dans le cadre de cette convention, les données parvenues à l'Administration entre 2012 et 2017 ont été encodées et jointes à la base de données CAPASOL 3, pour constituer la base de données CAPASOL 6 comptant 25 592 analyses.

L'encodage de données lors de CAPASOL 6 a été réalisé en grande partie à l'aide d'un script *R* mis au point lors de la convention CAPASOL 3. Ce script reprend un grand nombre d'opérations de réduction de doublons et de récupération de diverses informations. Il contient par exemple une procédure qui permet, lorsque le nom du laboratoire n'est pas renseigné, de le déduire à partir de la structure de l'identifiant du bulletin d'analyse. Ce script a été mis à jour pour correspondre aux nouvelles situations qui se sont présentées dans les fichiers disponibles lors de CAPASOL 6. Ce nouveau script a permis de constituer une base de données complète, consolidée et opérationnelle. Un résumé de ces étapes est donné dans les sections ci-dessous.

1. Arrêté du gouvernement wallon du 12 avril 1995 réglementant l'utilisation sur ou dans les sols des boues d'épuration ou des boues issues de centres de traitement de gadoues de fosses septiques (M.B. 12.04.1995)

Pour toute information complémentaire par rapport au script et à l'encodage de ces données en général, le lecteur est invité à consulter le rapport final de la convention CAPASOL 3².

1.1.1 Encodage des nouvelles données

Les données parvenues à l'Administration depuis 2012 ont été transmises à l'UCL sous forme de sept tables de données (fichiers en format Ms Excel). Selon l'organisme intermédiaire entre le producteur de boues et l'agriculteur³, la transmission du résultat d'analyse a pu avoir lieu de plusieurs manières différentes. Les résultats d'analyse peuvent avoir été transmis sous forme de fichiers standardisés type E1000XXXX.xlsx, soient encodés directement en ligne (bases de données TALISOL et boues industrielles - ANASOL), soient avoir été transmis sous forme d'un fichier Excel différent du type E1000. Au total, 10526 nouvelles données étaient contenues dans ces fichiers (voir Table 1.1).

Fichiers	Nombre de données
E1000 (111 fichiers)	2517
Extrait TALISOL	2918
Extrait ANASOL	4998
Autres (4 fichiers)	93
Total	10526

TABLE 1.1 – Nombre de données contenues dans les fichiers reprenant les résultats d'analyse parvenus à l'Administration entre 2012 et 2017.

1.1.2 Constitution de la base de données CAPASOL 6

Les 10 526 données parvenues depuis 2012 ont été jointes au 17 668 données de la base de données CAPASOL 3, pour un total de 28 194 données. Ensuite, 1759 doublons de localisations -résultat d'analyse ont été retirés en ne conservant que la donnée la plus ancienne à chaque fois. Pour les 26 435 données restantes, 843 doublons de numéros d'analyse ont été retirés. Pour ce faire, les doublons de numéro d'analyse ont tous été examinés à la main, car plusieurs cas pouvaient se présenter. Les mêmes règles que celles utilisées lors de CAPASOL 3 ont été suivies. Si pour deux données identifiées par le même numéro d'analyse :

- Les résultats d'analyses (pH-ETM) sont différents, alors le suffixe *bis* est accolé à un des deux numéros d'analyse.
- Les résultats d'analyse (pH-ETM) sont identiques, les parcelles agricoles sont différentes et les résultats d'analyse ont été communiqués à plusieurs années d'intervalle.

2. Vandeuken, Aubry ; Pereira, Benoît ; Sonnet, Philippe. Rapport final de la convention de recherche intitulée "Suivi et aide à l'administration concernant la mise en application (1) de l'outil de prédiction et de gestion de la capacité des sols à accepter l'épandage d'amendements organiques conforme à la réglementation (outil CAPASOL), et (2) du Décret « sols »." (acronyme : CAPASOL 3), 2014. 168 p.

3. L'intermédiaire est l'organisme qui se charge de mettre en contact l'agriculteur et le producteur de boues, ainsi que des démarches pour la valorisation des boues, en ce compris le contact avec le laboratoire pour l'analyse du sols de la parcelle et la réception du résultat d'analyse. C'est donc cet organisme qui transmet le résultat d'analyse à l'Administration.

Ce cas de figure représente typiquement le cas où un agriculteur réutilise le résultat d’analyse d’une parcelle pour une autre parcelle plusieurs années plus tard. C’est alors le lien résultat-parcelle communiqué en premier lieu qui est conservé.

- Les résultats d’analyse (pH-ETM) sont identiques, les parcelles agricoles sont différentes et les résultats d’analyse ont été communiqués en même temps, les deux données sont supprimées, car il n’est pas possible de retrouver la localisation de la parcelle analysée.

Ensuite, sur les 25 592 données restantes, 107 étaient localisées hors Région wallonne. Étant donné qu’il ne devrait pas y avoir de données localisées hors Région wallonne étant donné que ces analyses de sols sont réalisées dans le cadre d’une réglementation wallonne, il est probable que la grande majorité de ces localisations hors Région wallonne soient dues à des erreurs d’encodage. Deux principaux types d’erreurs facilement corrigeables ont été identifiées : une division/multiplication par 10 d’une (ou deux) des coordonnées, et une inversion des coordonnées. A chaque fois que la correction de ces erreurs donnait lieu à une localisation sur une parcelle agricole située en Région wallonne, les coordonnées ont été corrigées comme telles. Au total, 25 localisations ont pu être corrigées sur les 107 données situées hors Région wallonne.

Finalement, étant donné que les données sont censées être localisées en usage agricole, une vérification de leur localisation a été réalisée. Toutes les données situées à plus de 10 mètres d’une parcelle agricole ont été premièrement identifiées en utilisant la Carte d’Occupation des Sols de Wallonie (COSW⁴, version 2005). Ensuite, pour chacune de ces localisations, une vérification manuelle à l’aide d’une photo aérienne actuelle a été réalisée. Au terme de cette étape, 307 données ont été localisées hors domaine agricole.

La Table 1.2 reprend le bilan du nombre de données conservées au terme de chacune des étapes citées ci-dessus. Au terme de ces étapes, 25 203 données sont potentiellement utilisables pour la cartographie des concentrations de fond en ETM. La figure 1.1 reprend leur localisation en Wallonie.

Étapes	Nombre de données
Dans les fichiers sources	28 194
Après retrait des doublons totaux (pH-ETM-Localisation)	26 435
Après retrait des doublons de numéro d’analyse (nombre final)	25 592
Après retrait des localisations hors Région wallonne	25 510
Après retrait des localisations hors usage agricole	25 203

TABLE 1.2 – Évolution du nombre de données conservées après chaque étape de l’encodage. La base de données compte 25 592 données, car les données localisées hors Région wallonne ou hors usage agricole ont été conservées, bien qu’elles ne soient pas utilisables pour la cartographie des concentrations de fond.

4. DGARNE, PCNOSW, projet du gouvernement wallon (GW VIII / 2005 / Doc.1022 / 28.04 / B.L.)

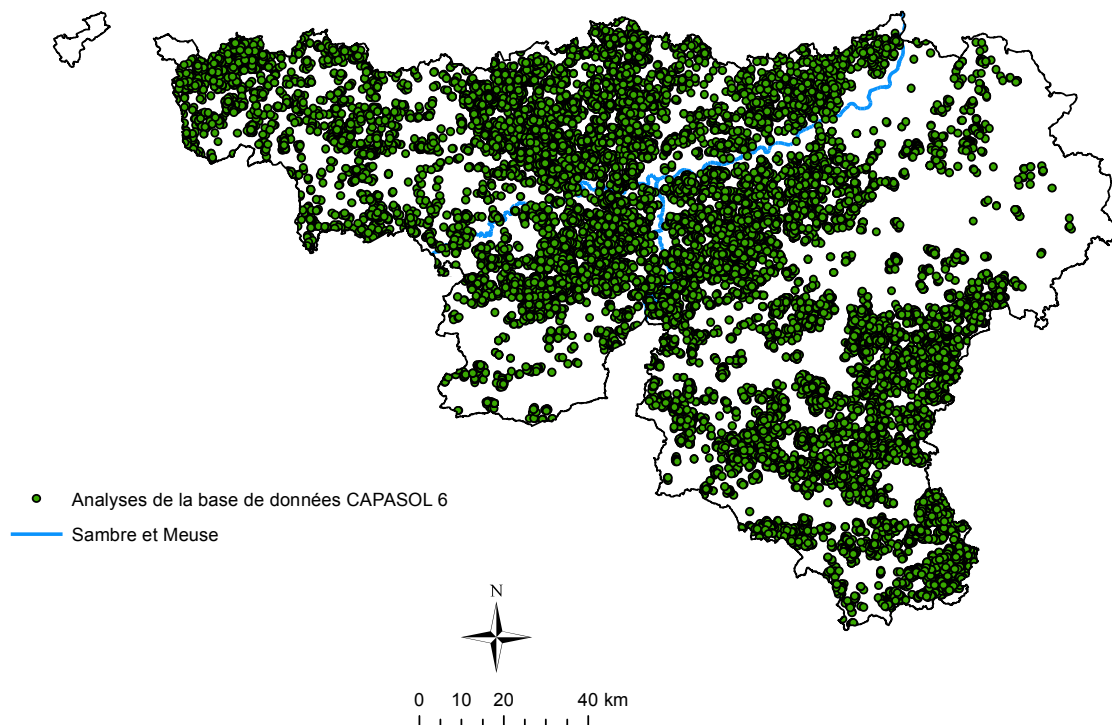


FIGURE 1.1 – Localisation des 25 203 résultats d’analyse de la base de données CAPASOL 6 utilisables pour la cartographie des concentrations de fond.

1.2 Campagne d’échantillonnage et d’analyse effectuée lors de la présente convention

En vue de l’amélioration du modèle des retombées atmosphériques de polluants à l’échelle régionale, une campagne d’échantillonnage et d’analyse a été réalisée lors de la présente convention selon le plan d’échantillonnage réalisé au cours de la convention CAPASOL 4⁵.

Des échantillons de sols agricoles, de forêts et de parcs urbains ont été prélevés par l’équipe de recherche de l’UCL sur 126 points répartis dans l’agglomération liégeoise (Figure 1.2). Sur chaque point, un échantillon de sol a été prélevé à la tarière dans l’horizon de surface. Les échantillons prélevés ont été séchés à l’air libre, tamisés à 2 mm avant d’être analysés par fluorescence des rayons X portable (pXRF) pour un grand nombre d’éléments majeurs (Al, Ca, Fe) et traces (As, Cd, Co, Cr, Cu, Hg, Ni, Pb, Zn).

Les résultats de cette campagne sont illustrés aux Figures 1.3, 1.4, 1.5, 1.6 et 1.7.

5. Pour plus de détail, nous renvoyons le lecteur au rapport final de la convention CAPASOL 4 : Pereira, Benoît ; Sonnet, Philippe ; Vandeuken, Aubry ; Xyrafis, Stratos. Rapport final de la convention de recherche intitulée : « Géochimie et écotoxicité des éléments traces métalliques dans les sols agricoles de Wallonie (acronyme : CAPASOL 4) », 2016.

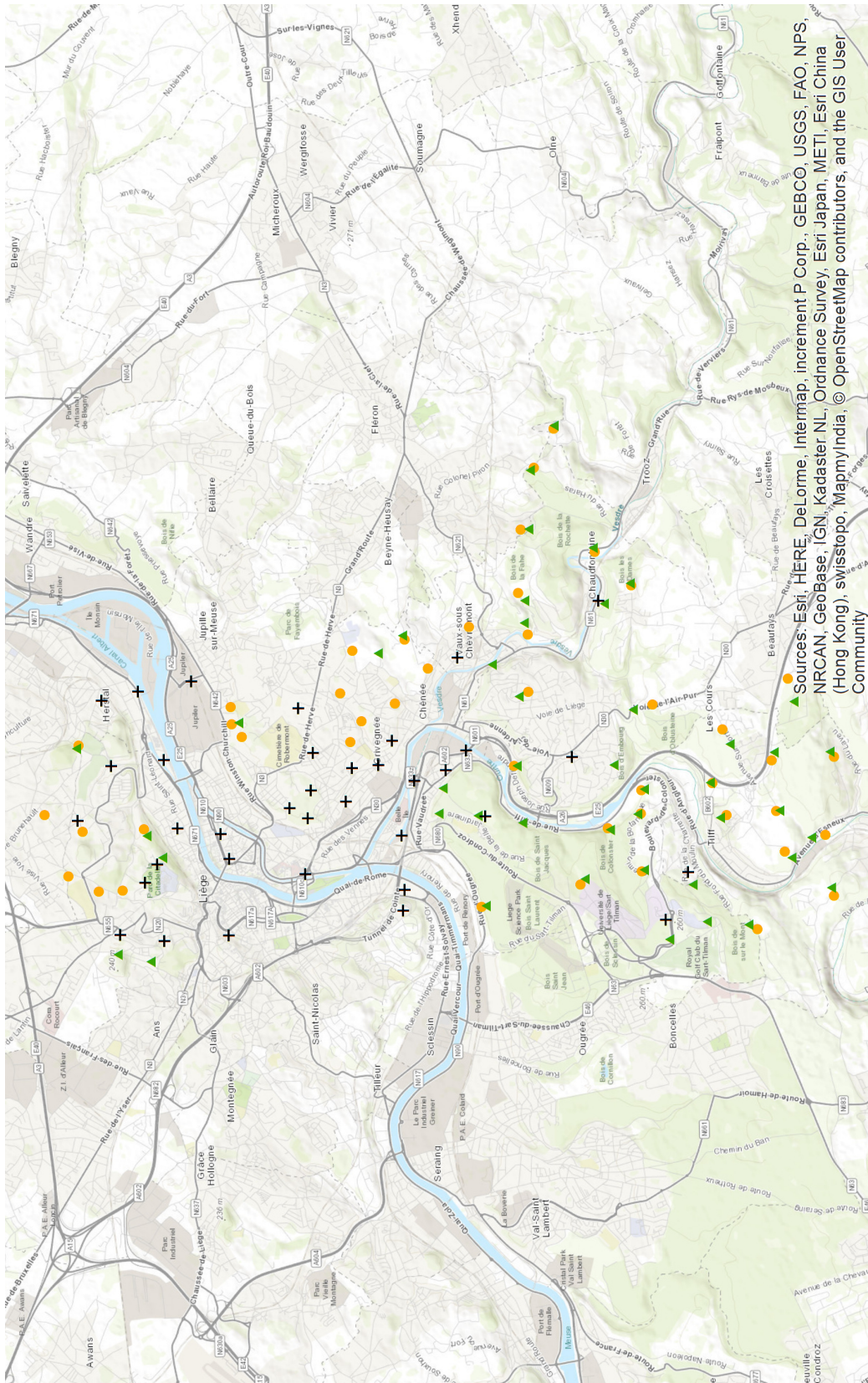


FIGURE 1.2 – Localisation des 126 points échantillonnés lors de CAPASOL 6 : 45 sols forestiers (triangle vert), 45 sols agricoles (rond orange) et 36 sols de parcs et divers (croix noire).

Pb (ppm)

- 34 - 100
- 100 - 150
- 150 - 200
- 200 - 300
- > 300

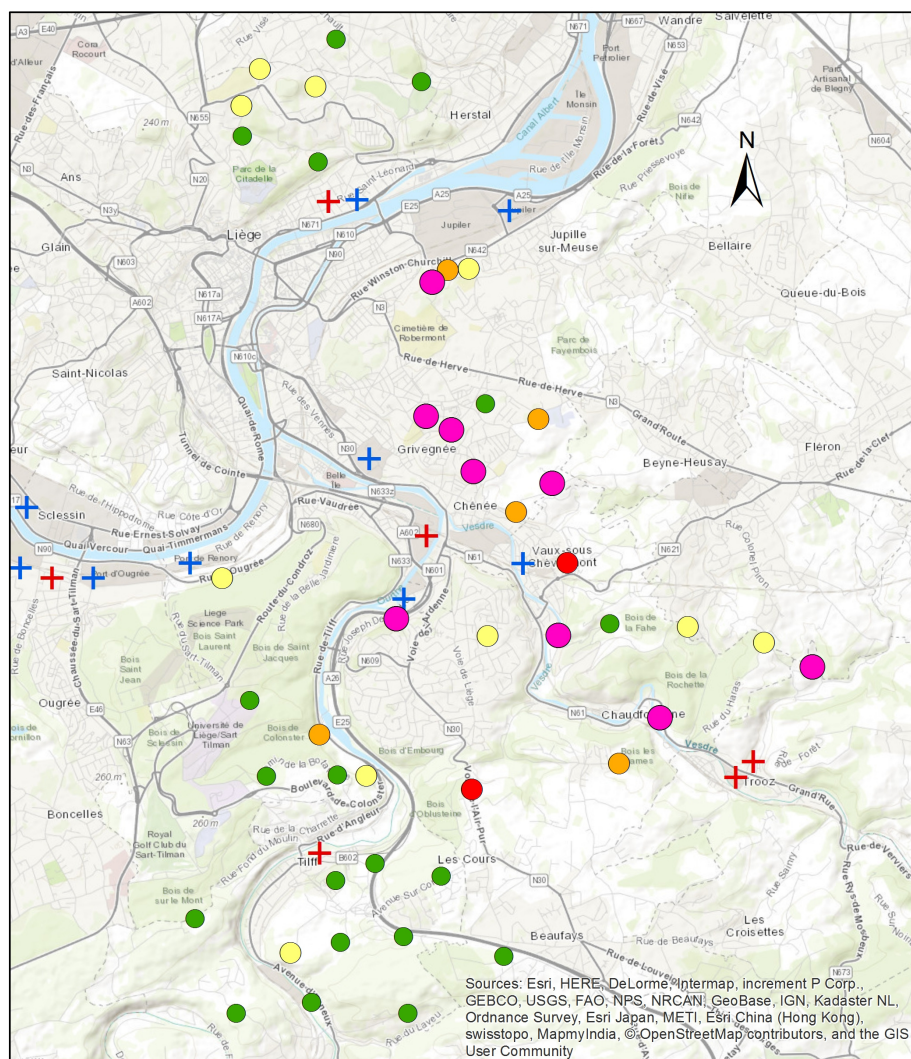


FIGURE 1.3 – Résultat de la campagne d'analyse du Pb dans les sols agricoles échantillonnés lors de la présente convention.

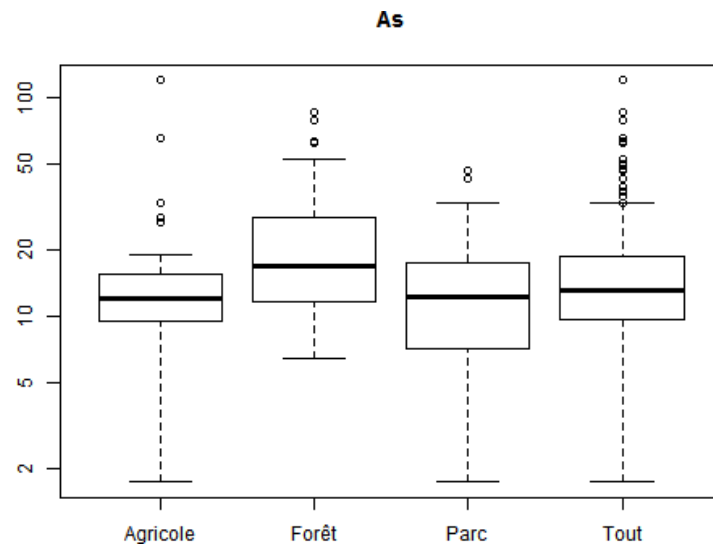


FIGURE 1.4 – Arsenic : comparaison des résultats (ppm) obtenus en fonction du type de sol.

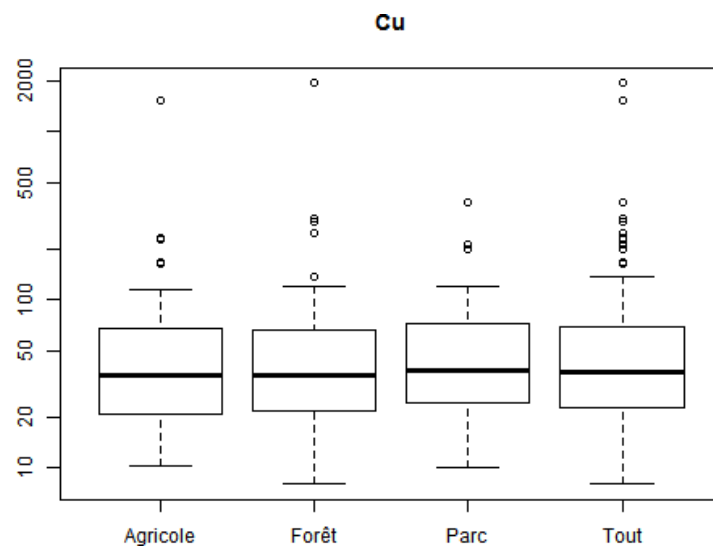


FIGURE 1.5 – Cuivre : comparaison des résultats (ppm) obtenus en fonction du type de sol.

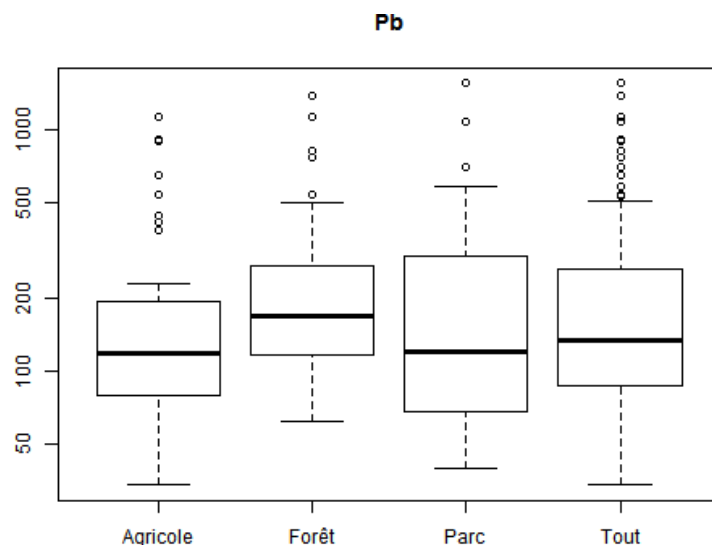


FIGURE 1.6 – Plomb : comparaison des résultats (ppm) obtenus en fonction du type de sol.

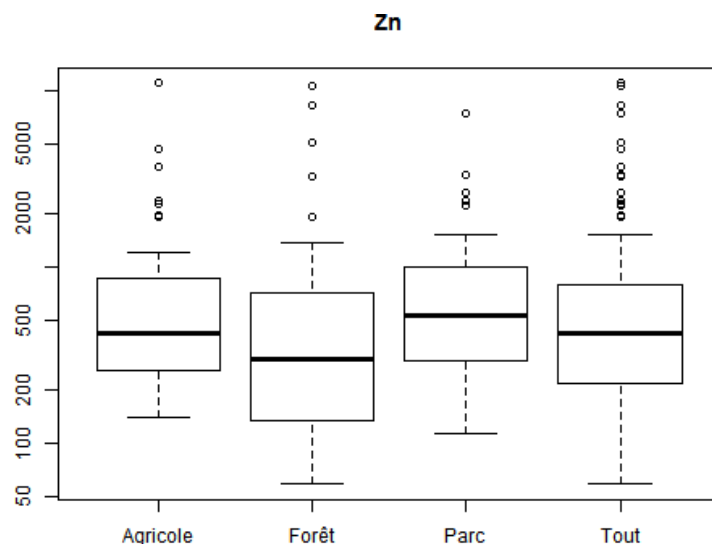


FIGURE 1.7 – Zinc : comparaison des résultats (ppm) obtenus en fonction du type de sol.

1.3 Données disponibles pour les cartes de concentrations de fond

La Table 1.3 reprend le nombre de données et les usages des sols des jeux de données utilisés pour l'établissement des cartes de concentration de fond. A notre connaissance, la Table reprend l'ensemble des jeux de données disponibles en Wallonie pour lesquels une certaine garantie existe par rapport à la qualité des données et qui concernent exclusivement des sols non touchés par de la pollution localisée⁶.

6. Pollution qui touche une aire relativement faible et due à une ou plusieurs source(s) bien identifiée(s) et souvent très proche(s). Il s'agit généralement d'apports massifs, souvent associés aux exploitations minières,

TABLE 1.3 – Données géochimiques utilisées pour la cartographie des concentrations de fond.

Etude / projet	Usage(s)	Nombre de données
BD CAPASOL 6	Agricole	25 592
Inventaire forestier	Forêt	824
POLLUSOL 1	Agricole, Forêt	161
POLLUSOL 2	Agricole, Forêt, Urbain, Potager	796
ALTERASOL	Agricole	120
Campagne CAPASOL 6	Agricole, Forêt, Urbain	126
URBAN SOIL	Potager	39
Thèse d'Amandine Liénard	Agricole, Forêt	247
Analyses pXRF CAPASOL 4	Agricole, Forêt	48
Analyses pXRF CAPASOL 5	Forêt, Urbain	21
Analyses « eau régale » CAPASOL 4 et 5	Agricole, Forêt, Urbain	32

Références des études citées dans la Table 1.3 :

CAPASOL 4 : Pereira, Benoît ; Sonnet, Philippe ; Vandeuren, Aubry ; Xyrafis, Stratos. Rapport final de la convention de recherche intitulée : "Géochimie et écotoxicité des éléments traces métalliques dans les sols agricoles de Wallonie (acronyme : CAPASOL 4).", 2016.

CAPASOL 5 : Vandeuren, Aubry ; Pereira, Benoît ; Sonnet, Philippe ; Delmelle, Pierre. Gestion du risque pour la santé humaine et les écosystèmes des contaminants métalliques présents dans les sols wallons (acronyme : CAPASOL 5), 2017. 101 p.

Inventaire forestier : Alderweireld M., Burnay F., Pitchugin M., Lecomte H. Inventaire Forestier Wallon. Résultats 1994 – 2012. SPW, DGO3, DNF, Direction des Ressources forestières, Jambes, 2015, 236 p.

POLLUSOL 1 : Ducarme, Francois ; Bock, Laurent ; Bogaert, Patrick ; Colinet, Gilles ; Delcarte, Emile ; Delvaux, Bruno ; Laroche, Jean ; Maesen, Philippe ; Marcoen, Jean-Marie ; Wibrin, Marie-Aline ; Sonnet, Philippe. Rapport final de l'étude POLLUSOL : Établissement et cartographie des teneurs bruits de fond en éléments traces métalliques et micro-polluants organiques dans les sols de la région wallonne, UCL, Louvain-la-Neuve, 2003, 124 pp.

POLLUSOL 2 : Pereira, Benoît ; Titeux, Hugues ; Schneider, Arnaud ; Sonnet, Philippe. Rapport final du projet POLLUSOL 2 - partie "sols" , 2012. 1126 p.

ALTERASOL : Chouters, Elisabeth ; Colinet, Gilles. Recherche d'intérêt général portant sur le développement et la mise en œuvre d'une méthodologie d'échantillonnage et de caractérisation des sols en vue de l'établissement des concentrations de fond dans les zones peu couvertes par les inventaires existants ainsi que dans les horizons en profondeur. ULiège-Gembloux Agro-Bio Tech, 2014. 88 p.

URBAN SOIL : Colinet, G., Barbieux, S., Leyh, R., Henquin, G., Lemtiri, A., & Liénard, A. Qualité des sols, pratiques agricoles et perception des services rendus par les jardins potagers : cas d'études en jardins collectifs. ULiège-Gembloux Agro-Bio Tech, 2018.

Thèse d'Amandine Liénard : Liénard, Amandine. Caractérisation multiscalaire de contaminations en éléments traces dans les sols : distribution spatiale et verticale, évaluation du transfert sol-plante-vers de terre et gestion potentielle par amendements organiques. 2016. Thèse de doctorat. Université de Liège, Liège, Belgique.

Chapitre 2

Correction des biais intralaboratoires et interlaboratoires et constitution de jeux de données pour la cartographie

La convention de recherche CAPASOL 3 a mis en évidence la présence de biais temporels intralaboratoires et de biais interlaboratoires dans la base de données CAPASOL¹. Une méthode de correction de ces biais basée exclusivement sur le jugement d'expert avait été mise au point. Dans le cadre de cette convention, cette méthode a été améliorée et objectivée en faisant notamment intervenir des procédures développées ces dernières années dans le cadre de la thèse de doctorat de B. Pereira². Dans ce Chapitre, cette méthode est développée ci-dessous après le rappel de ce qui avait été fait lors de la convention CAPASOL 3.

2.1 Méthode établie et appliquée lors de la convention CAPASOL 3

2.1.1 Principe général et fondement de la méthode

La méthode mise au point dans CAPASOL 3 est fondée sur la connaissance *a priori* des facteurs explicatifs des concentrations géochimiques dans un sol. Dans la plupart des situations, les concentrations géochimiques peuvent être expliquées par des facteurs géogéniques ou anthropiques, tels que la lithologie ou la contamination industrielle. Dans un même jeu de

1. Vandeuren, Aubry ; Pereira, Benoît ; Sonnet, Philippe. Rapport final de la convention de recherche intitulée "Suivi et aide à l'administration concernant la mise en application (1) de l'outil de prédiction et de gestion de la capacité des sols à accepter l'épandage d'amendements organiques conforme à la réglementation (outil CAPASOL), et (2) du Décret « sols »." (acronyme : CAPASOL 3), 2014. 168 p.

2. Voir notamment les articles : Pereira, B., Vandeuren, A., & Sonnet, Ph. (2015). Geochemical mapping based on multiple geochemical datasets : A general method, and its application to Wallonia (Southern Belgium). *Journal of Geochemical Exploration*, 158, 34-43. Pereira, B., Vandeuren, A., Govaerts, B. B., & Sonnet, Ph. (2016). Assessing dataset equivalence and leveling data in geochemical mapping. *Journal of Geochemical Exploration*, 168, 36-48.

données géochimique, les échantillons de sol qui possèdent les mêmes facteurs explicatifs présenteront des concentrations géochimiques comparables. Par « comparable », nous entendons un degré de similitude tel que les valeurs mesurées appartiendront à une même population (avec ses caractéristiques propres : valeur centrale, dispersion, etc.). Cette connaissance *a priori* des facteurs explicatifs des concentrations géochimiques se traduit dans la conception d'Unités Spatiales de Comparaison (USC). La création des Unités Spatiales de Comparaison consiste à diviser la zone d'étude en unités spatiales qui sont homogènes par rapport aux facteurs connus pour expliquer les concentrations dans le jeu de données géochimique. L'idée qui sous-tend la création d'entités (unités) spatiales repose sur le constat que les facteurs explicatifs ne sont pas disposés aléatoirement dans l'espace. Par exemple, deux échantillons prélevés à des endroits très proches auront plus de chance de reposer sur une même lithologie ou d'avoir été impactés par la même intensité de contamination atmosphérique que deux échantillons prélevés à des endroits très éloignés. Deux échantillons prélevés à des endroits très proches auront donc plus de chance de présenter des niveaux de concentration géochimiques semblables que deux échantillons prélevés à des endroits très éloignés.

Les projets de cartographie géochimique (POLLUSOL 1 et POLLUSOL 2) ont mis en évidence les principaux facteurs qui expliquent les concentrations dans les sols wallons : le type de matériau parental du sol, le mode de gestion du sol et les retombées atmosphériques de contaminants. Partant de ce constat, CAPASOL 3 a délimité des USC pour les éléments traces dans les sols wallons (Un exemple est donné à la Figure 2.1). Cette délimitation a été réalisée en négligeant le facteur explicatif « mode de gestion du sol », CAPASOL 3 ne s'intéressant qu'aux sols agricoles.

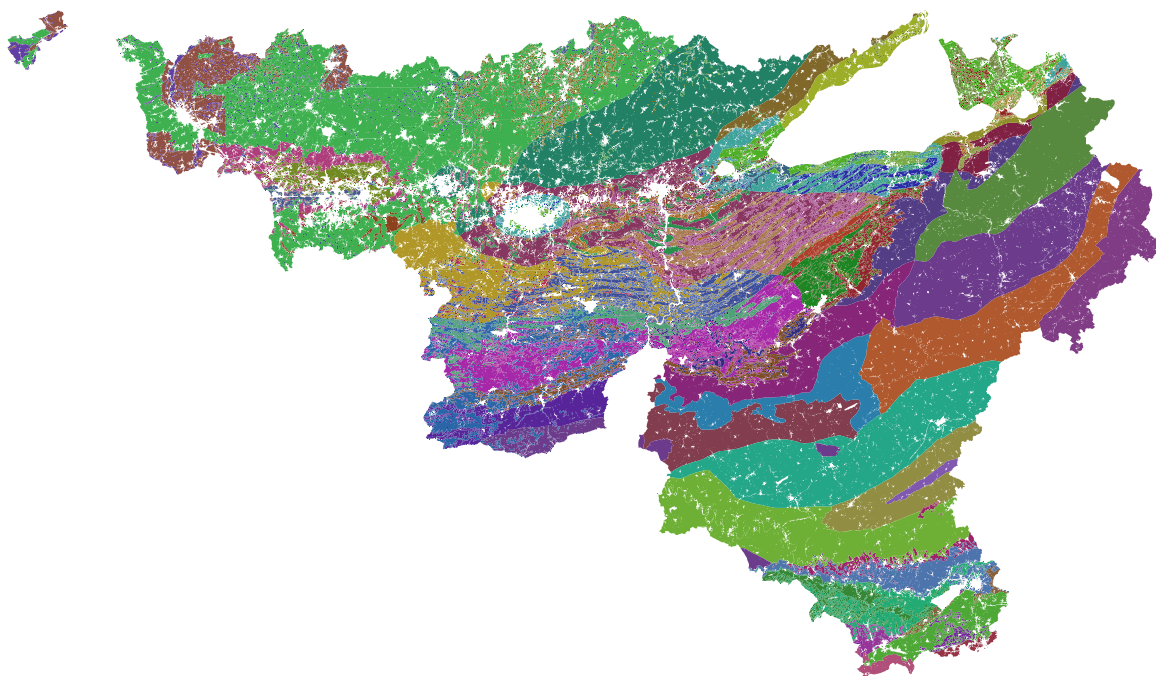


FIGURE 2.1 – Unités Spatiales de Comparaison (USC) pour le Zn (convention CAPASOL 3). Chaque couleur correspond à une USC. En blanc, les sols non classés en USC par manque d'information (ex : non cartographié dans la carte des sols de Wallonie) ou à cause de teneurs qui varient trop fort localement (ex : à proximité de la ville de Liège, à cause de retombées atmosphériques de proximité).

2.1.2 Résumé de la méthode

Etape 1 : biais temporels intralaboratoires : élimination des données biaisées

Lors de CAPASOL 3, un examen visuel des données de chaque laboratoire comparées année après année dans chacune des USC constituées lors du projet et pour lesquelles un nombre suffisant des données étaient disponibles avait permis de faire apparaître des biais temporels pour la plupart des laboratoires. Ces biais consistaient en des tendances monotones (croissantes/décroissantes année après année), des « sauts » (d'un groupe d'années à l'autre, une différence moyenne de quelques % était observée) ou des variations d'une année à l'autre sans qu'il soit possible de déterminer une tendance générale sur l'ensemble des années. Cette forte hétérogénéité de cas observés et le peu de données disponible par année/laboratoire/USC rendait difficile la mise en oeuvre d'une méthode systématique de corrections des biais temporels. La méthode retenue pour s'affranchir des biais temporels consistait à ne conserver les groupes d'années de laboratoires pour lesquelles les valeurs moyennes étaient semblables. Lorsqu'il n'était pas possible de comparer les données d'années en années, les données étaient toutes conservées. Cette méthode était donc entièrement basée sur le jugement d'expert, au cas par cas.

Au terme de cette étape, l'analyse des biais temporels intralaboratoire nous a permis de nous affranchir des biais temporels au sein des données produites par un même laboratoire.

Etape 2 : biais interlaboratoire : constitution de groupes de laboratoires non-biaisés

Cette étape a consisté à grouper les laboratoires qui ne présentaient pas de biais entre-eux. Le nombre de données par laboratoire et par USC était alors suffisant pour établir une méthode de regroupement systématique, qui avait été mise au point via une procédure complexe qui faisait intervenir une grande liste de critères (paramètres basés sur le jugement d'expert). Au terme de cette étape, des groupes de laboratoires considérés comme non biaisés entre eux ont été constitués.

Etape 3 : biais interlaboratoire : correction des biais interlaboratoires

L'étape suivante consiste à corriger les biais interlaboratoires entre groupes de laboratoires. Cette étape a été réalisée à l'aide d'une régression orthogonale robuste réalisée sur les médianes par USC de chacun des deux groupes de laboratoires. Les USC prises en compte sont celles qui comportent un nombre de données suffisant ($n > 10$ dans les deux groupes). Cette étape de correction de biais interlaboratoires est illustrée à la Figure 2.2. Les groupes de laboratoires 1-100-5-9-3 et 2-4 peuvent être corrigés l'un par rapport à l'autre en utilisant les paramètres de correction de la droite (pente : 0.88 et intercept : -3.9).

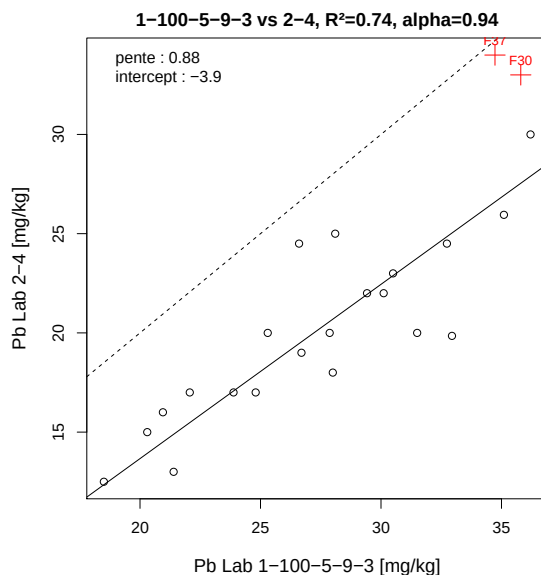


FIGURE 2.2 – Nuage de points des médianes de teneurs en Pb par USC des groupes de laboratoires « 1-100-5-9-3 » et « 2-4 ». Les paramètres des droites de régression orthogonale sont donnés en haut à gauche du graphique. Le coefficient de détermination robuste (R^2) entre les deux groupes de laboratoires est de 0.74. La méthode de régression robuste utilisée a estimé que les deux croix en rouge sont des outliers.

2.2 Méthode établie et appliquée lors de la convention CAPASOL 6

La méthode mise au point dans CAPASOL 6 est fondée sur les mêmes principes généraux que ceux établis lors de la convention CAPASOL 3 (*cf.* section 2.1.1). Elle comporte les mêmes trois étapes que dans CAPASOL 3, mais la façon de réaliser ces étapes a été améliorée. De nouvelles procédures statistiques adaptées pour la problématique ont notamment été utilisées afin de limiter l'emploi du jugement d'expert et d'objectiver au maximum les décisions qui sont prises sur les données. Notons que le recours au jugement d'expert reste toutefois indispensable dans les cas extrêmes où la configuration des données rend les résultats des tests statistiques discutables³. Les trois étapes de la méthode appliquée dans CAPASOL 6 sont expliquées dans les sections qui suivent, après la description de l'étape de préparation des données et un résumé des procédures statistiques appliquées pour comparer les jeux de données.

2.2.1 Préparation des jeux de données

Dans le cadre de ce travail, une « donnée géochimique » (*geochemical data*) correspond à une valeur de concentration chimique mesurée dans un matériau échantillonné et géoréférencé. Un « jeu de données » est un ensemble de données qui a été extrait d'une base de données géochimique dans un format « tableau de données ». Dans ce tableau, chaque ligne ou enregistrement (*record*) correspond à un échantillon qui a été analysé. Le tableau contient plusieurs colonnes, dont au minimum celles qui précisent la localisation géographique de l'échantillon et les valeurs analytiques.

Lorsque l'on souhaite réaliser une cartographie géochimique de la Wallonie, plusieurs bases de données de différentes origines coexistent (*cf.* section 1.3). Chaque base de données correspond à une façon particulière d'organiser les informations collectées. La préparation des jeux de données géochimiques consiste à extraire et organiser les informations contenues dans les bases de données pour constituer des jeux de données géochimiques remplissant les critères suivants. Tout d'abord, un jeu de données géochimique ne comprendra les valeurs de concentrations que pour un élément. Ensuite, un jeu de données géochimique sera constitué d'enregistrements répondant à plusieurs critères d'homogénéité et de similarité. Les valeurs de concentrations correspondront à des mesures réalisées selon un protocole analytique unique. Les enregistrements ne concerneront qu'un seul type de matériau (horizon de surface de sols) qui a été prélevé selon un protocole d'échantillonnage unique. Finalement, en cas de mesures effectuées sur de grandes périodes de temps (par exemple plusieurs années), chaque enregistrement devra inclure la date de l'analyse. Ces critères de similarité des enregistrements sont essentiels pour

3. Il arrive dans de rares cas que l'on ne rejette pas l'hypothèse d'équivalence alors que la p-valeur était inférieure à 0.05 ou inversement. Ceci est expliqué dans la section 2.2.2. Le recours au jugement d'expert s'avère parfois indispensable dans les cas extrêmes où la configuration des données rend les résultats des tests discutables. En effet, l'utilisation de procédures statistiques paramétriques, comme la régression BLS, requiert de remplir des conditions (par exemple : la normalité des données au sein de chaque USC) qui pourraient ne pas être satisfaites. De plus, la présence d'une valeur extrême/outlier dans une USC peut modifier grandement le nuage de point et la p-valeur du test BLS.

que l'on puisse appliquer les 3 étapes de la méthode CAPASOL 6 décrites dans les sections qui suivent.

2.2.2 Procédures statistiques appliquées pour comparer les jeux de données

Les procédures statistiques utilisées ont pour objectif principal l'évaluation de l'équivalence des jeux de données, ainsi que le nivellement des jeux de données présentant un biais. Des différences entre les jeux de données seront toujours présentes, notamment à cause de l'hétérogénéité naturelle du matériau d'échantillonnage ainsi que la variabilité analytique. On peut donc se poser la question suivante : à partir de quel niveau de différence les jeux de données doivent être considérés comme biaisés ? En pratique, le constat de biais entre jeux de données géochimiques existants se fait généralement par jugement d'expert basé sur la comparaison des données à partir de cartes, de graphiques ou de comparaison de quantiles. Dans CAPASOL 3, le constat de biais était réalisé par jugement d'expert à partir de la comparaison de médianes sur USC. Dans CAPASOL 6 le constat de biais est établi à l'aide des procédures statistiques décrites dans ce qui suit.

Les procédures statistiques utilisées pour la comparaison des jeux de données font l'hypothèse (« hypothèse nulle ») que les jeux de données sont équivalents. Cette hypothèse sera rejetée si la probabilité que cette hypothèse puisse être acceptée est trop faible (ici nous avons considéré une probabilité (p-valeur) de 0.05, ce qui signifie que l'hypothèse nulle à moins de 5% de chance d'être respectée). Rejeter l'hypothèse d'équivalence revient à mettre en évidence l'existence d'un biais entre le jeu de données (positif, négatif, proportionnel ou systématique, cf. Figure 2.3).

Les méthodes utilisées sont des méthodes adaptées au cas de jeux de données géochimiques couvrant un même périmètre géographique. Chaque méthode est basée sur une procédure statistique qui possède ses limitations propres.

En priorité, lorsque c'était possible, les jeux de données sont comparés par la « *méthode SCU*⁴ ». Cette méthode illustrée à la Figure 2.4 est basée sur la comparaison par régression BLS⁵ des moyennes des jeux de données sur des Unités Spatiales de Comparaison (cf. section 2.1.1). Si les jeux de données ne sont pas biaisés, les moyennes devraient être disposées autour de la droite 1 : 1. Le test associé à la régression BLS consiste à vérifier si l'intervalle de confiance autour des paramètres (pente et intercept) de la droite de régression inclut les valeurs de pente = 1 et d'intercept = 0. Si c'est le cas, la p-valeur du test est élevée et l'hypothèse de biais est rejetée. Nous avons estimé que cette méthode était applicable si les jeux de données pouvaient être comparés pour un minimum de 7 USC⁶ et si le coefficient de

4. SCU : Unités Spatiales de Comparaison en anglais, nom donné à la méthode décrite dans Pereira, B., Vandeuren, A., Govaerts, B. B., & Sonnet, Ph. (2016). Assessing dataset equivalence and leveling data in geochemical mapping. *Journal of Geochemical Exploration*, 168, 36-48.

5. BLS : *Bivariate Least Squares* : Francq, Bernard G and Govaerts, Bernadette B. Hyperbolic confidence bands of errors-in-variables regression lines applied to method comparison studies, *Journal de la Société Française de Statistique* (2014), 23-45.

6. Nous avons considéré ici qu'il est possible de comparer deux jeux de données sur une USC à condition que ces deux jeux de données possèdent au moins 5 données sur cette USC.

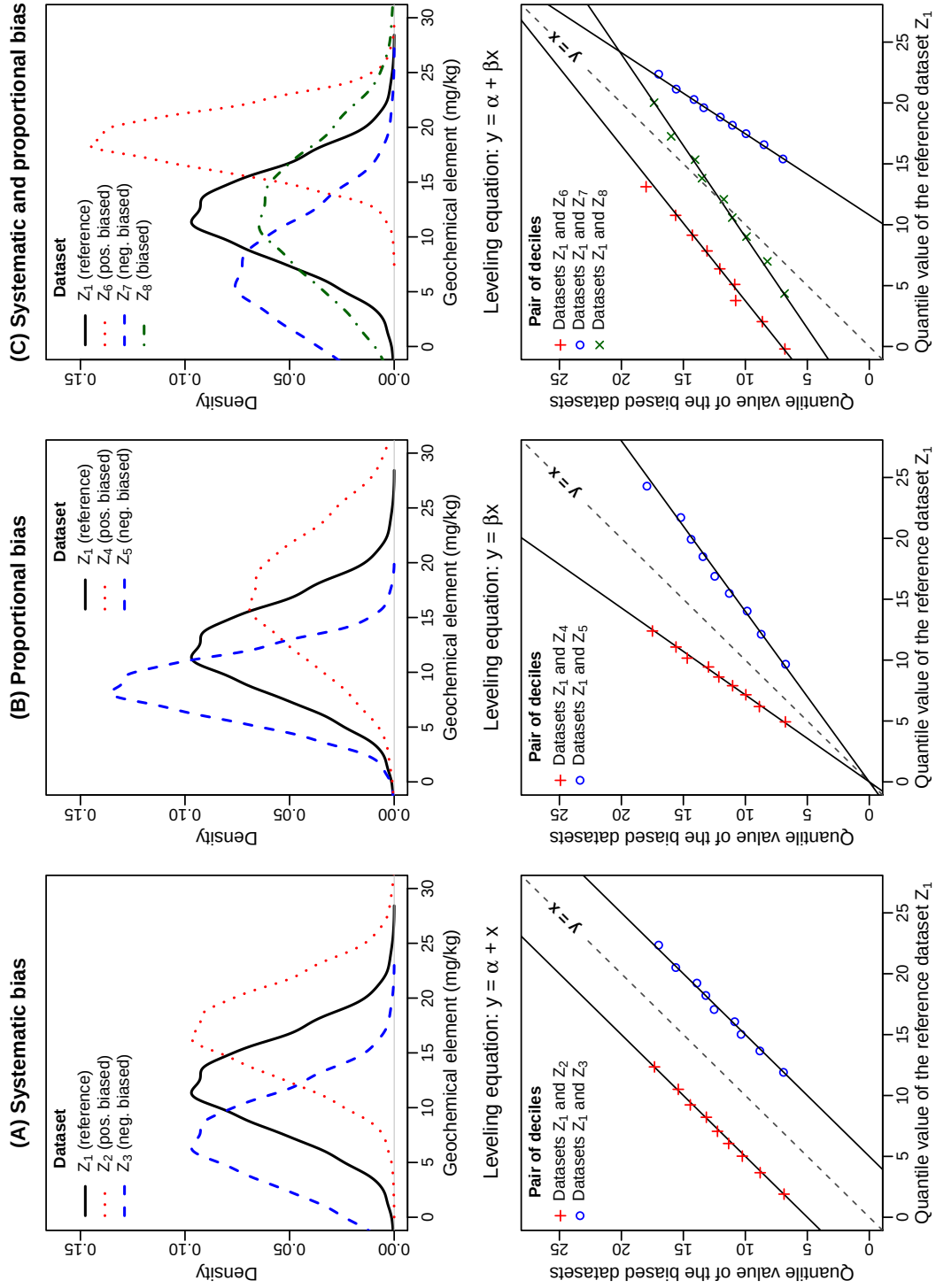


Figure 2.3 – Différents types de biais entre jeux de donnée géochimiques (voir Pereira B. et *al.*, 2016, JGE 168).

corrélation des moyennes des valeurs mesurées par USC était supérieur à 0.5. Si la p-valeur était inférieure à 0.05, les jeux de données étaient considérés comme biaisés (non équivalents).

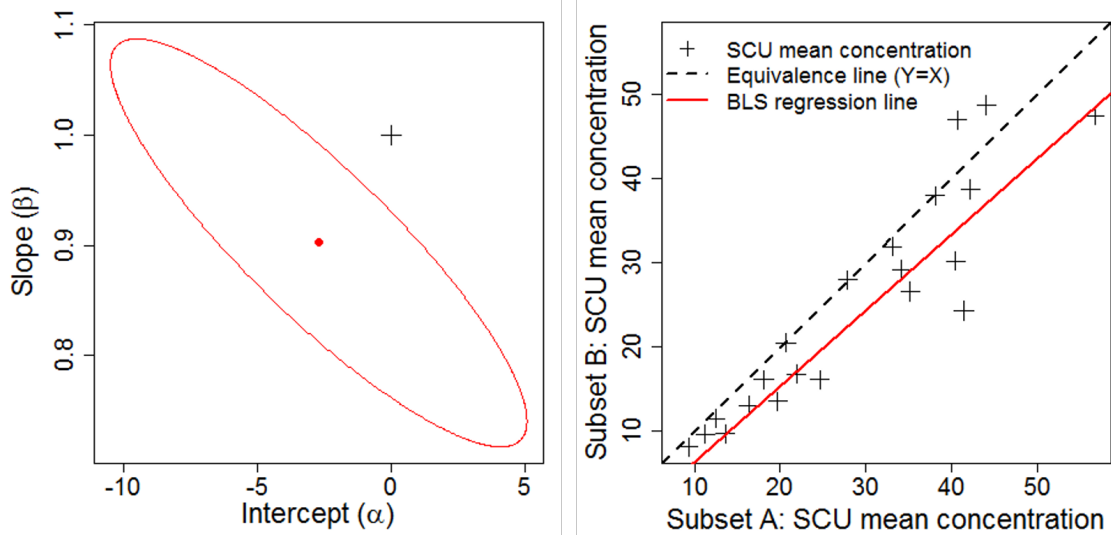


FIGURE 2.4 – Comparaison des valeurs mesurées dans les jeux de données (subset) A et B. Diagramme de gauche : la droite d'équivalence (croix noire) est comprise à l'extérieur de l'intervalle de confiance à 95% (ellipse rouge) : les deux jeux de données ne peuvent pas être considérés comme équivalents. Diagramme de droite : les croix représentent les moyennes des valeurs mesurées par USC. La droite de régression BLS (en rouge) tient compte de l'incertitude sur les moyennes des valeurs mesurées par USC (croix).

Le résultat de la comparaison des jeux de données par la méthode SCU et la droite de nivellement calculée par la régression sont ensuite validés par un examen visuel⁷. Cet examen visuel se base sur les recommandations pratiques pour l'usage de la méthode SCU établie sur base de la réalisation de simulations stochastiques⁸. Certains types de biais entre des jeux de données sont difficiles à niveler, et dans ces situations la transformation linéaire du jeu de données n'est pas efficace pour réduire l'ampleur du biais entre les jeux de données. Le premier type de « biais problématique » correspond aux situations où l'ampleur du biais entre les jeux de données est très importante. Graphiquement, l'ampleur du biais peut être représenté comme la distance entre les coordonnées des droites d'équivalence et de régression BLS sur le graphique de gauche de la Figure 2.4. Cette recommandation s'est traduit dans le cadre de CAPASOL 6 par le choix de ne pas corriger/niveler un jeu de données lorsque cela nécessite d'appliquer une transformation linéaire comprenant une pente très élevée ou très faible (supérieure à 2 ou inférieure à 0.5), ou un intercept très élevé ou très faible (supérieure ou inférieure à la moyenne du jeu de données). Le second type de biais problématique entre des jeux de données correspond

7. Il arrive dans de rares cas que l'on ne rejette pas l'hypothèse d'équivalence alors que la p-valeur était inférieure à 0.05 ou inversement. Comme indiqué précédemment, le recours au jugement d'expert s'avère parfois indispensable dans les cas extrêmes où la configuration des données rend les résultats des tests discutables. Par exemple, la présence d'une valeur extrême/outlier dans une USC peut modifier grandement le nuage de point et donc la p-valeur du test BLS.

8. Pereira, B. (2016). Homogenization of data : issues and methods in soil geochemical mapping (PhD dissertation), Université catholique de Louvain, Louvain-la-Neuve.

aux situations où le biais est faible, et qu'il est à la fois systématique et proportionnel (pente différente de 1 et intercept différent de 0). Les simulations stochastiques ont aussi montré que le test associé à la régression BLS pouvait manquer de performance en cas de faible biais entre des jeux de données. Le test a en effet tendance à indiquer un biais significatif entre deux jeux de données générés aléatoirement de façon équivalente. La génération aléatoire de données peut produire deux jeux de données présentant de faibles différences, différences qui modélisent une situation réelle étant donné la variabilité due à l'échantillonnage et à l'analyse. Ce constat s'est traduit dans le cadre de CAPASOL 6 par l'utilisation de jugement d'expert (notamment par l'inspection graphique) en cas de faible biais entre des jeux de données. Par exemple, dans certaines situations, le résultat du test associé à la régression BLS n'aura pas été suivi, ou encore il aura été choisi de ne pas corriger/niveler un jeu de données lorsque l'on est en présence de ce type de biais.

Lorsque la méthode SCU n'était pas applicable pour la comparaison des jeux de données, nous avons comparé les données d'une même USC par le test de Wilcoxon. Cette méthode a été appliquée pour autant qu'il existe une USC pour laquelle chacun des deux jeux de données comporte au moins 12 données. Cette procédure statistique non paramétrique est couramment utilisée pour déterminer si deux jeux de données possèdent des moyennes (médianes) équivalentes. En appliquant le test de Wilcoxon, nous faisons l'hypothèse que les données des deux jeux de données sur une même SCU proviennent d'une même distribution théorique (sans faire d'hypothèses sur les propriétés de celle-ci). A nouveau, si la p-valeur du test est inférieure à 0.05, les jeux de données étaient considérés comme biaisés (non équivalents). Le résultat de la comparaison des jeux de données est ensuite validé par un examen visuel (comparaison par diagramme « boxplot »).

2.2.3 Etape 1 : Traitement des biais temporels intralaboratoire de la BD CAPASOL

Dans la méthode appliquée dans CAPASOL 3, les données géochimiques de la BD CAPASOL produites par un même laboratoire ont été traitées ensemble, quelle que soit l'année de la mesure (production de la donnée). L'étude des biais temporels intralaboratoires a simplement consisté à éliminer les années de mesures pour lesquelles les valeurs moyennes étaient dissemblables de la majorité des données du même laboratoire. Dans CAPASOL 6, nous avons considéré chaque année de mesures réalisées par un même laboratoire comme étant un jeu de données unique (*cf.* section 2.2.1). Chaque jeu de données correspond à une année de mesure d'un même laboratoire, de telle sorte que la base de données CAPASOL comporte 76 jeux de données⁹. Notons que cette étape n'a pas été appliquée pour les autres jeux de données utilisés pour la cartographie des concentrations de fond (listés dans la Table 1.3). En effet, ces jeux de données ont été constitués dans un intervalle de temps restreint et/ou par des équipes universitaires qui ont mis en place des procédures de qualité pour éviter les problèmes de biais temporels.

9. Le nombre 76 correspond au nombre de jeux de données (année-laboratoire) de la base de données CAPASOL qui comporte au minimum 50 données. En effet, les jeux de données de moins de 50 données n'ont pas été pris en compte, car la petite taille du jeu de données rend impossible de tester l'éventuelle présence de biais temporel.

Pour chacun des 15 laboratoires de la base de données CAPASOL, pour chacun des sept ETM (Cd, Cr, Cu, Hg, Ni, Pb, Zn) mesurés, les jeux de données (une année de mesure d'un ETM) ont été comparés deux à deux sur base des tests de comparaison des jeux de données (cf. section 2.2.2). Lorsque c'était possible, le regroupement des années consécutives a été réalisé en priorité. On part du principe que les données obtenues lors de deux années consécutives ont davantage de chance d'avoir été obtenues par un même protocole/opérateur que lors de deux années non consécutives. Ensuite, lorsqu'il n'y a plus d'années consécutives regroupables, les années (ou groupes d'années) semblables sont regroupées, même si elles ne sont pas consécutives dans le temps. Le processus est répété en boucle jusqu'à ce qu'il n'y ait plus de jeux de données regroupables.

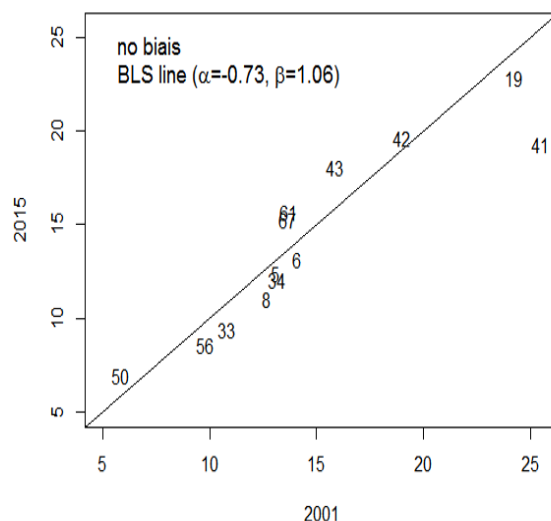


FIGURE 2.5 – Comparaison des données de Cu de deux jeux de données (années 2001 et 2015 du laboratoire 3 de la base de données CAPASOL) par la méthode SCU. Les axes correspondent à des concentrations mesurées, en mg/kg. Les numéros représentent les moyennes des valeurs mesurées par USC. La droite de régression BLS (en noir) tient compte de l'incertitude sur les moyennes des données par USC. Dans cet exemple, les paramètres de la droite de régression BLS (pente = 1.06 et intercept = -0.73) ne sont pas statistiquement différents de ceux de la droite d'équivalence (pente = 1 et intercept = 0).

2.2.4 Etape 2 : Constitution de groupes de jeux de données non-biaisés

Cette étape de regroupement de jeux de données non-biaisés a pour objectif de constituer des groupes que l'on corrige entre eux, de la même manière que ce qui a été fait lors de CAPASOL 3. L'étape de regroupement est nécessaire car elle permet d'une part de limiter la transformation (nivellement) des données et d'autre part, d'augmenter le nombre de données par groupe pour obtenir des équations de nivellement plus fiables car établies à partir d'un plus grand nombre de données. Dans cette étape, les jeux de données (constitués à l'étape précédente pour ceux correspondant à la base de données CAPASOL, ainsi que les autres jeux de données listés à la Table 1.3) ont été comparés deux à deux à l'aide des tests de

comparaison des jeux de données décrits à la section 2.2.2 (un exemple est donné à la Figure 2.6). Dans un premier temps, nous avons comparé entre eux les (groupes de) jeux de données de la base de données CAPASOL. Dans un second temps, nous avons comparé entre eux les (groupes de) jeux de données de la base de données CAPASOL. Le processus est répété en boucle jusqu'à ce qu'il n'y ait plus de jeux de données regroupables. Au terme de cette étape, des groupes de jeux de données considérés comme non biaisés entre eux ont été constitués.

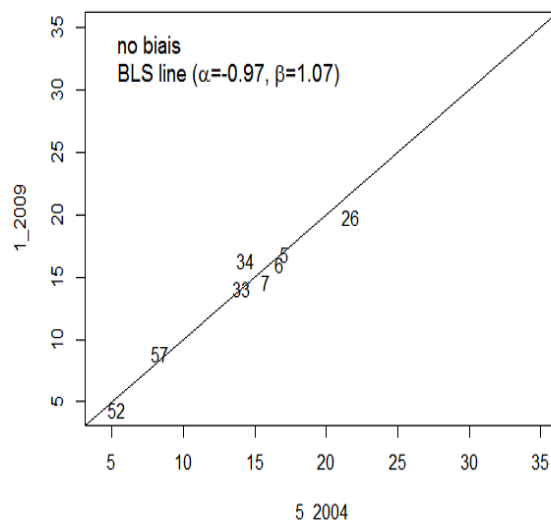


FIGURE 2.6 – Comparaison des données du Ni de deux jeux de données (correspondant aux codes « 5_2004 » et « 1_2009 ») par la méthode SCU. Les numéros représentent les moyennes des valeurs mesurées par USC. La droite de régression BLS (en noir) tient compte de l'incertitude sur les moyennes des données par USC. Dans cet exemple, les paramètres de la droite de régression BLS (pente = 1.07 et intercept = -0.97) ne sont pas statistiquement différents de ceux de la droite d'équivalence (pente = 1 et intercept = 0).

2.2.5 Etape 3 : Correction des biais entre les (groupes de) jeux de données biaisés

Cette étape de correction (ou nivellement / « leveling » en anglais) des biais entre les groupes de jeux de données non biaisés obtenus au terme des processus de regroupement des étapes précédentes a pour objectif la constitution d'un jeu de données unique et homogène pour la cartographie. Dans CAPASOL 6, lorsqu'un biais est mis en évidence par la méthode SCU, les paramètres de la droite de régression BLS sont utilisés pour le nivellement des jeux de données (Figure 2.7).

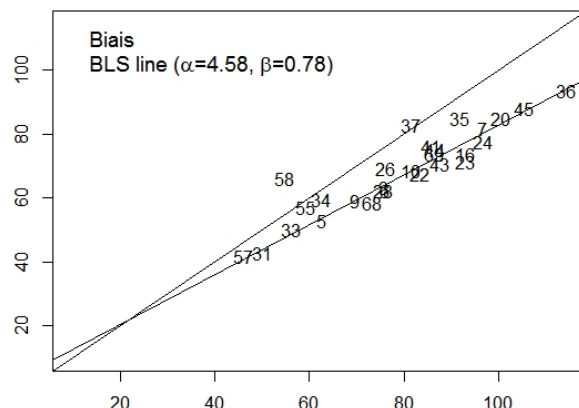


FIGURE 2.7 – Illustration de la comparaison des données du Zn de deux groupes de jeux de données par la méthode SCU. Les numéros représentent les moyennes des valeurs mesurées par USC. La droite de régression BLS tient compte de l'incertitude sur les moyennes des données par USC. Dans cet exemple, les paramètres de la droite de régression BLS (pente = 0.78 et intercept = 4.58) sont statistiquement différents de ceux de la droite d'équivalence (pente = 1 et intercept = 0, droite qui va du coin supérieur gauche au coin inférieur droit de la figure) et ont été utilisés pour le nivellement des données.

En cas de biais entre jeux de données géochimiques, il n'est généralement pas possible de déterminer le jeu de données présentant des valeurs « vraies » ou celui qui serait affecté par des « erreurs systématiques » (constantes ou proportionnelles à la valeur mesurée). Des différences, même considérées comme minimales, dans le matériel d'échantillonnage, la méthode d'échantillonnage, la préparation des échantillons ou dans les protocoles analytiques peuvent avoir un impact majeur sur les valeurs de concentrations observées dans un jeu de données¹⁰. Dans CAPASOL 6, il nous est donc impossible de nous prononcer sur le groupe de jeux de données présentant des valeurs « vraies » ou sur celui qui serait affecté par des « erreurs systématiques ». Cependant, le choix d'un jeu de données doit obligatoirement être réalisé : constituer un jeu de données unique et homogène pour la cartographie requiert de retirer les biais des jeux de données affectés par des erreurs systématiques. Le choix a été fait ici de considérer le plus grand (groupe de) jeu(x) de données comme celui comportant les valeurs « vraies ». Ce choix a été effectué afin de corriger par transformation linéaire un minimum de données, ce qui permet de conserver un maximum des données originales produites par des laboratoires.

La méthode SCU ne pouvait pas toujours s'appliquer pour le nivellement des jeux de données (pour différentes raisons, voir section 2.2.2 pour plus de détails à ce sujet). Lorsque le jeu de données provient de la base de données CAPASOL, celui-ci a été écarté et donc n'a pas été pris en compte pour la cartographie. Pour les autres jeux de données, lorsque les jeux de données ne pouvaient pas être comparés à l'aide des méthodes décrites à la section 2.2.2, il a généralement été pris la décision de ne pas les corriger, c'est-à-dire de les considérer comme « non biaisés ». Cette décision a été prise étant donné l'importance de ces jeux de données

10. C. Reimann & P. de Caritat, New soil composition data for Europe and Australia (...), Science of The Total Environment (2012), 239-252.

pour la cartographie (ex : données d’ALTERASOL, localisée dans une région non couverte par tous les autres jeux de données disponibles) et la bonne qualité présumée du jeu de données.

La Table 2.1 présente le nombre de données de la base de données CAPASOL disponibles pour la cartographie après les méthodes établies et appliquées lors de la convention CAPASOL3 et CAPASOL 6. Pour la plupart des éléments, il y a une différence importante du nombre de données rendues disponibles pour la cartographie. Cette différence n’est pas seulement expliquée par la taille plus importante de la base de données CAPASOL lors de CAPASOL 6 : elle s’explique aussi par la différence entre les procédures appliquées. Par exemple, lors de CAPASOL 3, il avait été fait le choix d’écarter les années de mesures qui ne ressemblaient pas au groupe d’année majoritaire pour un même laboratoire (*cf.* section 2.1.2). Dans CAPASOL 6, nous avons fait le choix de corriger les biais intralaboratoires, ce qui permet de conserver un plus grand nombre de données. Notons que pour le Hg, c’est l’inverse qui est constaté : il y a moins de données CAPASOL disponibles pour la cartographie dans CAPASOL 6. Ceci s’explique par le fait que les biais temporels intra- et interlaboratoires n’ont presque jamais pu être corrigés, car souvent il y avait une absence de corrélation linéaire entre les jeux de données. D’autre part, les décisions qui ont été prises par jugement d’expert lors de CAPASOL 3 ont été moins sévères que celles qui s’appuient sur les tests statistiques utilisés dans CAPASOL 6.

TABLE 2.1 – Nombre de données de la base de données CAPASOL disponibles pour la cartographie après les méthodes établies et appliquées lors de la convention CAPASOL 3 (colonne de gauche) et CAPASOL 6 (colonne de droite).

	CAPASOL 3	CAPASOL 6
Cd	6902	8194
Cr	7198	17 859
Cu	12 452	18 315
Hg	7322	6660
Ni	8751	20 418
Pb	11 289	18 290
Zn	12 787	19 651

2.3 Constitution des jeux de données « de référence »

2.3.1 Principe général

Dans ce qui suit, nous appellerons jeu de données « de référence » celui que l’on considérera comme présentant des valeurs « vraies ». Dans CAPASOL 6, deux jeux de données ont été constitués : un jeu de données correspondant au mode de gestion agricole, et un jeu de données correspondant au mode de gestion forestier (*cf.* section 4.1). Il y aura donc un jeu de données de référence par élément et par mode de gestion du sol. Comme pour dans l’étape de correction des biais entre les jeux de données décrite à la section précédente, le choix a été fait ici de considérer le plus grand (groupe de) jeux de données comme celui comportant les valeurs « vraies » afin de conserver un maximum des données originales produites par des laboratoires. En conséquence, en mode de gestion agricole le jeu de données de référence est constitué d’un

jeu de données provenant de la base de données CAPASOL, tandis qu'en mode de gestion forestier le jeu de données de référence est celui provenant de l'inventaire forestier¹¹.

Afin d'obtenir le plus grand jeu de données possible pour la cartographie des concentrations de fond, des droites de nivellement de l'effet « mode de gestion » ont été calculées à l'aide des méthodes présentées à la section 2.1.1. Lorsque le nivellement des données était possible (voir conditions détaillées à la section 2.1.1), les données ont été transformées afin de correspondre au mode de gestion recherché. Pour As et Cd, la comparaison des données selon le mode de gestion du sol n'a pas montré de différences significatives : un seul jeu de données pour la cartographie sera donc constitué pour ces deux éléments.

2.3.2 Statistiques descriptives des jeux de données disponibles pour la cartographie des concentrations de fond

Les Tables 2.2 et 2.3 présentent les statistiques descriptives des jeux de données disponibles pour la cartographie des concentrations de fond après les méthodes établies et appliquée lors de la convention CAPASOL 6.

Pour ce qui est du nombre de données disponibles (première ligne des Tables 2.2 et 2.3), nous pouvons observer que le jeu de données disponibles pour la cartographie de l'As est le plus petit. Ceci s'explique par le fait que l'As ne fait pas partie des éléments contenus dans la base de données CAPASOL ni dans l'inventaire forestier, qui sont les deux plus grandes sources de données disponibles pour la cartographie de concentrations de fond en Wallonie. Le jeu de données disponibles pour la cartographie du Pb dans les sols forestiers est le second plus petit : en effet, les jeux de données provenant de la base de données CAPASOL n'ont pas pu être corrigés/nivelés. Le jeu de données est donc essentiellement constitué de données provenant de l'inventaire forestier ainsi que des projets POLLUSOL 1 et 2.

Pour ce qui est des caractéristiques des distributions, une grande différence entre la moyenne et la médiane des jeux de données peut être remarquée (2e et 9e lignes des Tables 2.2 et 2.3). Les distributions sont aussi asymétriques : ceci s'observe par la différence plus importante entre le centile 75 et la médiane qu'entre le centile 25 et la médiane (voir les lignes 8 à 10 des Tables 2.2 et 2.3). Ces constats s'expliquent par les caractéristiques courantes des jeux de données géochimiques : distributions log-normales et présence de valeurs aberrantes (outliers).

La comparaison des caractéristiques des distributions en mode de gestion « forestier » avec le mode de gestion « agricole » permet d'observer que les sols forestiers ont généralement des teneurs plus élevées en Hg et Pb tandis que les sols agricoles présentent des teneurs plus élevées pour Cr, Cu, Ni et Zn. Ces constats sont en adéquations avec les tendances observées lors des projets POLLUSOL 1 et 2.

11. Alderweireld M., Burnay F., Pitchugin M., Lecomte H., 2015. Inventaire Forestier Wallon. Résultats 1994 – 2012. SPW, DGO3, DNF, Direction des Ressources forestières, Jambes, 236 p.

TABLE 2.2 – Statistiques descriptives des jeux de données disponibles pour la cartographie des concentrations de fond des sols agricoles après les méthodes établies et appliquées lors de la convention CAPASOL 6 (pour As et Cd, c’est le même jeu de données pour les sols forestiers, *cf.* texte).

	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
n	828	9441	19655	20127	7190	22212	19045	21480
moy	14.9	0.54	37.7	16.9	0.09	25.8	29.9	101.2
sd	23.5	2.97	17.2	24.4	0.04	172.4	34.1	422.5
mad	4.4	0.22	12.5	5.1	0.03	11.4	6.8	29.8
min	1.8	0.10	2.0	1.0	0.05	1.0	1.0	2.0
c5	6.1	0.10	18.6	8.7	0.05	9.0	16.5	41.1
c10	7.0	0.10	21.6	10.0	0.05	11.3	18.1	47.0
c25	7.8	0.12	26.3	12.0	0.07	15.5	21.1	56.2
med	10.4	0.25	33.1	15.0	0.08	21.5	24.9	75.0
c75	14.5	0.43	45.1	19.3	0.10	33.1	31.0	98.5
c90	20.9	0.79	58.2	23.7	0.12	42.0	39.2	127.3
c95	35.8	1.31	74.2	27.2	0.14	46.4	50.3	164.7
max	397.1	158.00	228.2	2234.3	0.84	25650.0	2031.9	25364.3

TABLE 2.3 – Statistiques descriptives des jeux de données disponibles pour la cartographie des concentrations de fond des sols forestiers après les méthodes établies et appliquées lors de la convention CAPASOL 6 .

	Cr	Cu	Hg	Ni	Pb	Zn
n	19506	19978	7190	22063	1682	21331
moy	29.6	13.1	0.17	21.3	80.3	82.1
sd	16.5	23.0	0.06	169.5	176.7	378.0
mad	11.9	5.4	0.04	11.3	20.5	29.6
min	2.0	1.0	0.05	1.0	1.0	2.0
c5	11.1	4.3	0.10	4.8	28.6	23.9
c10	14.2	5.7	0.11	7.0	32.0	29.9
c25	18.8	7.8	0.13	11.1	41.1	39.1
med	25.3	11.0	0.16	17.1	52.0	57.7
c75	36.7	15.7	0.19	28.5	71.1	81.1
c90	49.3	20.4	0.22	37.1	108.9	108.7
c95	64.4	24.3	0.26	41.4	179.2	142.3
max	212.3	1903.0	1.42	25132.9	3285.7	21900.0

Chapitre 3

Cartographie de l'impact des retombées atmosphériques de polluants

3.1 Contexte du travail et méthodologie

Un modèle estimant l'impact des retombées atmosphériques de polluants sur les sols de Wallonie a été mis au point dans le projet POLLUSOL 2 ¹. Le modèle a été établi grâce à une méthodologie en quatre étapes :

1. Sélection de 125 cheminées parmi les cheminées figurant dans la base de données « sources historiques de pollution » de SPAQuE et détermination des caractéristiques de l'auréole de retombées atmosphériques autour de chacune des 125 cheminées prise isolément (voir section 3.2) ;
2. Calcul de l'enrichissement du sol attribuable aux retombées atmosphériques, et ceci pour le plus grand nombre possible de points de contrôle (analyses géoréférencées d'horizon de surface de sol) ;
3. Estimation de la contribution de chacune des cheminées à l'enrichissement calculé aux points de contrôle. Cette contribution (coefficient de pondération) a été calculée pour chaque cheminée et chaque polluant au moyen de régressions modélisant les enrichissements en polluants aux points de contrôle à l'aide d'une combinaison linéaire de l'ensemble des auréoles de retombées ;
4. Pour chaque paramètre, calcul de l'enrichissement dû aux retombées atmosphériques en tout point de la Région wallonne par addition de l'ensemble des auréoles de pollution des cheminées en leur appliquant les coefficients de pondération.

Des cartes de l'impact des retombées atmosphériques sur les sols wallons ont été dressées pour 5 ETM : le zinc, le plomb, le cadmium, le cuivre et le mercure. Il a été démontré dans

1. Pereira, Benoît ; Titeux, Hugues ; Schneider, Arnaud ; Sonnet, Philippe ; et. al. Rapport final du projet POLLUSOL 2 - partie "sols". (2012) 1126 pages

la convention CAPASOL 4 que ces cartes présentaient des faiblesses importantes, car dans certaines zones de la Wallonie, le nombre d'analyses de sol disponibles (points de contrôle) n'était pas suffisant pour l'établissement d'un modèle fiable dans ces zones. C'est pourquoi ces zones ont fait l'objet d'un plan d'échantillonnage et d'analyse au cours de la présente convention (*cf.* section 1.2). Une des tâches de la présente convention est de préciser le modèle de retombées atmosphériques de POLLUSOL 2 à l'aide de ces nouvelles données. La même méthodologie en quatre étapes que celle établie lors de POLLUSOL 2 a été utilisée. Pour l'étape 1, nous avons repris les résultats de POLLUSOL 2 à l'identique, tandis que les étapes 2 à 4 ont fait l'objet d'importantes adaptations/améliorations. Ces différentes étapes sont décrites dans les sections qui suivent.

3.2 Sélection de cheminées et calcul des « auréoles de retombées unitaire »

Les cheminées à l'origine des retombées atmosphériques de polluants dans les sols considérées dans cette étude sont celles comprises dans la base de données « Sources Historiques de Pollution » de SPAQuE. Nous appelons « auréole de retombées unitaire » la matrice reprenant en tous points de la Région wallonne les valeurs de retombées attribuées à une cheminée particulière ayant émis 100t de particules par an durant 1 an. Le calcul de ces auréoles a été réalisé par A. Marijns et C. Passelecq (UMons) dans le cadre du projet POLLUSOL 2. Le temps de calcul pour la détermination d'une auréole pour une cheminée est relativement conséquent (de l'ordre d'une journée). Il n'a dès lors pas été possible de déterminer les auréoles de retombées unitaires pour l'ensemble des cheminées des sources de la base de données « Sources Historiques de Pollution » de SPAQuE (leur nombre dépasse 700). Il a donc été nécessaire d'effectuer une sélection parmi ces sources afin de ne s'intéresser qu'aux principales responsables des retombées d'ETM. A partir de statistiques de comparaison entre zones (corrélations entre les teneurs en polluants et le nombre d'industries de chaque type à proximité), deux catégories de sources ont été sélectionnées dans la base de données « Sources Historiques de Pollution » : les industries de la métallurgie des non-ferreux et les industries sidérurgiques. De cette façon, 120 industries/cheminées ont été sélectionnées : il s'agit de 28 cheminées de métallurgie des non-ferreux et de 92 cheminées de l'industrie sidérurgique. De plus, au cours des travaux de POLLUSOL 2, cinq cheminées supplémentaires liées à l'industrie de la transformation du charbon ont été sélectionnées afin de couvrir des zones où les résultats n'étaient pas satisfaisants (Verviers et est de Charleroi). Au total, 125 « auréoles de retombées unitaires » ont été modélisées, correspondant à 125 cheminées ayant émis une quantité identique (hypothèse de départ) de 100 tonnes de particules durant un an.

3.3 Calcul de l'enrichissement des sols en polluants par voie atmosphérique

3.3.1 Principe

Les « teneurs naturelles » d'un élément sont définies ici comme celles résultant de l'héritage du matériau parental, de processus pédogénétiques et de retombées atmosphériques naturelles, en dehors de tout apport d'origine anthropique. Les « teneurs habituelles » sont les teneurs habituellement rencontrées dans un sol. Dans le cas des métaux et métalloïdes, les teneurs habituelles résultent du fond pédogéochimique naturel (teneurs naturelles) et des contaminations anthropiques (contamination diffuse ou contaminations atmosphériques de proximité). L'enrichissement lié aux retombées atmosphériques sera calculé par différence entre la teneur habituelle mesurée en un point et la teneur naturelle attendue en ce point.

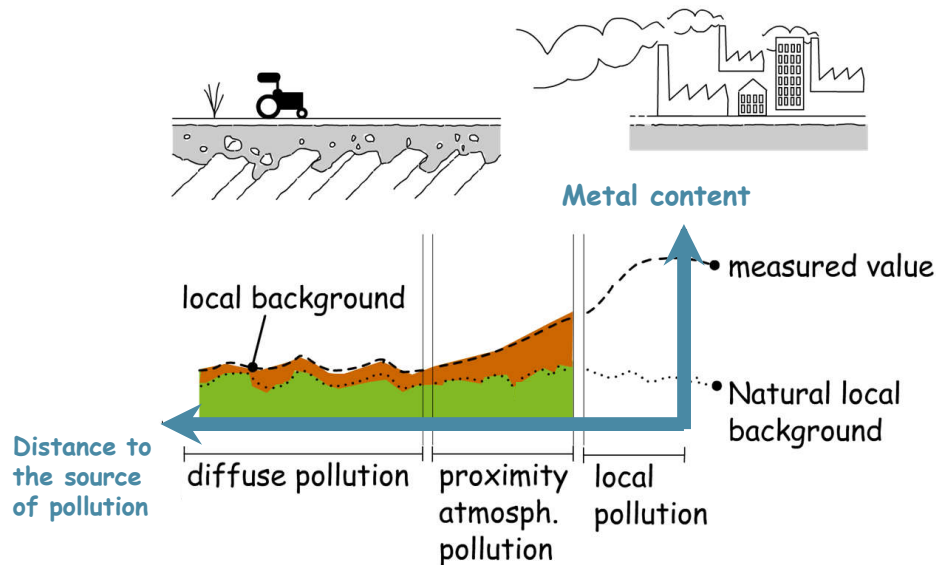


FIGURE 3.1 – Principe général du calcul de l'enrichissement des sols en polluants par voie atmosphérique. En dehors de la pollution locale, les teneurs mesurées (teneurs habituelles) résultent du fond pédogéochimique naturel local (teneurs naturelles), des contaminations diffuses et des contaminations atmosphériques de proximité. L'enrichissement des sols en polluants (représenté en brun) est la différence entre la teneur mesurée et la teneur naturelle (représentée en vert).

Prédiction des teneurs naturelles hors zone de pollution atmosphérique de proximité Pour le Hg, la teneur dans l'horizon de surface est directement considérée comme « l'enrichissement lié aux retombées atmosphériques », car en Wallonie, les teneurs dans les horizons profonds sont presque toujours inférieures aux limites de quantification. Pour le Cd, le constat est similaire au Hg, à l'exception des matériaux parentaux calcaires qui contiennent naturellement du Cd. Pour Cu, Pb et Zn, les teneurs naturelles ont été estimées à l'aide d'un modèle linéaire par la méthodologie suivante :

1. Sélection de profils de sols wallons situés à l'écart de toutes sources de pollution et non impactés par une anomalie géochimique. Pour effectuer cette sélection, nous nous sommes basés sur les résultats de l'étude POLLUSOL1.
2. Dans les profils de sols wallons sélectionnés à l'étape 1, sélection des horizons de sols situés à une profondeur supérieure à 40cm (afin de s'assurer que les horizons de sols n'ont pas été impactés par la pollution). Un jeu de données de 148 horizons de sols profonds prélevés dans 132 profils de sols wallons a ainsi été constitué. Dans ces 148 horizons de sols, les teneurs en Cu, Pb et Zn seront considérées comme « naturelles » ou non impactées par les apports anthropiques.
3. Pour chaque élément (Cu, Pb et Zn), sélection des meilleures variables prédictives des teneurs naturelles à l'aide de la procédure « regsubsets » du package « leaps » du logiciel R. Les variables prédictives étudiées sont les fractions granulométriques (% argile-sable-limon), les teneurs en Cr, Ni, Fe, Al, Mg, Ti, Zr. Ces variables, mesurées dans l'étude POLLUSOL1, ont été choisies car elles sont considérées comme indépendantes de la pollution dans les horizons de surface des sols.
4. Pour chaque élément, établissement d'un modèle de régression linéaire afin de prédire les teneurs naturelles à l'aide de la ou les variables explicatives sélectionnées à l'étape précédente. Cette étape a été réalisée par une méthode robuste à l'aide de la procédure « lmrob » du package « robustbase » du logiciel R.

Les résultats des modèles linéaires sont repris à la Table 3.1. Les coefficients de détermination (R^2) qui vont de 80% (Pb) à 90% (Cu) indiquent que les modèles linéaires sont très performants pour prédire les teneurs naturelles.

TABLE 3.1 – Modélisation des teneurs naturelles

	Cu (log)	Pb (log)	Zn (log)
n	140	140	144
R²	0.90	0.80	0.86
Constante	-0.61	-8.16	1.88
log(Fe)	/	1.35	/
log(Cr)	/	-0.918	/
log(Ni)	0.975	/	0.634

3.3.2 Calcul de l'enrichissement en polluant d'origine atmosphérique

L'enrichissement est calculé par différence entre les teneurs mesurées en surface et les valeurs naturelles attendues d'après les modèles sélectionnés ci-dessus (Table 3.1). Les enrichissements ont été calculés sur tous les points pour lesquels une analyse de sols était disponible, à l'exception des points qui répondaient à un des trois critères suivants :

1. les points situés sur des groupes de sol qui n'ont pas été prises en compte dans le travail de modélisation des teneurs naturelles. Il s'agit des points situés dans un groupe de sol qui n'est représentée par aucun des 132 profils de sols wallons utilisés pour l'établissement des modèles de régression linéaire ;
2. les profils de sols localisés sur une occupation autre qu'agricole, pour pouvoir négliger l'effet de l'occupation du sol² sur les teneurs mesurées en surface ;
3. les remblais, car ils sont liés par définition à un déplacement de sol, ce qui les rend inutilisables comme points de calibration spatialisés.

Pour le Cd et le Hg, le premier critère n'a pas été appliqué, étant donné que l'enrichissement du sol n'est pas calculé sur base de la modélisation des teneurs naturelles. En effet, dans les horizons profonds, ces éléments sont généralement situés à des concentrations inférieures aux limites de quantification, à l'exception des matériaux parentaux calcaires contenant naturellement du Cd. Les points situés sur des familles de sol avec des matériaux parentaux calcaires n'ont donc pas été pris en compte pour le Cd.

Pour le Cu et le Zn, en plus de ces trois critères, nous avons exclu les points pour lesquels nous ne possédions pas une analyse du Ni. En effet, la teneur en Ni est une des variables explicatives des teneurs naturelles en Cu et Zn.

Pour le Pb, de la même manière, nous avons exclu les points pour lesquels nous ne possédions pas une analyse de Cr. Pour la mesure du Fe, nécessaire lui aussi à la prédiction de la teneur naturelle en Pb, nous avons pris la moyenne des teneurs en Fe du groupe de sol correspondant, sur base des valeurs mesurées dans POLLUSOL 1.

2. L'effet de l'occupation du sol sur les teneurs observées en surface s'explique de nombreuses façons. Par exemple, des teneurs plus élevées dans les sols de forêts peuvent être liées à un potentiel d'interception plus important pour les polluants apportés via les poussières atmosphériques, ou à leur teneur en matière organique plus élevée, procurant une capacité de rétention plus élevée pour les polluants qui ont une grande affinité pour la matière organique. Pour les sols de potagers, cet effet peut être lié à des pratiques spécifiques aux potagers (apport d'ammendement, *etc.*), mais peuvent aussi s'expliquer par le fait que ce derniers sont souvent situés sur des substrats riches en déchets de construction (même lorsqu'il ne s'agit pas de remblais).

3.4 Estimation de la contribution des cheminées à l'aide des points de contrôle

Le principe est de calculer l'enrichissement du sol en polluant d'origine atmosphérique sur chacun des n points de contrôle (*cf.* section 3.3) par les valeurs de retombées relatives (auréoles de retombées unitaires) de chacune des 125 cheminées (*cf.* section 3.2).

Pour chaque polluant, nous avons utilisé un algorithme itératif qui minimise l'équation suivante :

$$\sum_{i=1}^n e - xR$$

Avec :

- e : le vecteur des concentrations en polluants attribuables à l'**enrichissement du sol** par voie atmosphérique aux n points de contrôle :
 $e = (b_1, b_2, \dots, b_n)^T$, avec b_i l'enrichissement du sol en polluant pour le $i^{\text{ème}}$ point de contrôle (et T indique la transposition).
- x : le vecteur des **coefficients de pondération** des cheminées fournissant le poids à accorder à chaque cheminée (pour tenir compte du fait qu'elle n'ont pas toutes émises les mêmes quantités) correspondant à la meilleure estimation des concentrations du vecteur e :
 $x = (p_1, p_2, \dots, p_{125}, \alpha)^T$, avec p_j le poids de la $j^{\text{ème}}$ source et ou α est le terme constant. α représente un éventuel enrichissement moyen des n points de contrôle par rapport aux points ayant servi à l'estimation des teneurs « normale » et qui n'est pas dû aux retombées des 125 cheminées .
- R : la matrice de dimension ($n \times 126$) contenant les **valeurs de retombées relatives** pour chacune des 125 cheminées aux localisations XY des n points de contrôles :
 $R = (r_1, r_2, \dots, r_n)^T$, avec $r_i = (r_{i1}, r_{i2}, \dots, r_{i125}, 1)$ les valeurs de retombées relative des 125 cheminées pour la $i^{\text{ème}}$ observation et 1 la constante pour l'ensemble des points (correspondant au paramètre α du vecteur x).

Pour résoudre cette équation, nous avons utilisé un algorithme itératif construit sur mesure à l'aide du logiciel libre « R ». L'algorithme a une contrainte imposée : aucun coefficient ne peut être inférieur à 0, car cela équivaldrait à considérer des « cheminées dépolluantes », c'est à dire des cheminées qui auraient émis une quantité négative de retombées.

3.5 Résultats

3.5.1 Qualité des modèles prédictifs

La table 3.3 fournit une indication de la qualité pour chaque modèle de retombées atmosphériques retenu. Il fournit d'une part l'erreur absolue moyenne du modèle prédictif et

d'autre part le coefficient de corrélation entre les retombées calculées aux points de contrôle et les retombées prédites par le modèle.

TABLE 3.3 – Modélisation des retombées atmosphériques de polluants : qualité des modèles prédictifs

Retombées atmosphériques modélisées	Erreur absolue moyenne (ppm)	Corrélation point de contrôle - modèle
Cd	0.21	0.76
Cu	2.6	0.46
Pb	5.3	0.82
Zn	12.0	0.88

L'erreur absolue moyenne est proportionnelle à l'ordre de grandeur de l'élément considéré, ce qui est un résultat attendu. D'une façon générale, les modèles prédictifs pour le Cd, le Pb et le Zn donnent de bons résultats (0.76 à 0.88 de corrélation entre les retombées estimées par le modèle et les retombées estimées aux points de contrôle), tandis que le modèle prédictif du Cu est moins fiable (0.46 de corrélation entre les retombées estimées par le modèle et les retombées estimées aux points de contrôle). A titre d'illustration, la figure 3.2 présente le résultat du modèle pour le Zn. Sur ce graphique, la droite représente la situation idéale dans laquelle le modèle prédit la même quantité de retombées atmosphériques que celle estimée aux points de contrôle. Dans l'ensemble, on observe que le modèle donne des résultats satisfaisants, les points étant situés généralement à proximité de cette droite.

Pour le Hg, aucune prédiction valable n'a été obtenue. Ce résultat s'explique sans doute par le fait que les valeurs mesurées aux points de contrôle sont généralement proches des limites de quantification de la méthode analytique. Nous ne présentons pas en détail les résultats pour ce polluant.

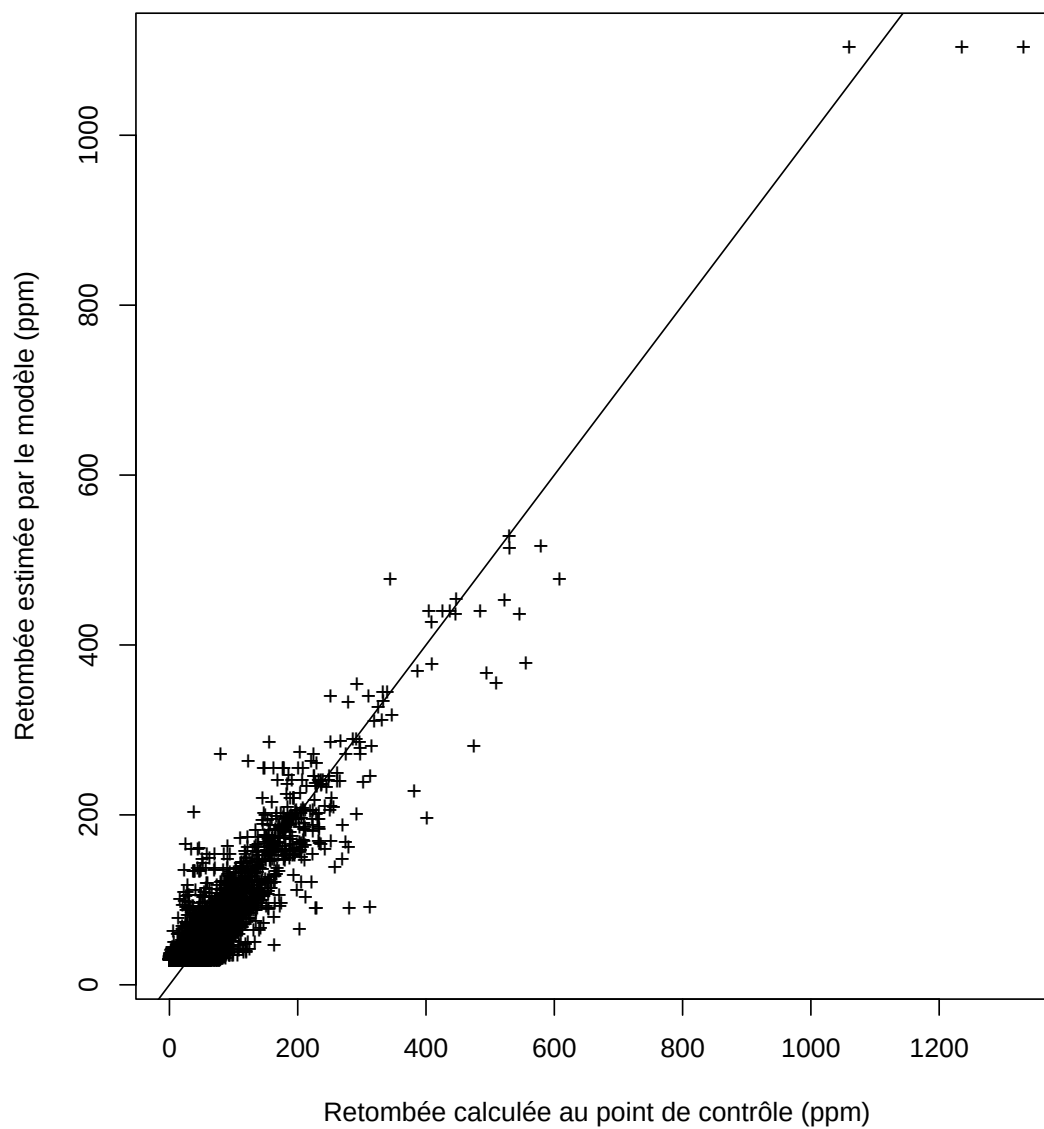


FIGURE 3.2 – Prédiction des retombées atmosphériques en Zn : résultat du modèle

3.5.2 Carte des retombées atmosphériques de polluants

Le modèle prédictif a été appliqué à la grille de points utilisés pour la cartographie de l'ensemble de la Région wallonne. Les cartes qui suivent présentent le résultat des modèles retenus pour la prédiction des retombées atmosphériques en Cd, Cu, Pb et Zn. L'augmentation des teneurs³ (en mg/kg MS) dans l'horizon de surface du sol attribuable aux retombées atmosphériques de polluants est cartographiée sur l'ensemble de la Wallonie. D'après le modèle, de 65 à 85 cheminées (selon l'élément) sur les 125 considérées (voir section 3.2) ont eu un impact significatif sur les teneurs observées dans les horizons de surface wallons.

La carte du Cd (Figure 3.3) suggère que les sols situés autour de la Meuse à l'aval de Namur et dans le pays de Herve ont été fortement touchés par les retombées atmosphériques : l'augmentation des teneurs dans l'horizon de surface du sol va de 0.75 ppm à 6.5 ppm. D'autres parties de la Wallonie ont aussi été touchées, mais pas de façon aussi importante. La carte est assez ressemblante à celle réalisée lors du projet POLLUSOL 2 : l'augmentation des teneurs en Cd est du même ordre de grandeur et dans les mêmes régions. On observe toutefois des petites différences : par rapport à ce qui avait été estimé lors de POLLUSOL 2, CAPASOL 6 suggère que les sols autour d'Andenne sont plus fortement touchés tandis que les sols de Charleroi auraient été plus faiblement touchés.

La carte du Cu (Figure 3.3) suggère que les sols situés dans le bassin industriel liégeois et dans le pays de Herve ont été touchés par les retombées atmosphériques. Dans ces régions, il est fréquent d'observer une augmentation des teneurs dans l'horizon de surface du sol supérieure à 10 ppm. D'autres parties de la Wallonie ont aussi été touchées, mais pas de façon aussi importante. La carte est assez ressemblante à celle réalisée lors du projet POLLUSOL 2, bien que CAPASOL 6 suggère que les sols de Charleroi auraient été plus faiblement touchés.

La carte du Pb (Figure 3.4) suggère qu'à l'échelle régionale, seuls les sols situés autour de la Meuse à l'aval de Namur et dans le pays de Herve ont été touchés par les retombées atmosphériques. Dans ces régions, on observe une augmentation des teneurs dans l'horizon de surface du sol allant de 20 ppm à 400 ppm. Par rapport à la carte réalisée lors du projet POLLUSOL 2, CAPASOL 6 suggère une augmentation plus faible des teneurs en Pb dans l'horizon de surface du sol, et les sols de Charleroi auraient été beaucoup plus faiblement touchés.

La carte du Zn (Figure 3.4) ressemble à celle du Pb. Dans les sols situés autour de la Meuse à l'aval de Namur et dans le pays de Herve, on observe une augmentation des teneurs dans l'horizon de surface du sol allant de 100 ppm à 1100 ppm. Par rapport à la carte réalisée lors du projet POLLUSOL 2, CAPASOL 6 suggère un même niveau d'impact des retombées atmosphériques, bien que les sols de Charleroi auraient été beaucoup plus faiblement touchés.

3. synonyme d'enrichissement

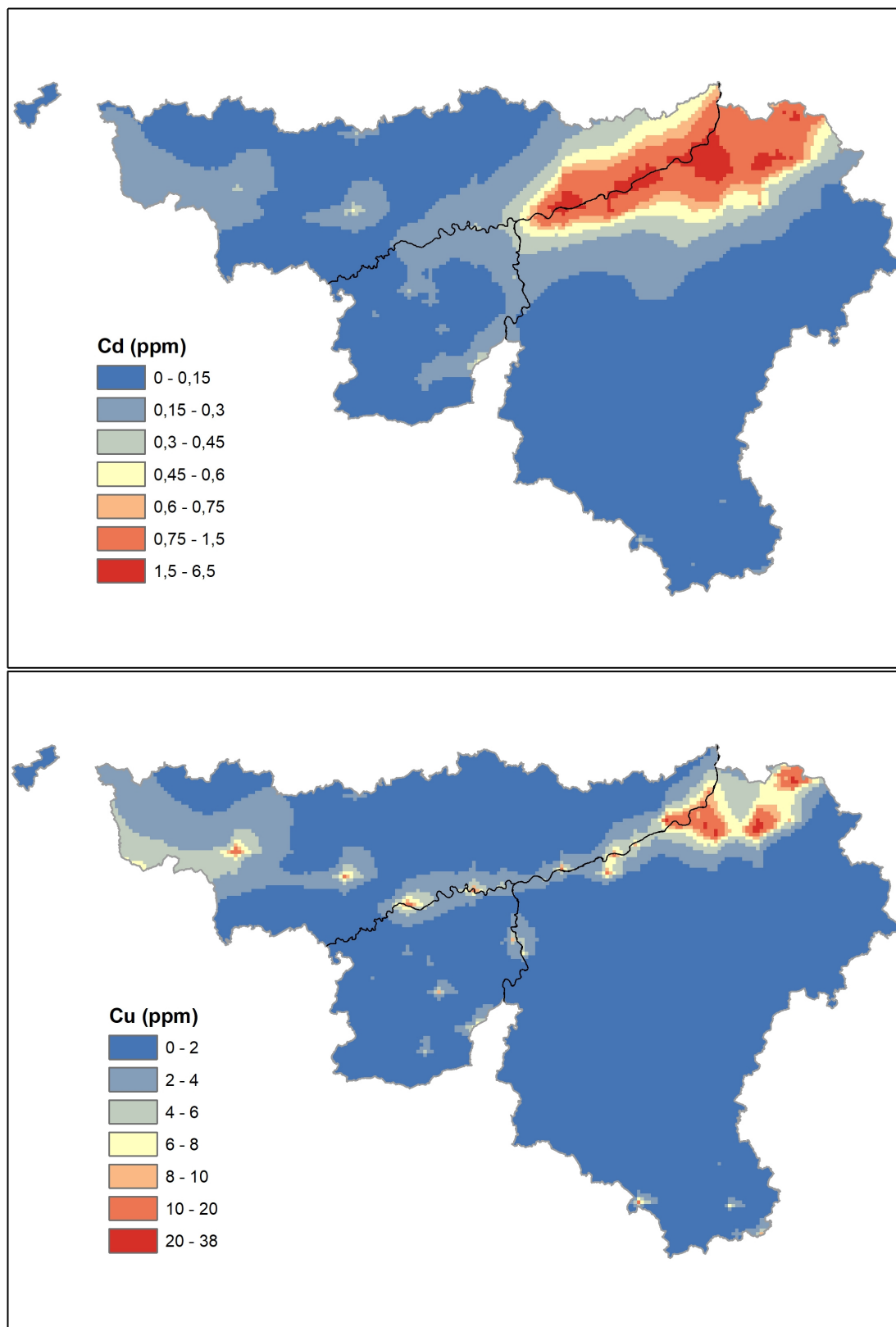


FIGURE 3.3 – Modélisation des retombées atmosphériques en Cd et Cu : prédiction de l'augmentation des teneurs dans l'horizon de surface du sol.

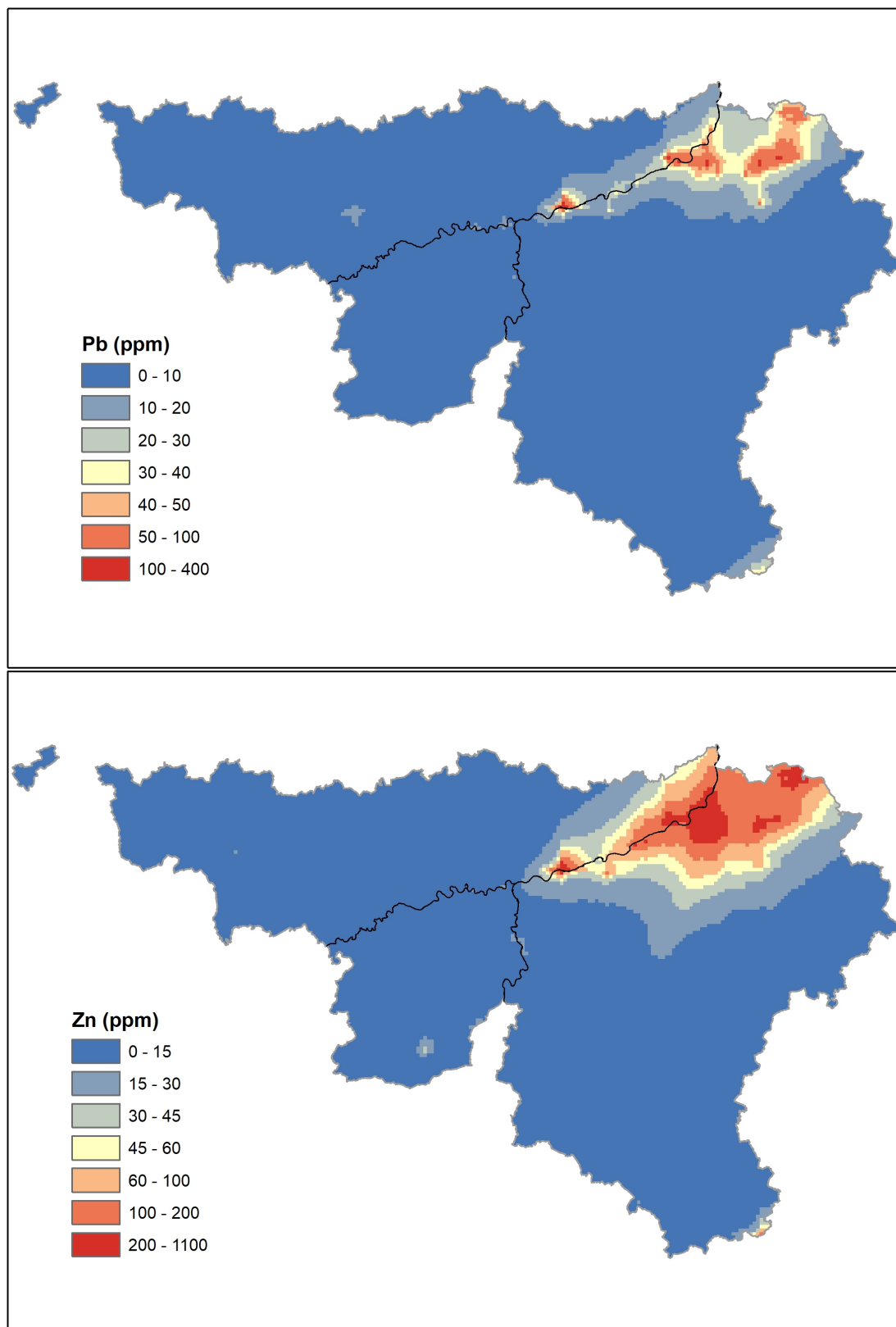


FIGURE 3.4 – Modélisation des retombées atmosphériques en Pb et Zn : prédiction de l'augmentation des teneurs dans l'horizon de surface du sol.

Chapitre 4

Cartographie des concentrations de fond

4.1 Introduction

Les cartes de concentrations de fond en polluants sont des outils de gestion pour les pouvoirs publics en matière d'aménagement du territoire, de gestion et de monitoring de l'environnement. Dans le cadre de la présente convention, la principale motivation pour l'établissement des cartes des concentrations de fond est la mise en oeuvre des dispositions du décret « sols »¹. La carte des concentrations de fond est, au sens du décret, un référentiel de teneurs en polluants dans les sols en Région wallonne. La concentration de fond est définie comme la *concentration ambiante d'un polluant dans le sol ; les concentrations ambiantes peuvent indiquer des variations géologiques naturelles ou l'influence d'une activité agricole, industrielle ou urbaine généralisée*. Selon le décret, l'établissement et la mise à jour de cette carte sont des tâches qui sont à effectuer par l'Administration². La présente convention a pour objectif de répondre à cette exigence du décret.

Les cartes de la présente convention ont été établies dans l'objectif de pouvoir être utilisées directement dans un outil cartographique d'aide à la détermination des concentrations de fond que l'Administration souhaite rendre disponible en ligne. Afin de préciser les besoins pour cet outil, l'Administration et les différents partenaires et membres du comité d'accompagnement de la convention ont été consultés. Ils ont été invités à répondre à un questionnaire³ et les réponses ont été débattues en réunion le 26 mars 2018⁴. Les résultats de cette consultation a permis à l'UCL d'élaborer des scénarios d'algorithmes possibles pour la conception des cartes des concentrations de fond. Ces scénarios ont été discutés en réunion le 18 mai 2018 avec

1. Décret relatif à la gestion et à l'assainissement des sols (M.B. 22.03.2018)

2. Décret « sols », Art. 10. L'administration établit et actualise la carte régionale des concentrations de fond sur la base des meilleures données disponibles, y compris celles fournies par les experts dans le cadre des études d'orientation ou de caractérisation.

3. Questionnaire réalisé par l'UCL envoyé par mail le 9 mars 2018.

4. Réunion intitulée « outil concentrations de fond : définition des besoins ». Lors de la réunion, quatre thèmes ont été abordés : 1. Usage du sol. 2. Format de l'information. 3. Incertitude sur la valeur renseignée. 4. Limitation éventuelle des valeurs renseignées.

l'Administration et les membres du comité d'accompagnement de la convention. Au cours de cette réunion, une discussion suivie d'une liste de décisions a permis de déterminer la meilleure façon de répondre au mieux aux besoins définis à la réunion du 26 mars 2018 étant donné les contraintes techniques. Ce travail de consultation a défini les options qui ont été prises et qui sont présentées dans ce qui suit :

Concernant la prise en compte du mode de gestion du sol dans la carte des concentrations de fond, il est prévu que l'outil cartographique d'aide à la détermination des concentrations de fond renseigne une seule valeur par sol, mais en indiquant le mode de gestion pris en compte. Au niveau des contraintes techniques, suite au constat que l'on ne disposait de jeux de données importants et complets au niveau spatial que pour les modes de gestion agricole et forestier, il a été choisi de ne réaliser des cartes que pour ces deux modes de gestion. Il a notamment été choisi de ne pas réaliser des cartes pour le mode de gestion « résidentiel » à partir des jeux de données « potager » car il a été estimé qu'un potager n'est qu'un cas particulier mode de gestion « résidentiel », et car il y a un manque de complétude spatiale des jeux de données « potager » disponibles. Il a aussi été choisi de ne pas réaliser des cartes pour le mode de gestion « récréatif » à partir des jeux de données « parc » de POLLUSOL 2 à cause du manque de complétude spatiale des jeux de données disponibles. Pour le mode de gestion « résidentiel » (en dehors du cas potager) et le mode de gestion « récréatif/commercial », il a été souligné que POLLUSOL 2 n'a pas mis en évidence d'effets importants de ces modes de gestion sur les teneurs rencontrées par rapport à un sol agricole. Sur base des sols de POLLUSOL 2, une pelouse ou un parc n'est pas nécessairement plus pollué qu'une parcelle agricole. Ceux-ci présentent des teneurs semblables, parfois légèrement plus faibles ou plus élevées, selon le polluant et l'usage du sol considéré. C'est pourquoi à la réunion du 18 mai 2018, compte tenu des données disponibles, il a été décidé que les jeux de données sur sols agricoles seront considérés comme représentatifs des concentrations de fond attendues dans les sols des usages 2 à 4, tandis que les jeux de données sur sols forestiers comme représentatif de l'usage 1 (*cf.* Table 4.1).

TABLE 4.1 – Correspondance entre usage du sol et jeu de données considéré comme représentatif des concentrations de fond (décision prise en réunion de comité de convention le 18 mai 2018).

Usage	Jeu de données
1	Sols forestiers
2	Sols agricoles
3	Sols agricoles
4	Sols agricoles
5	Sols agricoles

Concernant le format sur lequel l'information « concentration de fond » devrait être disponible, il est prévu que l'outil cartographique renseigne les concentrations de fond sur des mailles carrées du type de celles qui se trouvent dans l'outil INFO-FONDS développé dans le cadre de POLLUSOL 2. Pour la taille de la maille, $100m \times 100m$ a été choisi. Cette taille est identique à celle des précédentes cartes de concentration de

fond et est cohérente au vu de la densité spatiale des données disponibles. Il est aussi prévu à terme que le futur outil cartographique affiche les parcelles cadastrales, et les numéros de mailles pertinentes pour chaque parcelle.

Concernant la façon d’indiquer l’incertitude sur les valeurs renseignées par la carte des concentrations de fond, il est prévu que l’outil cartographique fournisse à la fois une valeur de concentration de fond (correspondant au centile 50 de la distribution théorique estimée en chaque centre de maille), et une valeur d’incertitude sur cette valeur sous la forme d’un centile élevé (90, 95 et 99). Les « cartes de centiles » sont définies ici comme des cartes qui renseignent une valeur qui a une probabilité théorique d’être dépassée correspondant à la valeur du centile. Par exemple, une carte du centile 95 fournit une valeur de concentration de fond qui a moins de 5% (100-95) de chance d’être dépassée. Afin d’aider l’Administration à définir le centile à indiquer simultanément à la concentration de fond dans l’outil cartographique, il est prévu que l’UCL fournisse des cartes pour plusieurs centiles élevés, en plus du centile 50, et que ces cartes soient comparées aux valeurs normatives wallonnes.

Concernant les gammes de valeurs renseignées par l’outil, il a été choisi que celles-ci ne seraient pas limitées (pas de censure), que ce soit au niveau des VS ou de toutes autres valeurs normatives⁵.

Concernant l’algorithme de prédiction (la détection des outliers, le krigeage et la prise en compte des limites floues), il est prévu que l’algorithme de prédiction des concentrations de fond soit établi de façon similaire à ce qui avait été réalisé par l’équipe de l’UCL (en collaboration avec le Pr. P. Bogaert de l’UCLouvain) lors de la convention CAPASOL 3.

Concernant la valeur renseignée en absence d’information sur le groupe de sol (c’est-à-dire lorsque l’information nécessaire à la prédiction est manquante), il a été choisi de compléter les informations disponibles afin que cette situation ne se présente pas. C’est déjà le choix qui avait été fait dans les précédentes cartes de concentrations de fond : cinq grands groupes de sols « par défaut »⁶ avaient été définis lorsque les sols (maille de prédiction) n’avaient pas pu être associé à un groupe de sols (par exemple, lorsque le sol était non cartographié sur la carte des sols de Wallonie). Dans le cadre de la présente convention, il a été choisi de procéder différemment. Lorsque l’information sur le groupe de sols est manquante, un travail d’estimation du groupe de sol le plus probable (parmi les groupes de sols définis précédemment) a été réalisé dans le cadre de CAPASOL 6 par l’UCL, afin d’avoir une couverture complète de la Région wallonne.

Notons que la définition de concentration de fond du décret « sols » est similaire à celle proposée par la norme ISO 19258 :2018⁷ : *background concentration concentration of an element or a substance characteristic of a soil type in an area or region arising from both natural sources and anthropogenic diffuse sources such as atmospheric deposition*. Afin de déterminer une concentration de fond, cette norme ISO 19258 :2018 propose un grand nombre de guidelines pour réaliser des campagnes d’échantillonnage et d’analyse ainsi qu’un traitement des

5. NB. Cela a été le cas dans des versions précédente de la carte.

6. Sur la façon de constituer ces groupes de sols, nous renvoyons le lecteur à CAPASOL 3.

7. ISO 19258 :2018 : *Soil quality - Guidance in the determination of background values*

résultats d'analyses. Les jeux de données retenus pour la cartographie (*cf.* Chapitre 1) satisfont dans l'ensemble les critères de qualité de la norme ISO, bien que la répartition spatiale des données disponibles ne soit pas idéale. Dans la présente étude, la prise en compte des biais entre les jeux de données existants (*cf.* Chapitre 2) suit les principes de la norme ISO, tout en allant dans des considérations/traitements plus avancés que ce qui est indiqué dans la norme. Le traitement des données (détection de outliers, etc.) et la présentation des résultats qui font l'objet du présent Chapitre suivent dans l'ensemble les guidelines de la norme ISO 19258 :2018, tout en répondant aux objectifs de la convention.

4.2 Groupe de sols et niveau d'appartenance

L'algorithme de prédiction des concentrations de fond prend en compte la carte des *Familles de Sols-ETM* constituée au cours de CAPASOL 2. Pour rappel, ces familles ont été constituées dans l'objectif de former des groupes sols présentant des concentrations en éléments traces métalliques semblables⁸. Dans ce qui suit, ces *Familles de Sols-ETM* seront appelées « groupes de sols ».

La valeur prédite sur chaque maille de $100m \times 100m$ dépendra de son « *appartenance* » à un groupe de sols. Le *degré d'appartenance* caractérise la probabilité pour une maille d'appartenir à un groupe de sols. La somme des degrés d'appartenance pour une maille doit toujours être égale à 1. Ce concept avait été introduit lors de CAPASOL 3 pour tenir compte des multiples sources d'incertitude intervenant lors du tracé des limites de certaines familles de sol. Il a été décidé d'incorporer l'incertitude spatiale sur ces limites, en ayant recours à une technique inspirée des méthodes de type « logique floue ». La « limite floue » n'est plus un simple trait délimitant deux unités spatiales, mais bien une bande (dont la largeur est à fixer par l'utilisateur) au sein de laquelle un point appartient en partie à l'une et en partie à l'autre des deux unités spatiales situées de part et d'autre de la bande (Figure 4.1).

8. Elles ont été utilisées comme Unités Spatiales de Comparaison dans l'étape de correction des biais dans les données (*cf.* section 2.1.1).

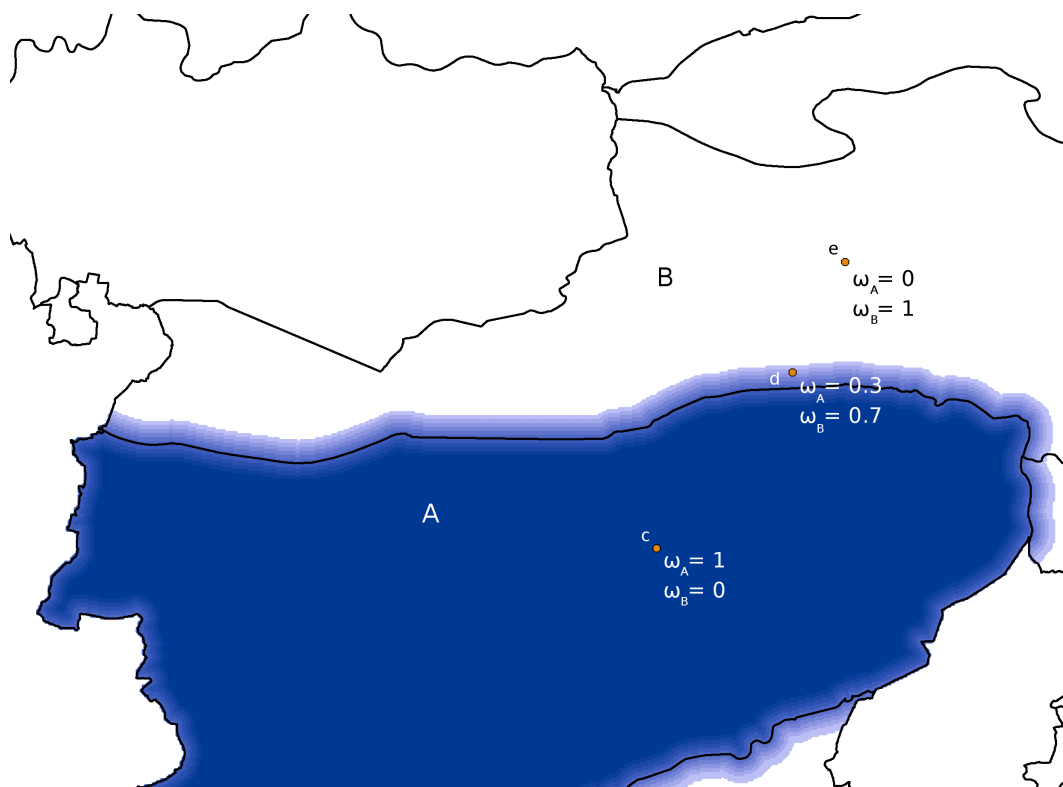


FIGURE 4.1 – Représentation du degré d'appartenance (ω) à une unité spatiale (unité A), symbolisé par un dégradé de bleu. Le point c situé au centre de l'unité spatiale A (intensité de bleu maximale) a un degré d'appartenance de 1 à A ($\omega_A = 1, \omega_B = 0$). Le point e situé au centre de l'unité spatiale B (intensité de bleu nulle) a un degré d'appartenance de 0 à l'unité spatiale A ($\omega_A = 0, \omega_B = 1$). Le point d a un degré d'appartenance de compris entre 0 et 1 pour les deux unités, car il est localisé dans la limite floue ($\omega_A = 0.3, \omega_B = 0.7$). Sa probabilité d'appartenance à l'unité spatiale A est donc de 0.3 (3 chances sur 10), tandis que sa probabilité d'appartenance à B est de 0.7 (7 chances sur 10).

Lors de la présente convention⁹, cette logique d'appartenance floue a été appliquée à toutes les mailles, et pas uniquement à celles situées aux limites de certaines familles de sol. Une maille de $100m \times 100m$ étant potentiellement couverte par plusieurs groupes de sols, il a été calculé pour chaque maille son pourcentage d'appartenance pour chaque groupe de sols.

Les 42 *Familles de Sols-ETM* (ou groupes de sols) de CAPASOL 2 ne couvrent pas 100% de la Wallonie. Afin qu'un maximum de mailles de prédiction soit classé dans un groupe de sols (et de permettre la prédiction sur toutes les mailles, *cf.* section 4.1), nous avons créé deux nouveaux groupes de sols. Ces deux groupes correspondent aux sols développés sur des alluvions (de la Sambre et de la Meuse jusqu'à Namur pour le premier groupe, et de la Meuse après Namur pour le second groupe). Ces deux nouveaux groupes de sols qui se justifient sur base de la nature du matériau parental permettent d'attribuer un groupe à une grande partie des mailles non classées. Ce travail de classification en 44 (42+2) groupes de sols est

9. Dans la grille de prédiction utilisée lors de CAPASOL 3, l'appartenance d'une maille à un groupe de sols était toujours de 100% en dehors des familles de sol avec des « limites floues ». L'appartenance d'une maille à un groupe de sols était fonction du groupe de sols situé au centre de la maille.

illustré dans la partie supérieure de la Figure 4.2. Sur cette Figure, le groupe de sols « NA » (en blanc) correspond soit aux sols qui ne sont classés dans aucun groupe, soit aux sols qui n'appartiennent majoritairement à aucun groupe (partage de 50% - 50%, généralement au niveau des limites floues).

Pour chaque maille non classée dans un groupe de sols après l'étape qui précède, un travail d'estimation du groupe de sols le plus probable a été réalisé. Pour cela, si un sol était classé dans un groupe à une distance de moins de 150m du centre de la maille non classée, la maille non classée s'est vue attribué une appartenance à 100% à ce groupe de sols. Si plusieurs groupes de sols sont situés à une distance inférieure à 150m du centre de la maille, seul le groupe de sols le plus proche a été retenu. Cette opération a permis d'attribuer un groupe de sols à environ 90% des mailles non classées. Cette opération a été répétée deux fois pour les mailles restantes, en considérant les sols classés à l'étape précédente, de façon à ce que l'ensemble des 1 689 976 mailles possèdent une appartenance à un groupe de sols. Ce travail d'attribution d'un groupe de sols aux mailles non classées est illustré dans la partie inférieure de la Figure 4.2. Dans cette Figure, le groupe de sols « NA » (en blanc) correspond aux sols qui n'appartiennent majoritairement à aucun groupe (partage de 50% - 50%, généralement au niveau des limites floues).

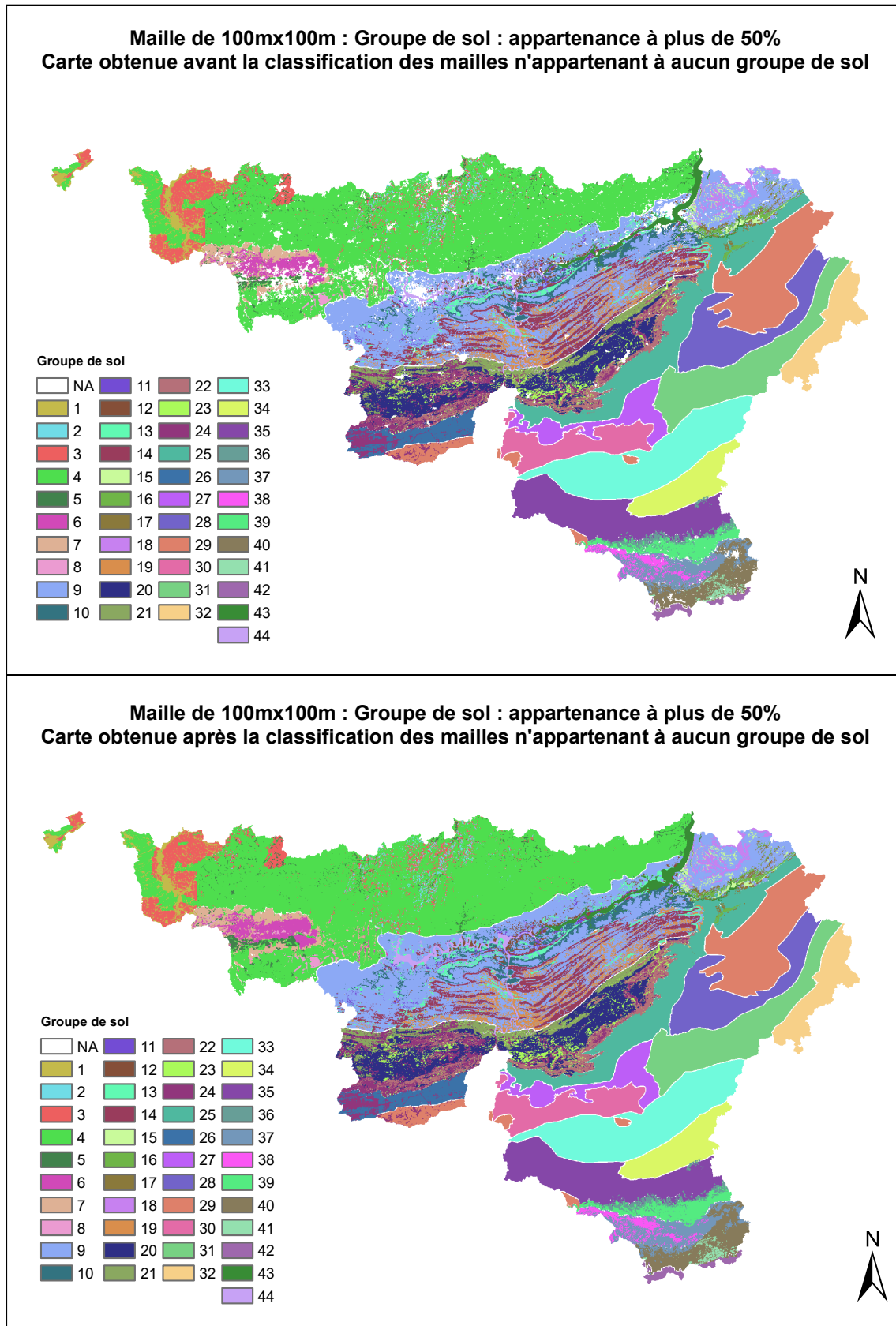


FIGURE 4.2 – Illustration du travail de classification des mailles dans le groupe de sols le plus probable.

4.3 Algorithme de prédiction des concentrations de fond

4.3.1 Introduction

L'algorithme utilisé pour obtenir les cartes des concentrations de fond a été mis au point au cours de la convention CAPASOL 3 et codé dans un script en langage R. Au cours de la convention CAPASOL 6, ce script a fait l'objet d'importantes améliorations, notamment au niveau de sa rapidité d'exécution et de sa documentation. Il a aussi été adapté pour tenir compte des caractéristiques des données (variogrammes) ainsi que des nouveaux groupes de sol (*cf.* section 4.2).

L'algorithme de prédiction des concentrations de fond permet d'obtenir, pour chaque point sur lequel on effectue la prédiction, un ensemble de valeurs probables caractérisées par une fonction de densité de probabilité. Sur base de cette fonction de probabilité obtenue, il est possible d'extraire n'importe quel centile de la distribution, ce qui rend son utilisation possible tant pour obtenir les cartes des concentrations de fond (centile 50) que pour obtenir les cartes d'incertitudes sur la valeur prédite sous la forme de concentration qui a moins de X % de chance d'être dépassée (centile $100-X$). Dans ce qui suit, les grands principes de l'algorithme de prédiction seront exposés. Nous renvoyons le lecteur vers le rapport final de CAPASOL 3 pour plus de détails sur celui-ci.

4.3.2 Description de l'algorithme

L'ensemble des opérations réalisées par l'algorithme depuis le jeu de données jusqu'à l'obtention des cartes comprend trois grandes étapes (illustrées à la Figure 4.3) :

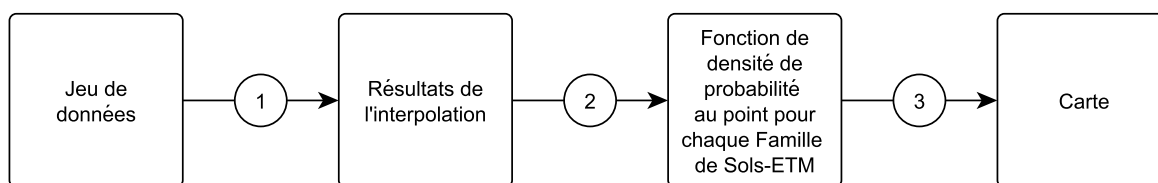


FIGURE 4.3 – Schéma simplifié de l'algorithme utilisé pour l'obtention des cartes de concentrations de fond.

Etape 1 : Préparation du jeu de données pour l'interpolation et le krigeage des données.

La préparation du jeu de données comprend premièrement le retrait des retombées atmosphériques de polluant (pour Cd, Cu, Pb et Zn, *cf.* Chapitre 3) et l'élimination des outliers après transformation logarithmique. Pour l'élimination des outliers, l'algorithme va retirer ce que l'on pourrait considérer comme les erreurs grossières, soit les valeurs qui ne pourraient pas raisonnablement correspondre à des concentrations de fond (*cf.* gammes de la Table 4.2). Il s'agit par exemple de valeurs correspondant à des erreurs d'encodage (ex : 0.05 ppm de Zn), ou de sols touchés par de la pollution locale ou à des anomalies géochimiques (ex : 50 000 ppm de Zn). L'algorithme ensuite retirer

les outliers univariés et multivariés pour chaque groupe de sols. Les valeurs éliminées sont celles qui ont une probabilité théorique d'occurrence de moins de 5 chances sur 10 000. Après cette étape d'élimination des outliers, les données sont standardisées (centrées et réduites) à partir des paramètres de distribution calculés pour chaque groupe de sols¹⁰. Les données sont ensuite interpolées par krigeage sur une maille de $100m \times 100m$ sur toute la Wallonie (soient 1 689 976 points). Les statistiques descriptives des jeux de données qui ont été krigés à cette étape sont présentées aux Tables 4.4 et 4.5.

TABLE 4.2 – Gamme de concentrations de fond considérée comme vraisemblable.

Elément	Concentration minimum (ppm)	Concentration maximum (ppm)
As	0.9	500
Cd	0.005	20
Cr	0.9	500
Cu	0.9	1000
Hg	0.005	10
Ni	0.9	500
Pb	0.9	1500
Zn	0.9	10000

10. Pour certains éléments comme l'As, certains groupes sols ont été réunis afin que chaque groupe puisse être caractérisé par un nombre suffisant de données.

TABLE 4.4 – Statistiques descriptives des jeux de données disponibles pour la cartographie des concentrations de fond des sols agricoles (et forestiers pour As et Cd) après détection et élimination des outliers.

	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
n	782	9385	19377	19585	7121	21653	17844	20027
moy	13.2	0.44	37.5	16.2	0.09	25.0	26.8	81.8
sd	11.2	0.75	16.3	5.9	0.03	12.1	10.0	41.0
mad	4.1	0.21	12.2	4.9	0.03	11.1	6.6	28.4
min	3.7	0.10	8.1	5.2	0.05	4.1	9.7	24.7
c5	6.6	0.10	19.0	9.0	0.05	9.9	16.8	44.0
c10	7.0	0.10	21.9	10.0	0.05	11.9	18.6	48.1
c25	7.9	0.13	26.4	12.0	0.07	15.9	21.1	57.0
med	10.3	0.26	33.1	15.0	0.08	21.7	24.9	74.7
c75	14.3	0.44	44.9	19.2	0.10	33.2	30.2	96.2
c90	19.6	0.85	57.5	23.4	0.12	42.0	36.5	118.9
c95	26.2	1.37	71.9	26.3	0.14	46.3	41.5	137.6
max	130.0	21.82	159.4	87.2	0.41	152.0	390.2	1530.7

TABLE 4.5 – Statistiques descriptives des jeux de données disponibles pour la cartographie des concentrations de fond des sols forestiers après détection et élimination des outliers.

	Cr	Cu	Hg	Ni	Pb	Zn
n	18691	19293	7118	21467	1541	19711
moy	28.6	12.4	0.16	20.6	62.0	66.4
sd	14.2	6.4	0.05	11.9	39.8	78.3
mad	11.5	5.2	0.04	11.1	20.8	28.2
min	2.0	1.0	0.05	1.2	15.8	9.5
c5	11.7	4.8	0.10	5.7	28.5	27.9
c10	14.4	5.7	0.11	7.7	31.2	31.9
c25	18.9	7.9	0.13	11.7	40.0	40.5
med	25.0	11.1	0.16	17.4	52.0	58.2
c75	35.7	15.6	0.19	28.8	70.0	79.7
c90	46.6	20.1	0.22	37.3	97.0	102.4
c95	55.7	23.5	0.25	41.5	131.0	122.9
max	146.3	100.3	0.70	144.9	452.1	6133.2

Etape 2 : Back-transformation des résultats d'interpolation en valeurs vraies (opération inverse de la préparation avant interpolation). Les données sont d'abord déstandardisées, puis une transformation exponentielle est appliquée (inverse de la transformation logarithmique de l'étape précédente). Une fonction de probabilité est alors générée pour chaque point pour lequel on effectue une prédiction, et pour chaque groupe de sols présent au point. Pour tous les points situés sur plusieurs groupes de sols (par exemple en raison des limites floues), on obtient donc plusieurs fonctions de probabilité par point (au moins deux), tandis que pour les points situés en dehors des limites

floues, on en obtient qu'une seule.

Etape 3 : Extraction d'une valeur unique des fonctions de probabilité en chaque point de prédiction, pour obtenir une carte. Pour chaque point, les fonctions de probabilité sont sommées (de manière pondérée, selon les degrés d'appartenance du point aux groupes de sols), pour obtenir une seule fonction de probabilité pour chaque point sur lequel on effectue la prédiction. La fonction de répartition est alors dérivée de cette fonction de probabilité pondérée, et n'importe quel quantile peut être extrait de cette fonction de répartition pour obtenir la carte que l'on recherche. Pour la carte de concentrations de fond, on choisit la valeur la plus « centrale » de la distribution en chaque point. Cela revient à sélectionner pour chaque point le centile 50 de la fonction de répartition. Pour les cartes d'incertitude, on prendra un centile élevé correspondant à la valeur qui a autant de probabilité d'être dépassée (ex : 10 % de chance pour le centile 90). La figure 4.4 illustre cette opération. C'est à la fin de cette étape que l'on ajoute les retombées atmosphériques de polluant pour Cd, Cu, Pb et Zn (*cf.* Chapitre 3).

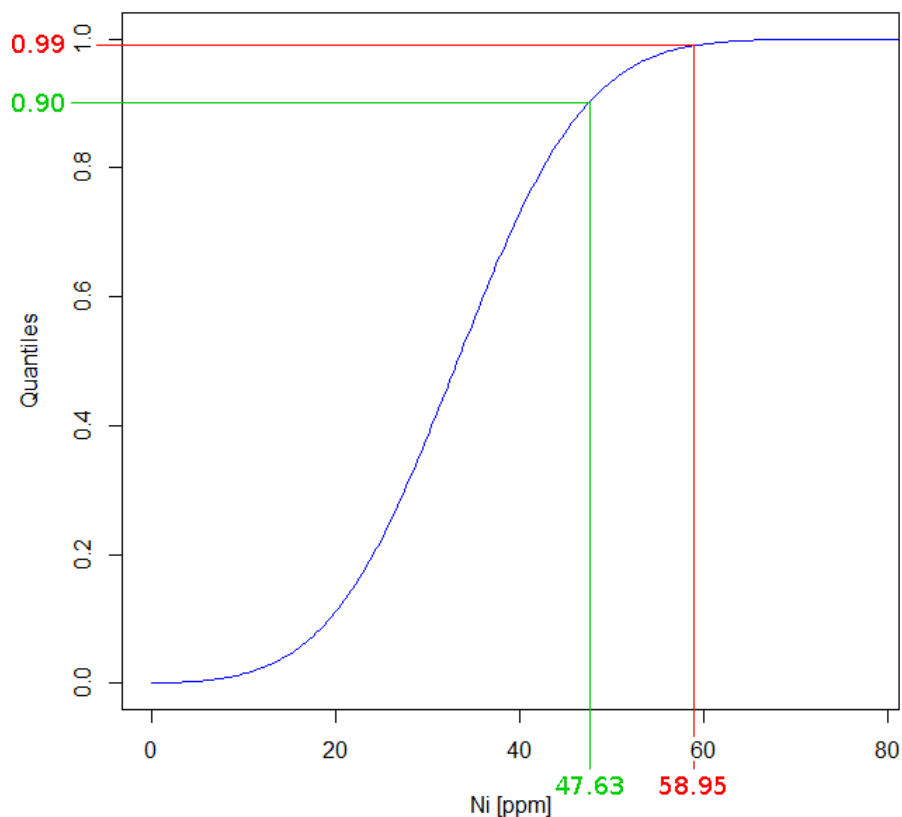


FIGURE 4.4 – Fonction de répartition des teneurs en Ni pour un point de la Wallonie. Cette courbe permet d’associer à chaque niveau de teneur (axe des abscisses) un risque de dépassement de cette teneur (axe des ordonnées). Par exemple, on peut observer que le centile 90 (en vert) correspond à la teneur de 47.63 ppm en Ni. Cela signifie qu’il y a moins de 10% de chance que la teneur en Ni en ce point dépasse 47.63 ppm. Par contre, le centile 99 (en rouge) en ce point correspond à la teneur de 58.95 ppm. Cela signifie qu’il y a moins d’1% de chance que la teneur en Ni en ce lieu dépasse 58.95 ppm.

4.4 Validation des cartes de concentration de fond

4.4.1 Cartes du centile 50

Plusieurs méthodes sont couramment employées pour valider une carte obtenue par interpolation spatiale. La méthode la plus utilisée est une approche de validation croisée appelée *leave-one-out*¹¹. Elle consiste à enlever chacune des observations temporairement du jeu de données et à prédire une valeur à l’endroit de l’observation retirée à partir des observations restantes. La valeur prédite pour cette observation est alors comparée à la valeur du jeu de

11. «*leave-one-out*» pourrait se traduire par «*mettre une donnée à part*»

données, et cette opération est répétée pour chacun des points du jeu de données. La comparaison des teneurs observées et estimées peut être effectuée à l'aide d'un diagramme de dispersion et le calcul d'un coefficient de corrélation (ex : Figure 4.6). Cette méthode permet aussi de calculer des statistiques sur les erreurs de prédiction : erreur moyenne, erreur médiane ou centile 95 de la distribution des erreurs (ex : « Err. abs. moy. (ppm) », « Err. abs. p50 (ppm) » et « Err. abs. p95 (ppm) » de la Table 4.6). Les erreurs sont calculées en valeurs absolues car étant donné que l'algorithme utilisé se base sur le krigeage (*cf.* section 2.2.2), l'erreur moyenne ou biais de la carte est toujours proche de 0. La médiane des erreurs absolues est une estimation centrale de l'erreur sur la carte, tandis que la moyenne des erreurs absolues représente l'espérance de l'erreur. L'erreur absolue moyenne (Err. abs. moy. ppm) est plus élevée que la médiane des erreurs absolues (Err. abs. p50 ppm). Cela s'explique par le fait que les valeurs les plus élevées du jeu de données sont plus difficiles à reproduire, ce qui est une propriété (ou un défaut) du krigeage utilisé dans l'algorithme de prédiction¹². Exprimée en pourcent, l'erreur absolue moyenne d'estimation (Err. abs. moy. %) est sans doute le meilleur indicateur de qualité de la carte, car il rend compte du fait que si l'erreur analytique de mesure est proportionnelle à la valeur mesurée, il devrait en être de même pour l'erreur de prédiction. Finalement, le centile 95 de la distribution des erreurs est un indicateur du niveau maximum de l'erreur de prédiction¹³.

La validation par leave-one-out a été appliquée selon deux protocoles différents :

Dans le premier protocole, les données utilisées pour la validation par leave-one-out correspondent à l'ensemble du jeu de données disponible pour la cartographie des concentrations de fond (présentées dans les Tables 2.2 et 2.3). Les résultats de la validation des cartes par cette méthode se trouvent aux Figures 4.5, 4.6 et 4.7. On observe des corrélations autour de 0.8 entre les teneurs observées et estimées, à l'exception du Hg (0.67-0.68). Pour le Hg, ceci s'explique par le fait que les valeurs mesurées en Wallonie sont généralement proches des limites de quantification, donc difficiles à reproduire. Pour le Pb, cette corrélation est plus faible dans les cartes des sols forestiers (0.76) que pour les sols agricoles (0.9). Des statistiques descriptives de la qualité des cartes obtenues pour les sols agricoles se trouvent sur la Table 4.6 et pour les sols forestiers sur la Table 4.7. L'erreur absolue moyenne d'estimation (Err. abs. moy. %) se trouve entre 12.1% et 28.2% (en pourcentage de la valeur de l'observation), ce qui peut être considéré comme très satisfaisant. C'est en effet du même ordre de grandeur que l'erreur aléatoire de mesure d'un laboratoire commercial (*cf.* projet POLLUSOL 2, Chapitre 2¹⁴).

Dans le second protocole, les données utilisées pour la validation par «leave-one-out» correspondent à un échantillonnage aléatoire du jeu de données (du premier protocole)

12. Dans certains cas, cela pourrait aussi être un résultat « heureux », car certaines observations élevées non reproductibles pourraient en réalité être des outliers qui sont passés au travers notre algorithme de détection des outliers, *cf.* section 2.2.2.

13. Cet indicateur (centile 95) a été choisi plutôt que le maximum car nous ne cherchons pas nécessairement à prédire correctement 100% du jeu de données, certaines observations outliers ayant pu passer au travers notre algorithme de détection des outliers (*cf.* section 2.2.2).

14. Pereira, Benoît ; Titeux, Hugues ; Schneider, Arnaud ; Sonnet, Philippe. Rapport final du projet POLLUSOL 2 - partie "sols" , 2012. 1126 p.

selon une grille régulière couvrant la Wallonie. Pour cette méthode, une grille avec des mailles de 7.5km¹⁵ de côté a été utilisée, et sur chaque maille, trois observations ont été sélectionnées aléatoirement¹⁶. Cette méthode tient compte du fait que les observations (jeu de données disponibles utilisé dans le premier protocole) ne sont pas réparties uniformément sur l'ensemble du territoire wallon. La stratification spatiale du jeu de données appliquée dans ce protocole permet d'évaluer la qualité des cartes en s'affranchissant du fait que le jeu de données disponible n'est pas réparti uniformément sur le territoire de la région wallonne. Les résultats de la validation des cartes par cette méthode se trouvent aux Figures 4.8, 4.9 et 4.10. On observe des corrélations entre les teneurs observées et estimées comparables à ce qui a été obtenu à l'aide du premier protocole. Des statistiques descriptives de la qualité des cartes calculées à partir de ce second protocole se trouvent sur les Tables 4.8 et 4.9. En général ces statistiques sont comparables à celles obtenues à partir du premier protocole. On note toutefois des meilleures statistiques pour l'As (erreurs plus faibles quel que soit l'indicateur).

Nous pensons que les statistiques obtenues à l'aide de ce second protocole sont plus représentatives de la qualité des cartes dans leur ensemble, bien que la validation par le premier protocole soit basée sur davantage de données.

En dernière analyse, nous nous sommes questionnés sur la répartition spatiale de ces erreurs. Sont-elles réparties uniformément sur la région ? Des parties de la Wallonie sont-elles moins bien prédites que d'autres ? Pour répondre à cette question, nous avons utilisé la méthode de validation du second protocole, mais au lieu de sélectionner aléatoirement des observations dans chaque maille, nous avons calculé l'erreur absolue médiane de la maille¹⁷. Cette erreur a ensuite été spatialisée sur la Wallonie, afin de permettre d'examiner graphiquement la présence de tendances spatiales. Les résultats cartographiques sont présentés en Annexe 1. Il en ressort que dans l'ensemble, les erreurs sont bien distribuées géographiquement : il n'y a pas une partie de la Wallonie qui serait moins bien prédite qu'une autre. On constatera aussi qu'il y a une tendance (plus ou moins marquée, selon la carte) à ce que les parties de la région présentant des teneurs plus élevées présentent des erreurs médianes de prédiction plus élevées. Les teneurs plus élevées sont plus difficiles à prédire, ce constat avait déjà été formulé lors de la convention CAPASOL 3. Ceci s'explique principalement par le fait que l'erreur analytique de mesure est plus grande lorsque la valeur est grande (comme expliqué précédemment, l'erreur analytique de mesure s'exprime généralement en pourcentage de la valeur mesurée). Par la méthode du «leave-one-out», lorsque nous cherchons à prédire une observation correspondant à une valeur élevée, on considère comme parfaitement «juste» cette observation (non influencée par l'erreur aléatoire de mesure), ce qui n'est pas le cas effectivement. Il est donc normal de faire des erreurs de prédiction plus grande lorsque l'observation que l'on cherche à prédire est elle-même «entachée» d'une erreur plus grande. Pour certains éléments comme Cd, Pb et Zn, ce constat s'explique aussi par le fait que dans les régions présentant des teneurs plus élevées, on observe

15. Sauf pour la carte de l'As, pour laquelle une grille de 15km a été utilisée étant donné la taille moins importante du jeu de données disponible.

16. Notons que toutes les observations sont retenues pour les mailles comportant moins de trois observations. Certaines mailles ne comportent aucune observation : ce cas est relativement fréquent pour la carte de l'As.

17. Nous avons calculé cette statistique que si la maille comportait au moins trois observations. Dans le cas contraire, la maille n'est pas prise en compte et non représentée cartographiquement.

des variations de teneurs plus importantes sur de courtes distances. Pour ces éléments, la carte pourrait être améliorée avec de nouvelles données dans les régions présentant des teneurs plus élevées.

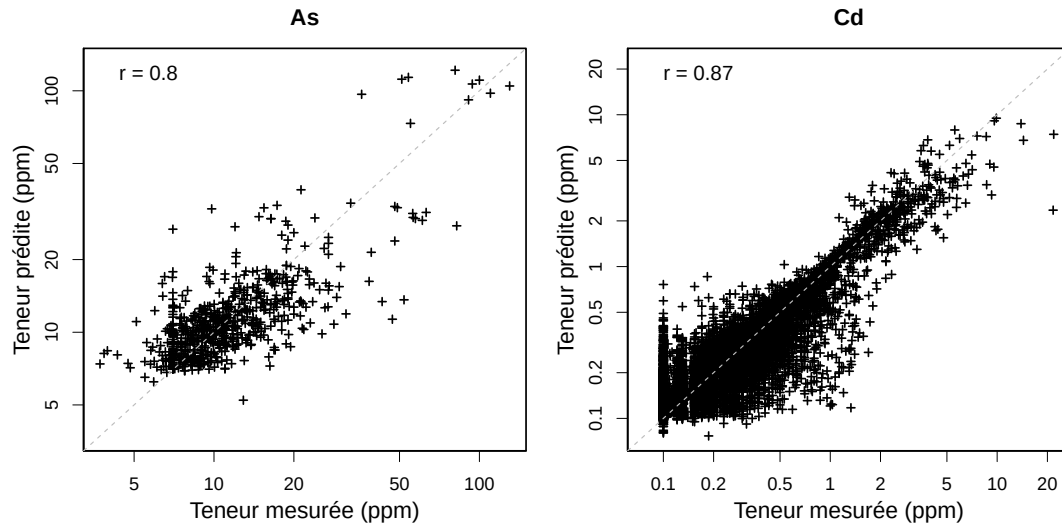


FIGURE 4.5 – Illustration graphique de la validation par leave-one-out des cartes de concentrations de fond pour As et Cd. En abscisse : les teneurs mesurées aux points retirés du jeu de données. En ordonnée : les teneurs prédites aux points correspondants.

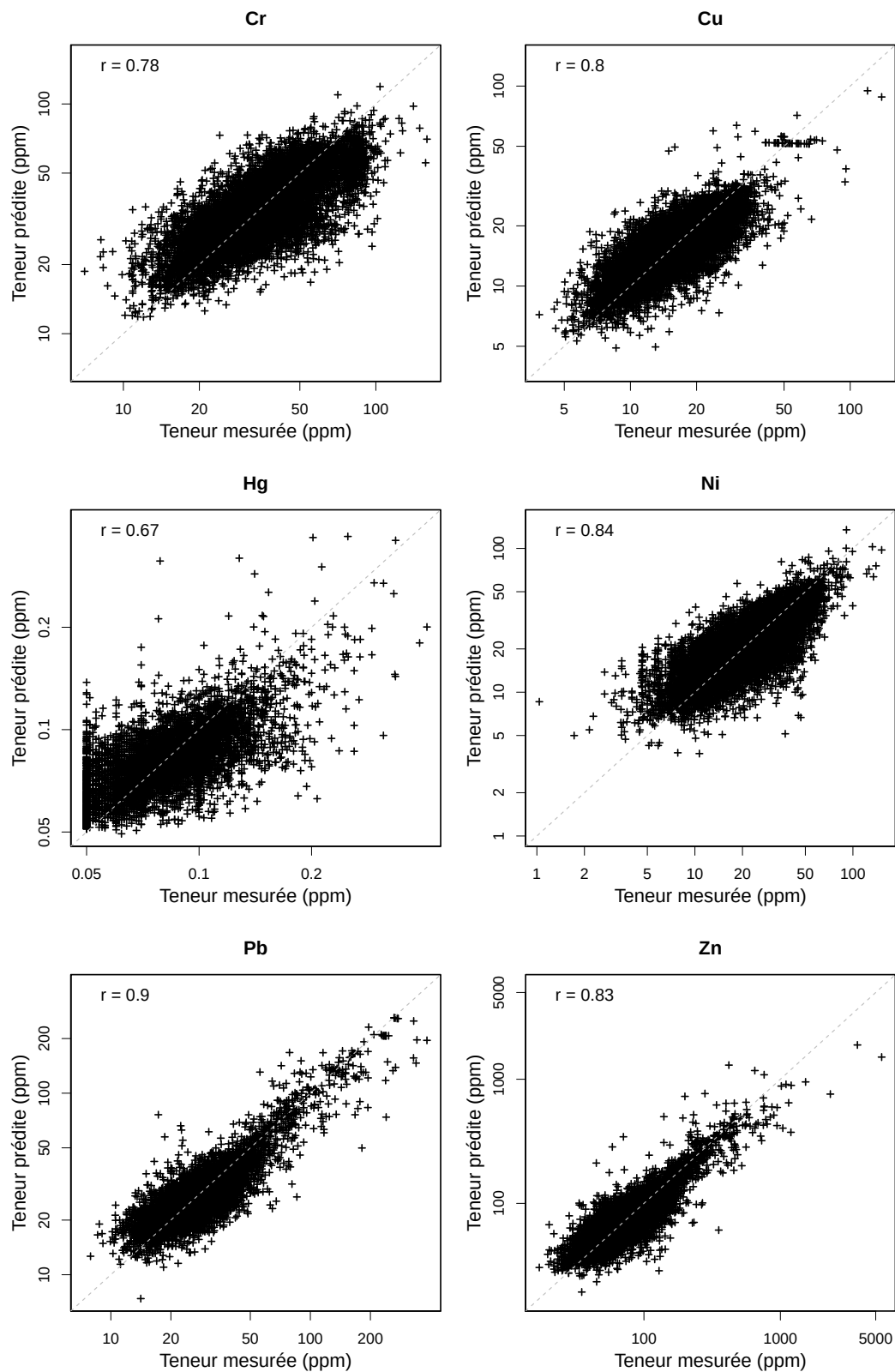


FIGURE 4.6 – Illustration graphique de la validation par leave-one-out des cartes de concentrations de fond dans les sols agricoles. En abscisse : les teneurs mesurées aux points retirés du jeu de données. En ordonnée : les teneurs prédites aux points correspondants.

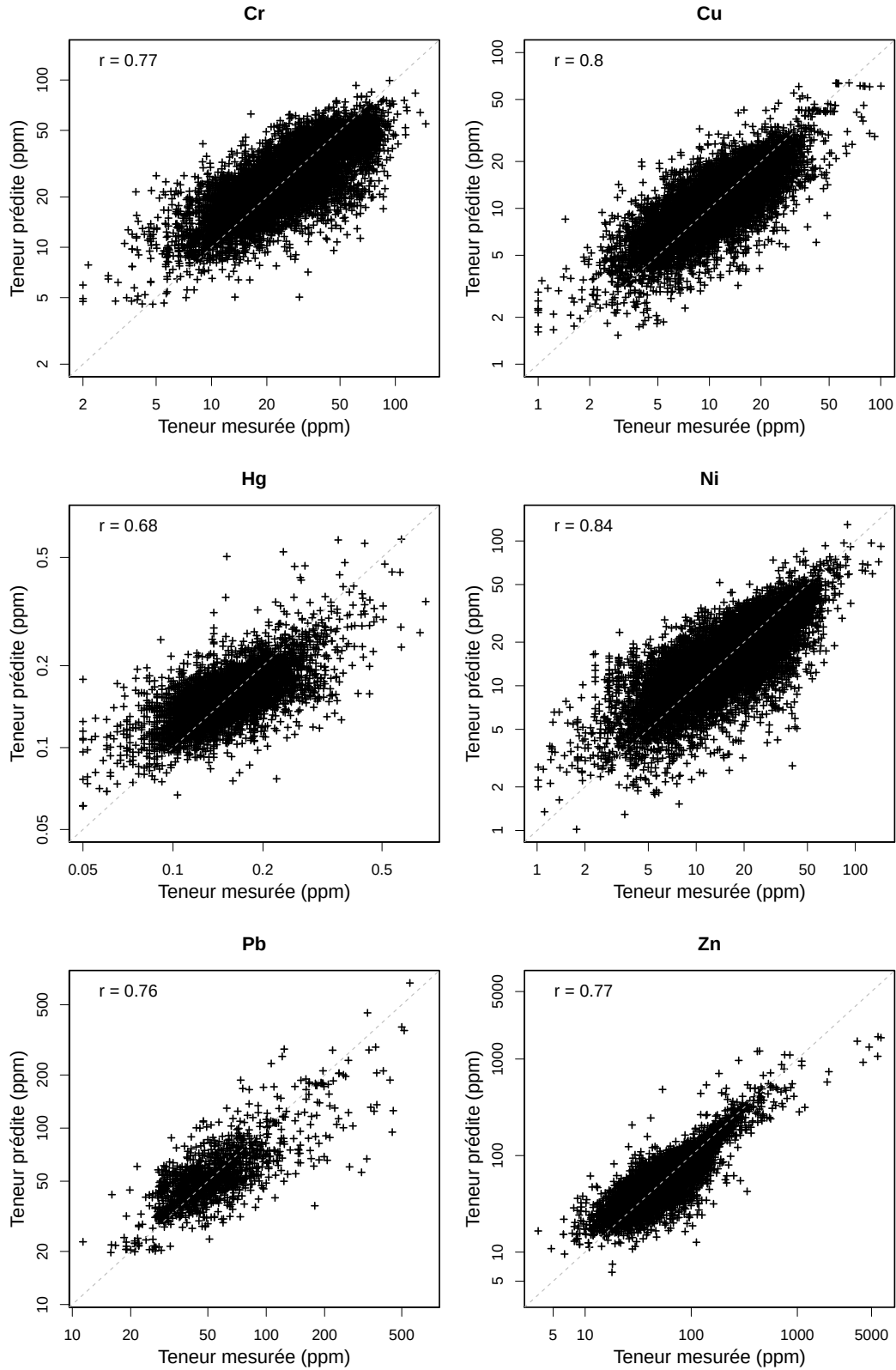


FIGURE 4.7 – Illustration graphique de la validation par leave-one-out des cartes de concentrations de fond dans les sols forestiers. En abscisse : les teneurs mesurées aux points retirés du jeu de données. En ordonnée : les teneurs prédites aux points correspondants.

TABLE 4.6 – Statistiques descriptives de la qualité de la prédiction spatiale obtenue par leave-one-out de l'ensemble du jeu de données disponible pour la cartographie des concentrations de fond dans les sols agricoles (et forestiers pour As et Cd).

	Cr	Cu	Hg	Ni	Pb	Zn	As	Cd
Cor. mesuré-prédit	0.78	0.80	0.67	0.84	0.90	0.83	0.80	0.87
Err. abs. moy. (ppm)	6.0	2.4	0.01	4.5	3.5	11.5	3.6	0.11
Err. abs. moy. (%)	16.6	14.8	16.8	20.5	12.1	13.2	24.5	28.2
Err. abs. p50 (ppm)	3.9	1.6	0.01	3.1	2.2	6.6	1.7	0.05
Err. abs. p95 (ppm)	19.5	7.0	0.04	13.4	10.3	33.2	13.0	0.35

TABLE 4.7 – Statistiques descriptives de la qualité de la prédiction spatiale obtenue par leave-one-out de l'ensemble du jeu de données disponible pour la cartographie des concentrations de fond dans les sols forestiers.

	Cr	Cu	Hg	Ni	Pb	Zn
Cor. mesuré-prédit	0.77	0.80	0.68	0.84	0.76	0.77
Err. abs. moy. (ppm)	5.9	2.5	0.03	4.4	16.2	12.3
Err. abs. moy. (%)	21.8	21.8	15.7	26.5	23.2	17.8
Err. abs. p50 (ppm)	3.8	1.7	0.02	3.0	8.4	6.5
Err. abs. p95 (ppm)	19.8	7.6	0.07	13.3	52.4	33.3

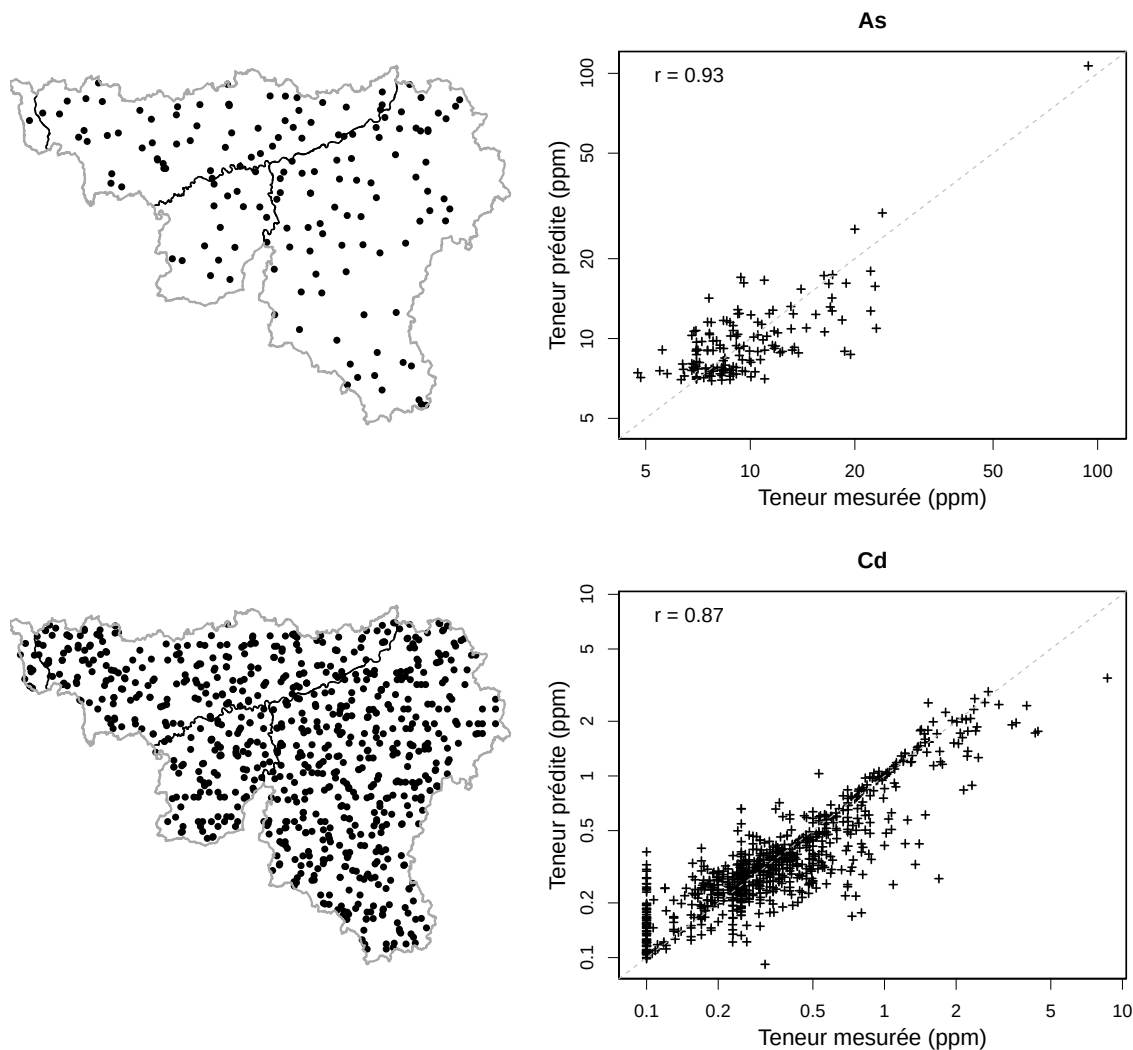


FIGURE 4.8 – Illustration graphique de la validation des cartes de concentrations de fond en As et Cd par leave-one-out stratifié géographiquement. Gauche : la localisations des observations retenues. Droite : comparaison des teneurs observées et estimées pour ces observations.

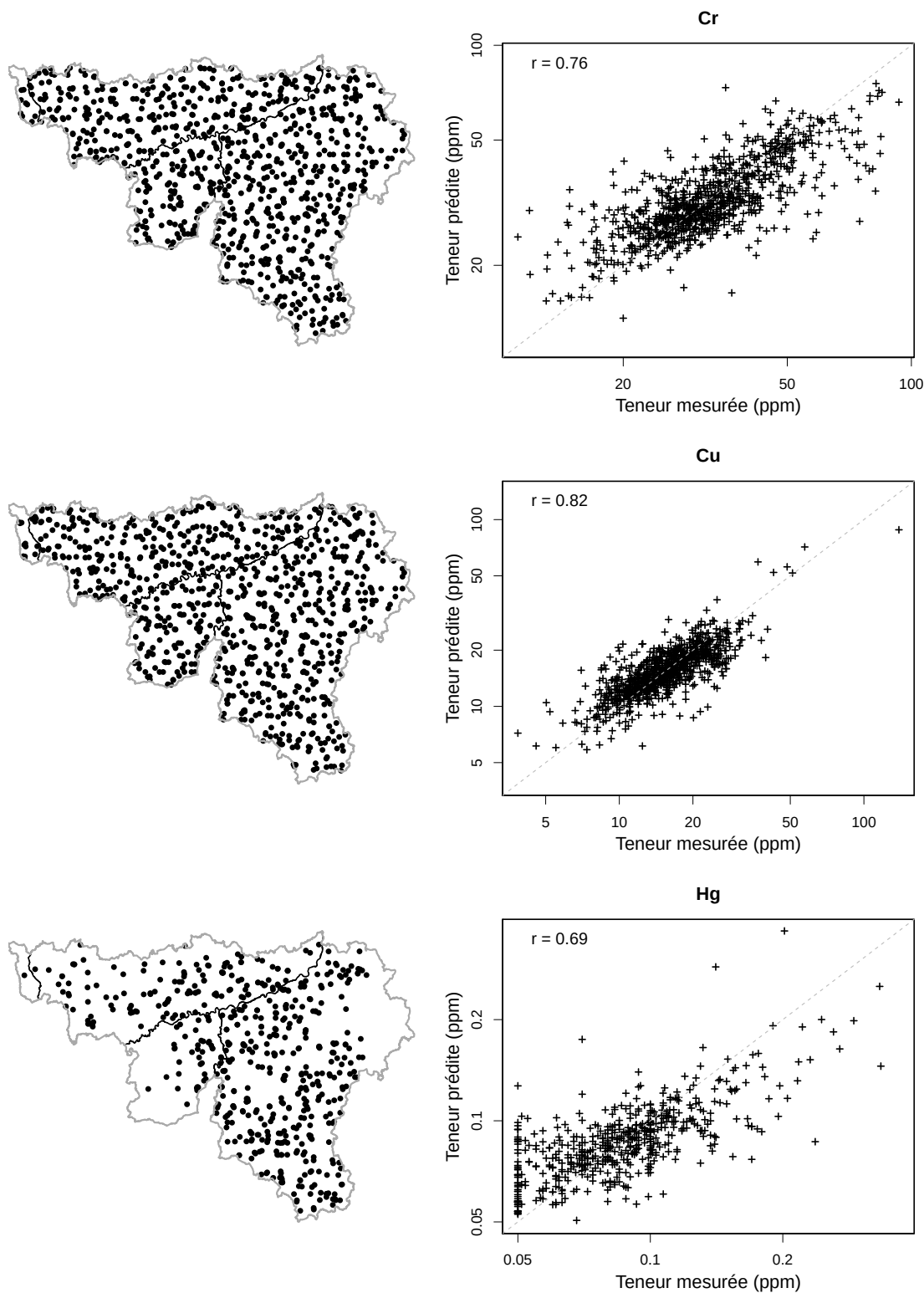


FIGURE 4.9 – Illustration graphique de la validation des cartes de concentrations de fond en Cr, Cu et Hg dans les sols agricoles par leave-one-out stratifié géographiquement. Gauche : la localisations des observations retenues. Droite : comparaison des teneurs observées et estimées pour ces observations.

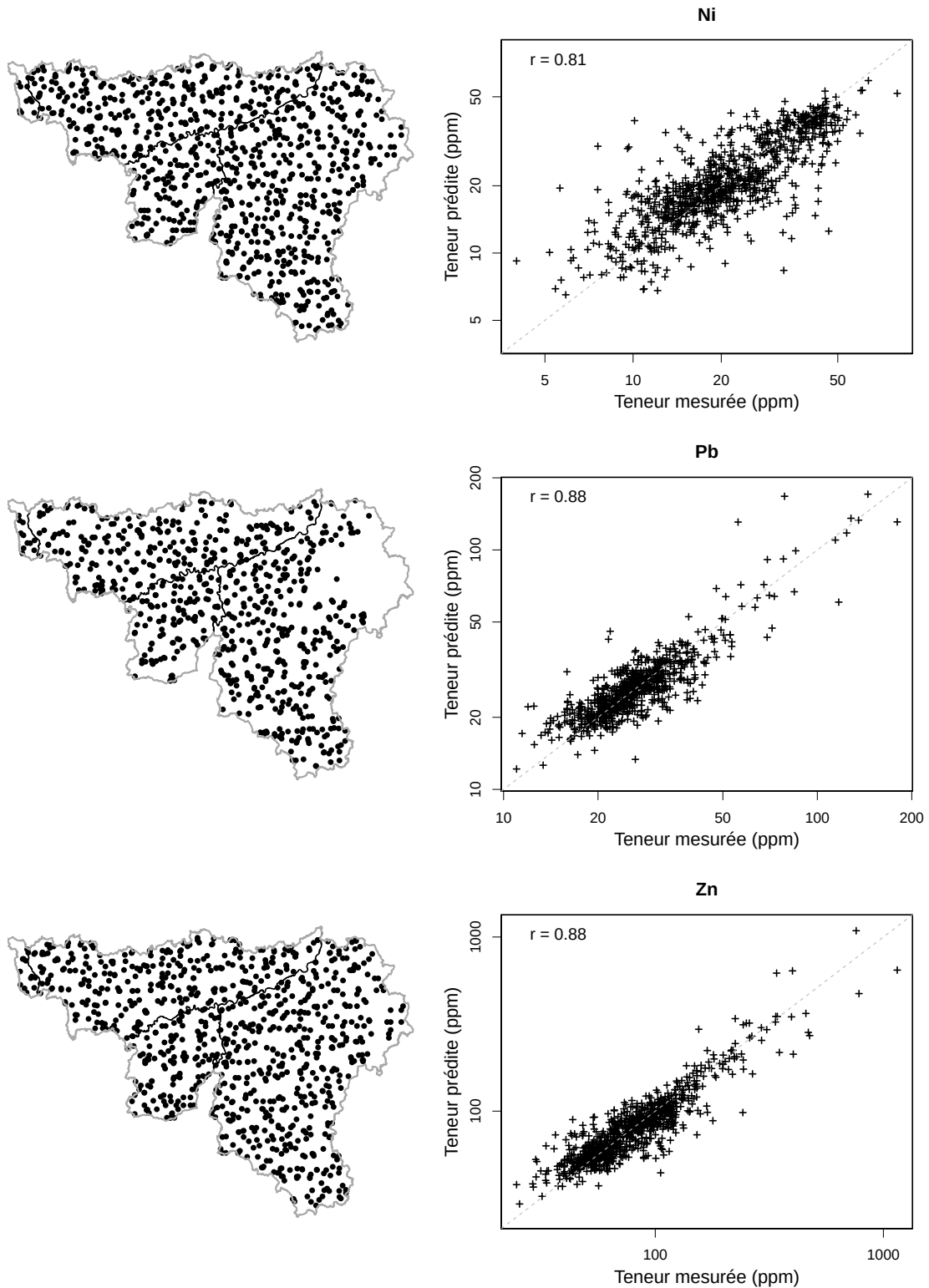


FIGURE 4.10 – Illustration graphique de la validation des cartes de concentrations de fond en Ni, Pb et Zn dans les sols agricoles par leave-one-out stratifié géographiquement. Gauche : la localisations des observations retenues. Droite : comparaison des teneurs observées et estimées pour ces observations.

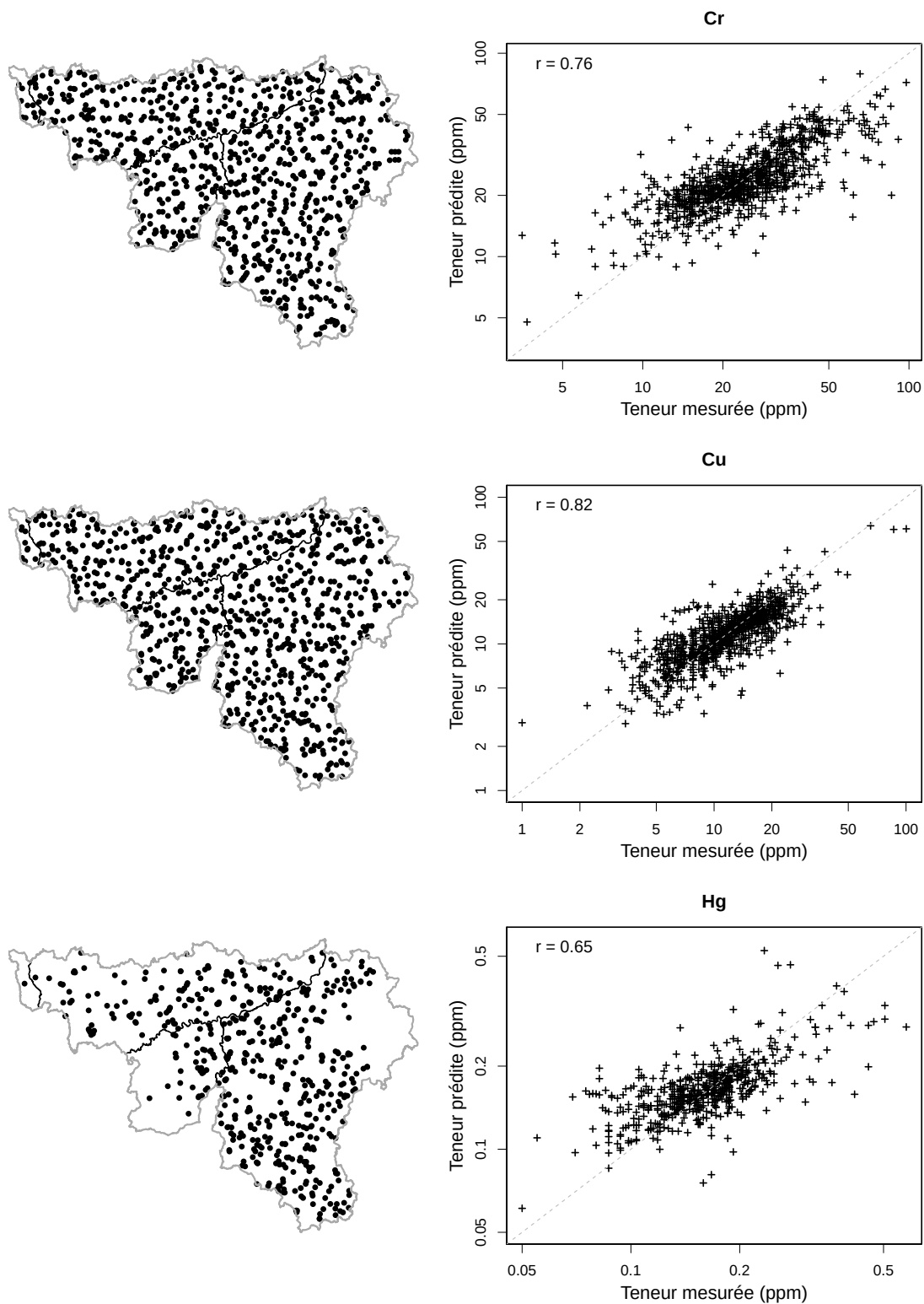


FIGURE 4.11 – Illustration graphique de la validation par leave-one-out stratifié géographiquement des cartes de concentrations de fond dans les sols forestiers pour Cr, Cu et Hg. Gauche : la localisations des observations retenues. Droite : comparaison des teneurs observées et estimées pour ces observations.

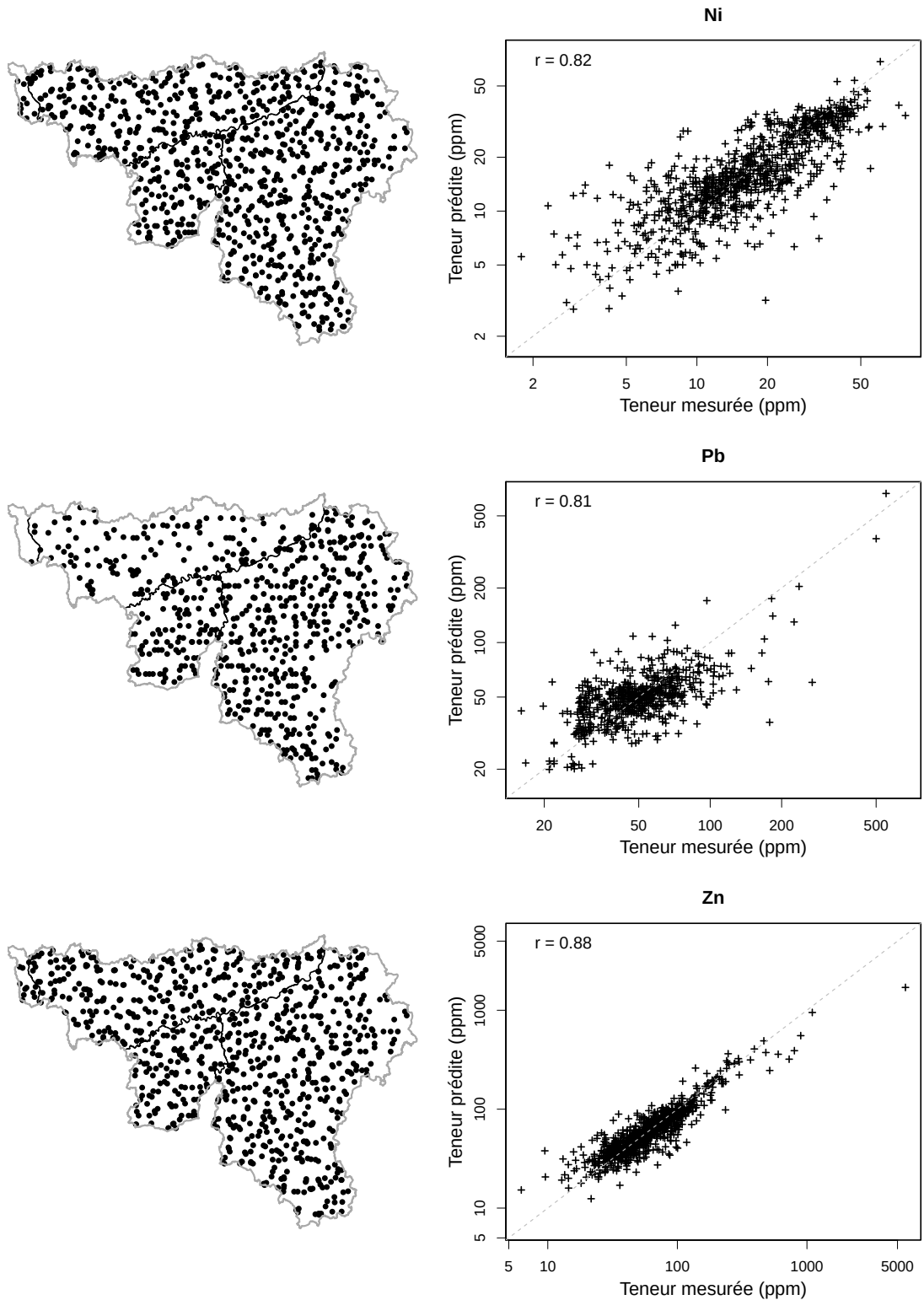


FIGURE 4.12 – Illustration graphique de la validation par leave-one-out stratifié géographiquement des cartes de concentrations de fond dans les sols forestiers pour Ni, Pb et Zn. Gauche : la localisations des observations retenues. Droite : comparaison des teneurs observées et estimées pour ces observations.

TABLE 4.8 – Statistiques descriptives de la qualité de la prédiction spatiale obtenue par leave-one-out du jeu de données utilisé pour la cartographie des concentrations de fond dans les sols agricoles (et forestiers pour As et Cd), après échantillonnage aléatoire de celui-ci selon une grille régulière couvrant la Wallonie.

	Cr	Cu	Hg	Ni	Pb	Zn	As	Cd
Cor. mesuré-prédit	0.76	0.82	0.69	0.81	0.88	0.88	0.93	0.87
Err. abs. moy. (ppm)	5.7	2.7	0.02	4.6	3.9	14.9	2.2	0.13
Err. abs. moy. (%)	17.1	16.8	19.2	22.8	13.2	15.1	20.6	26.6
Err. abs. p50 (ppm)	3.8	1.8	0.01	3.3	2.5	8.3	1.5	0.05
Err. abs. p95 (ppm)	18.0	8.6	0.06	14.2	11.4	38.7	6.6	0.46

TABLE 4.9 – Statistiques descriptives de la qualité de la prédiction spatiale obtenue par leave-one-out du jeu de données utilisé pour la cartographie des concentrations de fond dans les sols forestiers après échantillonnage aléatoire de celui-ci selon une grille régulière couvrant la Wallonie.

	Cr	Cu	Hg	Ni	Pb	Zn
Cor. mesuré-prédit	0.76	0.82	0.65	0.82	0.81	0.88
Err. abs. moy. (ppm)	5.9	2.7	0.03	4.6	13.4	18.2
Err. abs. moy. (%)	23.2	23.7	19.1	29.4	24.5	19.1
Err. abs. p50 (ppm)	4.0	1.7	0.02	3.1	8.4	7.4
Err. abs. p95 (ppm)	17.0	8.9	0.09	13.8	40.7	38.9

4.4.2 Carte des centiles 90, 95 et 99

Afin de fournir une indication d'incertitude sur les valeurs renseignées par la carte des concentrations de fond (centile 50), il est prévu que l'outil cartographique fournisse une carte d'un centile élevé (90, 95 ou 99). En effet, l'algorithme de prédiction des concentrations de fond fourni en chaque point sur lequel on effectue la prédiction, un ensemble de valeurs probables caractérisées par une fonction de probabilité (*cf.* section 4.3.2). Sur base de cette fonction de probabilité obtenue, nous avons extrait les centiles 90, 95 et 99 pour obtenir la concentration qui a moins de 10, 5 et 1 pourcents respectivement de chance d'être dépassée. Si la validation du centile 50 peut être réalisée en calculant des indicateurs à partir de la différence entre les valeurs observées et prédites, valider les valeurs des centiles élevés reviendrait à valider les formes et le calcul des fonctions de probabilité, ce qui est impossible. Cependant, dans cette section, nous allons examiner si les valeurs de centiles sont bien reliées à la réalité qu'elles sont censées représenter, c'est-à-dire une valeur qui a moins de X pourcent de chance d'être dépassée. Pour cela, notre analyse va se baser sur les résultats du leave-one-out. Pour rappel, avec cette méthode, chaque observation est enlevée temporairement du jeu de données et estimée à partir des observations restantes. En chaque point, on obtient donc une fonction de probabilité que l'on peut comparer à l'observation correspondante. Si la carte du centile X est cohérente, il est souhaitable que ce centile prédit soit, en moyenne, dépassé par la valeur de l'observation retirée 100- X pourcents de fois. Par exemple, la carte du centile 90 devrait être dépassée par 10% (=100-90) des observations utilisées dans le leave-one-out.

Les Tables 4.10, 4.11, 4.12 et 4.13 présentent le résultat de cette analyse, selon les deux protocoles de leave-one-out effectués à la section précédente. Selon ces Tables et en moyenne (dernière ligne de chaque Table), les centiles sont bien représentatifs d'une probabilité d'être dépassée à la valeur correspondante. Pour les sols agricoles, en moyenne, les centiles 90 sont dépassés par environ 9% des observations, les centiles 95 sont dépassés par environ 5.5% des observations, et les centiles 99 sont dépassés par environ 2% des observations. Les résultats obtenus avec les deux protocoles sont assez similaires, à l'exception d'As qui obtient des pourcentages plus faibles avec le protocole 2 qu'avec le protocole 1.

TABLE 4.10 – Pourcentage des observations qui dépassent les centiles 90, 95 et 99 des cartes de concentrations de fond calculées par leave-one-out du jeu de données « agricole » (protocole 1, détaillé à la section 4.4.1).

	Centile 90	Centile 95	Centile 99
Cr	9.4	5.9	2.5
Cu	9.5	5.6	2.1
Hg	8.8	5.2	2.0
Ni	9.0	5.0	1.7
Pb	10.2	6.3	2.4
Zn	9.5	5.6	2.3
As	10.0	6.4	3.1
Cd	9.1	5.4	2.5
Moy.	9.4	5.7	2.3

TABLE 4.11 – Pourcentage des observations qui dépassent les centiles 90, 95 et 99 des cartes de concentrations de fond (colonnes 1, 2 et 3 respectivement) calculées par leave-one-out du jeu de données « forestier » (protocole 1, détaillé à la section 4.4.1).

	Centile 90	Centile 95	Centile 99
Cr	8.5	5.1	2.1
Cu	8.9	5.1	1.7
Hg	8.2	4.8	1.8
Ni	8.9	4.6	1.4
Pb	10.0	5.6	1.9
Zn	9.4	5.6	2.3
Moy.	9.0	5.1	1.9

TABLE 4.12 – Pourcentage des observations qui dépassent les centiles 90, 95 et 99 des cartes de concentrations de fond (colonnes 1, 2 et 3 respectivement) calculées par leave-one-out du jeu de données « agricole » après échantillonnage aléatoire de celui-ci selon une grille régulière (protocole 2, détaillé à la section 4.4.1).

	Centile 90	Centile 95	Centile 99
Cr	8.7	4.6	2.1
Cu	10.0	6.1	2.1
Hg	7.7	5.0	1.5
Ni	8.2	4.9	1.8
Pb	9.3	6.6	1.9
Zn	11.3	6.0	2.7
As	8.0	4.7	1.3
Cd	8.1	4.1	1.6
Moy.	8.9	5.2	1.9

TABLE 4.13 – Pourcentage des observations qui dépassent les centiles 90, 95 et 99 des cartes de concentrations de fond (colonnes 1, 2 et 3 respectivement) calculées par leave-one-out du jeu de données « forestier » après échantillonnage aléatoire de celui-ci selon une grille régulière (protocole 2, détaillé à la section 4.4.1).

	Centile 90	Centile 95	Centile 99
Cr	8.2	4.7	1.5
Cu	8.6	5.2	0.9
Hg	6.1	2.8	1.5
Ni	8.2	4.9	1.2
Pb	8.9	5.6	1.8
Zn	9.3	5.2	1.6
Moy.	8.2	4.7	1.4

4.5 Cartes des concentrations de fond

Selon la norme ISO 19258 *Soil quality - Guidance in the determination of background values*, la valeur de concentration de fond est une *statistical characteristics (average, median, mode, etc.) of the total (natural pedogeochemical and anthropogenic) content of a substances in soil*. Dans ce qui suit, nous avons considéré que la concentration de fond est la valeur du centile 50 (médiane) de la distribution des valeurs de concentrations attendue calculée en chaque maille (obtenue à partir de la fonction de répartition, *cf.* section 4.3.2).

4.5.1 Statistiques descriptives des concentrations de fond en Wallonie

Les cartes produites fournissent une valeur de concentrations de fond sur des mailles de $100m \times 100m$ couvrant l'ensemble de la Wallonie (soient 1 689 976 mailles, *cf.* 4.3.2). Sur base de ces cartes, des statistiques descriptives des concentrations de fond pour les sols agricoles et forestiers wallons ont été calculées (Tables 4.14 et 4.15). Pour le calcul des statistiques selon le mode de gestion du sol (sols agricoles ou forestiers), seules les mailles correspondant au mode de gestion ont été prises en compte, sur base de la localisation du centre de la maille sur la Carte Numérique d'Occupation du Sol de Wallonie (910 144 mailles pour les sols agricoles, 486 925 mailles pour les sols forestiers).

TABLE 4.14 – Sols agricoles de Wallonie : statistiques descriptives des concentrations de fond.

	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
moy	9.6	0.42	32.9	15.5	0.08	22.4	27.9	84.8
sd	4.5	0.40	9.2	4.0	0.02	9.2	11.6	50.6
mad	1.8	0.12	5.7	3.7	0.01	6.1	6.5	27.0
min	4.6	0.07	11.4	4.7	0.05	1.9	8.8	21.7
c5	7.2	0.14	22.6	10.4	0.06	11.4	18.6	47.8
c10	7.4	0.17	24.5	11.3	0.07	13.5	19.6	50.8
c25	7.7	0.23	27.0	12.6	0.07	16.1	21.3	56.3
med	8.9	0.29	30.2	14.6	0.08	19.2	25.4	72.3
c75	10.0	0.42	36.0	18.1	0.09	27.5	30.4	96.2
c90	12.1	0.77	47.7	20.8	0.10	37.4	36.4	120.0
c95	14.7	1.20	53.2	22.5	0.11	40.0	46.3	173.0
max	180.0	8.61	113.0	114.0	0.46	125.0	412.0	2270.0

TABLE 4.15 – Sols forestiers de Wallonie : statistiques descriptives des concentrations de fond.

	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
moy	10.9	0.34	28.4	13.0	0.17	20.6	52.3	70.3
sd	3.8	0.31	8.7	4.2	0.04	10.0	14.2	38.7
mad	1.5	0.10	8.6	3.9	0.03	12.8	8.9	23.5
min	5.1	0.08	3.7	1.2	0.07	1.1	13.6	9.0
c5	7.6	0.14	16.8	5.4	0.12	5.3	31.9	27.5
c10	8.4	0.16	18.5	7.8	0.13	6.9	38.4	35.3
c25	9.0	0.20	21.9	10.4	0.15	12.4	45.2	50.1
med	9.6	0.26	26.6	13.4	0.17	19.7	51.1	67.3
c75	11.8	0.38	34.5	15.7	0.19	29.6	57.3	81.4
c90	16.2	0.56	40.8	17.7	0.21	33.7	68.5	98.8
c95	17.2	0.74	43.9	19.0	0.23	35.4	75.5	117.0
max	171.0	9.65	82.9	65.8	0.75	101.0	585.0	1770.0

4.5.2 Présentation des résultats cartographiques

Les pages qui suivent présentent l'ensemble des résultats cartographiques obtenus lors de la mise à jour des cartes de concentrations de fond. Le code couleur utilisé est le même pour les cartes des sols agricoles et forestiers, afin de rendre possible la comparaison des cartes. Cependant, pour les deux classes de couleurs les plus élevées, la borne utilisée est la Valeur Seuil du décret « sols » correspondante (« naturel » pour les sols forestiers, et « agricole » pour les sols agricoles). Les Valeurs Seuils utilisées dans le cadre de ce rapport sont celles reprises à la Table 4.16.

TABLE 4.16 – Valeurs Seuils (VS) prises en compte dans CAPASOL 6.

	naturel	agricole	résidentiel	récréatif ou commercial	industriel
As	30	30	40	40	65
Cd	1.8	1.8	3	10	20
Cr	57	57	78	140	288
Cu	53	53	156	490	600
Hg	1.1	1.1	1.75	5	5
Ni	87	87	146	350	350
Pb	120	200	200	390	1840
Zn	196	196	415	3000	3000

Les résultats obtenus sont en cohérence avec les cartes produites lors de POLLUSOL 2 et CAPASOL 3. Pour As, Cr et Ni, c'est la nature du matériau parental qui explique les grandes variations de teneurs observées à l'échelle régionale. Par exemple, pour As, les teneurs les plus élevées sont observées sur des sols ardennais développés sur des schistes, phyllades

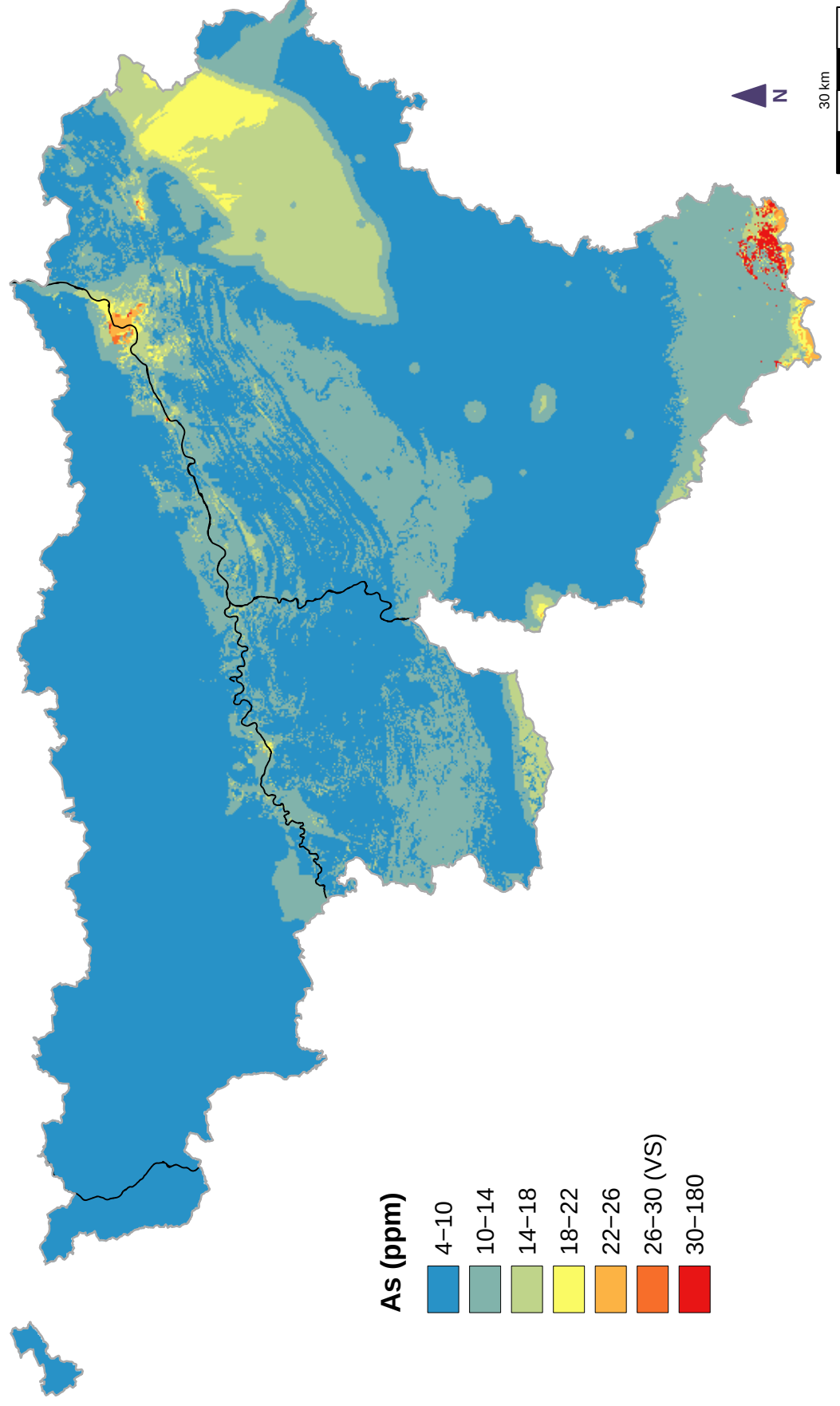
et grès datant principalement du Cambrien et sur les sols de Lorraine développés sur les argiles et marnes datant du Jurassique inférieur. Pour Cd, les retombées atmosphériques de contaminants sont responsables de la plupart des variations de teneurs observées à l'échelle régionale. C'est la raison pour laquelle la carte des concentrations de fond ressemble très fort à la carte des retombées atmosphériques de Cd (*cf.* section 3.5.2). Pour Pb, Zn et Cu, ce sont à la fois les retombées atmosphériques de contaminants et la nature du matériau parental qui expliquent les grandes variations de teneurs observées à l'échelle régionale. Pour Hg, les teneurs élevées sont rassemblées dans le sillon Sambre-et-Meuse et autour de Verviers, illustrant l'influence des retombées atmosphériques de contaminants sur les teneurs observées à l'échelle régionale. Notons que pour cet élément, c'est le premier projet de cartographie géochimique de la Wallonie qui permet de montrer une structure spatiale des teneurs aussi nette. Pour les éléments fortement influencés par les retombées atmosphériques de contaminants, il s'agit des premières cartes qui font apparaître aussi clairement l'influence de l'activité industrielle de Sclaigneaux sur les teneurs observées au niveau régional. Les teneurs prédites dans le bassin industriel liégeois sont aussi beaucoup plus proches de la réalité. Ce constat s'explique principalement grâce aux nouvelles cartes d'impact des retombées atmosphériques de polluants utilisés dans l'algorithme de prédiction des concentrations, et à la prise en compte des nouvelles données disponibles pour le projet¹⁸.

Remarquons finalement que pour l'As, la carte a été réalisée à partir d'un jeu de données comprenant beaucoup moins d'observations que pour les autres cartes (*cf.* Table 4.4). Ce manque de connaissance des teneurs en As des sols wallons a pour conséquence une moins grande fiabilité de la carte en certains endroits par manque de disponibilité de données¹⁹. Les fortes variations de teneurs que l'on peut visualiser sur la carte principalement en Lorraine et dans le centre de Liège mériteraient d'être étudiés plus en détail, notamment en raison des niveaux observés proches ou supérieurs aux VS et de l'importance de ce polluant dans les études de risque. Le même type de constat peut être réalisé pour la carte des concentrations de fond en Pb dans les sols forestiers : celle-ci a été réalisée à partir d'un jeu de données comprenant environ 10 fois moins d'observations que les autres cartes (à l'exception de l'As), rendant en conséquence la carte moins fiable en certains endroits. Or les niveaux observés sont souvent proches des VS et ce polluant présente aussi une grande importance pour les études de risque.

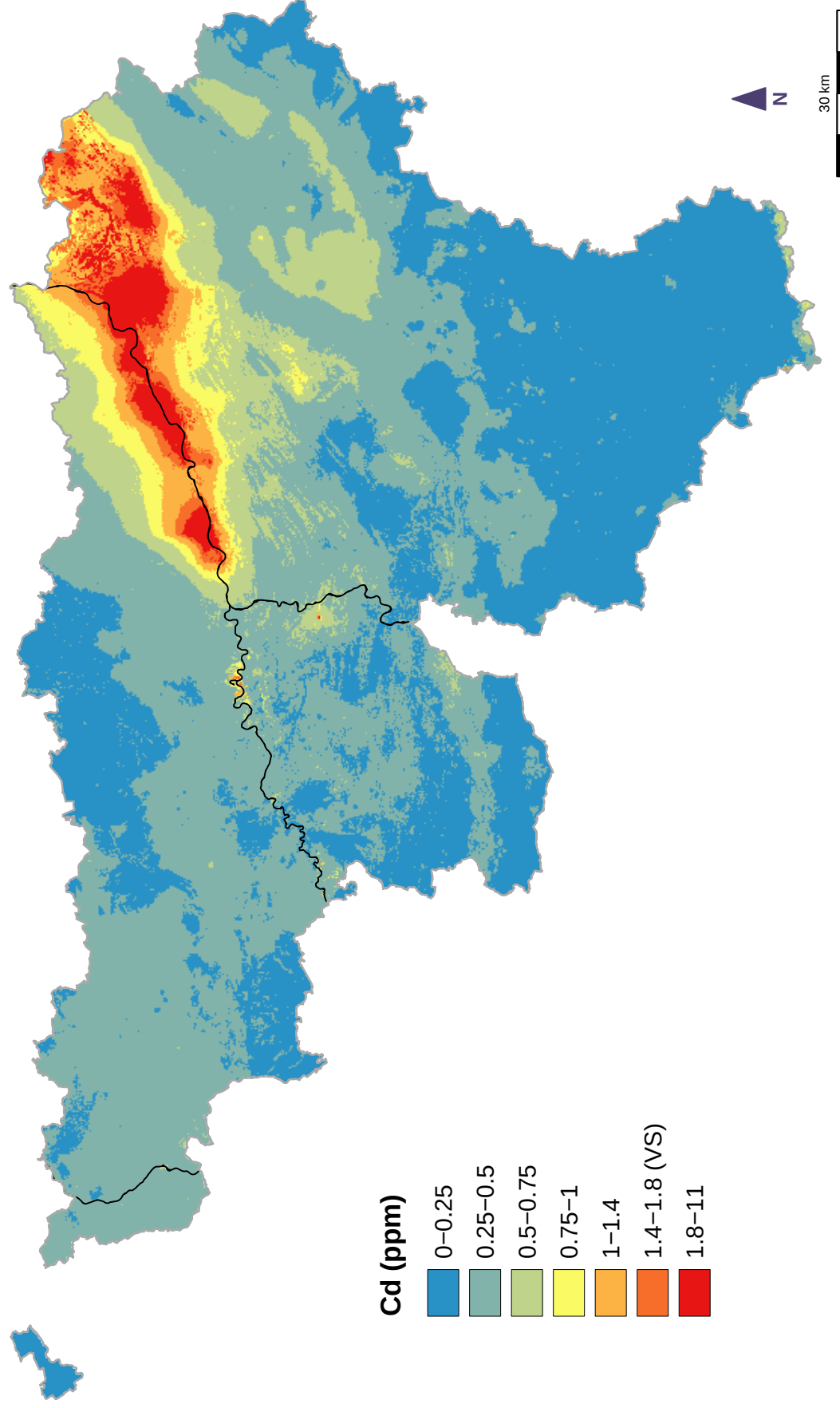
18. Pour Sclaigneaux, il s'agit principalement des données de la thèse d'Amandine Liénard (Liénard, Amandine. Caractérisation multiscalaire de contaminations en éléments traces dans les sols : distribution spatiale et verticale, évaluation du transfert sol-plante-vers de terre et gestion potentielle par amendements organiques. 2016. Thèse de doctorat. Université de Liège, Belgique.)

19. Le manque d'observations peut parfois conduire à ce que l'on visualise des structures spatiales artificielles créées par le krigeage de données, comme celles qui sont visibles à l'extrême-Est de la Wallonie.

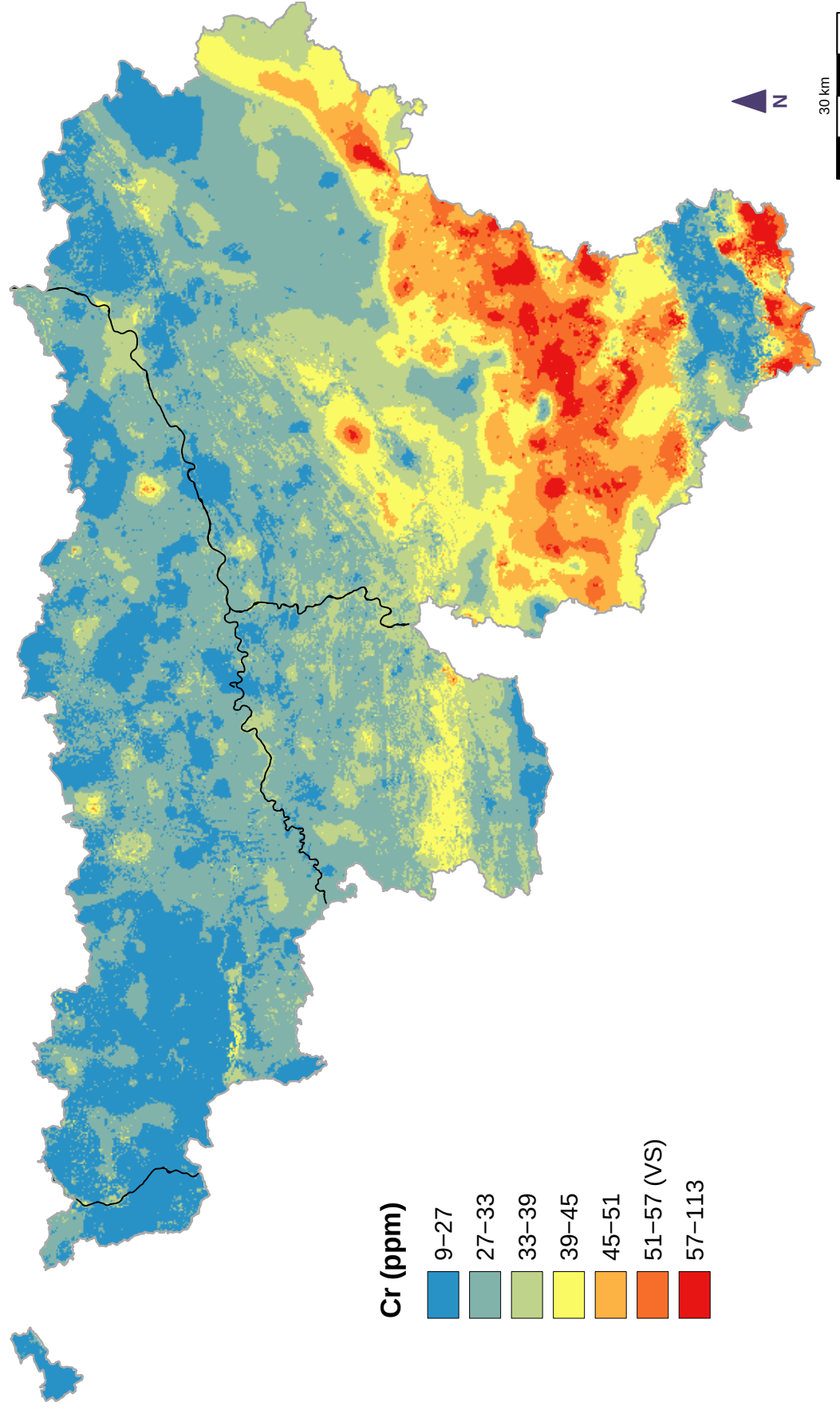
Teneur attendue dans un sol agricole ou forestier



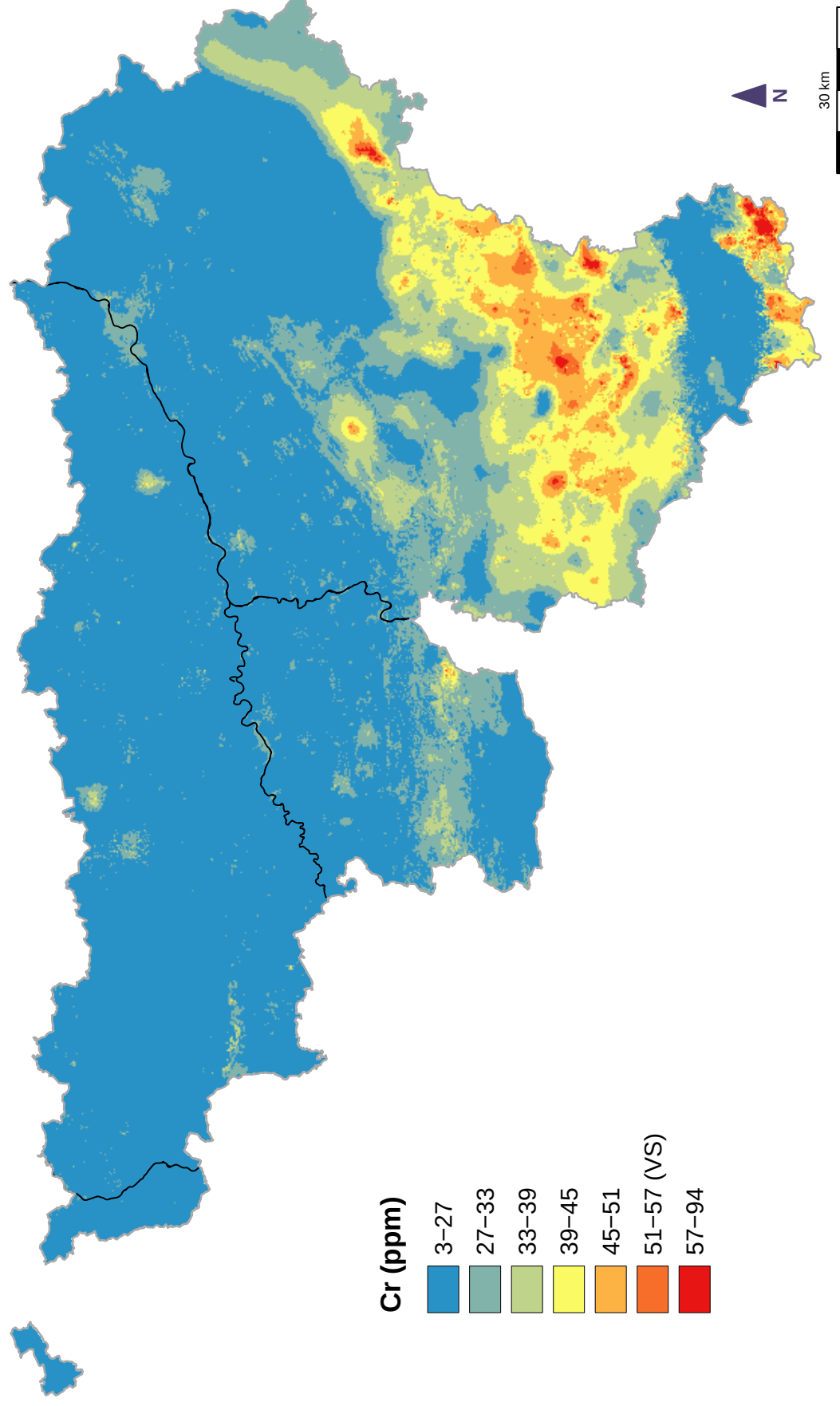
Teneur attendue dans un sol agricole ou forestier



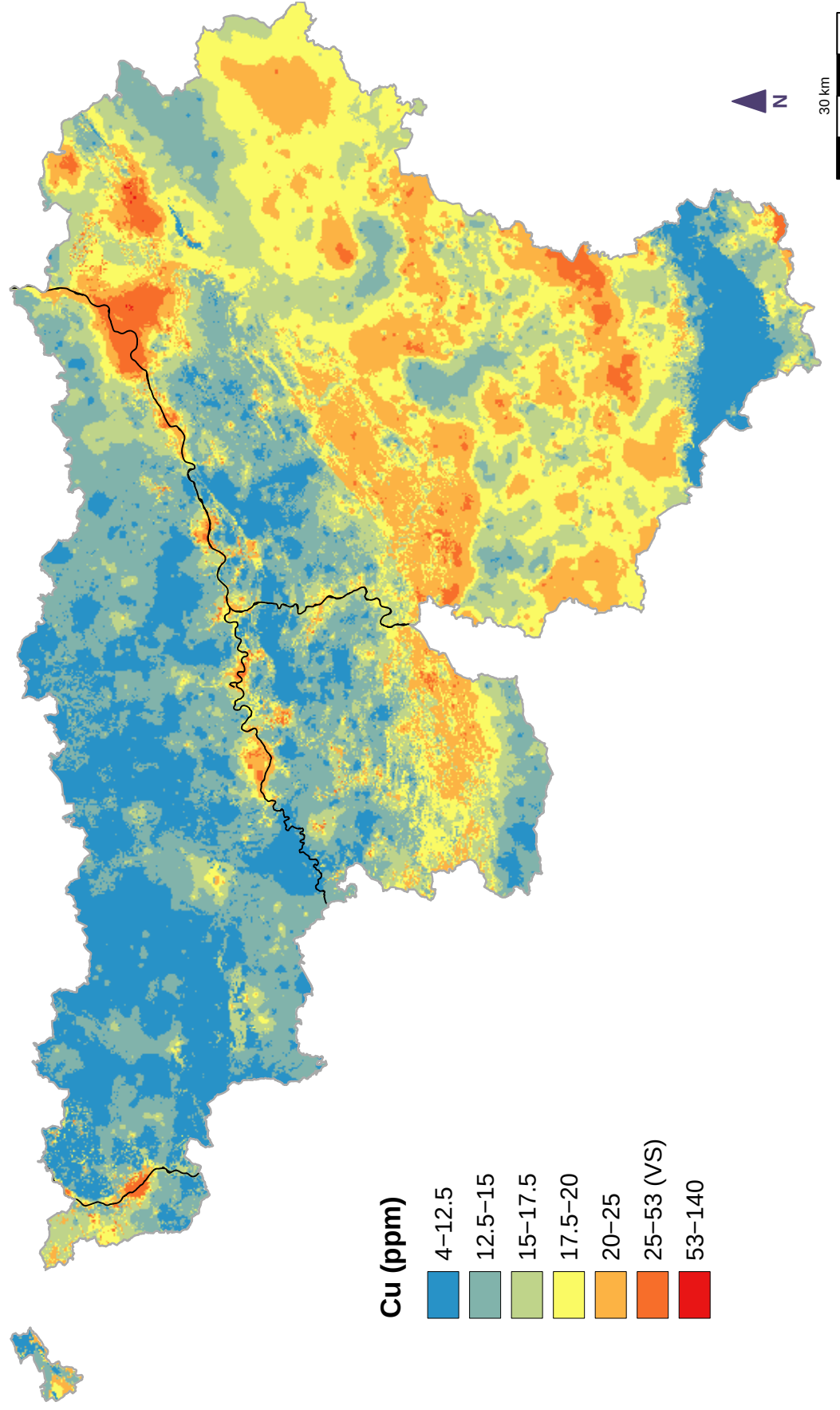
Teneur attendue dans un sol agricole



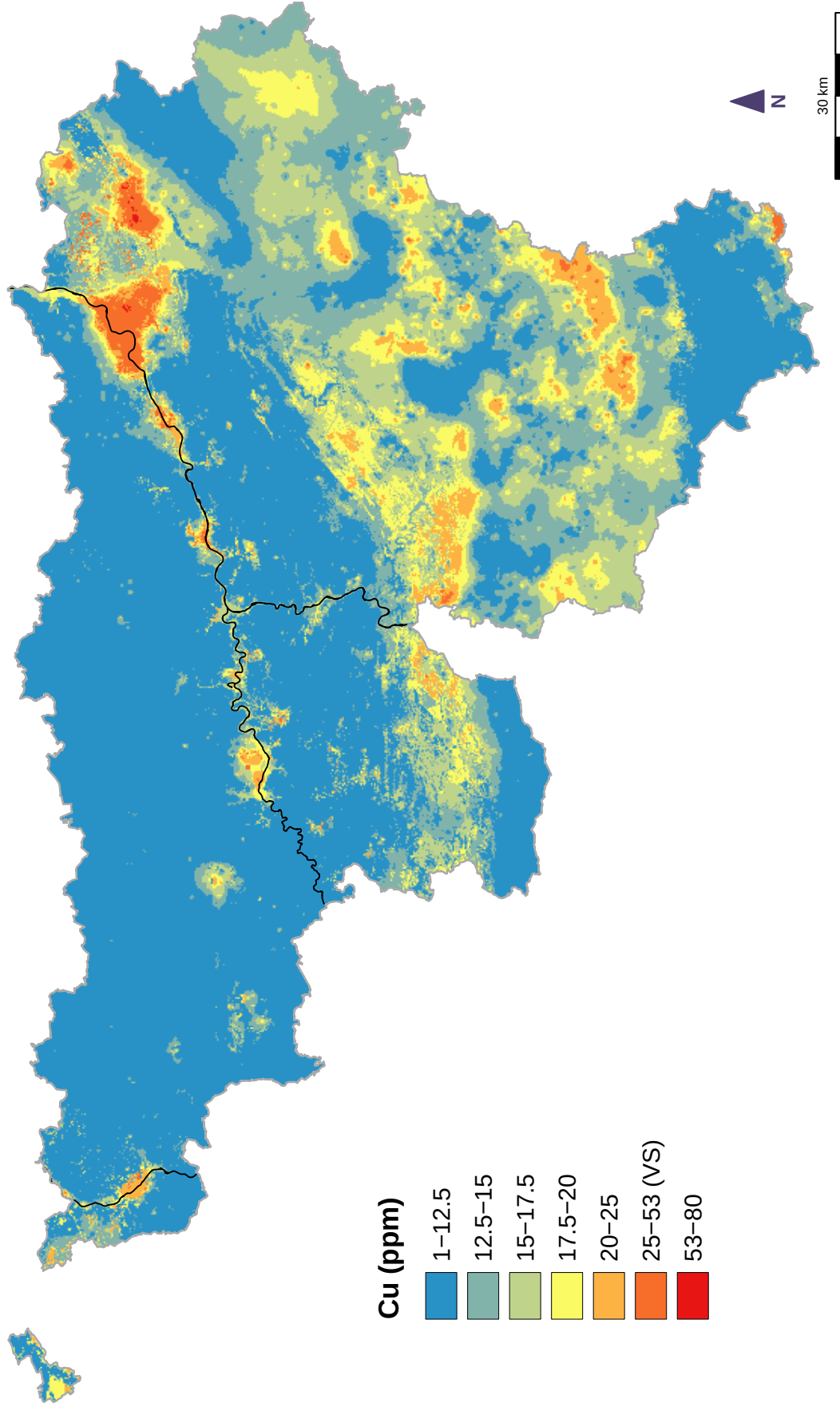
Teneur attendue dans un sol forestier



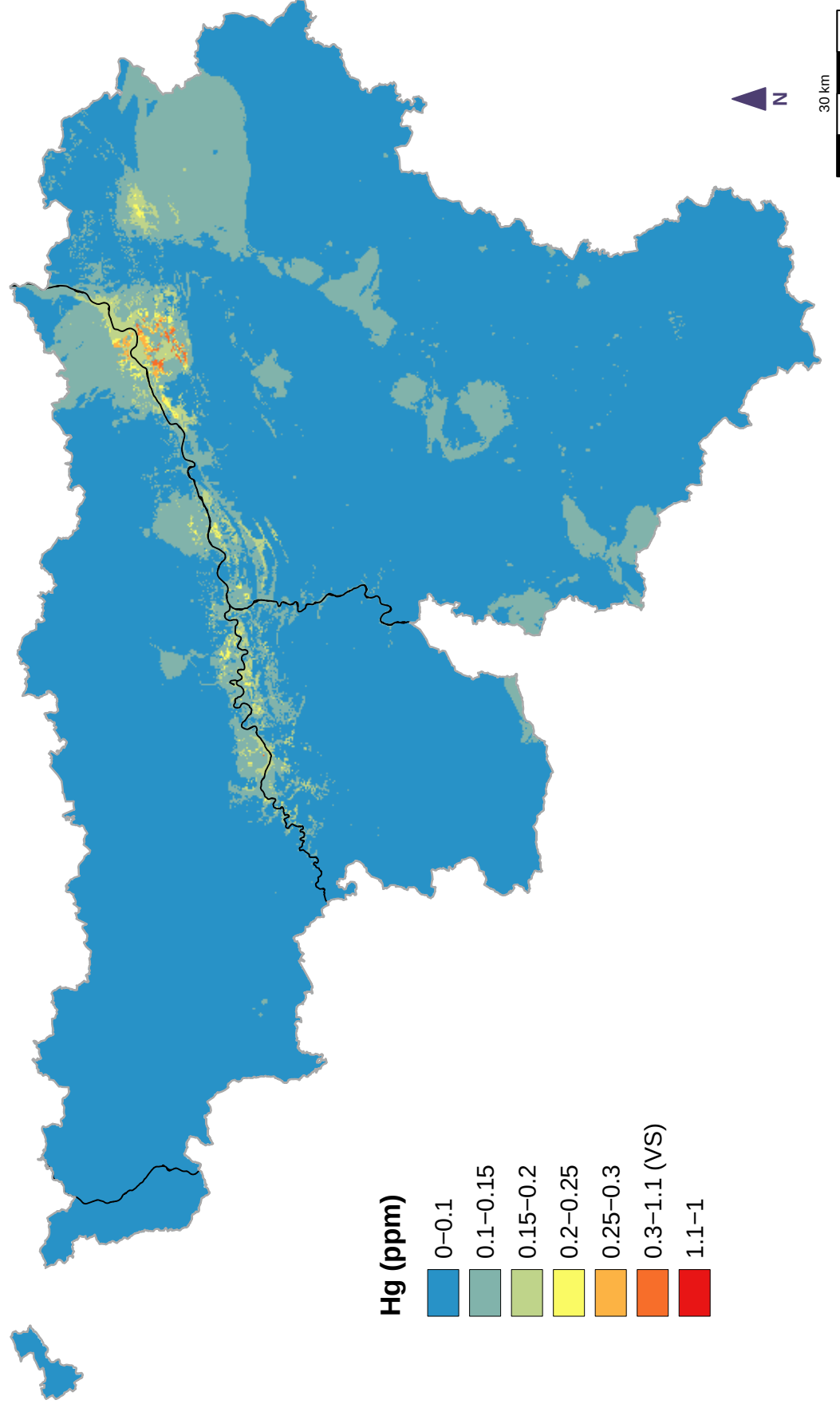
Teneur attendue dans un sol agricole



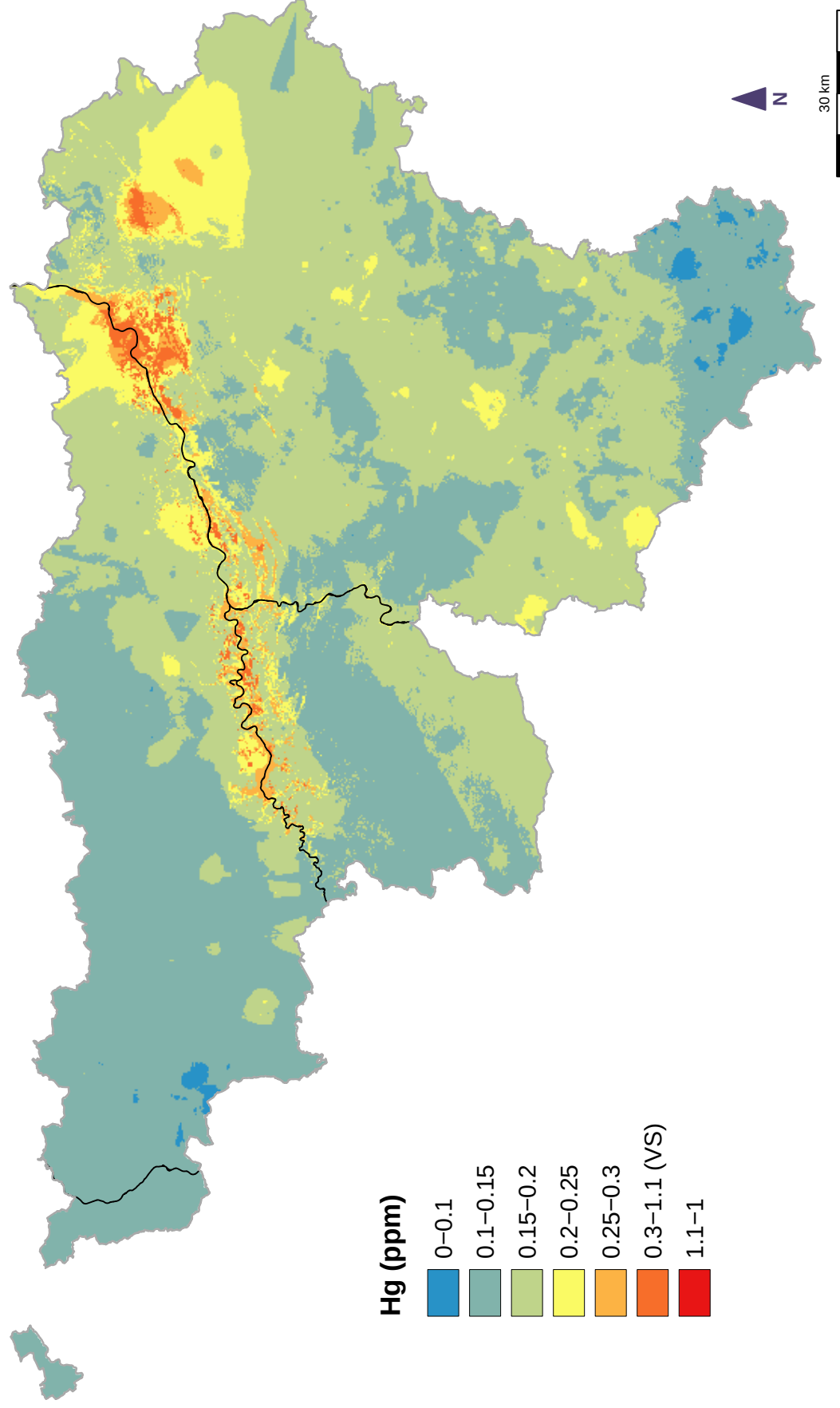
Teneur attendue dans un sol forestier



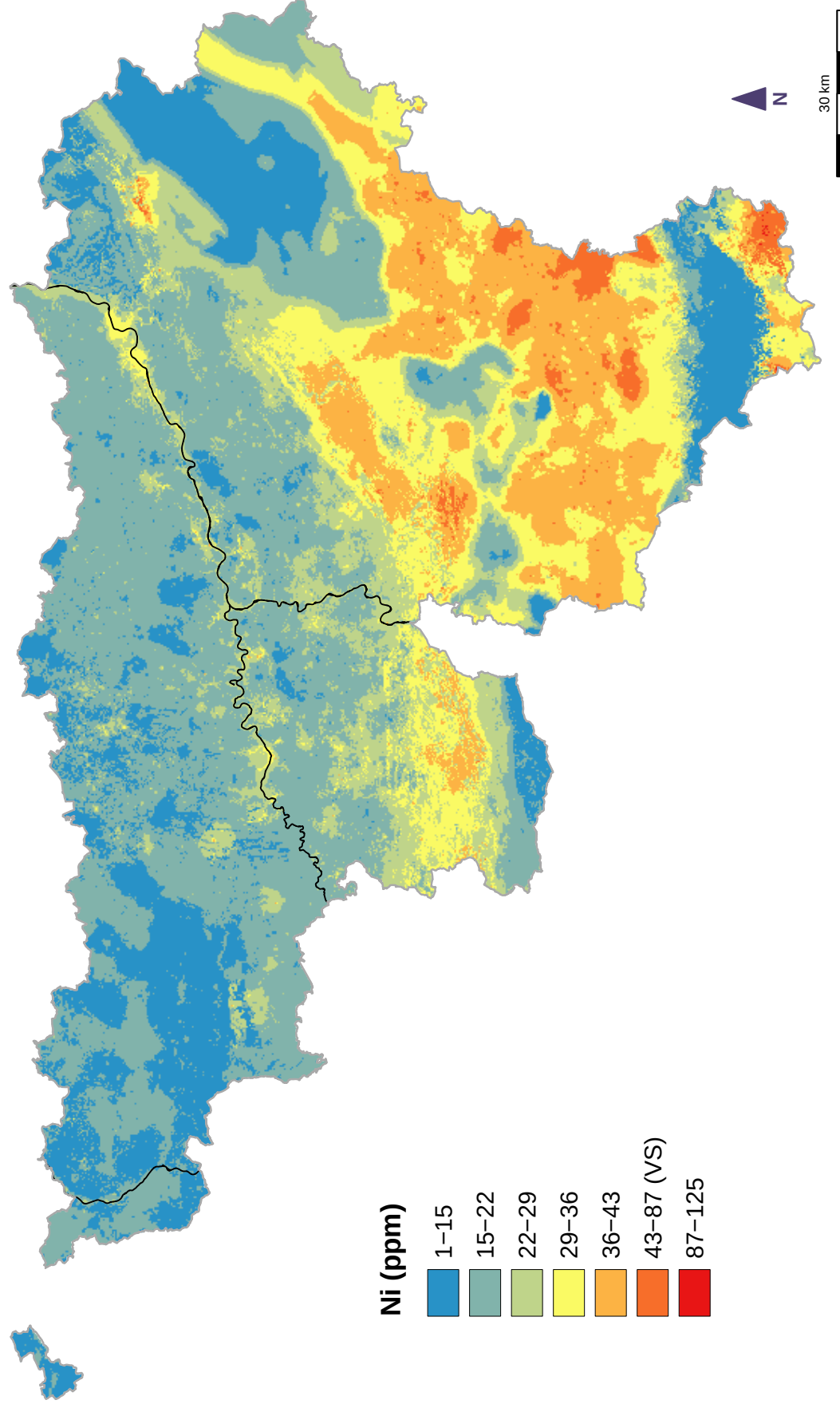
Teneur attendue dans un sol agricole



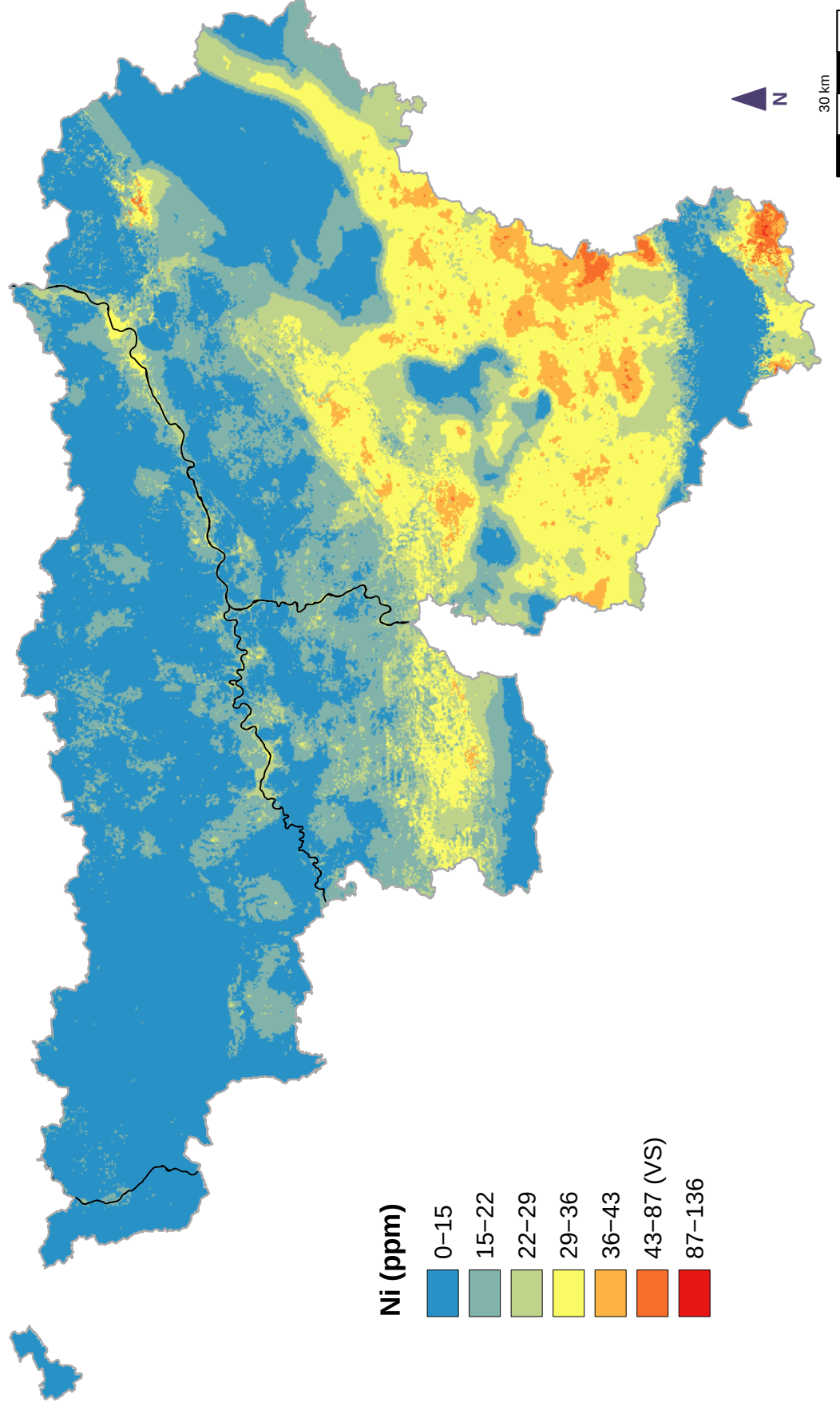
Teneur attendue dans un sol forestier



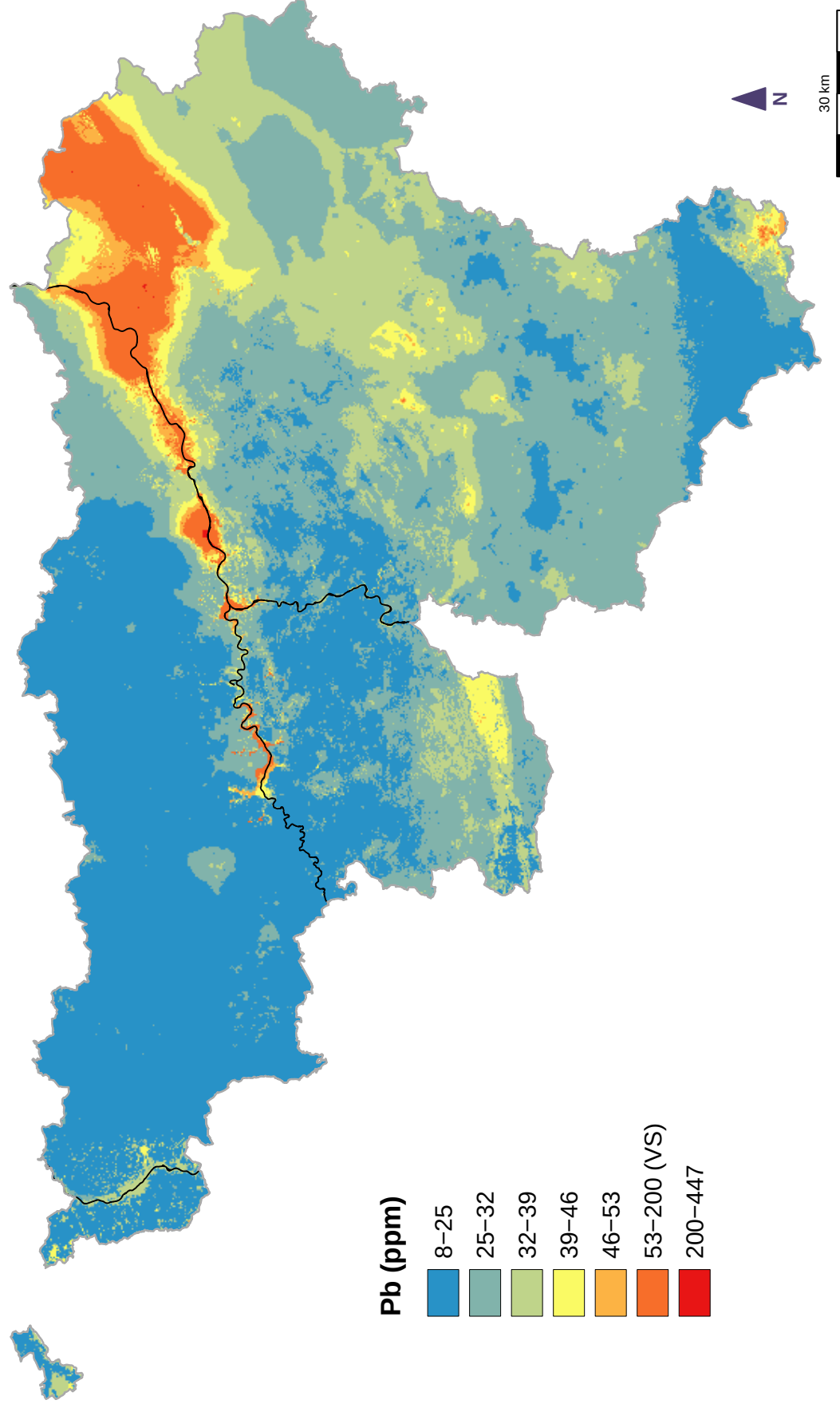
Teneur attendue dans un sol agricole



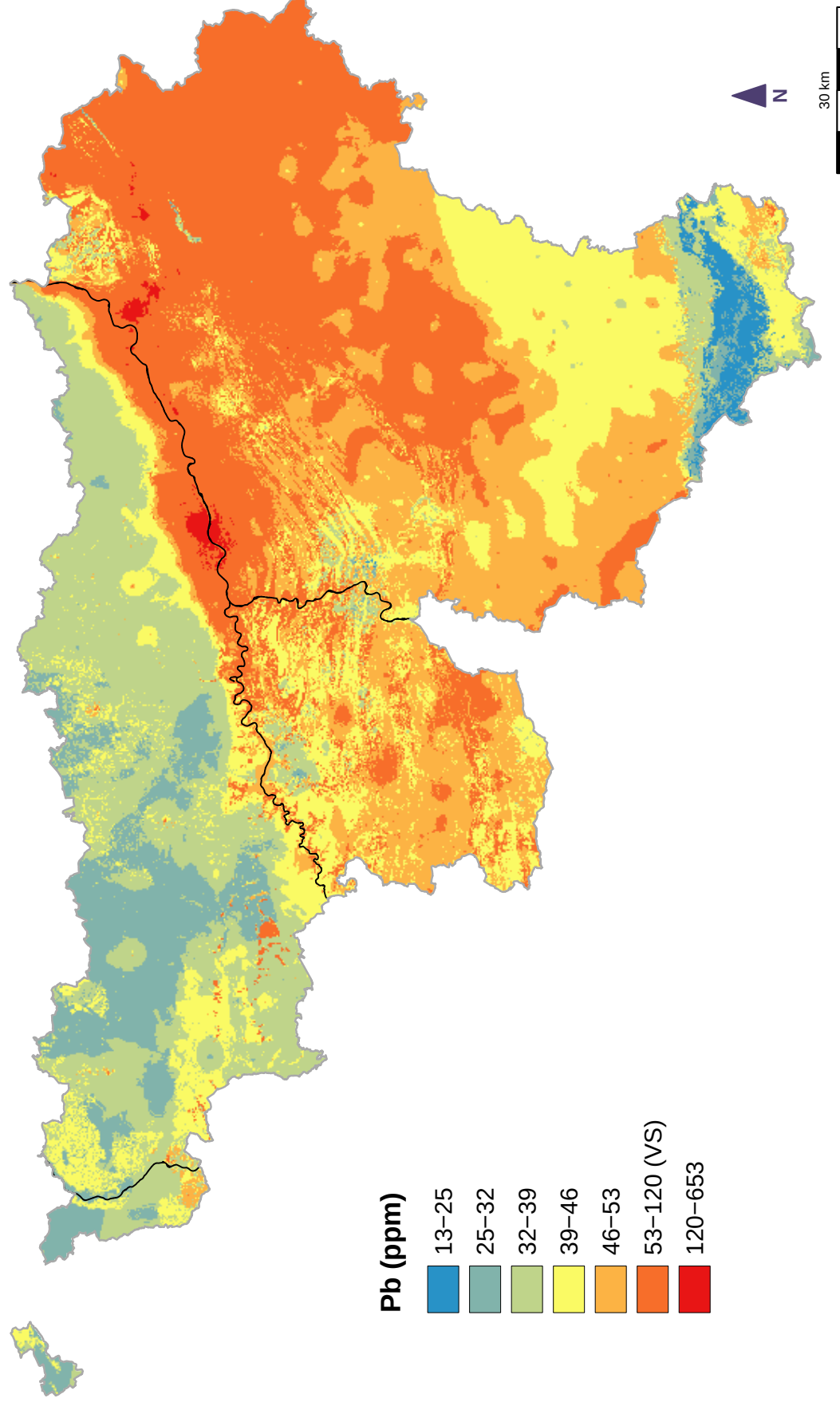
Teneur attendue dans un sol forestier



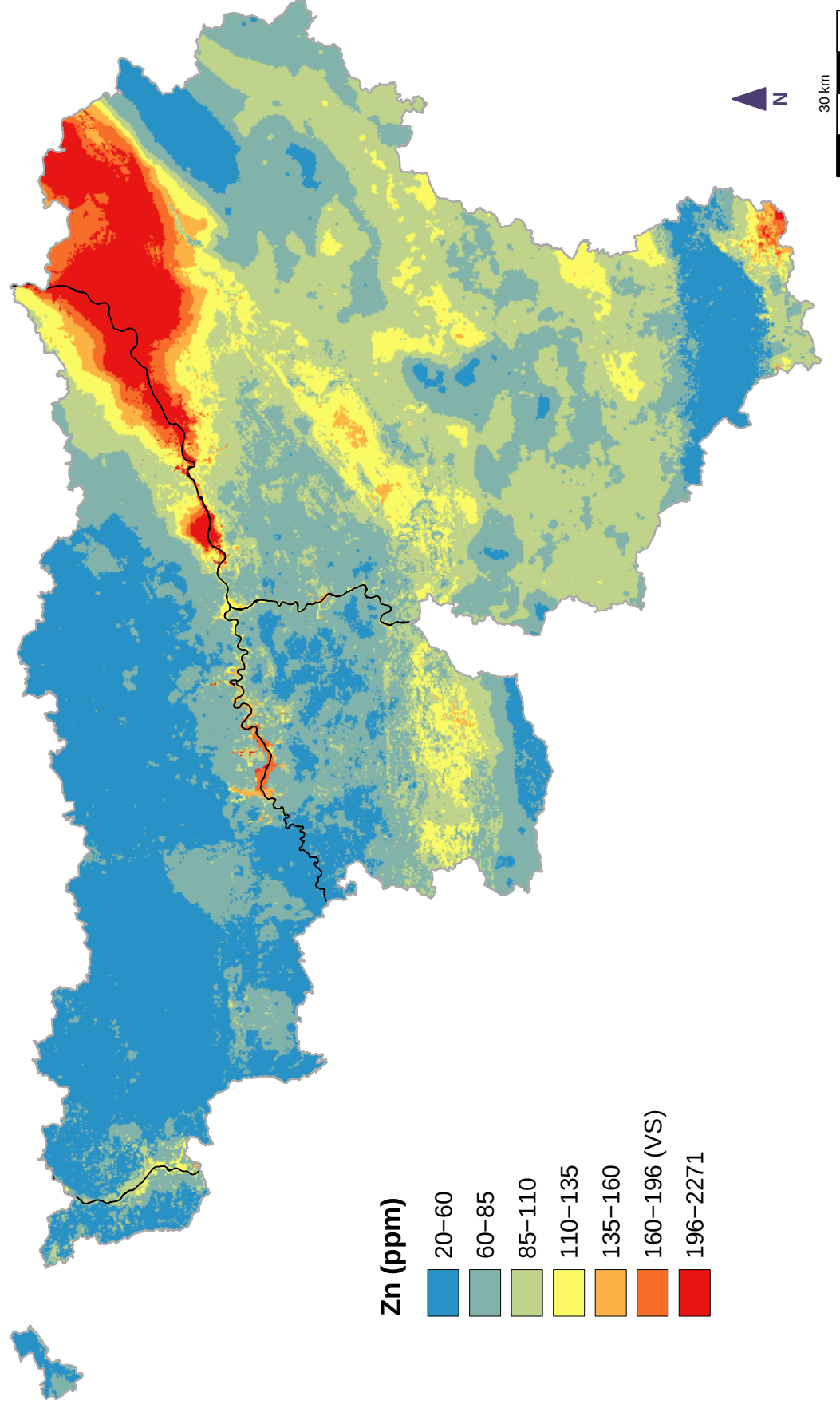
Teneur attendue dans un sol agricole



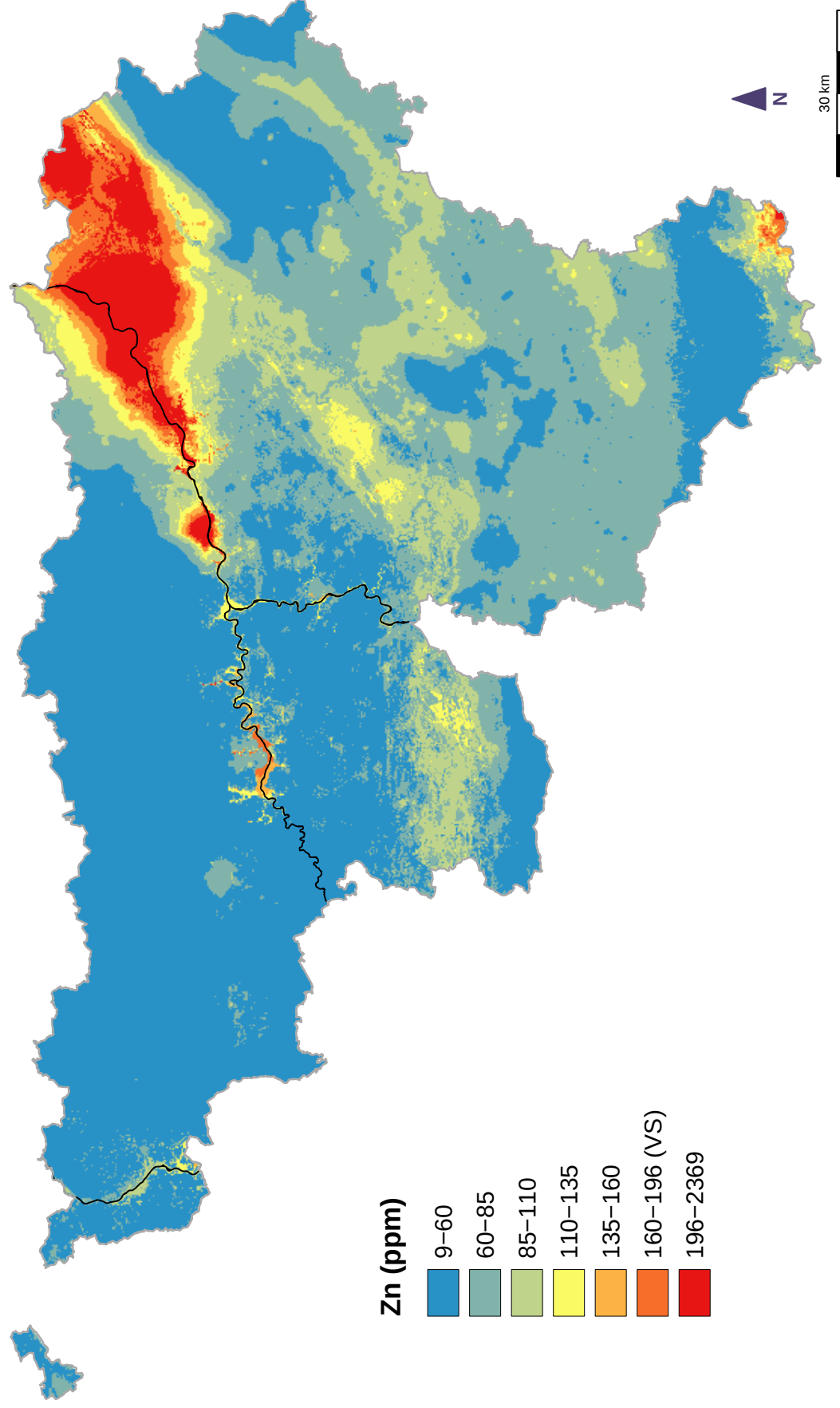
Teneur attendue dans un sol forestier



Teneur attendue dans un sol agricole



Teneur attendue dans un sol forestier



4.6 Carte d'incertitude sur les valeurs renseignées par la carte des concentrations de fond

Les cartes de centiles 90, 95 et 99 produites sont présentées en Annexe 2. Elles ont pour objectif de fournir une indication de l'incertitude sur les valeurs de la carte des concentrations de fond (*cf.* section 4.1).

Lorsque des valeurs élevées sont observées sur les cartes de centiles, cela s'explique souvent par le fait que la distribution des teneurs attendues à cet endroit est caractérisée par une fonction de densité de probabilité correspondant à des niveaux de teneurs élevées (valeur du centile 50 élevé, voir cartes présentées à la section précédente). Cependant, on peut aussi observer des centiles 90, 95 et 99 élevés à des endroits où les centiles 50 sont faibles. Des valeurs élevées sur les cartes de centiles 90, 95 et 99 peuvent aussi s'expliquer par une variance élevée des fonctions de densité de probabilité calculées à ces endroits. Cette variance élevée s'explique soit par une variance de krigeage élevée en raison de l'absence d'observations connues à proximité de l'endroit, soit en raison de fortes variations des teneurs dans le groupe de sols correspondant. Certains groupes de sols peuvent présenter des teneurs plus variables que d'autres, comme c'est le cas des sols développés sur des produits d'altération de calcaires. Les cartes de centiles 90, 95 et 99 reflèteront alors cette variabilité élevée. Les fonctions de densité de probabilité calculées sur ces groupes de sols auront une variance élevée, et donc des valeurs de centiles 90, 95 et 99 plus élevées (voir Figure 4.4 pour la traduction d'une fonction de densité de probabilité en concentration pour un centile).

Les valeurs faibles observées sur les cartes de centiles 90, 95 et 99 peuvent être expliquées par des raisonnements similaires à ceux que nous venons d'exposer. Par exemple, lorsque le groupe de sols présente des teneurs peu variables, comme c'est en général le cas des sols développés sur des loess en région limoneuse, cela se traduira par des fonctions de densité de probabilité à faible variance, et donc à des valeurs de centiles 90, 95 et 99 plus faibles.

4.7 Comparaison des cartes produites avec des valeurs normatives wallonnes

Suite à une demande exprimée par l'Administration et les membres du comité d'accompagnement de la convention, nous avons comparé les cartes produites aux valeurs normatives du décret « sols » et de l'arrêté réglementant l'utilisation des boues²⁰. Pour chaque maille de $100m \times 100m$, l'usage du sol pris en compte est celui du centre de la maille renseigné sur la Carte d'Occupation du Sol de Wallonie (reclassée selon les 5 usages du décret « sols » par l'Administration). Dans cette analyse, les mailles dont le centre de la maille est localisé sur un endroit non classé (route, lac, rivière, etc.) n'ont pas été prises en compte.

20. Arrêté du Gouvernement wallon du 12 avril 1995 réglementant l'utilisation sur ou dans les sols des boues d'épuration ou des boues issues de centres de traitement de gadoues de fosses septiques (M.B. 12.04.1995)

4.7.1 Arrêté réglementant l'utilisation des boues (M.B. 12.04.1995)

Pour la comparaison aux normes de l'arrêté, seules les mailles classées en usage agricole ont été considérées (910 144 mailles sur 1 689 986, soit 53.9% de la Wallonie). Pour ces mailles, les cartes de concentrations de fond pour les sols agricoles ont été comparées aux normes de l'arrêté (Table 4.17). Les résultats de cette comparaison sont présentés dans la Table 4.18. Les cartes des concentrations de fond produites (centile 50) anticipent un dépassement de la norme en Zn (200ppm, *cf.* Table 4.17) sur 3.4% des sols agricoles de Wallonie (*cf.* Table 4.18, dernière ligne, première colonne). C'est la norme qui dépasserait le plus fréquemment les valeurs des cartes des concentrations de fond. Outre Zn, les cartes des concentrations de fond dépassent les normes en Cd, Ni et Pb sur respectivement 1.3%, 0.3% et 0.4% des sols agricoles de Wallonie.

TABLE 4.17 – Conditions réglementaires de l'arrêté du G.W. du 12 avril 1995 pour l'épandage de boues sur un sol.

Matière	Unités	Conditions réglementaires pour l'épandage	
pH (eau)	-	$\geq$	6
Cadmium (Cd)	mg/kg M.S.	$\leq$	2
Cuivre (Cu)	mg/kg M.S.	$\leq$	50
Nickel (Ni)	mg/kg M.S.	$\leq$	50
Plomb (Pb)	mg/kg M.S.	$\leq$	100
Zinc (Zn)	mg/kg M.S.	$\leq$	200
Mercure (Hg)	mg/kg M.S.	$\leq$	1
Chrome (Cr)	mg/kg M.S.	$\leq$	100

M.S. = de matières sèches.

Les cartes de centiles 90, 95 et 99 indiquent un risque de dépassement des normes plus grand pour Ni, et dans une moindre mesure pour Cd et Zn. Cela signifie que pour le Ni, les variances des fonctions de densité de probabilité calculées (*cf.* section 4.6) induisent un risque plus important de dépassement des normes que pour les autres éléments, et ce malgré les pourcentages de dépassements des cartes des concentrations de fond (centile 50) plus faibles pour Ni que pour Zn, Cd et Pb. Cette interprétation mérite cependant d'être nuancée, car on ne peut pas traduire directement les chiffres de la Table 4.18 en probabilité de dépassement de norme. Estimer la probabilité de dépassement de la norme pour chaque maille de la carte revient à rechercher la valeur du centile correspondant à la valeur de la norme, en suivant le principe illustré à la Figure 4.4. La valeur du centile correspondant à la norme fournit la probabilité théorique de dépassement de la valeur de la norme. Comparer les risques de dépassement de normes entre éléments serait donc réalisable, mais uniquement par la comparaison des probabilités de dépassement de normes calculées à chaque maille (et non par la comparaison des valeurs de pourcentages de dépassement de centiles 90, 95 et 99).

TABLE 4.18 – Pourcentage des sols agricoles de la Wallonie qui présentent des teneurs supérieures aux normes de l'arrêté réglementant l'utilisation des boues (voir texte).

	Centile 50	Centile 90	Centile 95	Centile 99
Cd	1.3	4.6	7.2	21.8
Cr	0.0	0.5	1.1	3.0
Cu	0.0	0.3	0.5	1.9
Hg	0.0	0.0	0.0	0.0
Ni	0.3	9.2	17.2	32.4
Pb	0.4	1.3	1.8	3.6
Zn	3.4	5.6	6.7	12.5

4.7.2 Décret « sols » (M.B. 22.03.2018)

Pour la comparaison aux normes du décret « sols », la concentration de fond prise en compte pour les usages 2 à 5 est celle de la valeur prédite pour un sol agricole (*cf.* section 4.1). Pour l'usage 1 (naturel) la valeur prédite pour un sol forestier qui est prise en compte. Les Valeurs Seuils utilisées sont celles reprises à la Table 4.16. Les résultats de cette comparaison sont présentés dans les Tables 4.19, 4.20, 4.21, 4.22, 4.23, 4.24, 4.25, 4.26 ainsi que cartographiquement dans les pages qui suivent.

Pour les cartes de centile 50, les cartes des concentrations de fond produites anticipent un dépassement des VS du Zn sur 2.69% de la Wallonie (*cf.* Table 4.26, dernière ligne, deuxième colonne), soit environ $431.3km^2$ ($15976km^2 \times 2.69/100$). C'est la VS de l'usage agricole qui est la plus fréquemment dépassée. On s'attend en effet à ce qu'environ 3.6% des sols agricoles de Wallonie présentent des concentrations en Zn supérieures à la VS (*cf.* Table 4.26, deuxième ligne, première colonne). A l'échelle de la région wallonne, c'est le polluant pour lequel on s'attend à observer le plus fréquemment un dépassement des VS. Selon les résultats cartographiques présentés dans cette section, ces dépassements sont surtout attendus le long de la Meuse et au nord de la province de Liège. Outre le Zn, à l'échelle de la région wallonne, on s'attend aussi à observer fréquemment un dépassement des VS pour Cd et Cr (pour respectivement 1.4% et 1.3%, représentant une surface de l'ordre de $215km^2$). Pour le Cd, ces dépassements sont attendus aux mêmes endroits que pour le Zn. Pour le Cr, ces dépassements auront lieu principalement au sud de l'Ardenne et de la Lorraine belge. Comme pour le Zn, il s'agit surtout de dépassements de la VS agricole. Dans une moindre mesure, on s'attend aussi à observer des dépassements de VS pour As au sud de la Lorraine belge et pour Pb le long de la Meuse. Pour les autres éléments (Cu, Hg et Ni), on ne s'attend pas à observer des dépassements de VS.

Les valeurs de la carte du centile 90 dépassent les VS sur 12.3% de la Wallonie pour Cr et sur 6.8% de la Wallonie pour As. Il s'agit des deux cartes de centiles 90 qui dépassent le plus fréquemment les VS. Pour le Cr, ces dépassements ont lieu principalement sur des sols en usage naturel et agricole localisés sur le massif ardennais et au sud de la Lorraine belge. Pour l'As, ces dépassements ont lieu principalement sur des sols en usage naturel en Ardenne, sur des matériaux parentaux (schistes, phyllades et grès) datant principalement du Cambrien.

Ces matériaux parentaux présentent une grande variabilité dans les teneurs en As rencontrées, avec en général des teneurs en As plus élevées que sur le reste de la Wallonie (à l'exception des sols situés au sud de la Lorraine belge). Pour Pb, Cd et Zn, les valeurs des cartes de centiles 90 dépassent les VS sur 3.1% à 4.5% de la Wallonie. Ces dépassements sont surtout localisés le long de la Meuse et au nord de la province de Liège. Pour les autres éléments (Cu, Hg et Ni), les valeurs des cartes de centiles 90 ne dépassent presque jamais les VS.

Les valeurs de la carte du centile 95 dépassent les VS sur 18.6% de la Wallonie pour Cr et sur 10.4% de la Wallonie pour As. Il s'agit des deux cartes de centiles 95 qui dépassent le plus fréquemment les VS. Ces dépassements ont lieu principalement sur des sols en usage naturel et agricole localisés aux mêmes endroits que pour les cartes de centiles 90. Pour Cd et Zn, les centiles 95 dépassent les VS sur respectivement 7.5% et 4.7% de la Wallonie. Ces dépassements surviennent surtout sur les usages résidentiels et agricoles. Pour Pb, les centiles 95 dépassent les VS sur 6.8% de la Wallonie, principalement sur des sols en usage naturel (forêts) situés à proximité de la Sambre et de la Meuse ou au nord de la province de Liège. Comme pour les cartes du centile 90, les valeurs des cartes de centiles 95 de Cu, Hg et Ni ne dépassent presque jamais les VS.

Les valeurs de la carte du centile 99 du Cr dépassent les VS sur 38.9% de la Wallonie. Ces dépassements concernent près de la moitié des sols en usage agricole et naturel de la Wallonie. Les valeurs de la carte du centile 99 d'As et Cd dépassent les VS sur environ 23% de la Wallonie. Pour As, ces dépassements concernant 37.6% des sols en usage naturel de Wallonie et environ 15% des sols en usage agricole, résidentiel et commercial. Les dépassements de VS sont principalement localisés en Lorraine belge, en Calestienne, en Ardenne dans le Cambrien et dans le sillon Sambre-et-Meuse. Pour Cd, ces dépassements concernent environ 20% des sols en usages naturel, agricole et résidentiel et sont principalement localisés dans le sillon Sambre-et-Meuse, en Condroz et dans le nord de la province de Liège. Pour Pb, les centiles 99 dépassent les VS sur 14.2% de la Wallonie. Cela concerne presque la moitié (42.3%) des sols en usage naturel (forêts) de Wallonie.

La Table 4.27 fournit le pourcentage de la Wallonie qui dépasse au moins une VS selon l'usage du sol, calculé par comparaison des VS aux cartes produites dans la convention (centiles 50, 90, 95 et 99). Les cartes des concentrations de fond produites (centile 50) anticipent un dépassement d'au moins une VS sur environ 4.3% de la Wallonie, soit environ $687km^2$ (*cf.* Table 4.27, dernière ligne, première colonne). Les valeurs des cartes de centiles 90, 95 et 99 dépassent au moins une VS sur respectivement 25.1%, 37.5% et 68.2% de la Wallonie. Quelle que soit la carte de centile, les dépassements de VS pour les sols classés en usage industriel sont peu fréquents. Lorsqu'ils sont observés, ces dépassements concernent surtout As et Zn, et dans une moindre mesure Cd. A l'inverse, les dépassements de VS pour les sols classés en usage naturel sont les plus fréquents. Pour cet usage du sol, les valeurs des cartes de centiles 90, 95 et 99 dépassent au moins une VS sur respectivement 30.3%, 50.0% et 87.2% de la Wallonie. Lorsqu'ils sont observés, ces dépassements concernent surtout Cr, Pb et As, et dans une moindre mesure Cd.

Notons finalement que les dépassements des VS du Hg sont très peu fréquents, pour toutes les cartes produites.

TABLE 4.19 – As : pourcentage de la Wallonie qui dépasse les VS selon l’usage du sol, et selon l’usage attribué à la maille (dernière ligne).

	Surf. (km^2)	Centile 50	Centile 90	Centile 95	Centile 99
Naturel	4869	0.15	13.33	19.69	37.57
Agricole	9101	0.26	4.56	6.90	17.75
Résidentiel	1505	0.16	1.26	4.03	14.65
Récréatif/Commercial	221	0.06	1.96	4.52	16.12
Industriel	278	0.04	0.06	0.43	5.22
Usage à la maille	15976	0.21	6.81	10.38	23.25

TABLE 4.20 – Cd : pourcentage de la Wallonie qui dépasse les VS selon l’usage du sol, et selon l’usage attribué à la maille (dernière ligne).

	Surf. (km^2)	Centile 50	Centile 90	Centile 95	Centile 99
Naturel	4869	0.82	2.58	4.77	21.41
Agricole	9101	1.87	5.42	8.87	24.85
Résidentiel	1505	1.15	6.73	10.43	19.07
Récréatif/Commercial	221	0.00	0.71	2.56	7.18
Industriel	278	0.00	0.00	0.03	1.22
Usage à la maille	15976	1.42	4.52	7.52	22.60

TABLE 4.21 – Cr : pourcentage de la Wallonie qui dépasse les VS selon l’usage du sol, et selon l’usage attribué à la maille (dernière ligne).

	Surf. (km^2)	Centile 50	Centile 90	Centile 95	Centile 99
Naturel	4869	0.08	10.15	19.48	44.96
Agricole	9101	2.18	15.88	21.89	42.24
Résidentiel	1505	0.01	1.32	2.55	11.44
Récréatif/Commercial	221	0.00	0.00	0.06	0.66
Industriel	278	0.00	0.00	0.00	0.00
Usage à la maille	15976	1.27	12.27	18.65	38.85

TABLE 4.22 – Cu : pourcentage de la Wallonie qui dépasse les VS selon l’usage du sol, et selon l’usage attribué à la maille (dernière ligne).

	Surf. (km^2)	Centile 50	Centile 90	Centile 95	Centile 99
Naturel	4869	0.01	0.26	0.59	3.24
Agricole	9101	0.01	0.22	0.44	1.52
Résidentiel	1505	0.00	0.01	0.02	0.19
Récréatif/Commercial	221	0.00	0.00	0.00	0.00
Industriel	278	0.00	0.00	0.00	0.00
Usage à la maille	15976	0.01	0.21	0.43	1.87

TABLE 4.23 – Hg : pourcentage de la Wallonie qui dépasse les VS selon l’usage du sol, et selon l’usage attribué à la maille (dernière ligne).

	Surf. (km^2)	Centile 50	Centile 90	Centile 95	Centile 99
Naturel	4869	0.00	0.01	0.06	0.30
Agricole	9101	0.00	0.00	0.00	0.02
Résidentiel	1505	0.00	0.00	0.00	0.00
Récréatif/Commercial	221	0.00	0.00	0.00	0.00
Industriel	278	0.00	0.00	0.00	0.00
Usage à la maille	15976	0.00	0.00	0.02	0.11

TABLE 4.24 – Ni : pourcentage de la Wallonie qui dépasse les VS selon l’usage du sol, et selon l’usage attribué à la maille (dernière ligne).

	Surf. (km^2)	Centile 50	Centile 90	Centile 95	Centile 99
Naturel	4869	0.00	0.11	0.27	1.70
Agricole	9101	0.01	0.19	0.38	1.76
Résidentiel	1505	0.00	0.00	0.01	0.11
Récréatif/Commercial	221	0.00	0.00	0.00	0.00
Industriel	278	0.00	0.00	0.00	0.00
Usage à la maille	15976	0.01	0.14	0.30	1.53

TABLE 4.25 – Pb : pourcentage de la Wallonie qui dépasse les VS selon l’usage du sol, et selon l’usage attribué à la maille (dernière ligne).

	Surf. (km^2)	Centile 50	Centile 90	Centile 95	Centile 99
Naturel	4869	0.23	8.73	19.67	42.28
Agricole	9101	0.01	0.43	0.76	1.27
Résidentiel	1505	0.05	2.31	3.56	5.44
Récréatif/Commercial	221	0.00	0.22	1.26	4.62
Industriel	278	0.00	0.00	0.00	0.00
Usage à la maille	15976	0.08	3.13	6.78	14.19

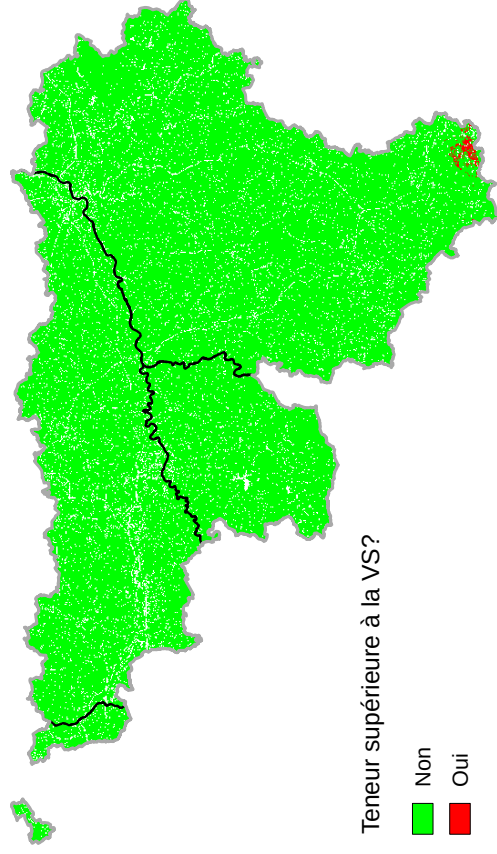
TABLE 4.26 – Zn : pourcentage de la Wallonie qui dépasse les VS selon l’usage du sol, et selon l’usage attribué à la maille (dernière ligne).

	Surf. (km^2)	Centile 50	Centile 90	Centile 95	Centile 99
Naturel	4869	1.35	3.02	4.15	10.88
Agricole	9101	3.63	5.80	7.11	13.58
Résidentiel	1505	2.29	4.45	5.05	6.43
Récréatif/Commercial	221	0.00	0.09	0.27	2.23
Industriel	278	0.00	0.39	1.31	5.04
Usage à la maille	15976	2.69	4.65	5.82	11.78

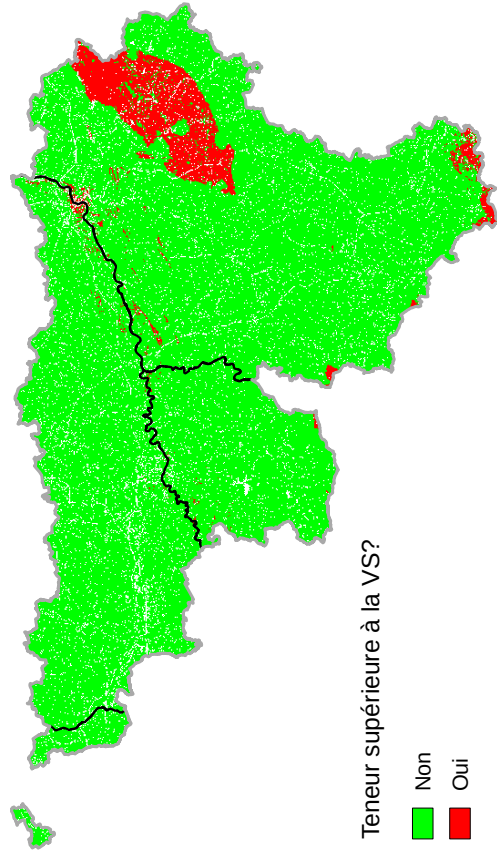
TABLE 4.27 – Pourcentage de la Wallonie qui dépasse au moins une VS selon l’usage du sol, et selon l’usage attribué à la maille (dernière ligne).

	Surf. (km^2)	Centile 50	Centile 90	Centile 95	Centile 99
Naturel	4869	1.80	30.70	50.50	87.60
Agricole	9101	6.10	26.10	36.20	66.20
Résidentiel	1505	3.00	9.40	15.80	37.20
Récréatif/Commercial	221	0.10	2.70	6.20	18.90
Industriel	278	0.00	0.40	1.70	7.90
Usage à la maille	15976	4.30	25.20	37.60	68.30

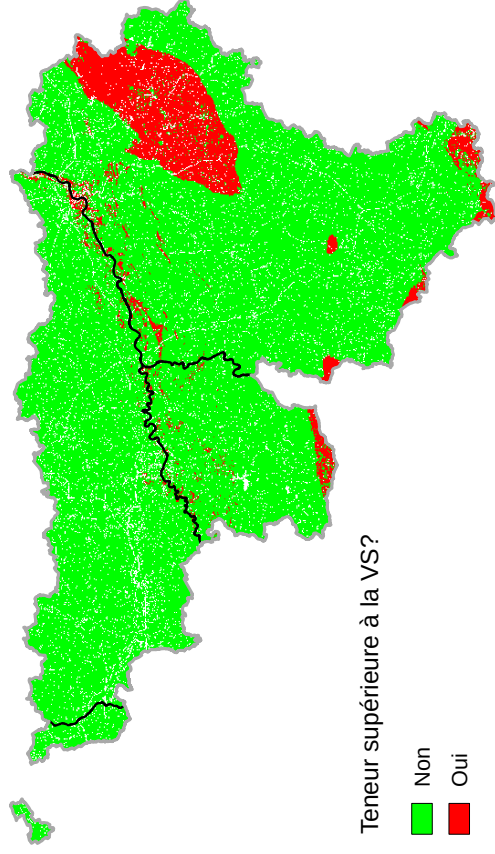
As : teneur attendue



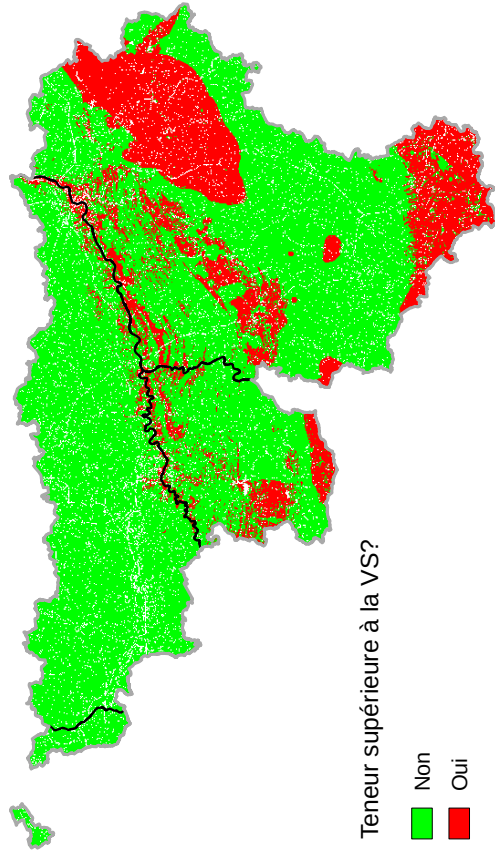
As : teneur qui a moins d'1 chance sur 10 d'être dépassée



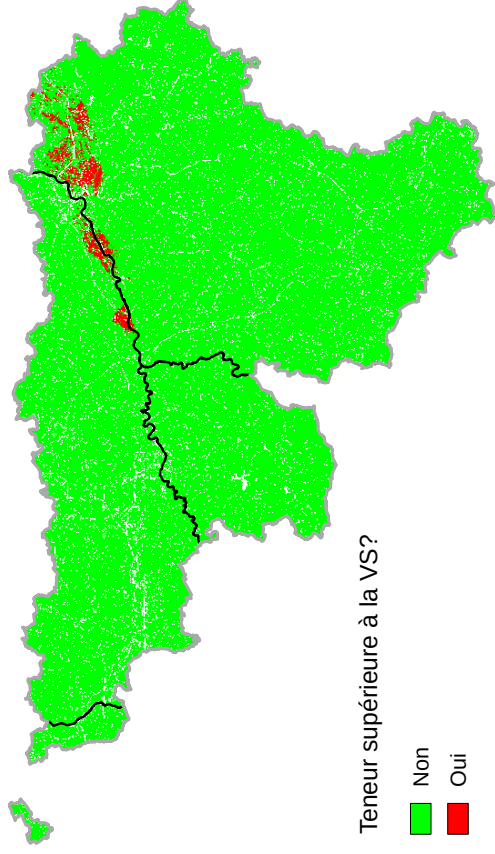
As : teneur qui a moins d'1 chance sur 20 d'être dépassée



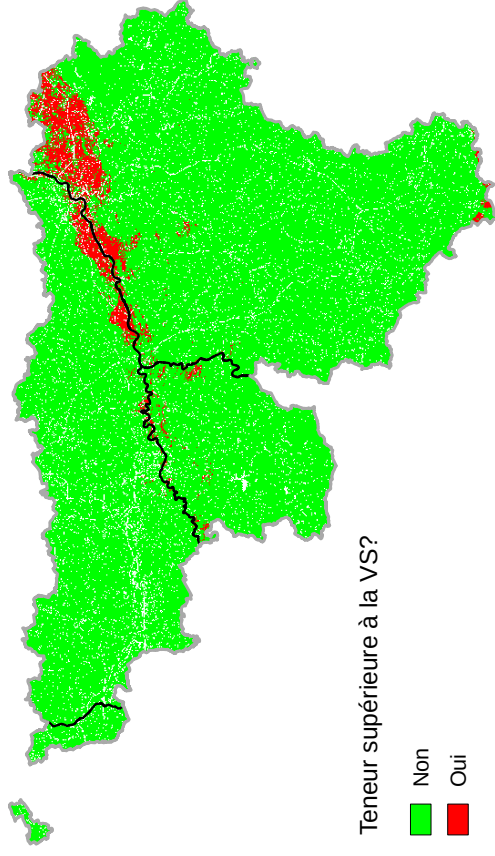
As : teneur qui a moins d'1 chance sur 100 d'être dépassée



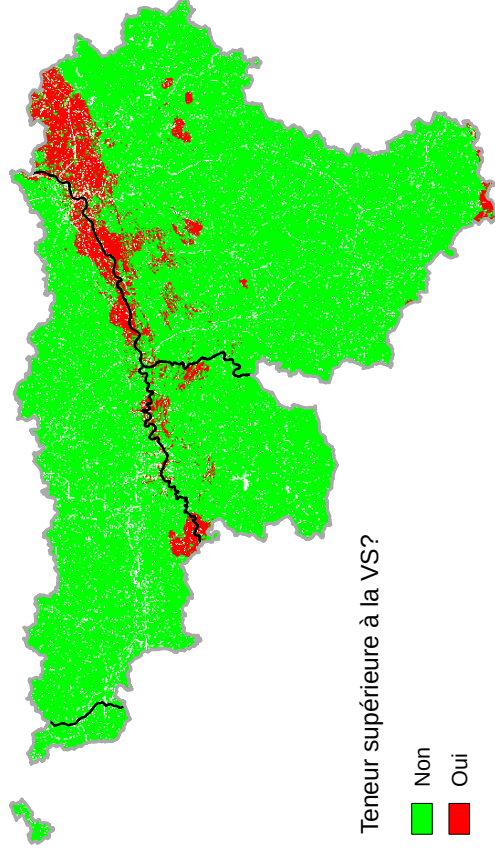
Cd : teneur attendue



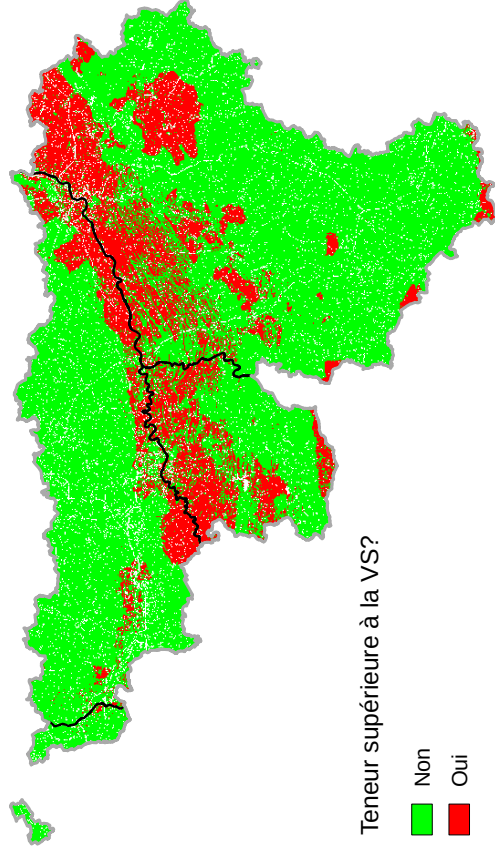
Cd : teneur qui a moins d'1 chance sur 10 d'être dépassée



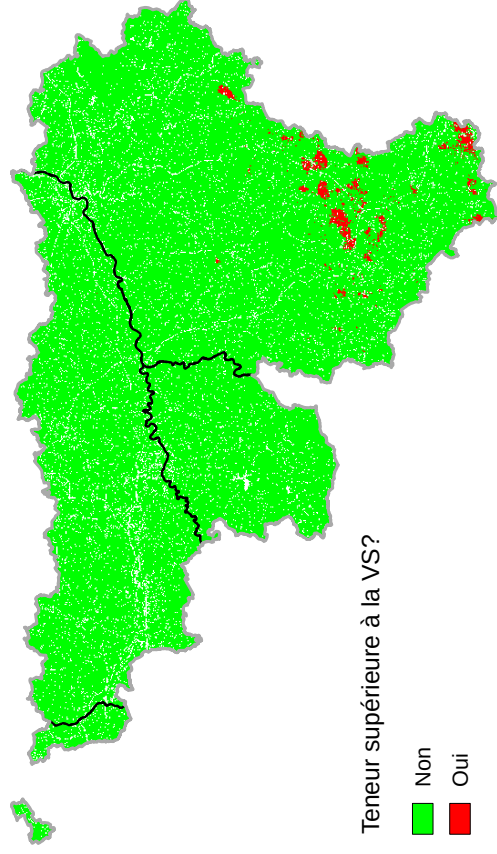
Cd : teneur qui a moins d'1 chance sur 20 d'être dépassée



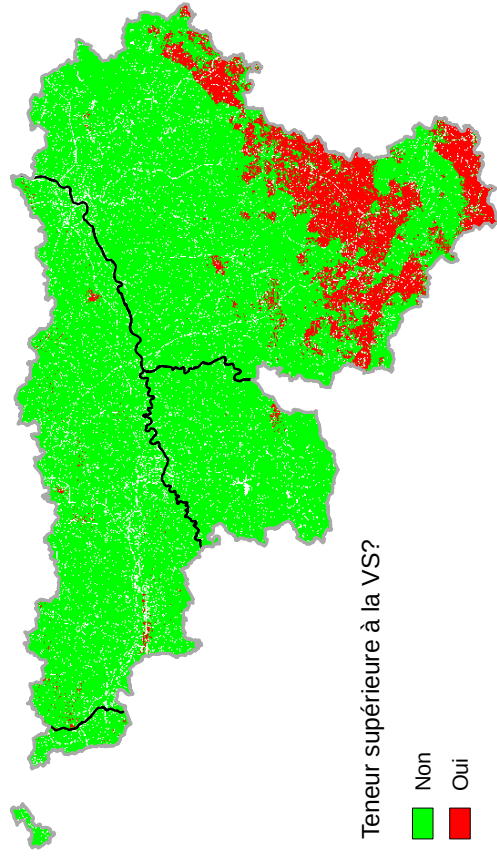
Cd : teneur qui a moins d'1 chance sur 100 d'être dépassée



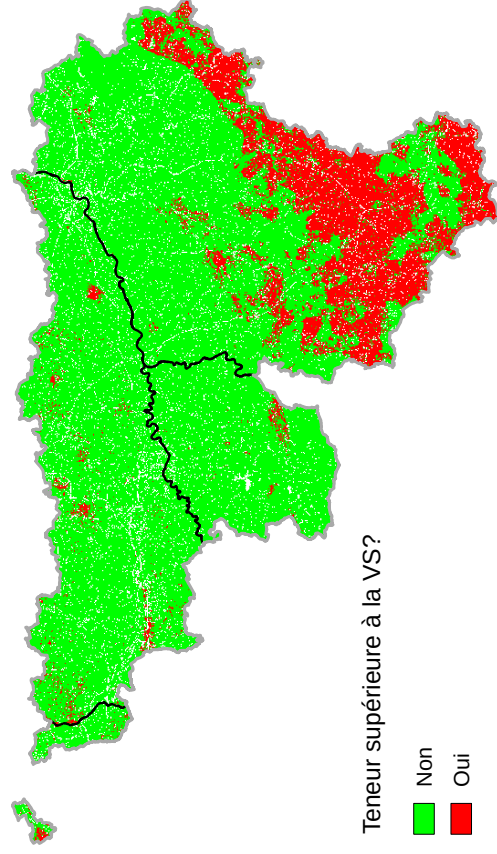
Cr : teneur attendue



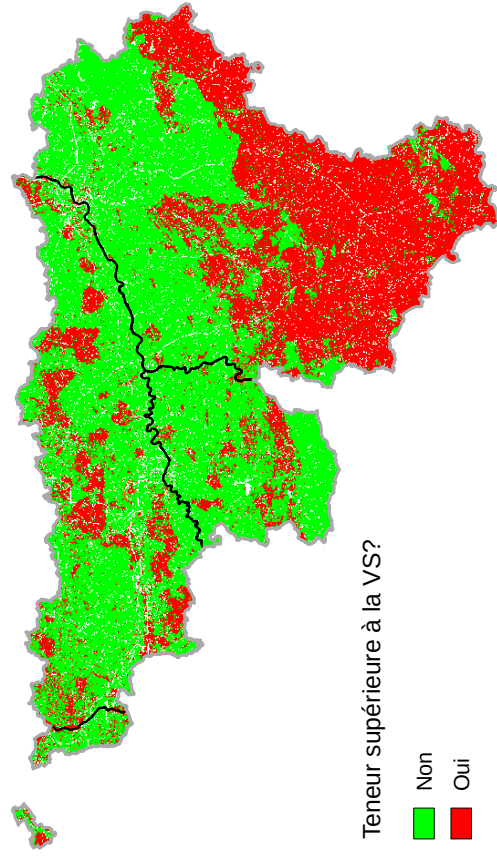
Cr : teneur qui a moins d'1 chance sur 10 d'être dépassée



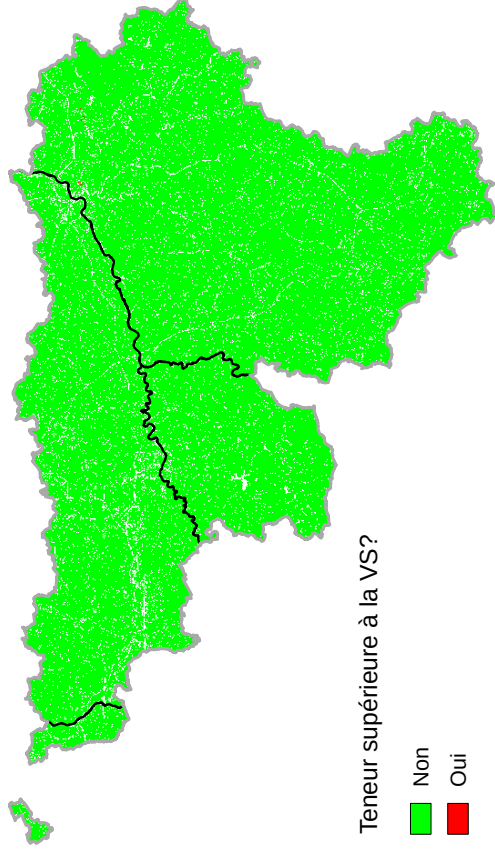
Cr : teneur qui a moins d'1 chance sur 20 d'être dépassée



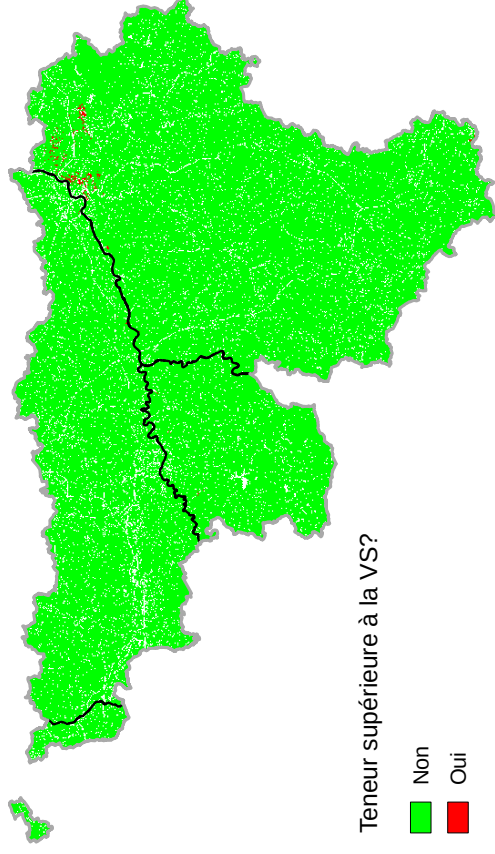
Cr : teneur qui a moins d'1 chance sur 100 d'être dépassée



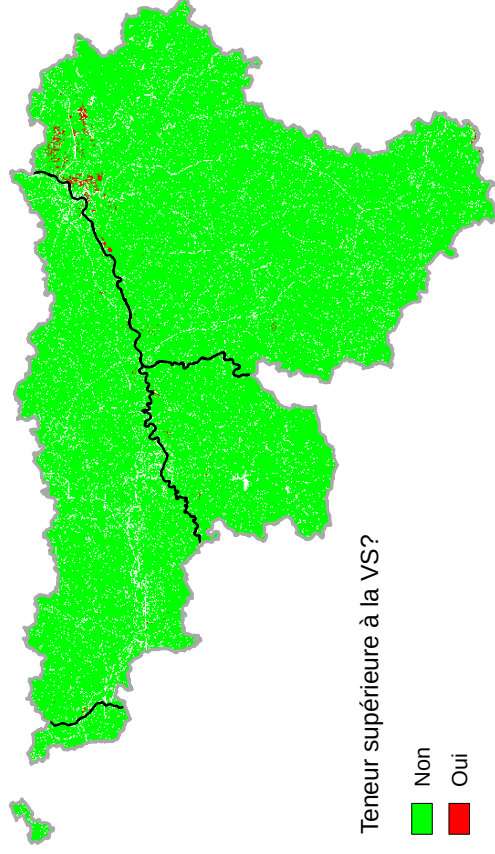
Cu : teneur attendue



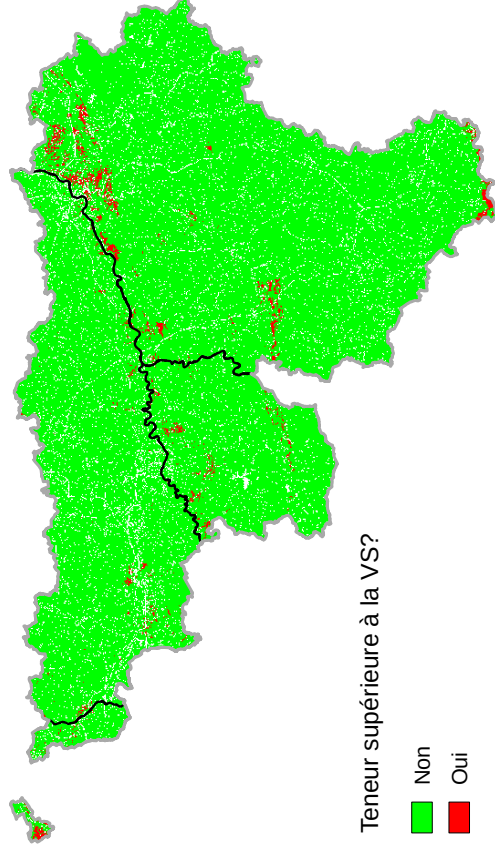
Cu : teneur qui a moins d'1 chance sur 10 d'être dépassée



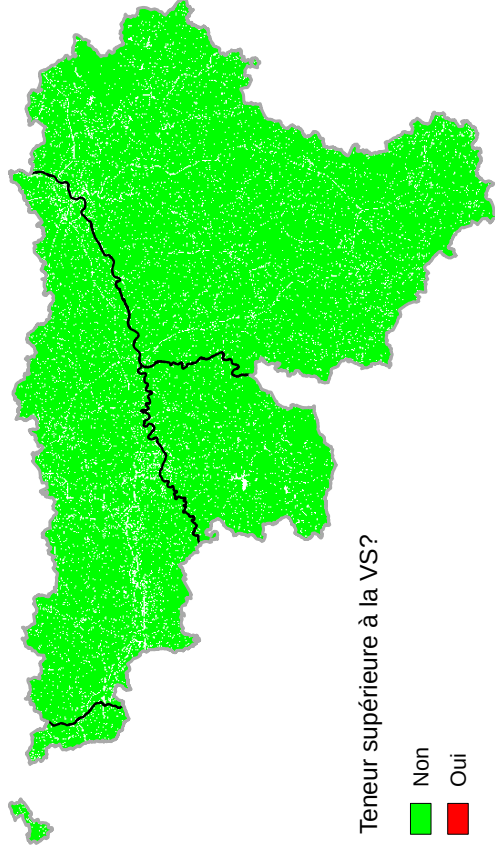
Cu : teneur qui a moins d'1 chance sur 20 d'être dépassée



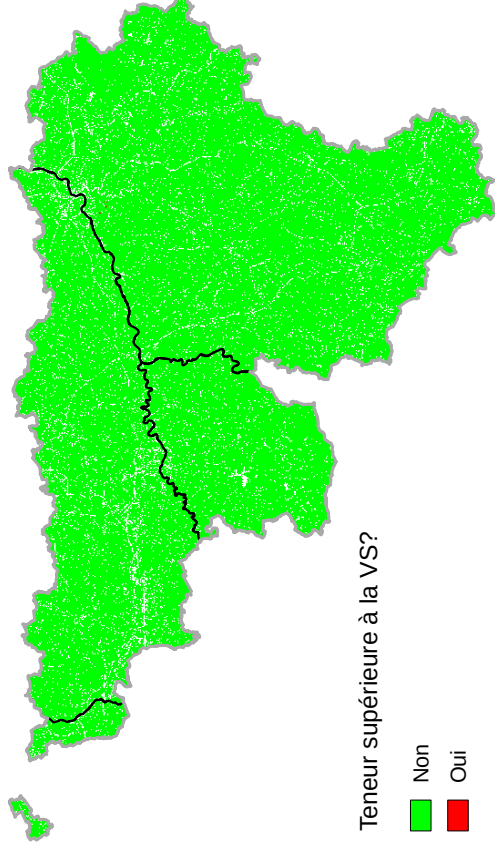
Cu : teneur qui a moins d'1 chance sur 100 d'être dépassée



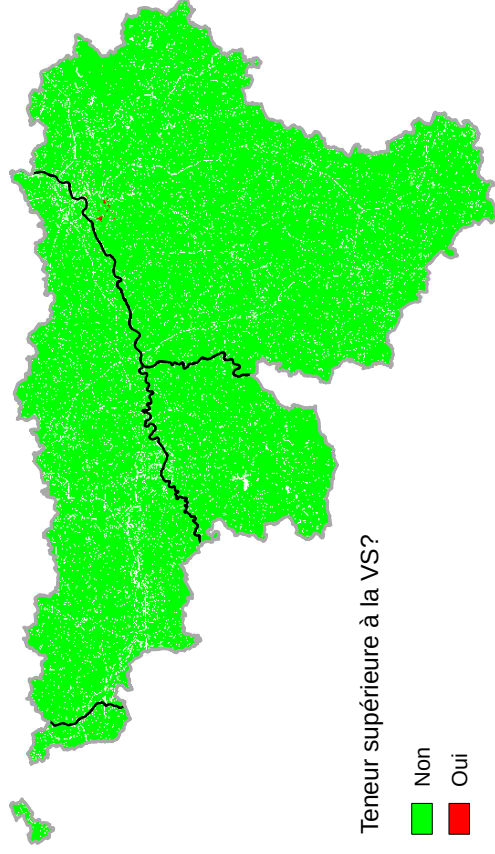
Hg : teneur attendue



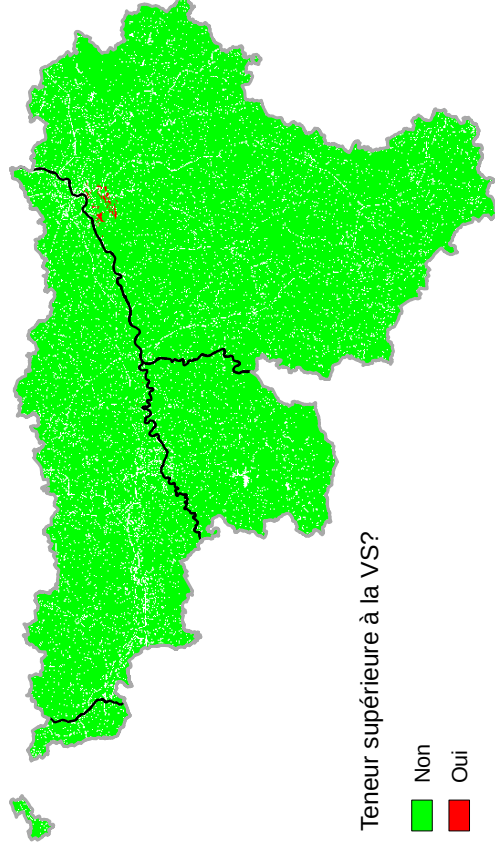
Hg : teneur qui a moins d'1 chance sur 10 d'être dépassée



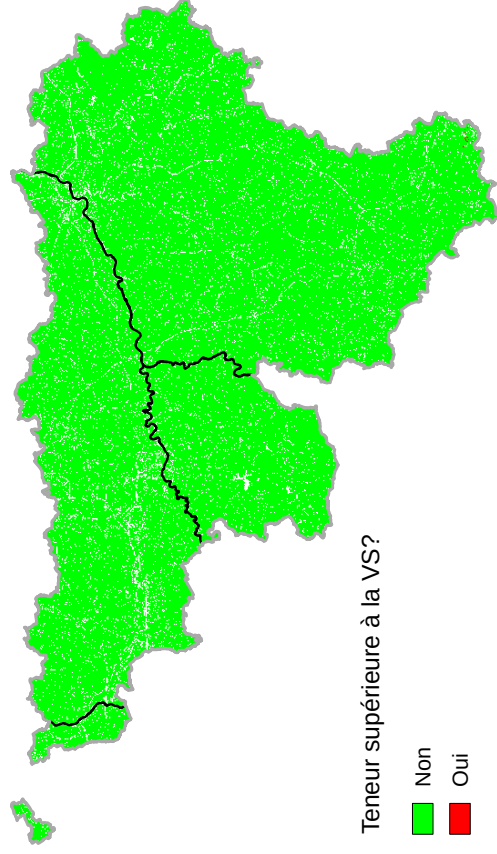
Hg : teneur qui a moins d'1 chance sur 20 d'être dépassée



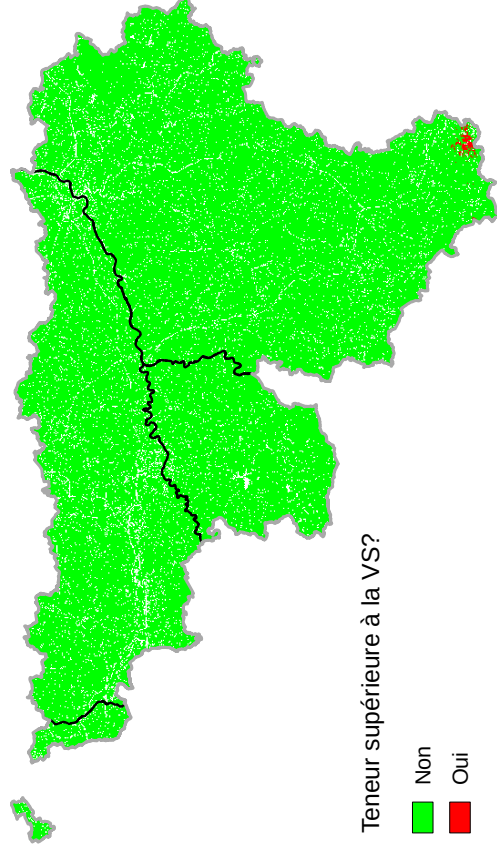
Hg : teneur qui a moins d'1 chance sur 100 d'être dépassée



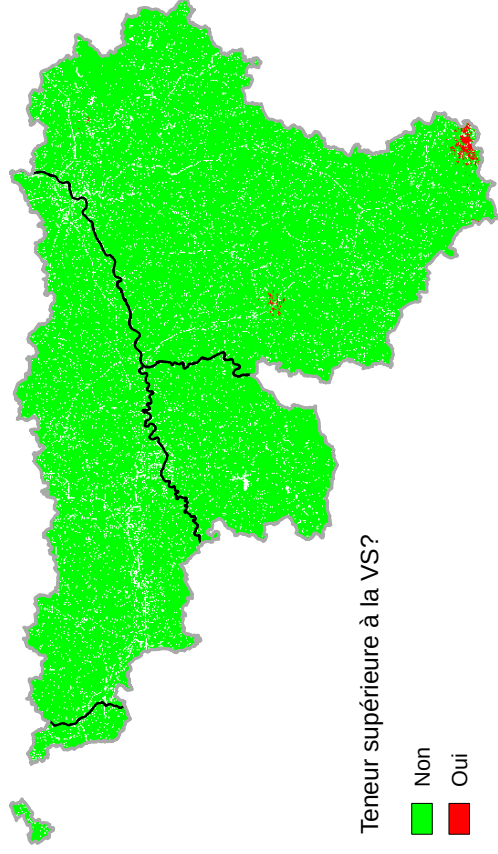
Ni : teneur attendue



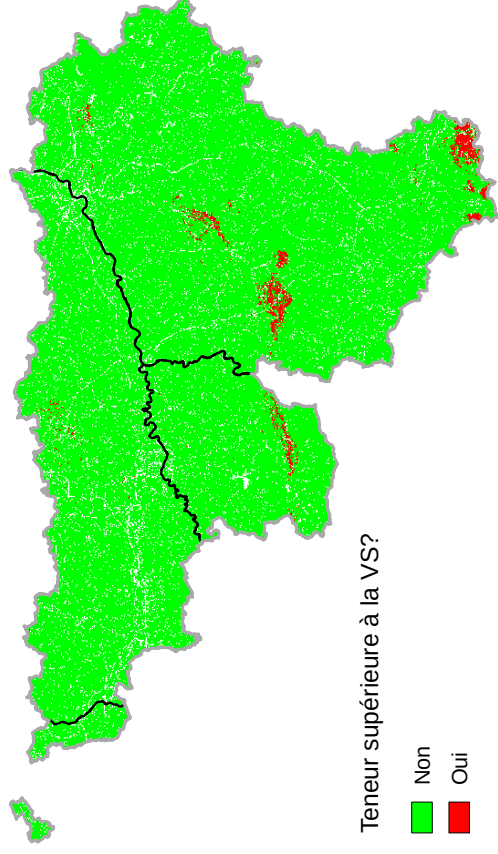
Ni : teneur qui a moins d'1 chance sur 10 d'être dépassée



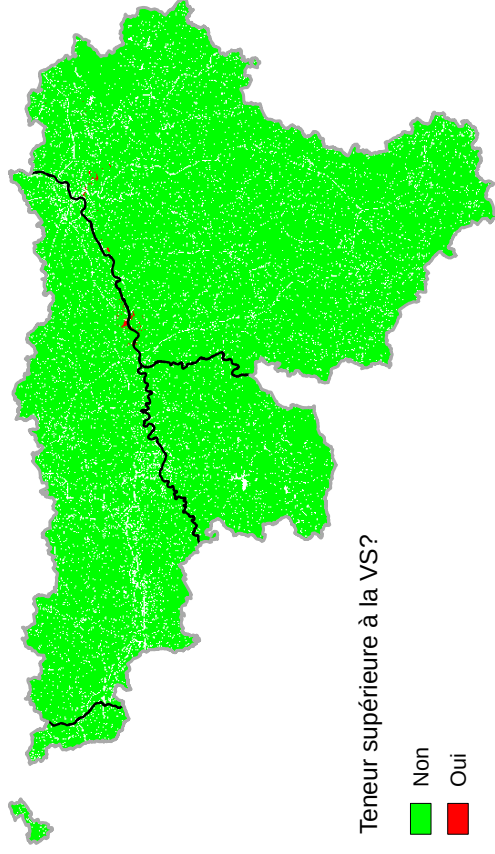
Ni : teneur qui a moins d'1 chance sur 20 d'être dépassée



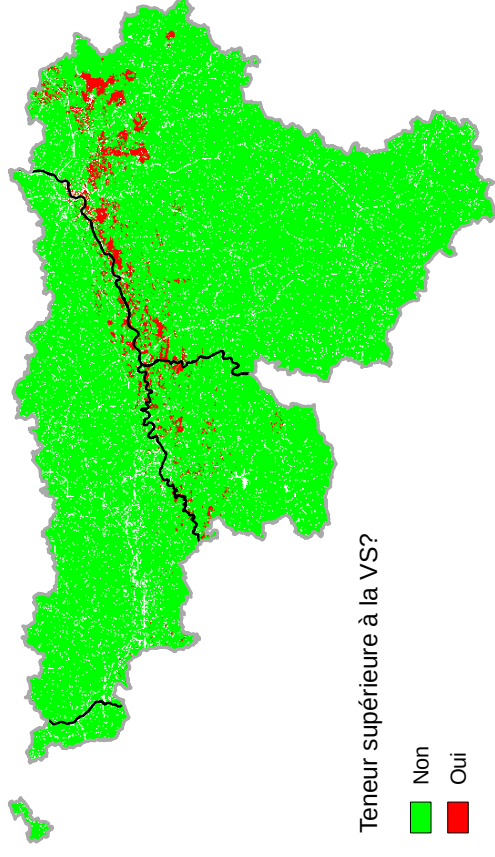
Ni : teneur qui a moins d'1 chance sur 100 d'être dépassée



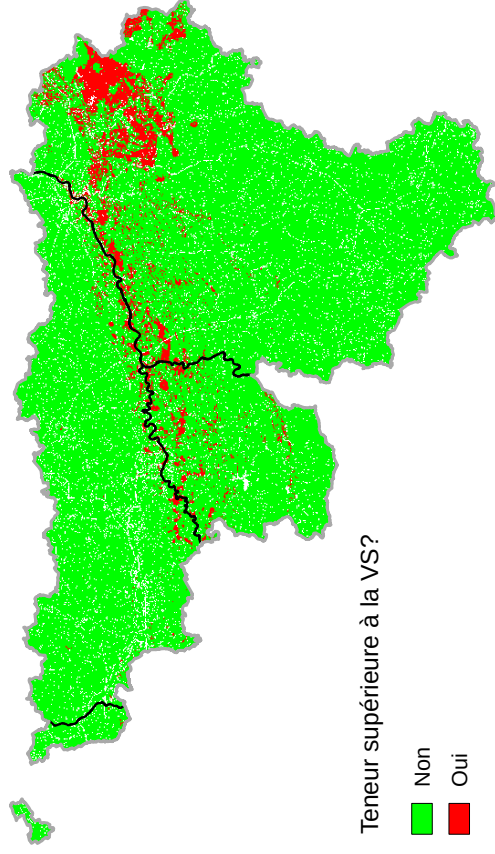
Pb : teneur attendue



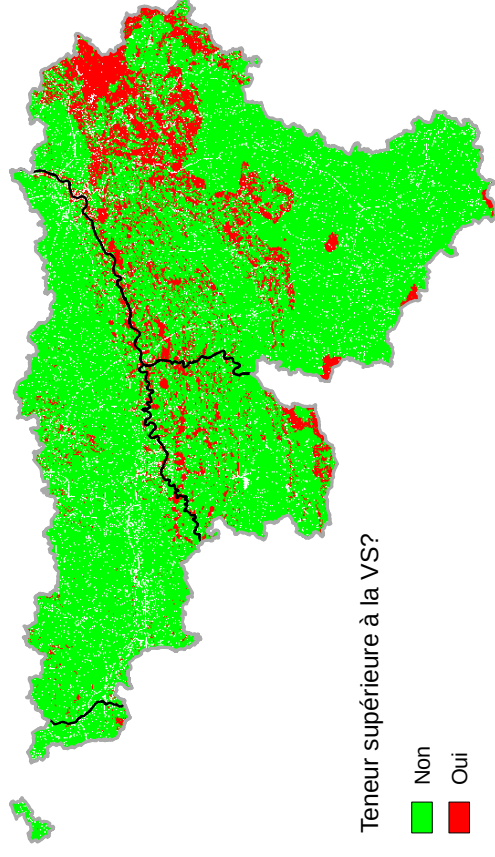
Pb : teneur qui a moins d'1 chance sur 10 d'être dépassée



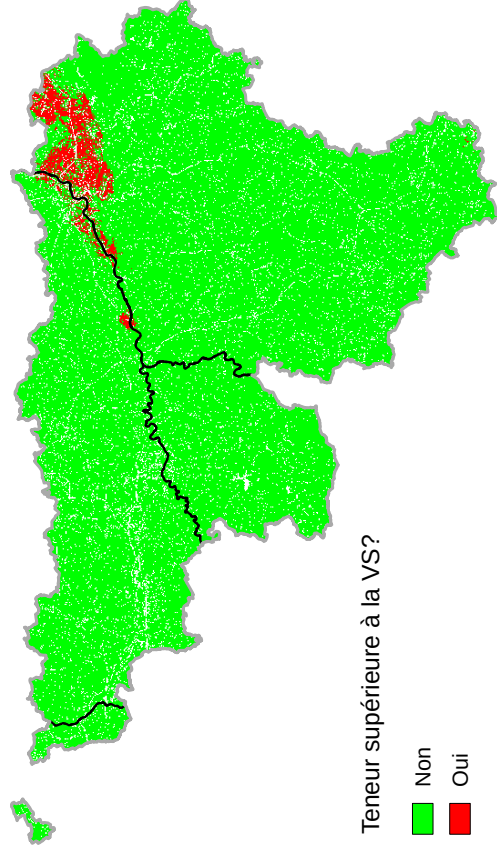
Pb : teneur qui a moins d'1 chance sur 20 d'être dépassée



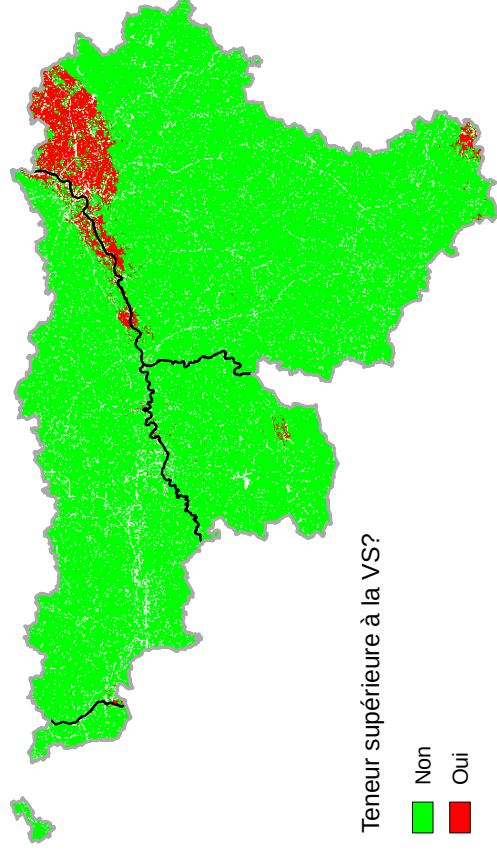
Pb : teneur qui a moins d'1 chance sur 100 d'être dépassée



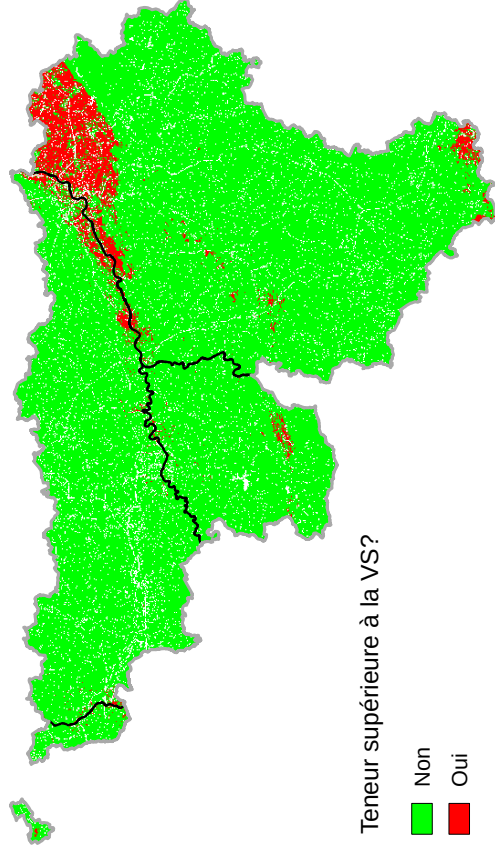
Zn : teneur attendue



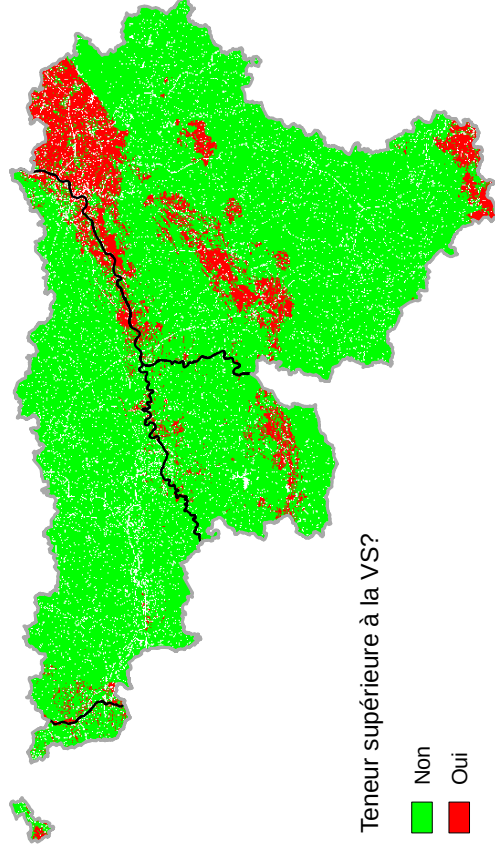
Zn : teneur qui a moins d'1 chance sur 10 d'être dépassée



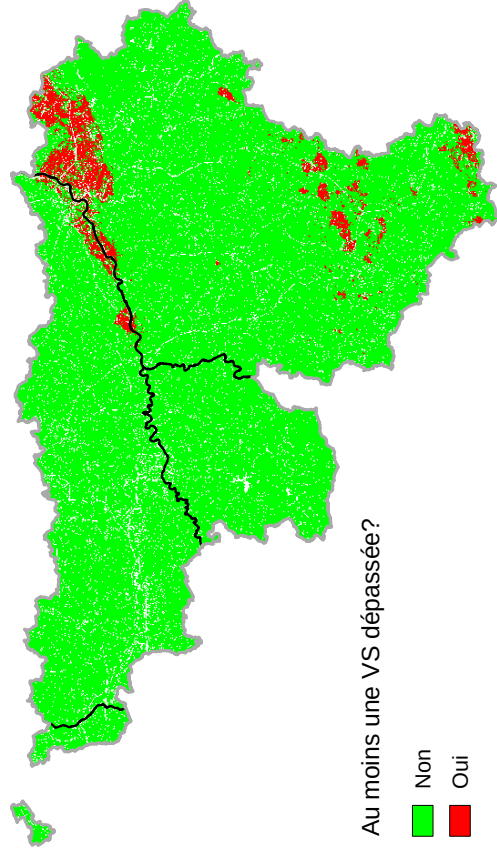
Zn : teneur qui a moins d'1 chance sur 20 d'être dépassée



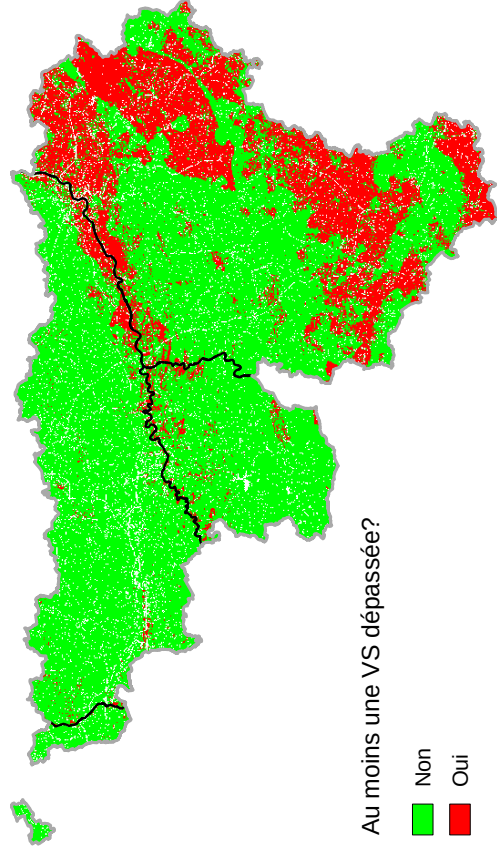
Zn : teneur qui a moins d'1 chance sur 100 d'être dépassée



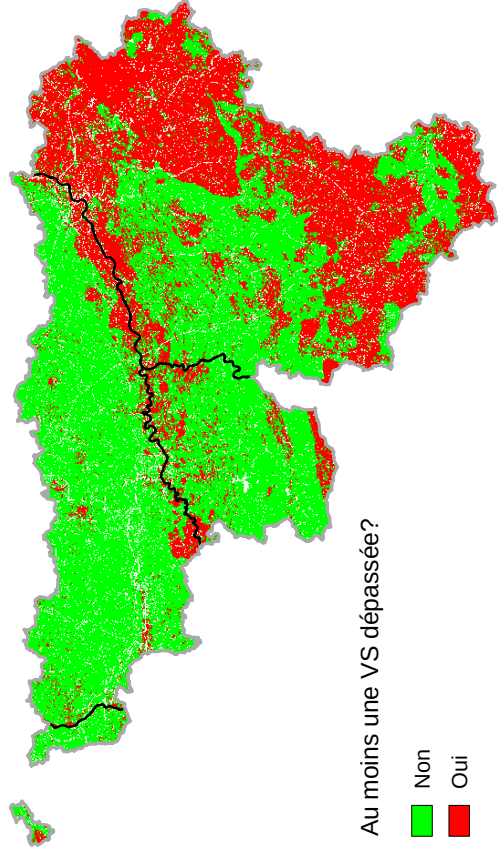
Teneur attendue



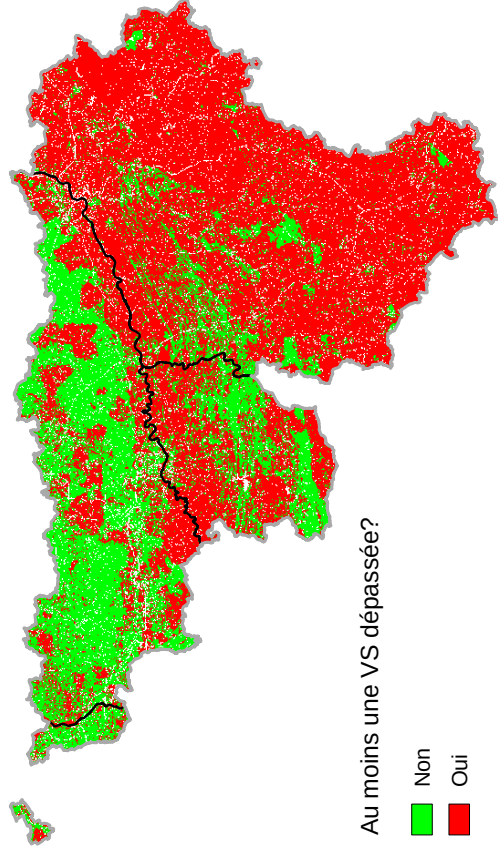
Teneur qui a moins d'1 chance sur 10 d'être dépassée



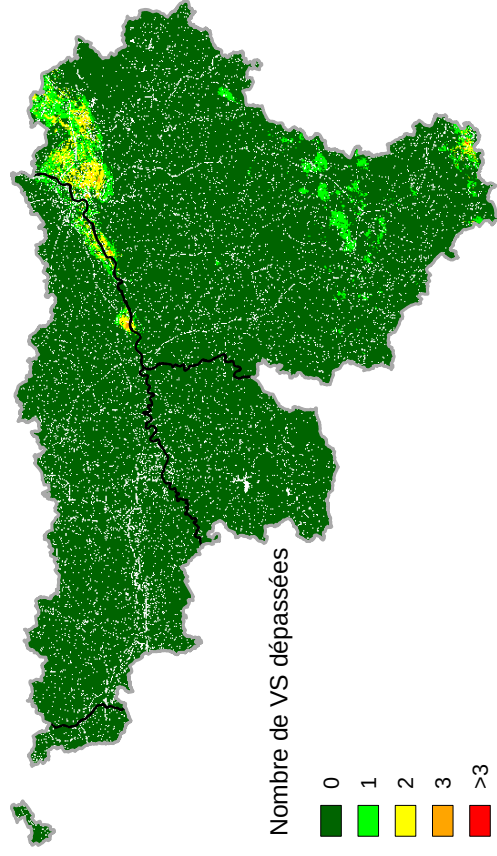
Teneur qui a moins d'1 chance sur 20 d'être dépassée



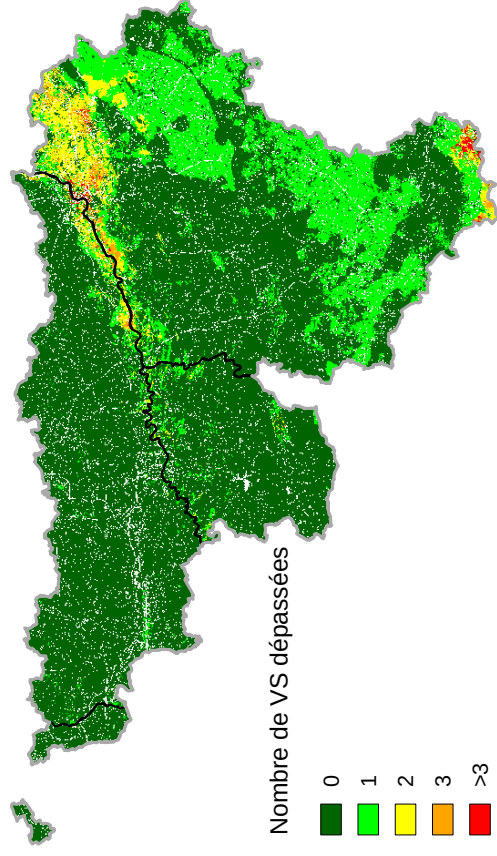
Teneur qui a moins d'1 chance sur 100 d'être dépassée



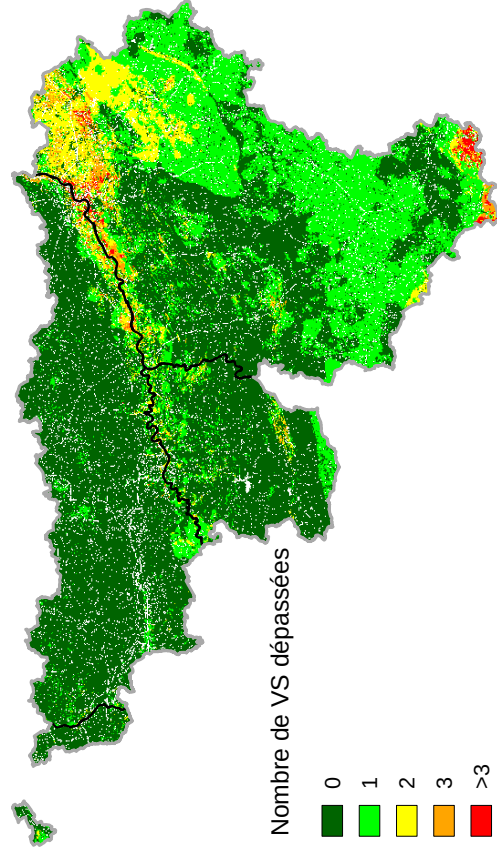
Teneur attendue



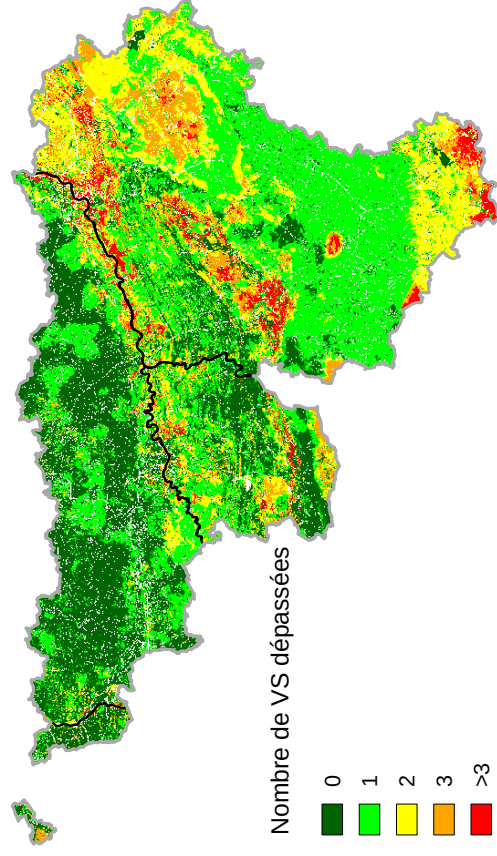
Teneur qui a moins d'1 chance sur 10 d'être dépassée



Teneur qui a moins d'1 chance sur 20 d'être dépassée



Teneur qui a moins d'1 chance sur 100 d'être dépassée



Chapitre 5

Etude de sensibilité du modèle S-Risk et applications de l’outil « concentration de fond limite »

5.1 Etude de sensibilité du modèle S-Risk et stratégies de gestion à l’échelle régionale des jardins potagers et parcelles agricoles

Avertissements par rapport à cette section

Le travail exposé dans cette section a été présenté à l’Administration lors de la réunion de suivi de la convention du 7 septembre 2017. A ce moment, la convention SANISOL qui porte spécifiquement sur les jardins potagers et parcelles agricoles n’avait pas encore commencé. Certains résultats détaillés dans cette section pourraient donc ne pas être en phase avec les avancées actuelles de la convention SANISOL.

Les paramètres de S-Risk utilisés dans cette section correspondent à ceux de l’outil « concentration de fond limite », c’est à dire aux paramètres de S-Risk qui nous ont été communiqués lors de la convention CAPASOL 5. De légères différences peuvent donc apparaître entre la version actuelle de S-Risk et les résultats de cette étude. Les résultats présentés ici sont cependant représentatifs de ce qui pourrait être obtenu avec la version actuelle de S-Risk.

Introduction

Ce travail expose les résultats d’une étude de sensibilité et d’impact des paramètres du modèle d’évaluation des risques pour la santé humaine S-Risk en vue de son utilisation comme outil de gestion des potagers contaminés à l’échelle régionale. En effet, depuis plusieurs années, la Direction de la Protection de Sols (DPS) du Service Public de Wallonie fait face à un nombre croissant de requêtes et sollicitations de la part des propriétaires/utilisateurs de jardins

potagers communautaires et privés ayant constaté une contamination aux métaux dans le sol de leur parcelle. Parmi ces sollicitations, le cas du potager communautaire des Coins de Terre de Bressoux, dont la contamination a été constatée dans le courant de l'année 2017, a fait ressentir l'absence de réponse concertée face à cette problématique grandissante.

Cette demande du Service Public de Wallonie était par ailleurs concomitante avec la parution d'une nouvelle version du décret réglementant la gestion des sites et sols pollués, dans laquelle une des modifications concerne le cas des jardins potagers. Dorénavant, le constat d'un dépassement de la Valeur Seuil (VS) pour un jardin potager collectif ou privé, ou pour une parcelle agricole via une analyse de fertilité, ne déclenche plus systématiquement d'étude d'orientation, comme la précédente version du décret l'imposait. Ce contexte rend possible une gestion spécifique des sols de potagers.

La convention CAPASOL 5 avait permis d'aboutir à la réalisation d'un outil cartographique de détermination de la concentration de fond limite en métaux dans le sol. Celui-ci était basé sur les équations du modèle S-Risk pour l'évaluation des risques pour la santé humaine, et sur le PNEC Calculator¹ pour l'évaluation du risque pour les écosystèmes. L'Administration s'est posé la question de savoir si cet outil pouvait éventuellement être adapté pour répondre à la problématique des jardins potagers et parcelles agricoles contaminés. Pour obtenir des éléments de réponses à cette interrogation, l'étude de sensibilité et d'impact des paramètres du modèle a été réalisée dans le cadre de cette note. L'étude a porté sur les paramètres du scénario RES-vg (résidentiel avec jardin potager) de S-Risk et sur des scénarios hérités de celui-ci. L'objectif est donc de déterminer les paramètres/voies d'exposition ayant le plus d'influence sur l'exposition totale des cibles calculées pour les activités potagères et agricoles ainsi que pour le consommateur de denrées produites sur une parcelle agricole ou potagère contaminée, et tenter d'anticiper le nombre de cas en Wallonie pour lesquels un indice de risque supérieur à 1 serait obtenu selon ces différents scénarios.

Les polluants concernés par cette étude sont les éléments traces métalliques. La note aborde en premier lieu les résultats de l'étude de sensibilité du modèle S-Risk avant de proposer différentes pistes de réflexion en vue d'établir une stratégie de gestion des potagers contaminés à l'échelle régionale.

Termes et abréviations utilisées

S-Risk : modèle de calcul de risque pour la santé humaine² utilisé dans cette étude ;

VTR : Valeur Toxicologique de Référence ;

Teneurs ambiantes : voies d'exposition qui ne dépendent pas du terrain étudié (teneurs en contaminants dans l'air ambiant, l'eau de distribution et les denrées alimentaires du

1. Predicted No-Effect Concentration : méthodologie inspirée des rapports EU RAR (Risk Assessment Reports) faisant suite à l'application de la législation REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals; Regulation EC No 1907/2006) et développée par la société ARCHE Consulting : www.arche-consulting.be – contact : koen.oorts@arche-consulting.be

2. C. Cornelis, A. Standaert, H. Willems, 2016, Final report S-Risk - Technical guidance document Study accomplished under the authority of OVAM 2013/MRG/R/76 - <https://s-risk.be/sites/s-risk.be/files/SRisk%20model%20equations.pdf>

commerce). Elles représentent l'exposition moyenne ambiante aux contaminants pour une personne vivant en Wallonie, qui s'additionne à l'exposition aux contaminants présents dans le sol de la parcelle ;

Approche « Wallonie » : S-Risk utilisé avec les paramètres en vigueur en Wallonie ;

Approche « Flandre » : S-Risk utilisé avec les paramètres en vigueur en Flandre ;

Concentrations $_{IR=1}$: concentration dans le sol induisant une exposition pour la santé humaine équivalente à la Valeur Toxicologique de Référence. C'est donc la concentration pour laquelle l'indice de risque calculé est égal à 1 ;

RES-vg : scénario « résidentiel avec jardin potager » de S-Risk.

5.1.1 Evaluation des risques pour un jardin potager

La Table 5.1 présente les résultats des calculs des concentrations $_{IR=1}$ pour le scénario RES-vg de S-Risk, selon l'approche « Flandre » (ligne 1) et « Wallonie » (ligne 2). Les valeurs calculées pour l'approche « Wallonie » sont bien inférieures à celles de l'approche « Flandre ». Cette différence est principalement due à la liste des VTR qui diffère entre les deux approches. Pour l'approche « Wallonie », la concentration $_{IR=1}$ calculée pour As est quasiment égale à 0. Pour Pb et Hg, la concentration $_{IR=1}$ est négative, car l'exposition aux teneurs ambiantes est supérieure à la valeur de la VTR. Pour Zn, la concentration est de 220 ppm, ce qui est très faible par rapport à la valeur calculée avec l'approche « Flandre » (4050 ppm).

Au cours de la convention POLLUSOL 2, des cartes de teneurs attendues en métaux avait été réalisées en prenant en compte l'occupation du sol comme variable prédictive du modèle géostatistique utilisé pour la cartographie. Toutefois, comme la toute grande majorité des données relevées en usage résidentiel provenaient de jardins potagers, toutes les zones situées en usage résidentiel en Wallonie ont été prédites comme s'il s'agissait de jardins potagers. Les valeurs prédites dans ces zones donnent donc une idée de la concentration que l'on pourrait retrouver dans les jardins potagers en Wallonie. Les concentrations $_{IR=1}$ de la Table 5.1 ont donc été comparées aux valeurs prédites sur les mailles (500m $\times$ 500m) situées en usage résidentiel dans les cartes de teneurs attendues de POLLUSOL 2³. La proportion des mailles « potagers » de POLLUSOL 2 présentant des concentrations prédites supérieures aux concentrations $_{IR=1}$ pour les deux approches sont donnés dans les lignes 3 et 4 de la Table 5.1. Pour l'approche « Wallonie », la concentration $_{IR=1}$ est dépassée dans 100% des mailles pour As, Hg et Pb, et de 46% pour Zn, ce qui est également très élevé. Pour l'approche « Flandre », seule la concentration $_{IR=1}$ calculée pour As est dépassée fréquemment (63% des mailles). Pour les autres métaux, le pourcentage de dépassement est faible.

3. Pereira, B.; Titeux, H.; Schneider, A. & Sonnet, Ph. (2012), Rapport final du projet POLLUSOL 2 - partie sols, Technical report, SPAQuE, 1126 p.

TABLE 5.1 – Concentrations_{IR=1} pour les approches « Flandre » et « Wallonie » et pourcentage des « potagers » pour lesquels la concentration dépasse les concentrations_{IR=1} calculées.

	As	Cd	Cr ₃	Cr ₆	Cu	Hg	Ni	Pb	Zn
C_{IR=1} Flandre (ppm)	10.4	16	144	25.9	2360	1.35	151	277	4050
C_{IR=1} Wallonie (ppm)	0.158	7.77	244	25.9	1850	< 0	135	< 0	220
Dépassement Flandre (%)	63	0.002	0	0	0	0	0	8	0
Dépassement Wallonie (%)	100	0.2	0	0	0	100	0	100	46

Comme il a été expliqué ci-dessus, les teneurs ambiantes ont été prises en compte dans les résultats du Table 5.1. Toutefois, si l'on ne prend pas en compte les teneurs ambiantes, la concentration_{IR=1} de Zn est bien plus élevée (5470 ppm au lieu de 220 ppm, voir rapport final de la convention CAPASOL 5). Par ailleurs, pour Hg, les concentrations de fond retrouvées en Wallonie sont en grande majorité très basses et proches des limites de quantification (voir Chapitre 4). Les sections qui suivent vont donc traiter plus particulièrement Pb et As.

5.1.2 Contributions à l'indice de risque pour les concentrations _{IR=1} pour le scénario RES-vg de S-Risk : cas de Pb et As

Les graphiques de la Figure 5.1 illustrent les contributions de chacune des voies d'exposition à l'indice de risque pour As et Pb en scénario « résidentiel avec jardin potager ». Etant donné que pour l'approche « Wallonie » les teneurs ambiantes induisent déjà une exposition supérieure à la VTR pour Pb et que pour As, la teneur dans le sol induisant une exposition équivalente à la VTR est quasiment nulle, l'exercice a été réalisé à partir des paramètres de l'approche « Flandre ». Pour ces deux Figures, le diagramme de gauche représente les contributions pour une valeur d'indice de risque = 1, et les courbes du graphique de droite représentent l'évolution des contributions des voies d'exposition lorsque l'indice de risque varie de 0 à 3. On peut constater que l'ingestion de sol/poussière est la voie d'exposition prépondérante pour Pb et As. Les contributions des teneurs ambiantes sont élevées pour les deux métaux, mais décroissent rapidement lorsque l'indice de risque augmente. A noter que pour As, pour l'approche « Wallonie », la VTR qui a été retenue est relative à un effet sans seuil (voir rapport CAPASOL 5), ce qui implique que les teneurs ambiantes ne sont pas prises en compte dans le calcul de l'indice de risque.

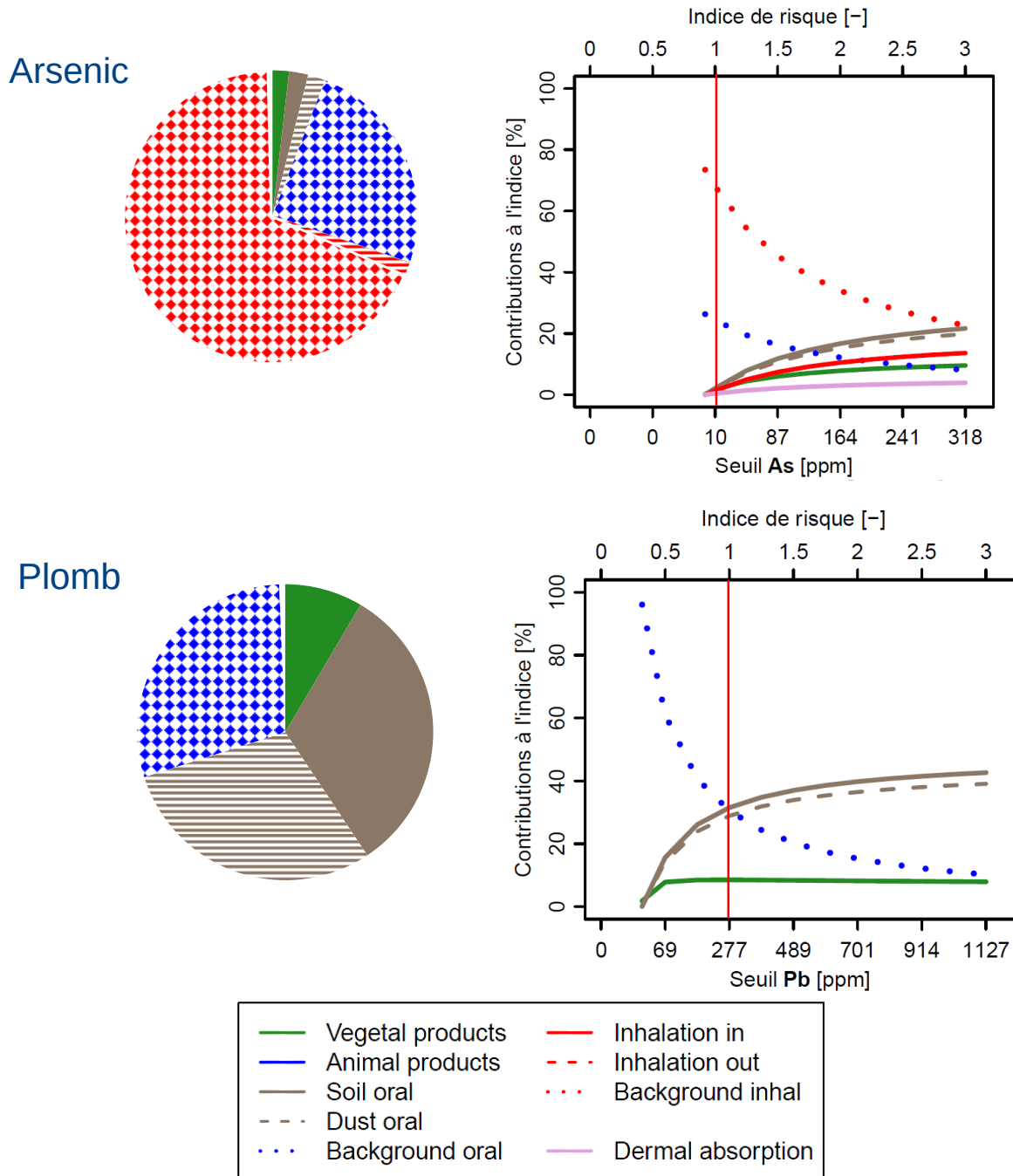


FIGURE 5.1 – Contributions des voies d'exposition à l'indice de risque pour As et Pb en scénario RES-vg pour un indice de risque de 1 (à gauche) et variant de 0 à 3 (à droite).

5.1.3 Impact de l'ingestion de légumes dans le calcul de l'indice de risque pour le scénario RES-vg

Les graphes de la Figure 5.1 montrent que pour Pb et As, c'est bien l'ingestion de sol et poussières qui est la voie d'exposition ayant le plus d'impact sur l'indice de risque. L'exposition

via ingestion de légumes est donc de moindre importance pour ces deux métaux. Afin de mettre en évidence l'impact de la consommation de légumes produits sur la parcelle sur l'indice de risque, deux scénarios ont été simulés :

Scénario 1 : *RES-vg sans denrées* est une adaptation du scénario RES-vg pour lequel toutes les voies d'exposition de celui-ci ont été conservées à l'exception de l'ingestion de légumes provenant de la parcelle. Ce scénario ferait état du cas hypothétique où la personne consommerait à la fois des légumes achetés dans le commerce et des légumes issus de son potager (selon les proportions du scénario RES-vg) mais pour lesquels les concentrations en métaux des légumes issus de son potager seraient nulles.

Scénario 2 : *RES-vg uniquement denrées*. C'est donc une adaptation du scénario RES-vg pour lequel aucun(e) voie/paramètre d'exposition de celui-ci n'a été conservé à l'exception de l'ingestion de légumes provenant de la parcelle et des teneurs ambiantes. Ce scénario ferait état du cas hypothétique où la personne ne serait exposée qu'à l'ingestion de légumes provenant d'une parcelle pour laquelle elle n'a pas accès, et où toutes les autres sources d'exposition ne seraient que l'exposition aux teneurs ambiantes.

Les concentrations_{IR=1} ont été calculées pour les scénarios *RES-vg sans denrées* et *RES-vg uniquement denrées* sont donnés dans les Tables 5.2 et 5.3 (lignes 1 et 2) pour les approches « Flandre » et « Wallonie ». De la même manière que pour le Table 5.1, les lignes 3 et 4 de ces Tables donnent la proportion des mailles classées « potagers » pour lesquelles la concentration prédite lors de POLLUSOL 2 dépasse ces concentrations_{IR=1}.

TABLE 5.2 – Concentrations _{IR=1} calculées pour le scénario *RES-vg sans denrées* pour les approches « Flandre » et « Wallonie », et pourcentage des « potagers » pour lesquels la concentration dépasse les concentrations _{IR=1}.

	As	Cd	Cr ₃	Cr ₆	Cu	Hg	Ni	Pb	Zn
C_{IR=1} Flandre (ppm)	13.7	124	278	25.9	16100	6.92	210	316	28500
C_{IR=1} Wallonie (ppm)	1.82	89.2	420	25.9	10000	< 0	197	< 0	3250
Dépassement Flandre (%)	25	0	0	0	0	0	0	6	0
Dépassement Wallonie (%)	100	0	0	0	0	100	0	100	0

TABLE 5.3 – Concentrations _{IR=1} calculées pour le scénario *RES-vg uniquement denrées* pour les approches « Flandre » et « Wallonie », et pourcentage des « potagers » pour lesquels la concentration dépasse les concentrations _{IR=1}.

	As	Cd	Cr ₃	Cr ₆	Cu	Hg	Ni	Pb	Zn
C_{IR=1} Flandre (ppm)	72	19.8	301	404	2760	1.7	537	3920	4740
C_{IR=1} Wallonie (ppm)	0.224	9.07	580	113	2260	< 0	430	< 0	236
Dépassement Flandre (%)	0.004	0	0	0	0	0	0	0	0
Dépassement Wallonie (%)	100	1	0	0	0	100	0	100	42

Les concentrations_{IR=1} calculées pour le scénario *RES-vg sans denrées* sont plus élevées que celles calculées pour le scénario RES-vg, mais ne sont pas d'un ordre de grandeur différent

(voir Table 5.1), à part pour Cd, Cu et Zn. La proportion de mailles classées « potagers » dépassant ces concentrations_{IR=1} est donc logiquement du même ordre que le nombre de mailles dépassées par les concentrations_{IR=1} faisant intervenir la consommation de végétaux pour tous les métaux, excepté pour Zn. Pour le scénario *RES-vg sans denrées*, on peut voir que les concentrations_{IR=1} calculées sont nettement plus hautes pour une partie des métaux incluant As et Pb, alors que pour d'autres métaux tels que Zn et Cu, ces valeurs sont du même ordre que pour le scénario RES-vg. La proportion de mailles « potagers » de POLLUSOL 2 pour lesquelles la teneur prédite est supérieure aux concentrations_{IR=1} calculés pour l'approche « Flandre » est nulle pour chaque métal (à l'exception de 0.004% pour As). Pour l'approche « Wallonie », les teneurs ambiantes étant supérieures aux VTR pour Pb et Hg, la proportion de mailles supérieures aux concentrations_{IR=1} selon ce scénario restent donc de 100%. Pour As, la concentrations_{IR=1} est encore beaucoup trop faible (0.224 ppm) et est donc dépassée en tout point de la Wallonie.

5.1.4 Eléments de discussion pour élaborer une stratégie de gestion des potagers et parcelles agricoles contaminés

Pour As et Pb, l'ingestion de sols et poussières constitue la voie d'exposition qui a le plus d'impact sur la concentrations_{IR=1} (cf. sous-section 5.1.2). Pour Cd, Cu et Zn, l'ingestion de végétaux constitue la voie d'exposition ayant le plus d'impact. Il convient donc de tenir compte de l'ensemble des voies d'exposition lorsqu'il s'agit d'évaluer le risque pour la santé humaine d'une personne exerçant une activité potagère ou agricole sur un terrain. Pour évaluer le risque pour la santé d'éventuel consommateur de denrées qui ne fréquenterait pas le site (qui ne participerait donc ni aux activités potagères et agricoles), seule l'ingestion de denrées produites sur la parcelle agricole/de potager doit être prise en compte, en plus des teneurs ambiantes.

Gestion régionale des activités potagères et agricoles

Evaluer un risque pour une personne exerçant une activité potagère sur un terrain revient à répondre à la question suivante : « L'utilisateur peut-il exercer l'activité potagère sans risque pour sa santé ? ». Dans ce contexte, l'ensemble des voies d'exposition doit être prise en compte. Comme on a pu le voir dans les points qui précèdent, ne pas tenir compte de l'ingestion de sol/poussières reviendrait à ne pas tenir compte de la voie d'exposition principale pour une partie des métaux, en ce compris Pb et As. Pour ces deux derniers métaux, un travail de recherche visant à limiter l'incertitude sur les paramètres ayant une influence sur l'ingestion de sol/poussières est hautement recommandé pour éviter de surestimer le risque calculé pour ces métaux et qui serait le fruit d'une approche trop précautionneuse. Améliorer les modèles de transfert sol-plante pour les rendre plus adaptés aux sols wallons est surtout important pour Cd, Zn et Cu.

Evaluer un risque que présenterait une contamination constatée sur une parcelle agricole reviendrait à répondre à la question suivante : « La contamination de la parcelle agricole

représente-t-elle un risque pour son exploitant/pour l'écosystème ? ». Dans ce contexte, l'ensemble des voies d'exposition doit être pris en compte, moyennant une adaptation des paramètres définissant « l'activité agricole ». En effet, l'agriculteur travaillant sa terre remue par exemple des particules de poussières et sol qui peuvent être mises en suspension puis ingérées, et sont donc à prendre en compte dans l'évaluation des risques pour l'activité en elle-même. Par ailleurs, contrairement aux potagers qui représentent une surface assez faible dans la majorité des cas, les parcelles agricoles remplissent un rôle écosystémique important simplement du fait qu'elles recouvrent de larges surfaces. La question qui se pose est alors : « Le sol de la parcelle remplit-il ses fonctions écosystémiques ? ». Il est donc souhaitable que l'approche inclue une évaluation du risque pour les écosystèmes. Ceci pourrait par exemple être réalisé via les équations du PNEC Calculator (voir rapport CAPASOL 5).

Production de denrées provenant d'un potager ou d'une parcelle agricole contaminés

Evaluer un risque pour une personne consommant des denrées produites sur une parcelle agricole ou un potager contaminé revient à répondre à la question suivante : « Le légume répond-il aux critères de consommation/commercialisation ? ». Dans ce cas, seule l'ingestion via les légumes devrait être prise en compte, puisque le consommateur ne ferait que consommer les légumes produits sur le potager, ne participerait pas au travail de la terre et ne se rendrait pas sur le site de production de la denrée. L'exposition aux teneurs ambiantes doivent être prises en compte, car elles rendent compte de l'exposition ambiante, qui ne dépendent pas du terrain étudié (teneurs en contaminants dans l'air, l'eau et les denrées alimentaires du commerce).

5.1.5 Comment évaluer la qualité d'une production végétale ?

Afin de déterminer si un légume répond aux critères de consommation ou de commercialisation vis-à-vis d'une substance, il s'agit de comparer la concentration en cette substance mesurée dans le légume avec une teneur limite tolérée/recommandée. Par exemple, la législation européenne EC 1881/2006⁴ propose une série de normes de commercialisation pour trois métaux, Hg, Pb et Cd, pour certaines catégories de denrées. Pour Hg, il n'existe pas de teneur limite pour les légumes/céréales. Pour tous les autres métaux, la Commission européenne ne recommande aucune valeur limite. Pour tenter de répondre à ces manquements, nous proposons ici trois voies de détermination de teneurs limites dans les légumes/céréales.

Normes de commercialisation existantes

La première option de détermination de teneurs limites consiste à utiliser des normes de commercialisation/consommation existantes, en se basant sur la littérature et les textes de loi étrangers. Parmi celles-ci, les normes de Pb et Cd de la législation européenne EC 1881/2006

4. Commission Regulation (EC) No 1881/2006 of 19 December 2006 setting maximum levels for certain contaminants in foodstuffs (Text with EEA relevance)

peuvent être utilisées. Des valeurs peuvent être recherchées ailleurs qu'en Europe. En Chine, par exemple, un organisme étatique s'est chargé de proposer une liste de teneurs limites en métaux dans les denrées alimentaires en dessous desquelles celles-ci sont considérées comme propre à la consommation⁵. Pour As, par exemple, la valeur limite dans les aliments est de 1 ppm (poids frais), mais de 0.3 ppm dans les denrées destinées aux enfants. Pour Zn, la teneur limite est de 20 ppm.

5.1.6 Calcul de nouvelles valeurs limites

Les valeurs limites dans les légumes pourraient être calculées par exemple par le modèle S-Risk à partir du scénario *RES-vg uniquement denrées*, tel que défini à la section 5.1.3. Ce scénario fait état d'une situation où toutes les voies d'exposition sont nulles, exceptées l'ingestion de légumes et les teneurs ambiantes. Les concentrations calculées dans les légumes pour la concentration_{IR=1} sont donc les teneurs maximales qui pourraient être présentes dans les légumes avant que l'ingestion de polluant via leur consommation ne donne lieu à une exposition totale équivalente à la VTR. Les graphes de la Figure 5.2 donnent les histogrammes de concentrations en Cd calculées dans les 22 légumes repris dans S-Risk selon cette méthode pour les approches « Flandre » et « Wallonie ». Comme on peut le constater, les normes de commercialisation EC 1881/2006 sont très basses comparées aux valeurs calculées par S-Risk. Le même constat peut être tiré pour Pb (Figure 5.3). Pour cet élément, les résultats de l'approche « Wallonie » n'ont pas pu être calculés étant donné que l'exposition aux teneurs ambiantes seules est supérieure à la VTR (voir section 5.1.1).

5. Appendix A9 : National Standards – Natural and Health Foods

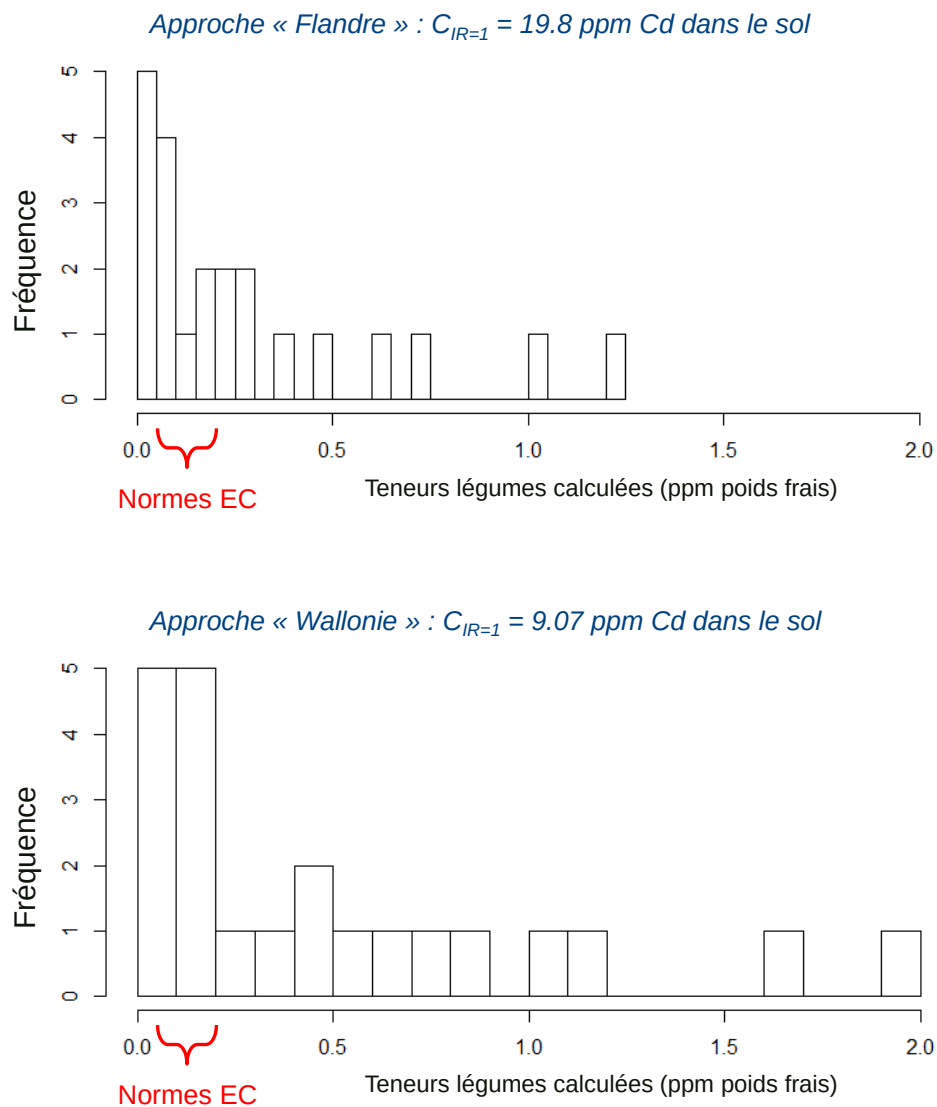


FIGURE 5.2 – Concentrations en Cd calculées dans les légumes pour une concentration_{IR=1} selon le scénario RES-vg uniquement denrées. L'accolade rouge donne la gamme des valeurs de commercialisation de la législation européenne EC 1881/2006.

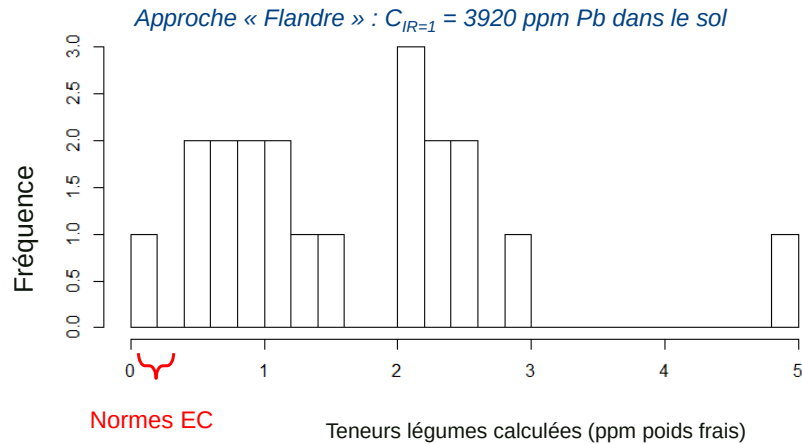


FIGURE 5.3 – Concentrations en Pb calculées dans les légumes pour une concentration $_{IR=1}$ selon le scénario RES-vg uniquement denrées. L'accolade rouge donne la gamme des valeurs de commercialisation de la législation européenne EC 1881/2006.

Valeurs mesurées dans les légumes du commerce wallon

Les teneurs limites dans les légumes pourraient également être établies à partir de la distribution de concentrations en polluants mesurées dans les légumes du commerce wallon (centile 95, par exemple). Cette manière de procéder partirait de l'hypothèse qu'il n'existe pas de problème de santé généralisé en Wallonie, et que les concentrations en polluants trouvés dans les légumes commercialisés actuellement ne présentent aucun risque pour la santé.

5.2 Outil « concentration de fond limite » : applications possibles

Une application informatique permettant un diagnostic rapide du risque pour la santé humaine et les écosystèmes a été mise au point dans la convention CAPASOL 5. Cette application, appelée outil « concentration de fond limite » se base sur les équations des modèles de prédiction de risque de S-Risk pour la santé humaine, et à partir du PNEC Calculator pour les écosystèmes. Il se présente sous la forme d'un prototype avec une interface simple permettant d'encoder une série de paramètres (sols, VTR, etc.) et d'obtenir un résultat (indice de risque) via le calcul d'équations de modèles de risques codées en langage informatique R. Ce prototype pourrait potentiellement être adapté pour être utilisé comme outil d'aide à la gestion d'un grand nombre de thématiques régionales, en dehors du cadre du décret « sols ». Dans le cadre de CAPASOL 6, un travail d'évaluation des éventuels contextes d'utilisation de l'outil « concentration de fond limite » a été réalisé. Au travers d'échanges informels, il a été question d'étudier la faisabilité de rendre l'outil utilisable dans différents contextes de gestion à l'échelle de la région, comme le contexte de la contamination des jardins potagers

contaminés, de l'épandage de matières fertilisantes sur parcelles agricoles, de l'épandage de boues de dragage/régalage et du remblayage de terres excavées.

Parmi les problématiques à l'échelle régionale pour lesquelles l'outil pourrait trouver des applications, celle de la contamination des jardins potagers est apparue dès le début de CAPASOL 6. Le 23 juin 2017, l'UCL a organisé une réunion de concertation sur la gestion de la problématique des potagers contaminés en Wallonie, suite à la sollicitation de l'ASBL « les Coins de Terre » par rapport à la découverte d'une contamination sur le site de Bressoux (Liège). Cette réunion a eu notamment pour objet d'examiner la possibilité de mettre en place un outil d'évaluation des risques propre à la problématique des jardins potagers contaminés à partir de l'outil « concentration de fond limite » et des éventuelles adaptations/améliorations à lui apporter. Cette réunion organisée au sein de l'Administration a permis de réunir un grand nombre d'acteurs wallons potentiellement impliqués dans cette problématique (cabinet du Ministre, SPAQuE, ISSeP, ULg-GxABT). Il a notamment été discuté de la question des données disponibles sur la problématique du transfert de polluants du sol vers les plantes (et des mesures déjà disponibles dans les fruits et légumes produits en Wallonie) pour voir dans quelle mesure de nouvelles données doivent être récoltées pour disposer de l'ensemble de l'information nécessaire. Une deuxième réunion a eu lieu sur la thématique le 7 septembre 2017 au sein de l'Administration. Il a notamment été discuté des différentes approches pour gérer la problématique des jardins potagers à l'échelle régionale, et des résultats de notre enquête sur le potager de Bressoux. Ces différentes concertations ont donné lieu à la convention SANISOL : « détermination des teneurs limites en polluants dans le sol en vue d'assurer la qualité commerciale des productions végétales en Wallonie et la gestion des risques pour le producteur - Développement d'un outil pilote permettant de fournir via internet des recommandations particulières de gestion et d'utilisation à tout producteur de biomasse alimentaire produite sur un sol contaminé en Wallonie ».

Nous avons aussi eu un nombre important d'échanges informels sur la faisabilité de rendre l'outil utilisable dans le contexte de la gestion des remblais faiblement contaminés et du remblayage de terres excavées. Ces échanges ont eu lieu lors de rencontres bilatérales ou de colloques nationaux/internationaux avec des acteurs publics/privés exerçant dans le domaine. Nous nous sommes notamment rendus le 23 novembre 2017 à *INTERSOIL* (Bruxelles) qui avait pour thème « Sols, Terres & Sédiments : même combat ? », ainsi qu'à Intersol 2018 (Paris du 27/03/2018 au 29/03/2018). De ces échanges est ressorti le fait que cette thématique fait beaucoup de bruit ces dernières années, avec notamment de nouvelles réglementations qui entrent en vigueur en Wallonie. Autant cette problématique a été soulignée comme cruciale par nos interlocuteurs, autant ceux-ci soulignent (systématiquement) le besoin que l'outil d'évaluation des risques (qui serait éventuellement développé pour simplifier la gestion des remblais et terres excavées) se situe dans un cadre réglementaire clair. Il semble en effet régner une grande inquiétude à ce sujet. Il conviendrait aussi d'examiner si l'utilisation d'une version adaptée de l'outil « concentration de fond limite » pourrait avoir lieu en restant en cohérence avec d'autres outils existants, comme S-Risk Wal. Dans le cadre de son éventuelle utilisation dans ce contexte, les bureaux d'études insistent notamment sur le fait que le type d'utilisateur de l'outil est déterminant (tout public ou réservé à une liste restreinte d'utilisateur, faisant partie de l'Administration ou du secteur privé). Nous relayons aussi la demande d'un outil

simple d'évaluation des risques permettant de gérer le risque des concentrations de fond élevées en métaux ainsi que le remblayage de terres excavées. Le besoin de procédures simples, efficaces et de bon sens revient aussi souvent dans les discussions, notamment lorsqu'il s'agit de la procédure simplifiée d'évaluation des risques (pour nos interlocuteurs des bureaux d'études wallons).

Nous avons aussi eu des échanges sur la faisabilité de rendre l'outil utilisable dans le contexte de l'épandage de matières fertilisantes. Pour l'épandage de boues dans les sols agricoles, l'outil semble particulièrement approprié vu qu'il intègre déjà les 7 polluants inorganiques repris par l'arrêté réglementant l'utilisation des boues⁶. Il pourrait rendre possible d'objectiver la délivrance d'une dérogation pour l'épandage de boues de stations d'épuration en cas de dépassement des normes de l'arrêté. Les équations les plus appropriées à cette problématique étant étudiées par la convention SANISOL, transformer l'outil pilote « concentration de fond limite » en outil opérationnel devrait pouvoir être possible à moindre coût. Pour réaliser cet outil d'aide à la gestion, il conviendrait d'abord de déterminer son l'utilisateur privilégié. Cela pourrait être l'Administration, le laboratoire qui réalise l'analyse et/ou la société intermédiaire de valorisation. Les sociétés intermédiaires de valorisation réalisent le lien entre le producteur de boues, les agriculteurs et l'administration régionale de contrôle. Elles ont des contacts privilégiés avec les laboratoires d'analyses : les laboratoires fournissent les analyses de boues et de sols à la société intermédiaire pour constituer les documents réglementaires demandés par l'administration régionale de contrôle (SPW/DSD/DPS). Le rôle central des sociétés intermédiaires dans la filière de valorisation des boues fait de celles-ci un interlocuteur privilégié dans l'éventualité où un outil de ce type devait être mis au point par la Région.

Au début de nos échanges sur l'outil « concentration de fond limite », la question du coût d'utilisation de l'outil revenait fréquemment dans les discussions. L'utilisation de l'outil « concentration de fond limite » demande en effet de disposer des résultats d'analyse de son sol afin de pouvoir calculer un risque, et le prix de cette analyse peut être considéré comme élevé par l'utilisateur (citoyen). Nous avons donc continué les échanges autour des travaux de recherche relatifs à l'établissement de la technique de mesure de concentration en métaux dans le sol par fluorescence des rayons X. Cette mesure permet en effet la réalisation d'un grand nombre d'analyses à prix réduit. Ces échanges ont notamment eu lieu à la conférence internationale Goldschmidt 2017 à Paris du 14/08/2017 au 18/08/2017, durant laquelle nous avons présenté les travaux sur la fluorescence des rayons X réalisés dans le cadre de CAPASOL 5⁷. Nous nous sommes aussi rendus à la conférence Intersol 2018 à Paris du 27/03/2018 au 29/03/2018 pour présenter nos travaux sur la mesure de sols, plante, litière et champignons par fluorescence des rayons X⁸. En conclusion de ces échanges, il est apparu que la principale limitation de cette technique de mesure est qu'elle n'est apte qu'à la mesure des polluants inorganiques dans les sols. De plus, pour certains polluants inorganiques, les limites

6. Arrêté du Gouvernement wallon du 12 avril 1995 réglementant l'utilisation sur ou dans les sols des boues d'épuration ou des boues issues de centres de traitement de gadoues de fosses septiques (M.B. 12.04.1995)

7. Pereira, Benoît ; Vandeuren, Aubry ; Sonnet, Philippe ; Delmelle, Pierre. Can portable X-ray fluorescence measurements replace standard soil assays in soil protection regulations? A case study in Wallonia, Belgium.

8. Vandeuren, Aubry ; Baillieux Thibaut ; Pereira, Benoît ; Sonnet, Philippe ; Delmelle, Pierre. Using portable XRF devices for quick and low-cost measurement of metal content in soils, plants and mushrooms in peri-industrial areas of Wallonia.

de quantifications peuvent être élevées (Cd, Hg). On notera toutefois que pour ces éléments, cette technique rend possible un screening de la pollution (si la concentration est élevée, on obtiendra une mesure correcte), et elle est parfaitement adaptée pour la mesure des polluants inorganiques les plus sensibles dans les études de risque (As, Pb). Le grand avantage de la fluorescence des rayons X qui a été souligné par nos interlocuteurs est que cette technique permet de rapidement reproduire la mesure de son sol, et d'ainsi éviter de baser l'analyse de risque sur une valeur élevée potentiellement non représentative du potager (outlier). Sa rapidité et son faible coût rend possible de caractériser son potager à partir de plusieurs échantillons de sol. Cette meilleure caractérisation du sol serait particulièrement adaptée pour réaliser une étude de risque sur un potager de grande taille, ou localisé sur des terres/remblais hétérogènes.

Conclusions et perspectives

Tout en étant indispensables à la mise en œuvre des dispositions du décret du 5 décembre 2008 relatif à la gestion des sols, les cartes de concentrations de fond en polluants produites dans le cadre de cette convention constituent des outils précieux de gestion et de monitoring de l'environnement pour les pouvoirs publics. Elles ont été élaborées à partir de l'ensemble des jeux de données pertinents et disponibles en 2018 en Wallonie (Chapitre 1). Ces jeux de données ont été nettoyés et rassemblés pour constituer des jeux de données géochimiques de qualité à partir des meilleures méthodes disponibles actuellement pour la correction des biais intralaboratoires et interlaboratoires (Chapitre 2). De nouvelles cartes d'impact des retombées atmosphériques de polluants ont été produites, permettant d'optimiser le modèle géostatistique utilisé pour prédire les concentrations de fond (Chapitre 3). Les cartes de concentrations de fond pour les huit polluants inorganiques du décret « sols » ont été établies à partir d'algorithmes de prédiction améliorés et documentés, tout en bénéficiant du fruit de l'expérience de travaux de recherche antérieurs. Des cartes de centiles 90, 95 et 99 ont été établies afin de fournir une indication de l'incertitude sur les valeurs de la carte des concentrations de fond. Finalement, les cartes produites bénéficient d'un important travail de validation, permettant d'apprécier leur niveau de précision (Chapitre 4).

On observe ces dernières années un l'intérêt croissant pour l'agriculture urbaine, et la sensibilisation des citoyens à la qualité de leur sol est plus importante qu'auparavant. De ce fait, l'établissement d'un outil de diagnostic de risque pour la santé humaine des concentrations de fond élevées adapté aux spécificités des sols wallons est devenu un enjeu régional important. Tous les efforts de CAPASOL 6 ont été rassemblés sur cette thématique, notamment par une participation aux travaux de la convention SANISOL⁹. Des travaux préalables à SANISOL ont été réalisés, dont une étude de sensibilité du modèle S-Risk appliqué aux sols agricoles et potagers wallons (Chapitre 5). L'application « concentration de fond limite » mise au point dans la convention CAPASOL 5 s'est révélée être une bonne base pour établir l'outil développé dans SANISOL. D'autres contextes d'utilisation pourraient également bénéficier de l'outil « concentration de fond limite » (Chapitre 5).

Au terme de la convention, la question de l'As dans les sols wallons reste ouverte. Le peu de données disponibles au niveau régional rend les concentrations de fond relativement méconnues

9. SANISOL : « Détermination des teneurs limites en polluants dans le sol en vue d'assurer la qualité commerciale des productions végétales en Wallonie et la gestion des risques pour le producteur - Développement d'un outil pilote permettant de fournir via internet des recommandations particulières de gestion et d'utilisation à tout producteur de biomasse alimentaire produite sur un sol contaminé en Wallonie ».

dans certaines parties de la Wallonie, comme le sud de la Lorraine où les sols présentent souvent des teneurs élevées au regard des normes et des modèles de risque. De plus, certains groupes de sols caractérisés par peu de données présentent une grande variabilité de concentrations. Mieux comprendre l'origine et la nature des phases porteuses de l'As (cristallines ou non) fournirait à la fois des informations utiles pour une meilleure prédiction des concentrations de fond ainsi que d'une application plus raisonnée des modèles de risque aux endroits présentant des concentrations élevées en As dont l'origine géogénique a été mise en évidence.

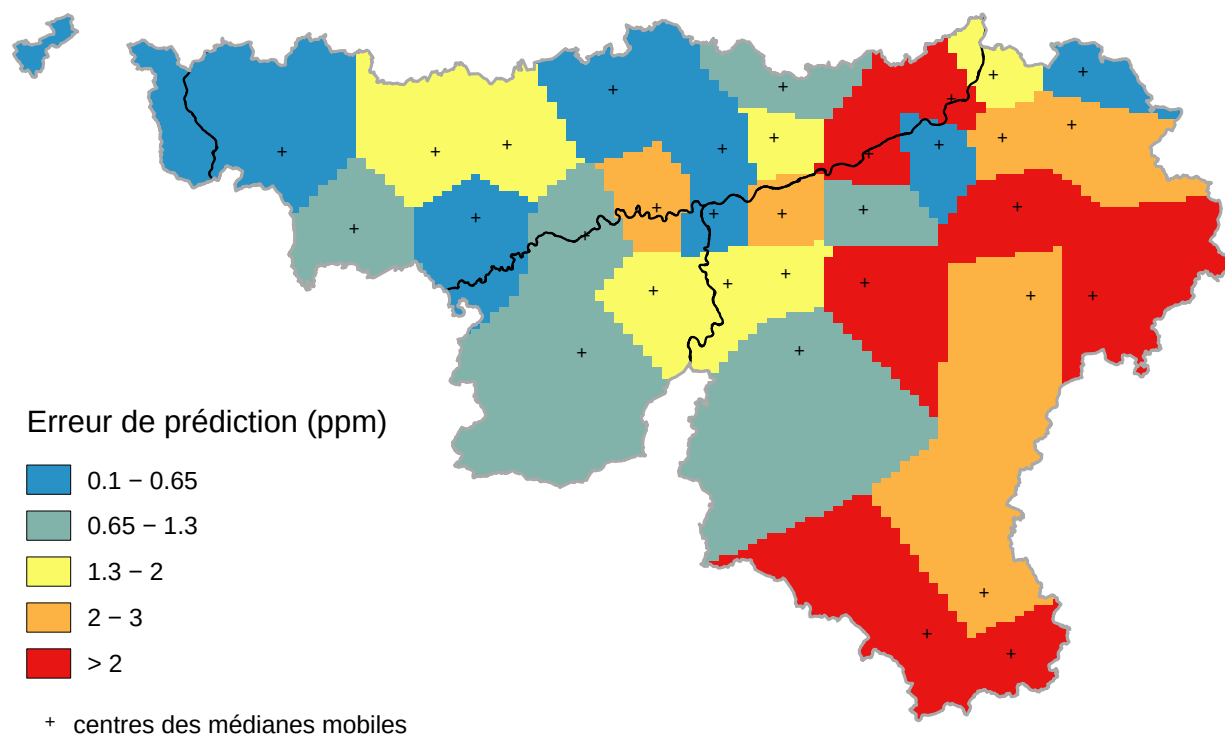
Dans les prochaines années, il est probable qu'émerge le besoin d'une meilleure connaissance des certains polluants autres que ceux étudiés dans CAPASOL 6. Par exemple, le molybdène et le manganèse sont souvent associés à la problématique des sols agricoles et des jardins potagers contaminés. L'étude de ces polluants ainsi que leur spatialisation au niveau régional devraient être grandement facilitées par l'utilisation des méthodes et algorithmes établis dans le cadre de la présente recherche.

L'établissement d'un outil de diagnostic de risque des concentrations de fond élevées adapté aux spécificités des sols wallons est un sujet de recherche étudié par les travaux en cours de la convention SANISOL pour la problématique des sols agricoles et des jardins potagers. D'autres applications de l'outil « concentration de fond limite » pourraient être mises au point, parfois sans nécessiter de développement important, comme pour la problématique de l'épandage de boues de station d'épuration. D'autres contextes d'applications de l'outil pourraient être envisagés, tout en gardant à l'esprit qu'il est nécessaire que celui-ci s'articule correctement autour des réglementations et des éventuels outils complémentaires déjà existants.

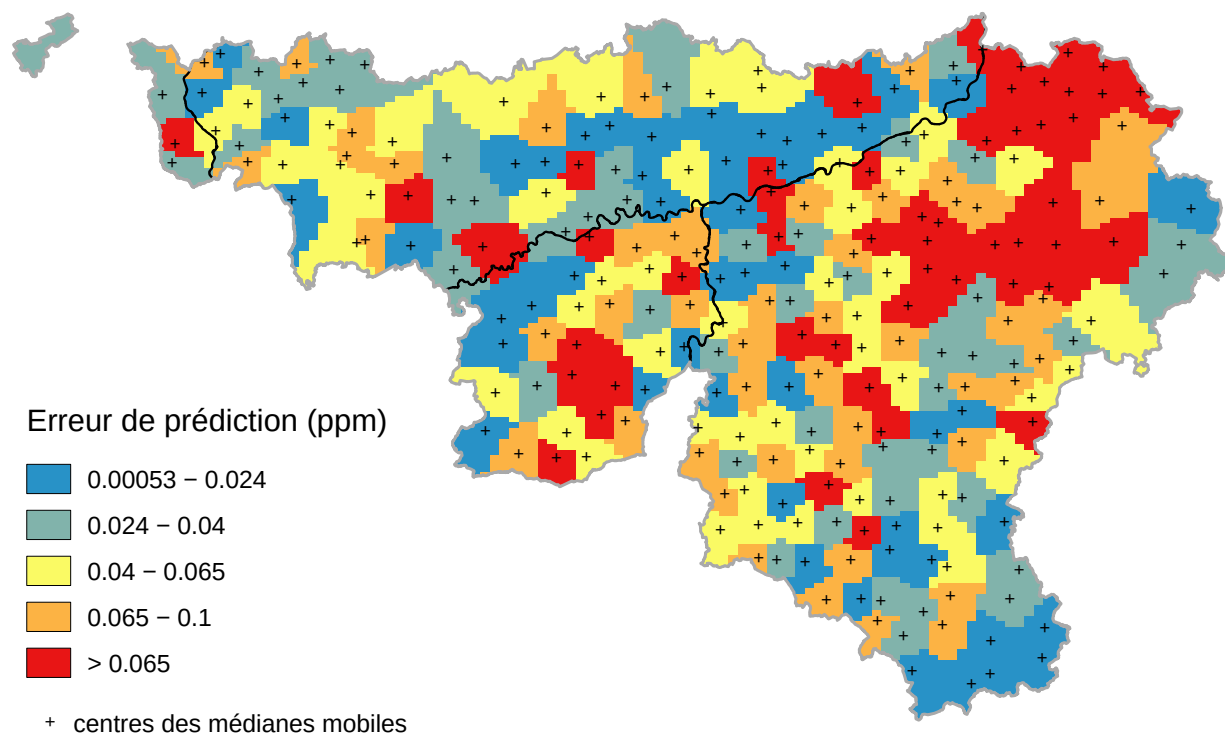
Annexe 1 : Cartographie de la répartition des erreurs de prédiction

La méthodologie utilisée pour l'obtention des cartes présentées dans les pages qui suivent est expliquée à la section 4.4 de ce rapport. Les erreurs de prédiction sont calculées par leave-one-out stratifié géographiquement, et sur chaque maille la médiane des erreurs absolues est calculée. Le centre géographique des observations comprises dans chaque maille est représenté sous l'appellation « centres des médianes mobiles » dans les cartes de cette annexe. L'interpolation spatiale des erreurs a consisté à attribuer à chaque point de la Wallonie la valeur de l'erreur correspondant au « centre de médiane mobile » le plus proche. Cinq classes de couleurs ont été utilisées, sur base de la distribution des erreurs médianes : min-centile 20, centile 20-centile 40, centile 40-centile 60, centile 60-centile 80 et centile 80-max.

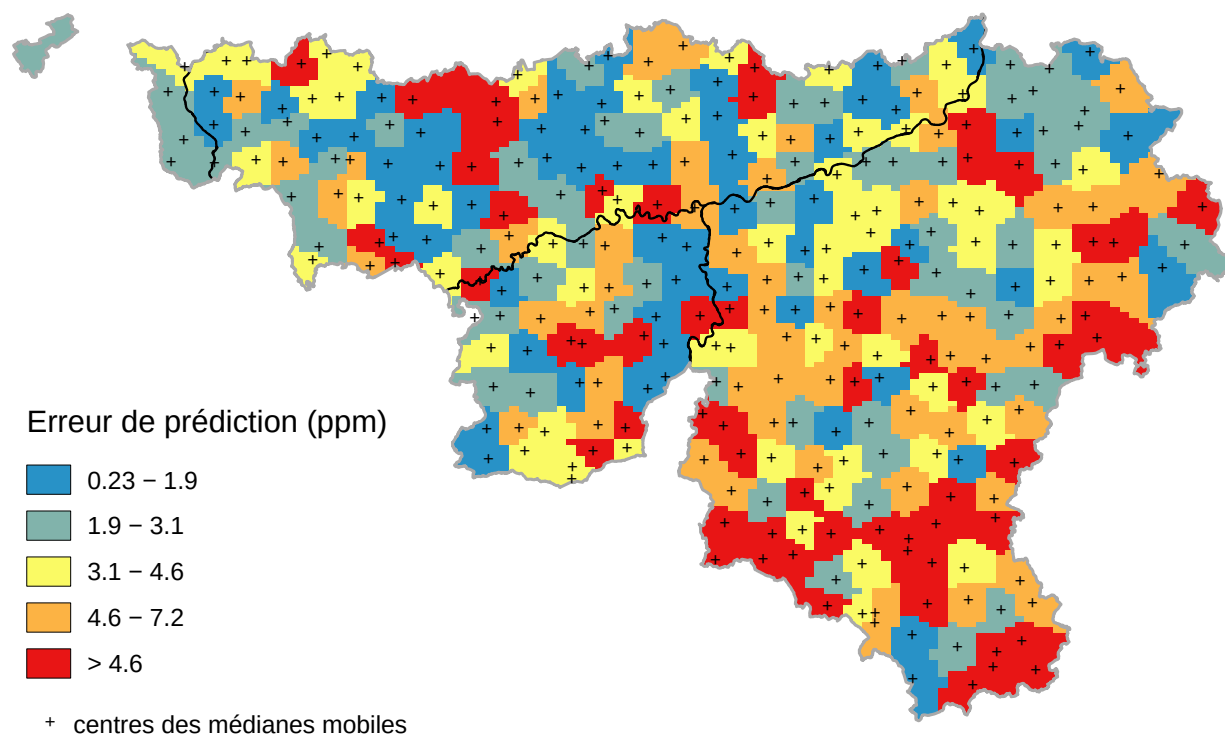
As : erreur de prédiction médiane dans les sols agricoles



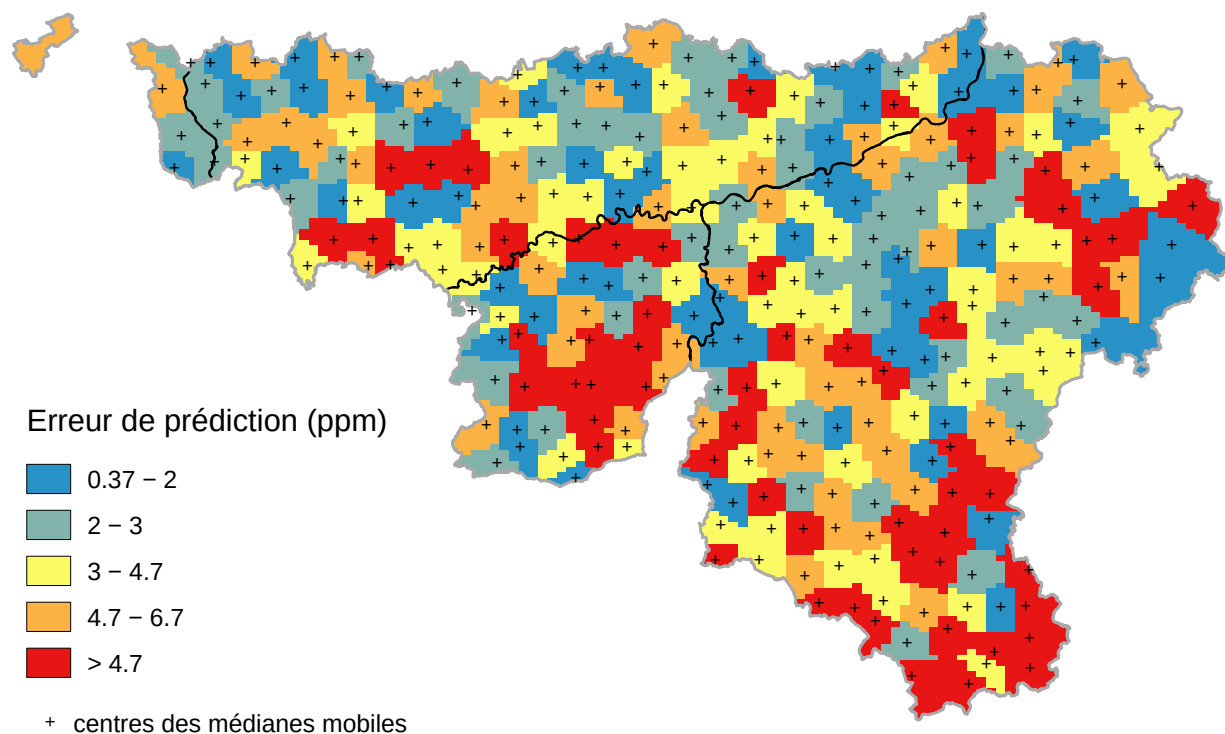
Cd : erreur de prédiction médiane dans les sols agricoles



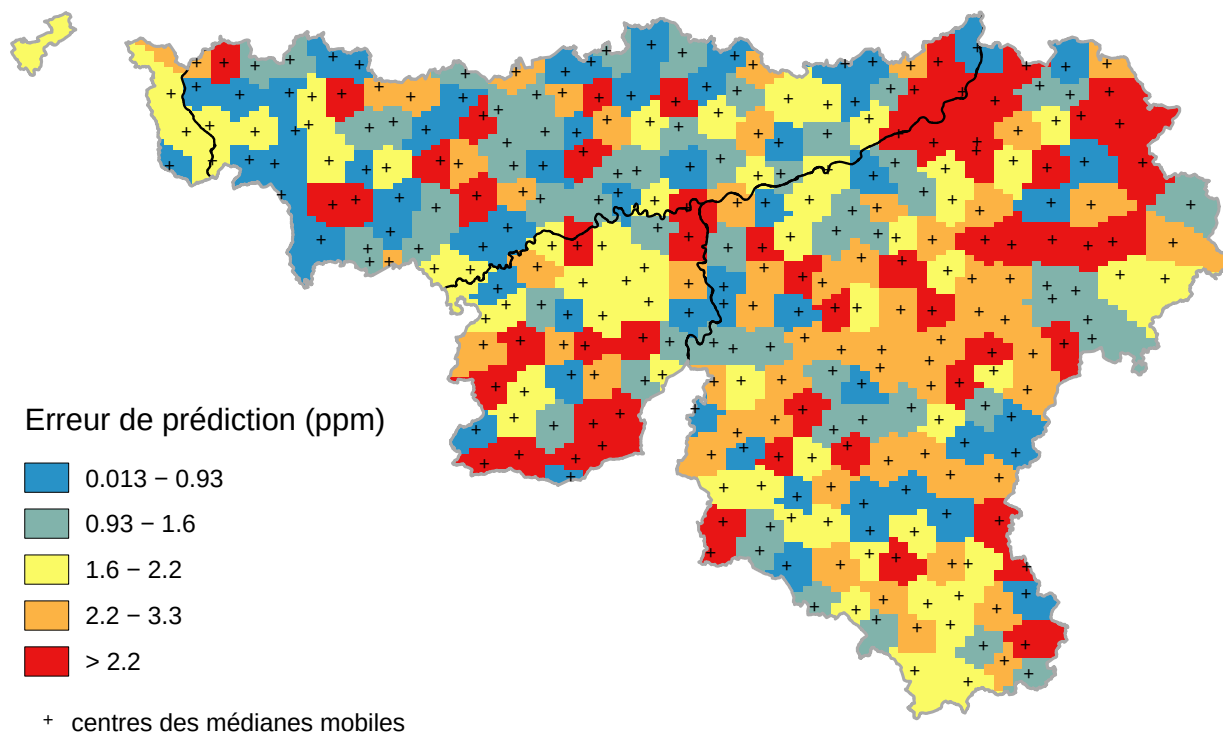
Cr : erreur de prédiction médiane dans les sols forestiers



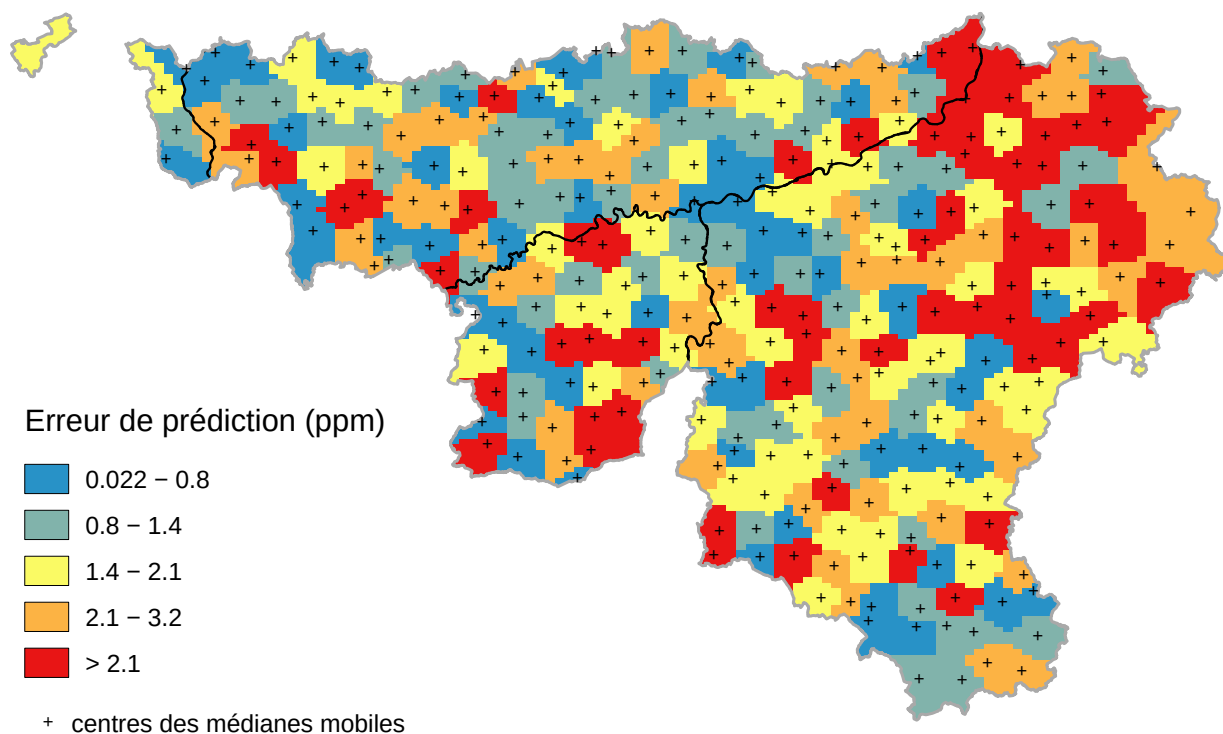
Cr : erreur de prédiction médiane dans les sols agricoles



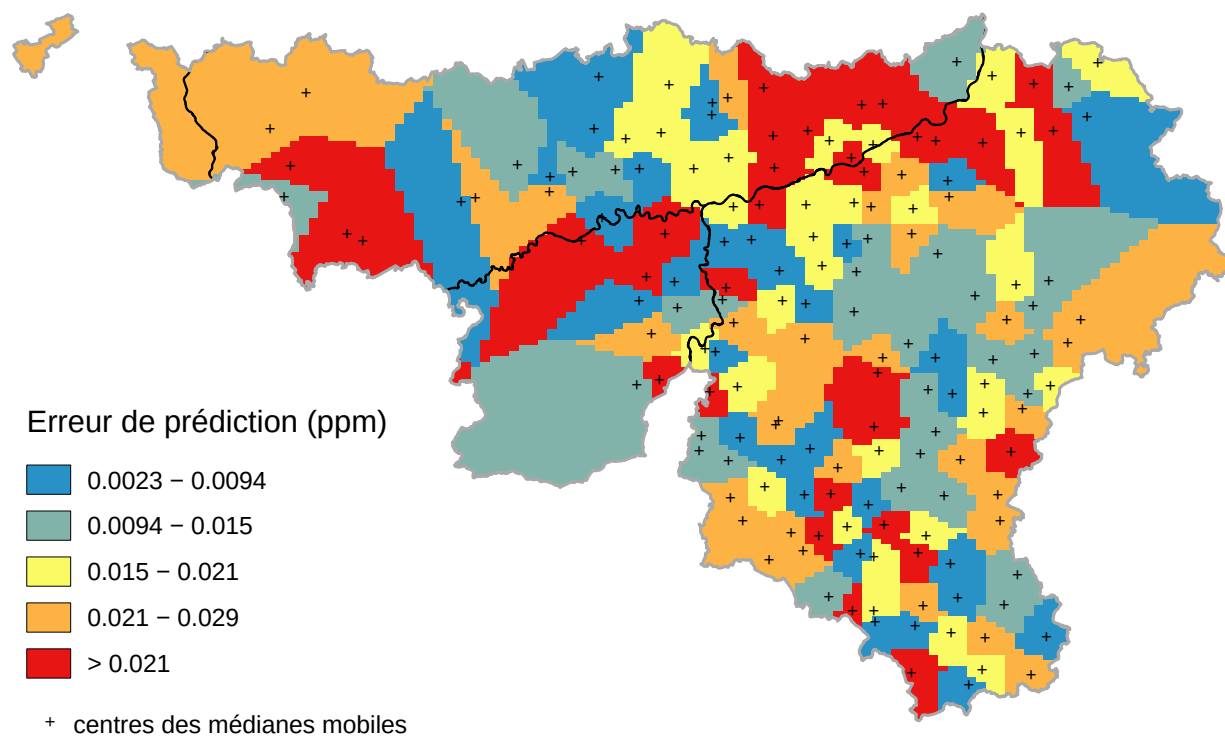
Cu : erreur de prédiction médiane dans les sols forestiers



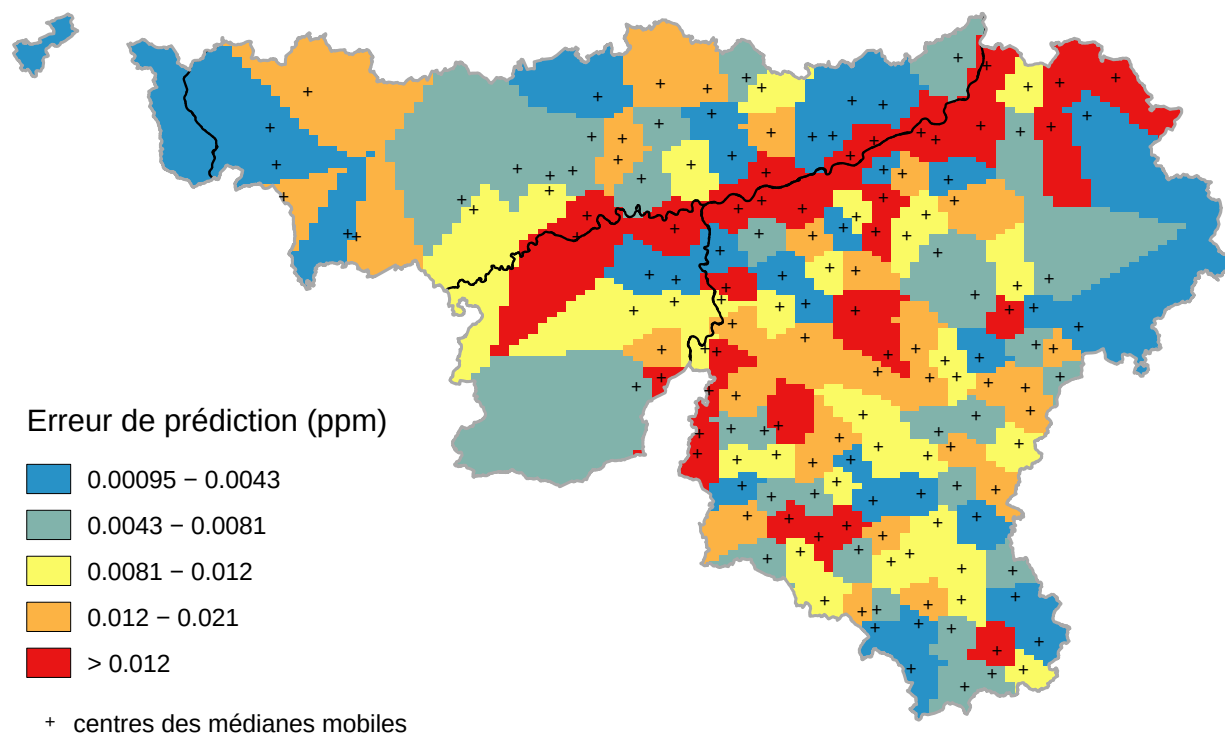
Cu : erreur de prédiction médiane dans les sols agricoles



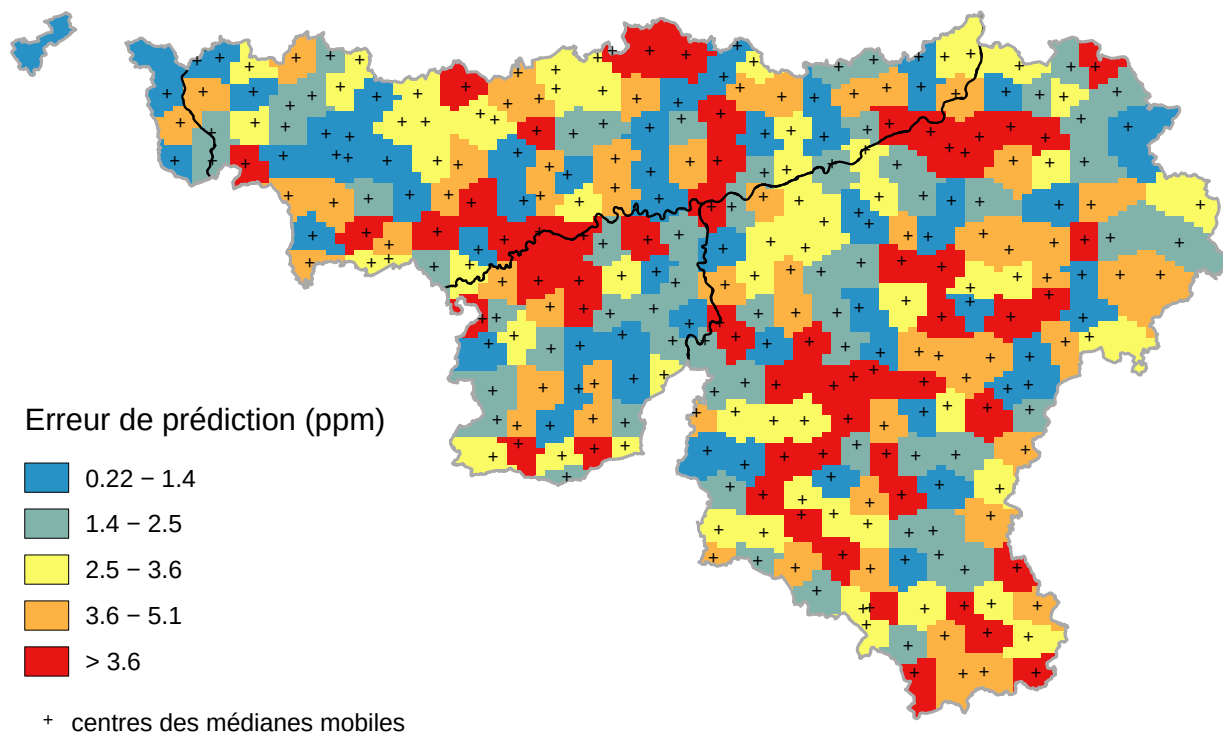
Hg : erreur de prédiction médiane dans les sols forestiers



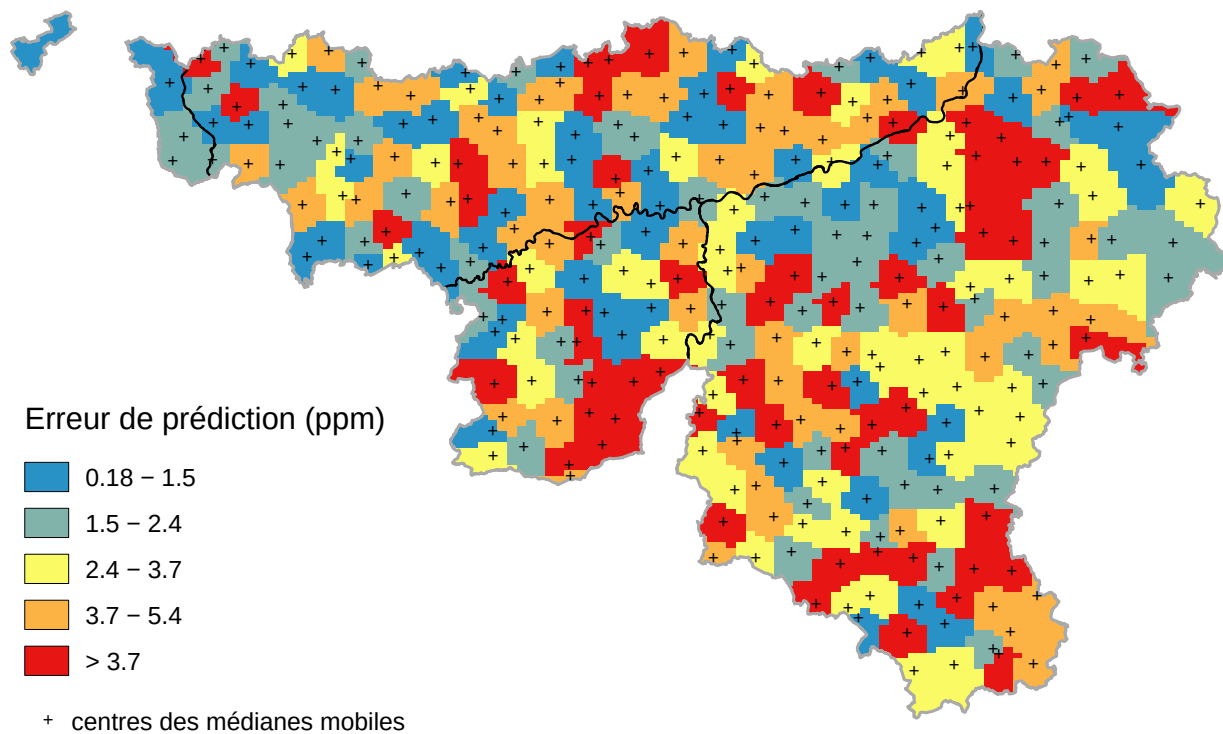
Hg : erreur de prédiction médiane dans les sols agricoles



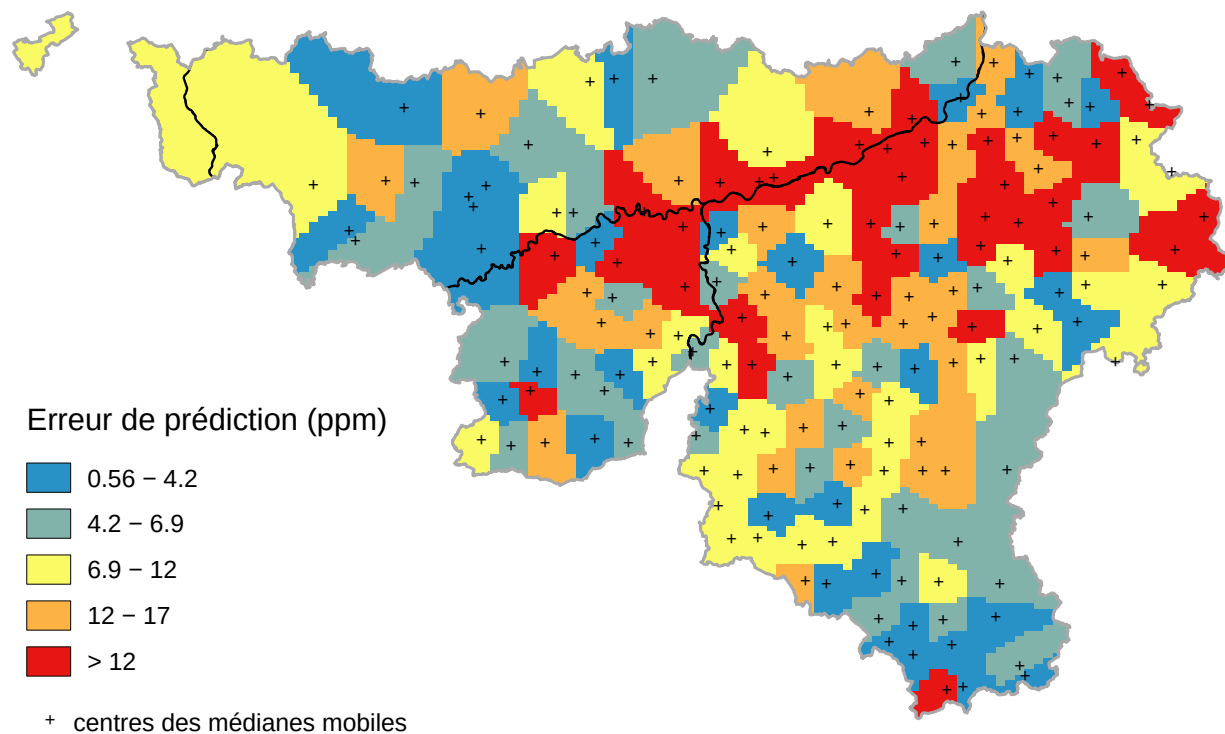
Ni : erreur de prédiction médiane dans les sols forestiers



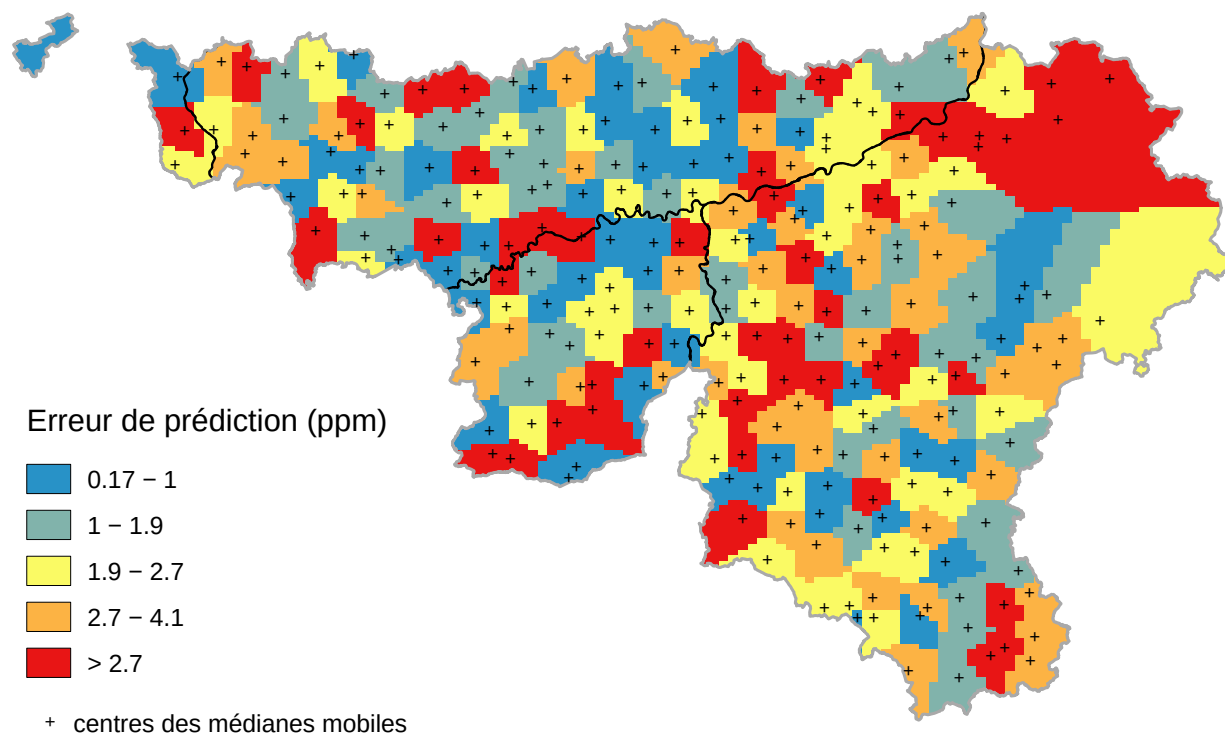
Ni : erreur de prédiction médiane dans les sols agricoles



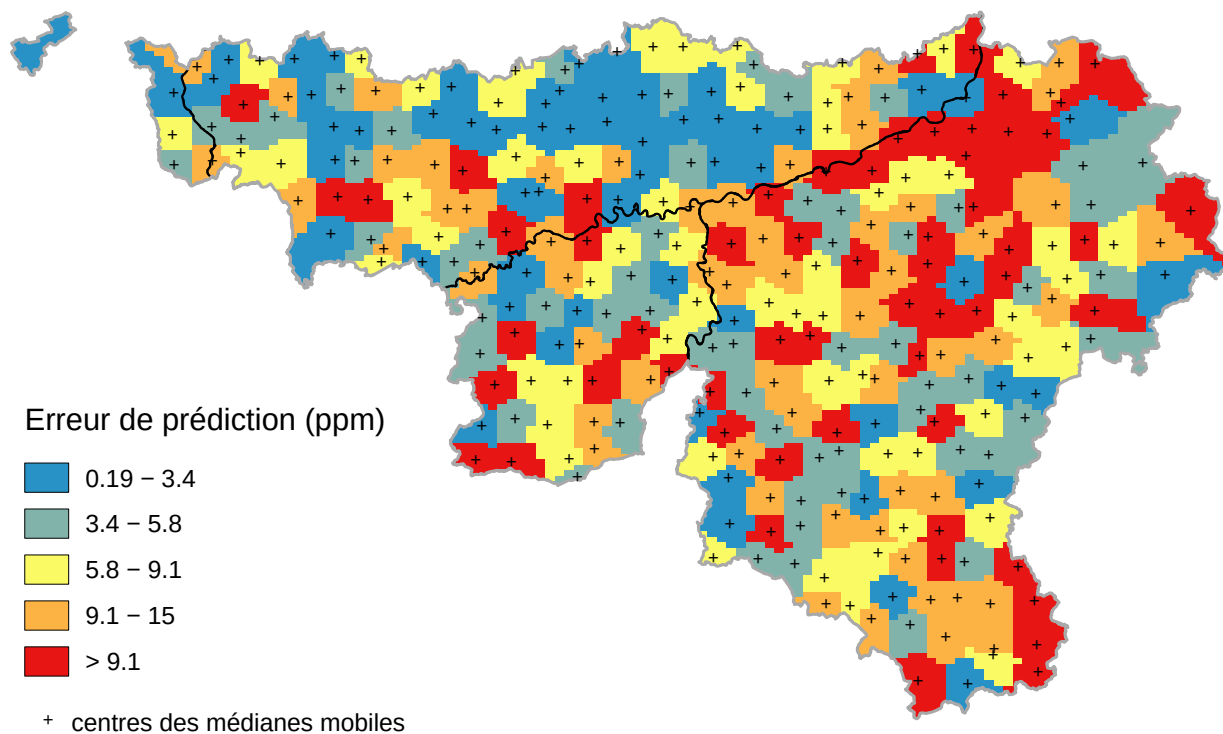
Pb : erreur de prédiction médiane dans les sols forestiers



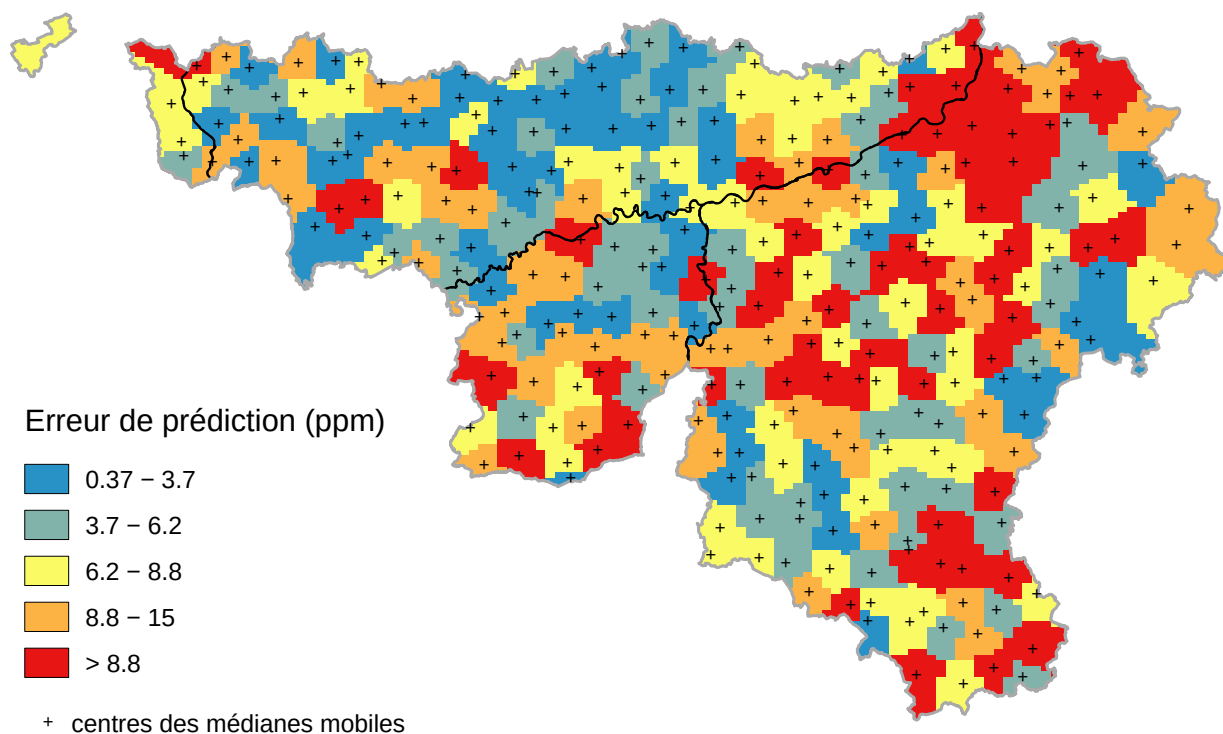
Pb : erreur de prédiction médiane dans les sols agricoles



Zn : erreur de prédiction médiane dans les sols forestiers



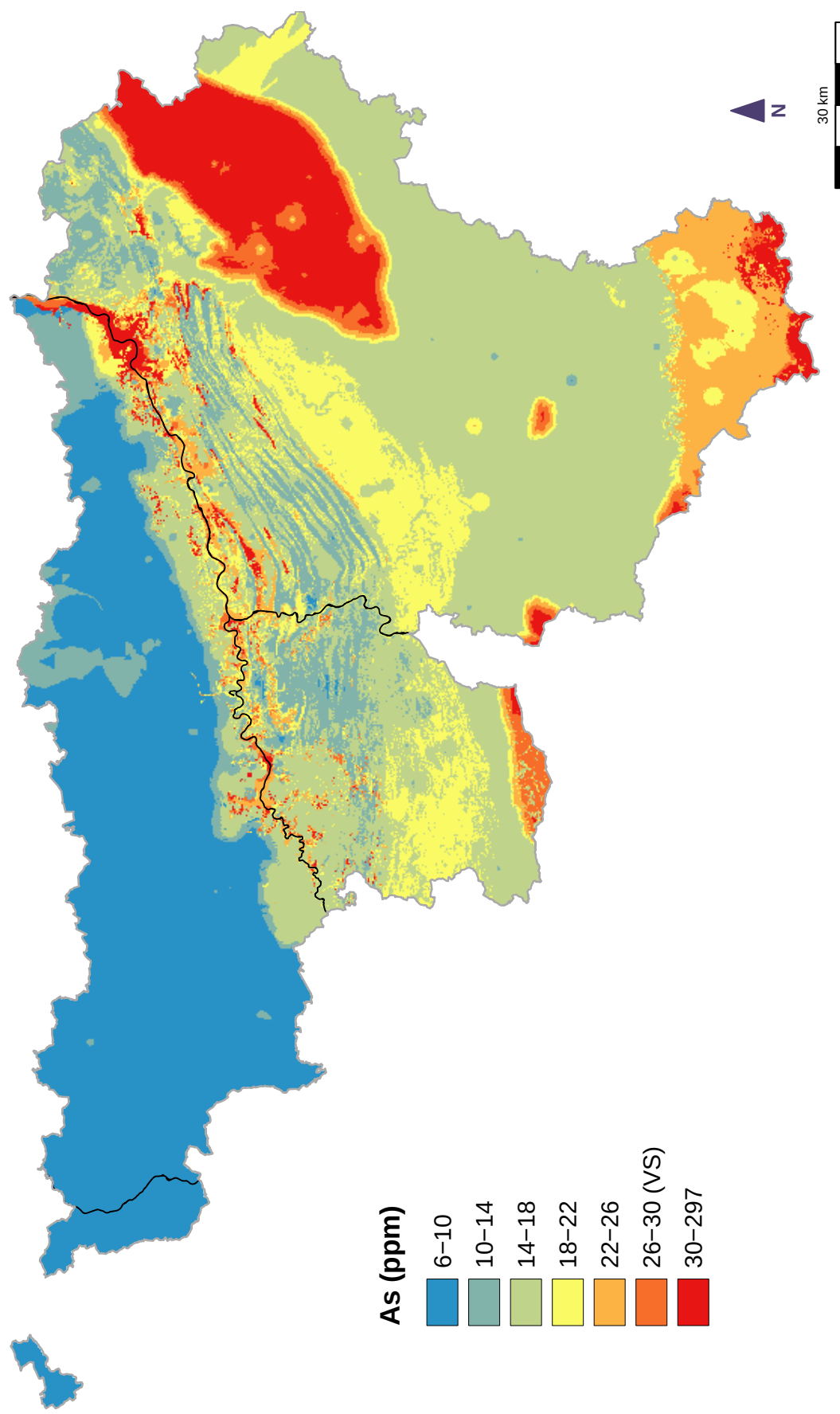
Zn : erreur de prédiction médiane dans les sols agricoles



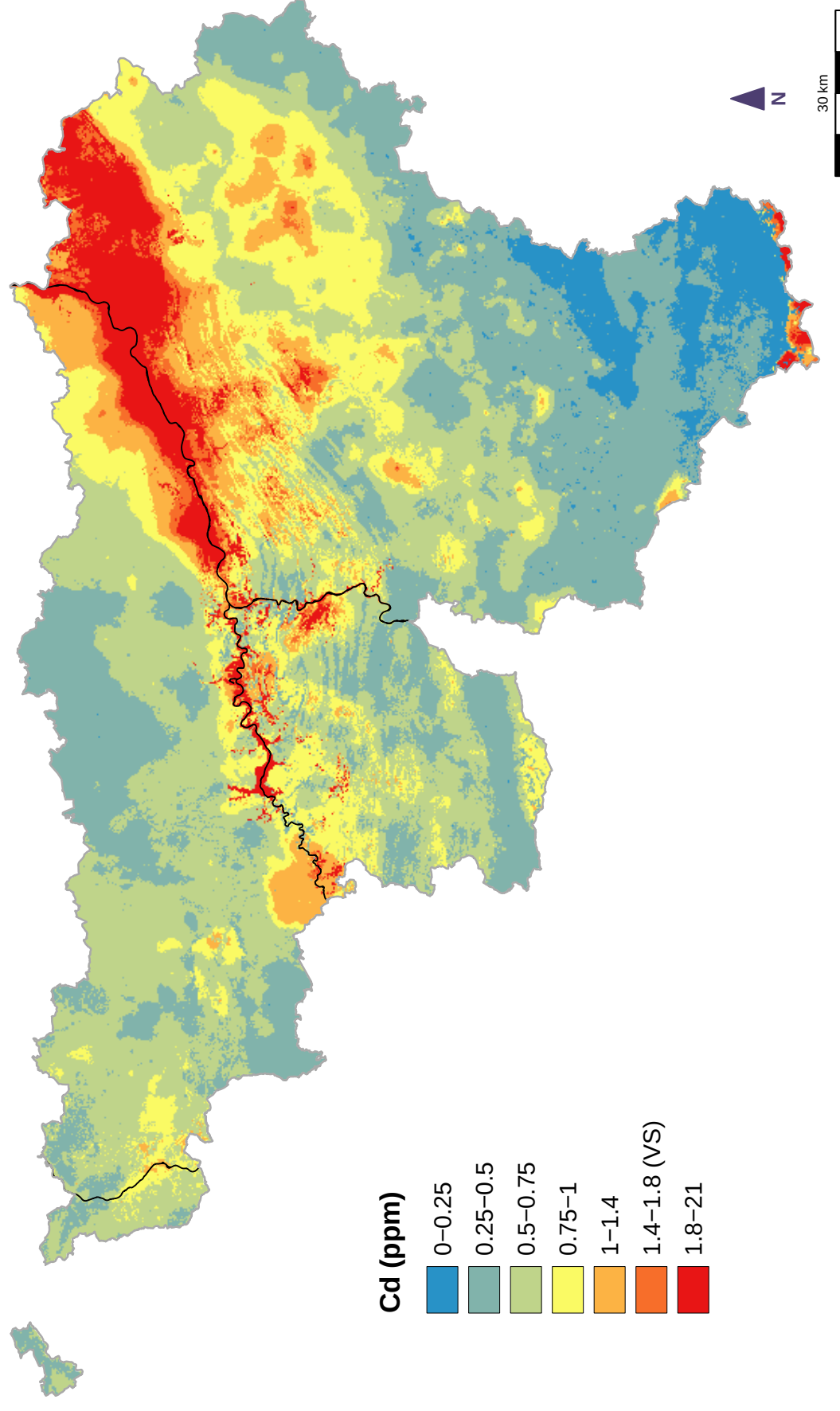
Annexe 2 : Cartes de l'incertitude sur les valeurs de la carte des concentrations de fond

Les cartes de centiles 90, 95 et 99 présentées dans cette annexe ont pour objectif de fournir une indication de l'incertitude sur les valeurs renseignées par la carte des concentrations de fond. Ces cartes renseignent une valeur qui a une probabilité théorique d'être dépassée correspondant à la valeur du centile. Dans ce qui suit se trouve les cartes des centiles 90, 95 et 99 afin de représenter la concentration qui a moins de 10, 5 et 1 pourcent respectivement de chance d'être dépassée. Les codes couleurs utilisées pour dresser les cartes de centiles sont les mêmes que ceux utilisés dans les cartes de concentration de fond (*cf.* section 4.5) afin de rendre possible la comparaison des cartes.

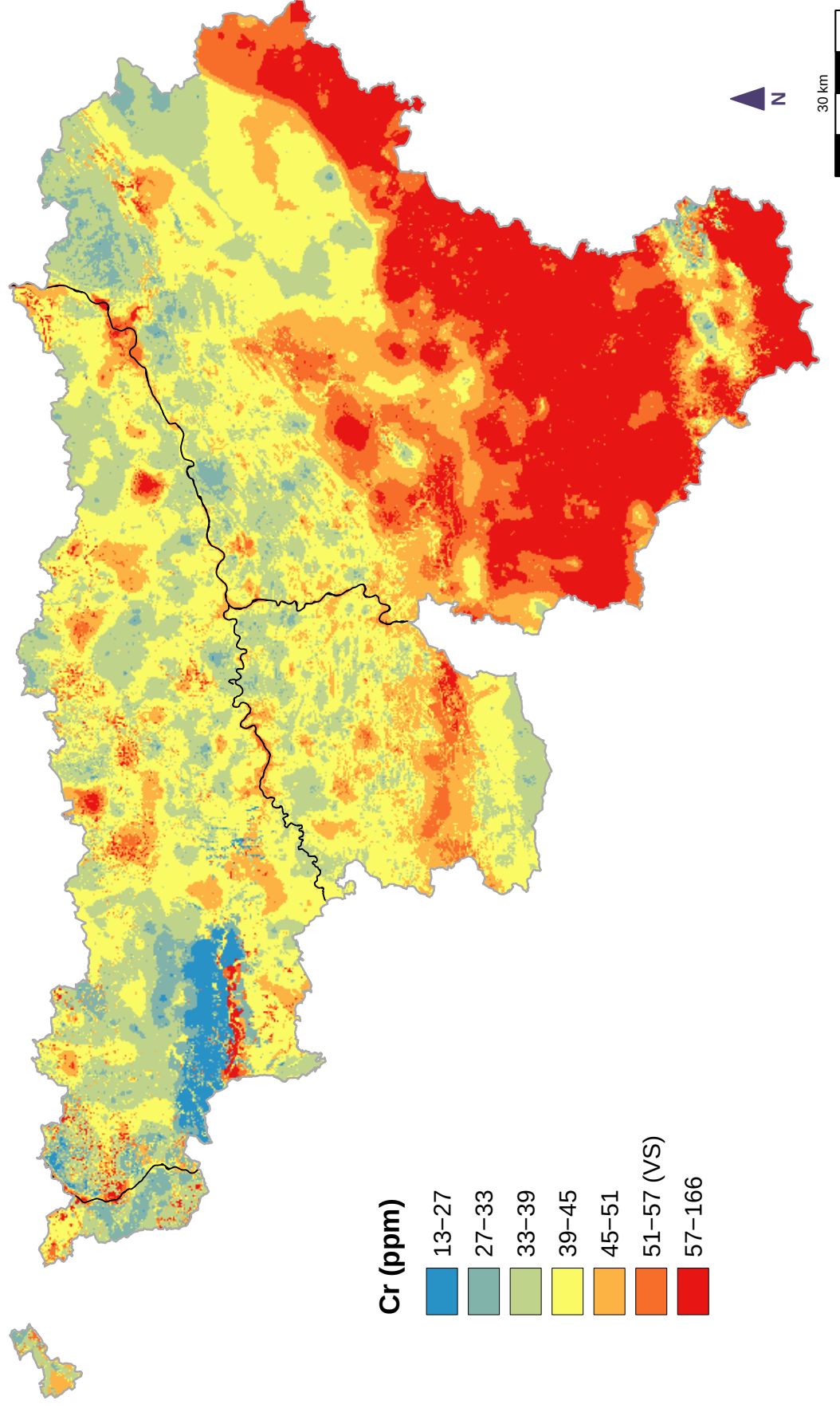
Teneur qui a moins d'1 chance sur 10 d'être dépassée dans un sol agricole ou forestier



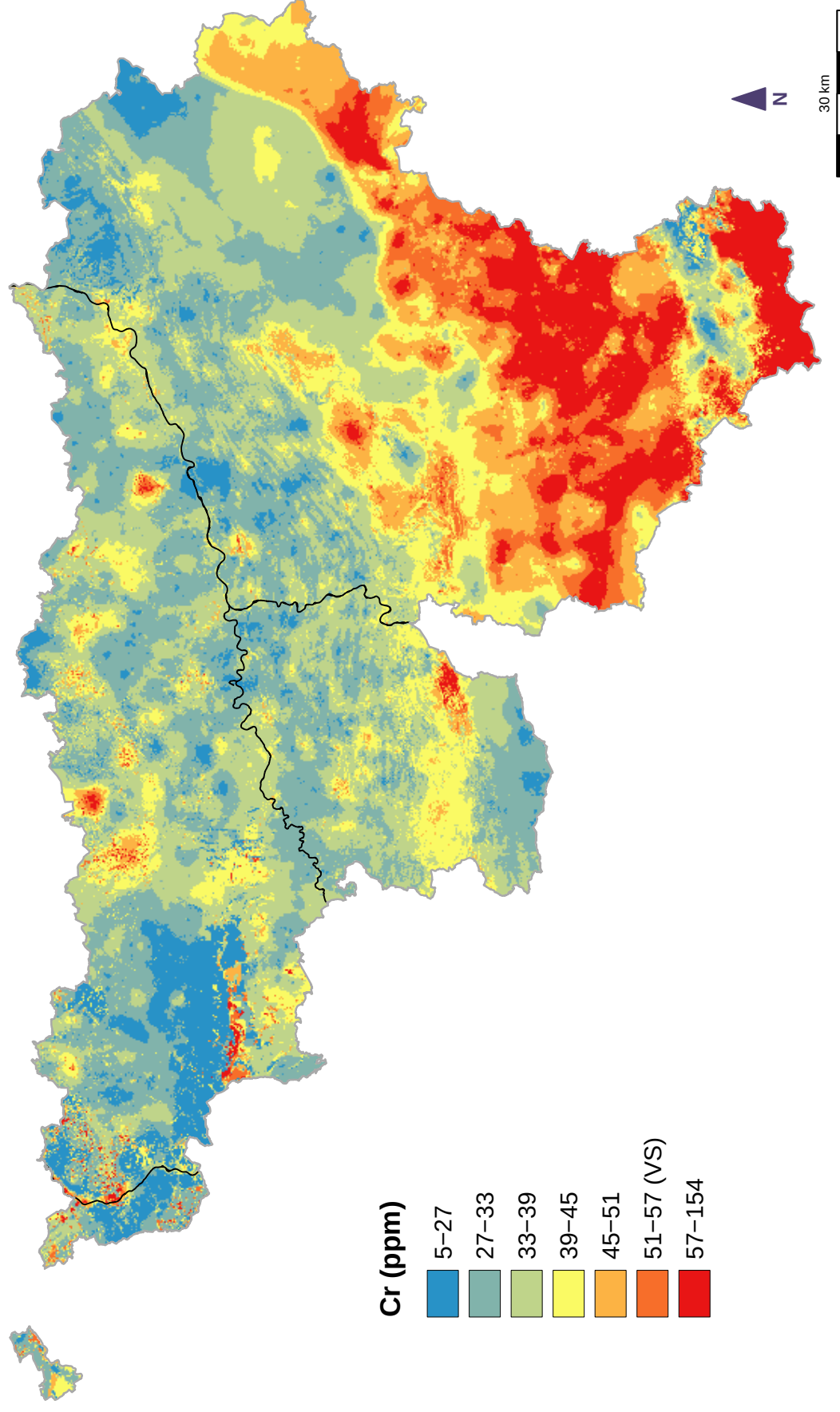
Teneur qui a moins d'1 chance sur 10 d'être dépassée dans un sol agricole ou forestier



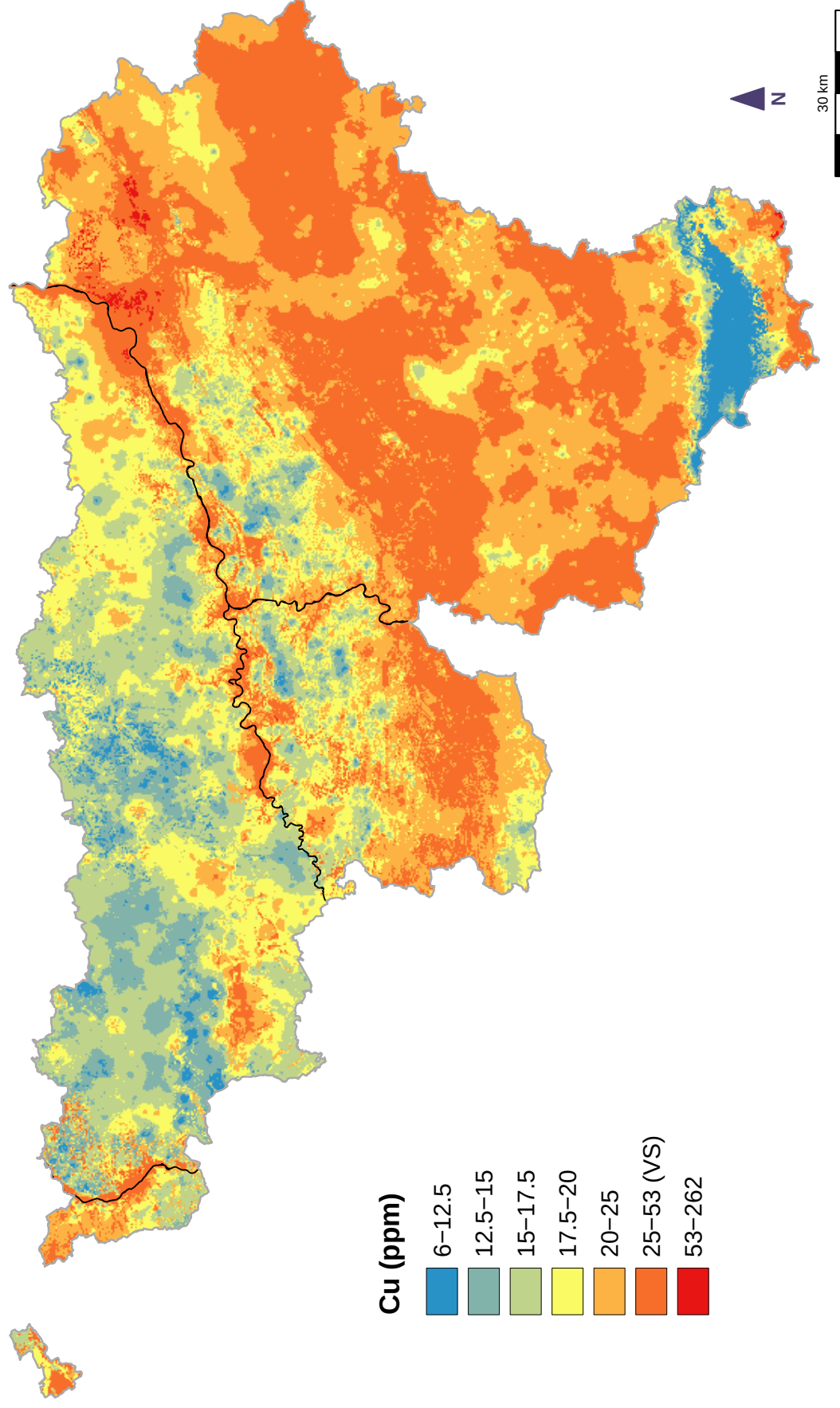
Teneur qui a moins d'1 chance sur 10 d'être dépassée dans un sol agricole



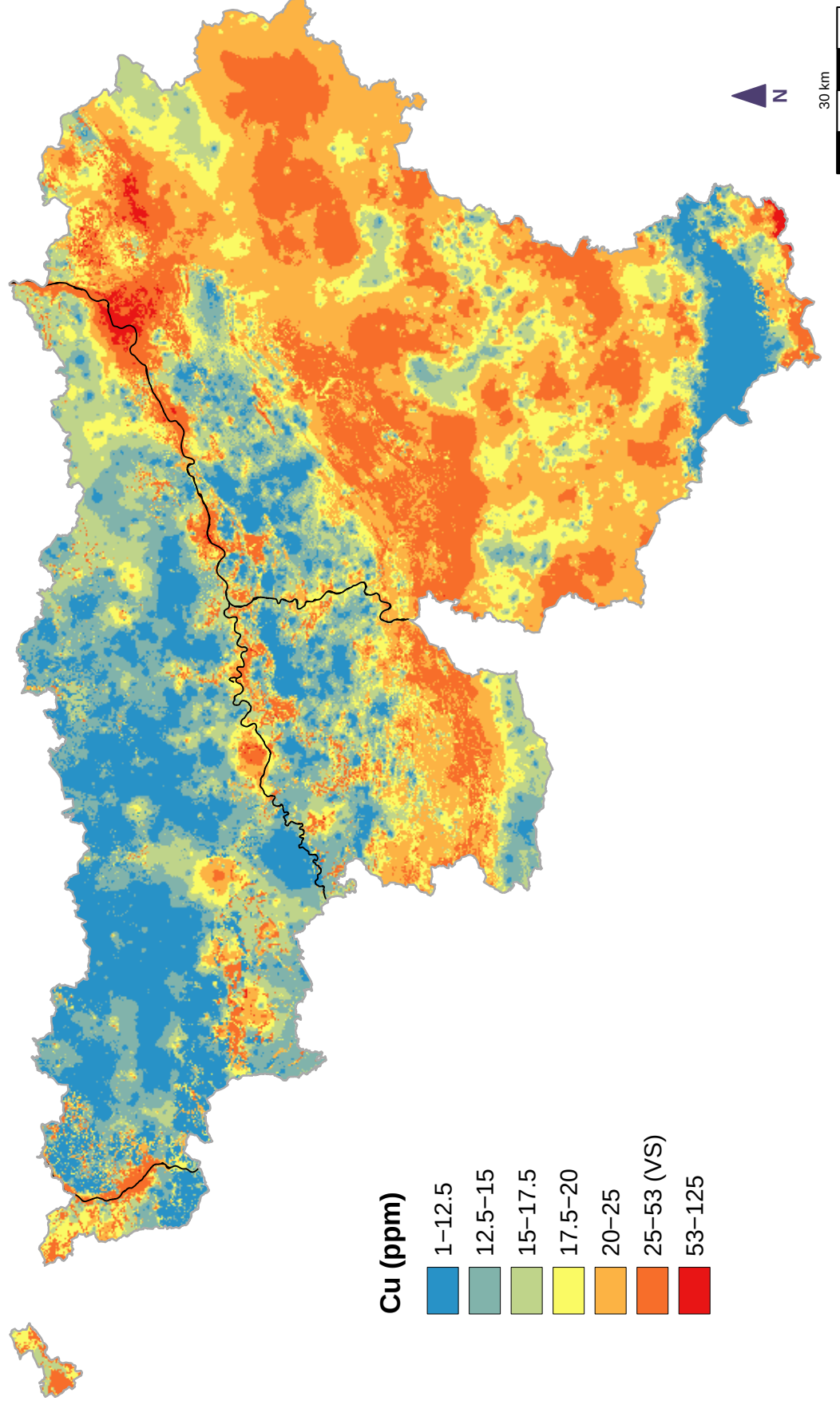
Teneur qui a moins d'1 chance sur 10 d'être dépassée dans un sol forestier



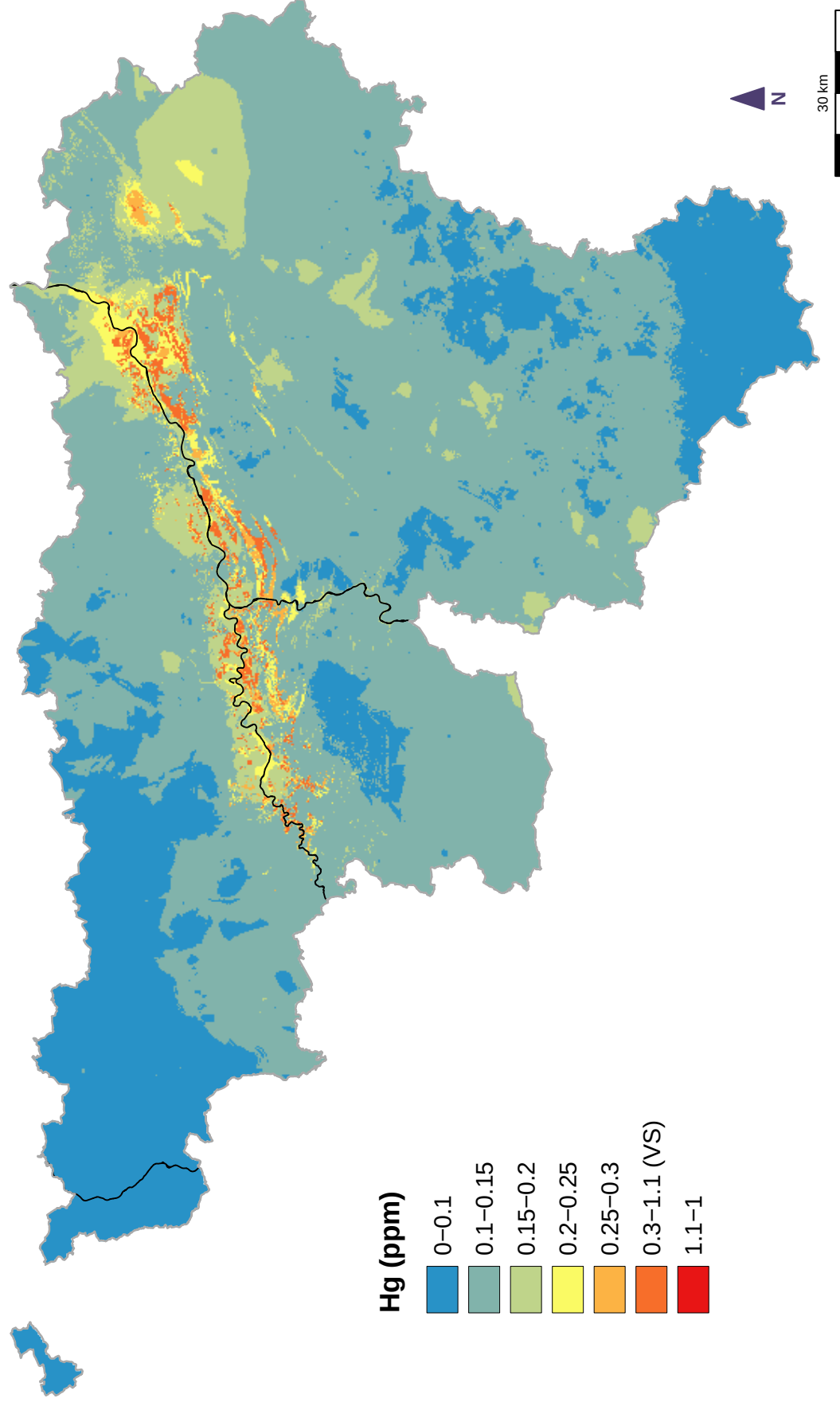
Teneur qui a moins d'1 chance sur 10 d'être dépassée dans un sol agricole



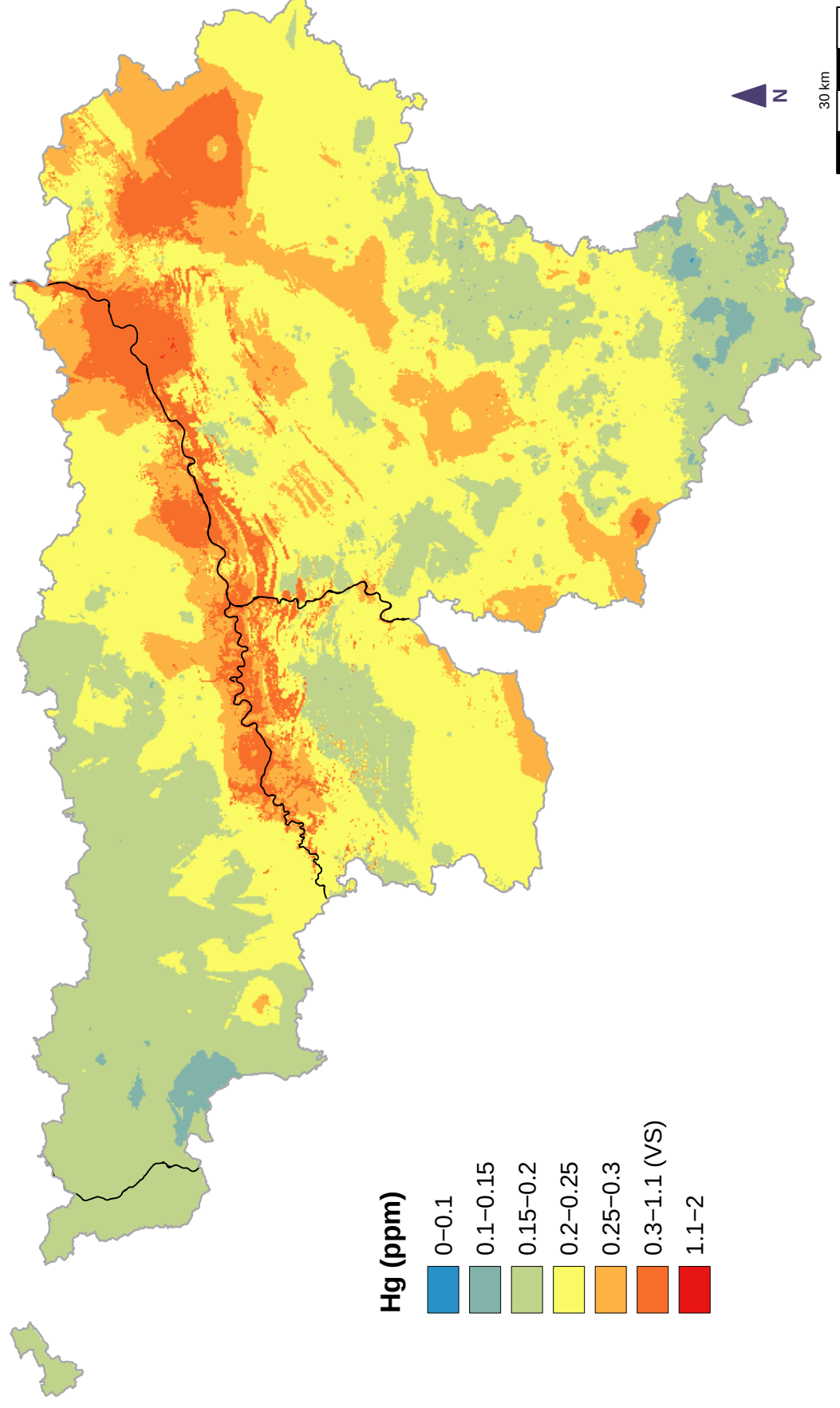
Teneur qui a moins d'1 chance sur 10 d'être dépassée dans un sol forestier



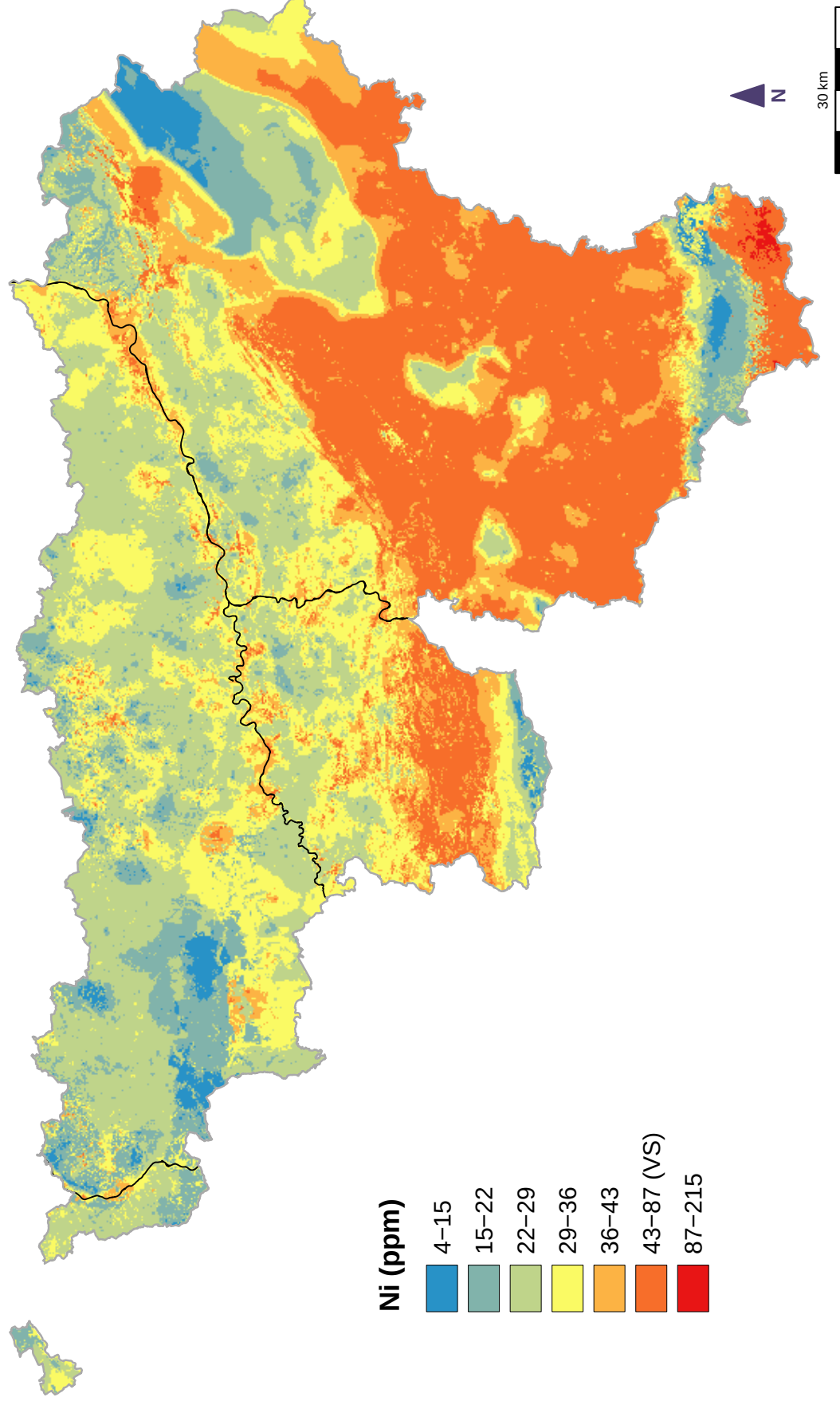
Teneur qui a moins d'1 chance sur 10 d'être dépassée dans un sol agricole



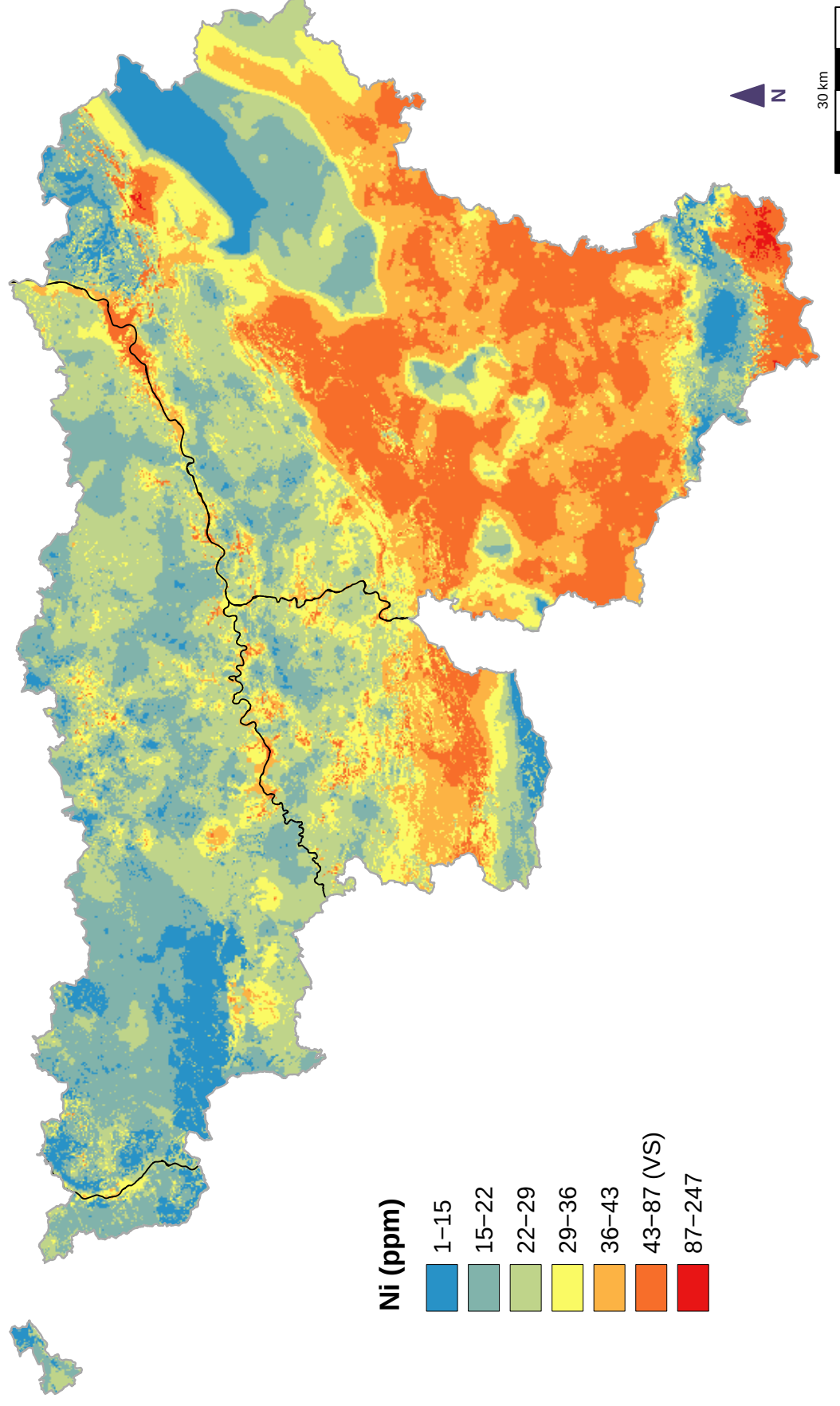
Teneur qui a moins d'1 chance sur 10 d'être dépassée dans un sol forestier



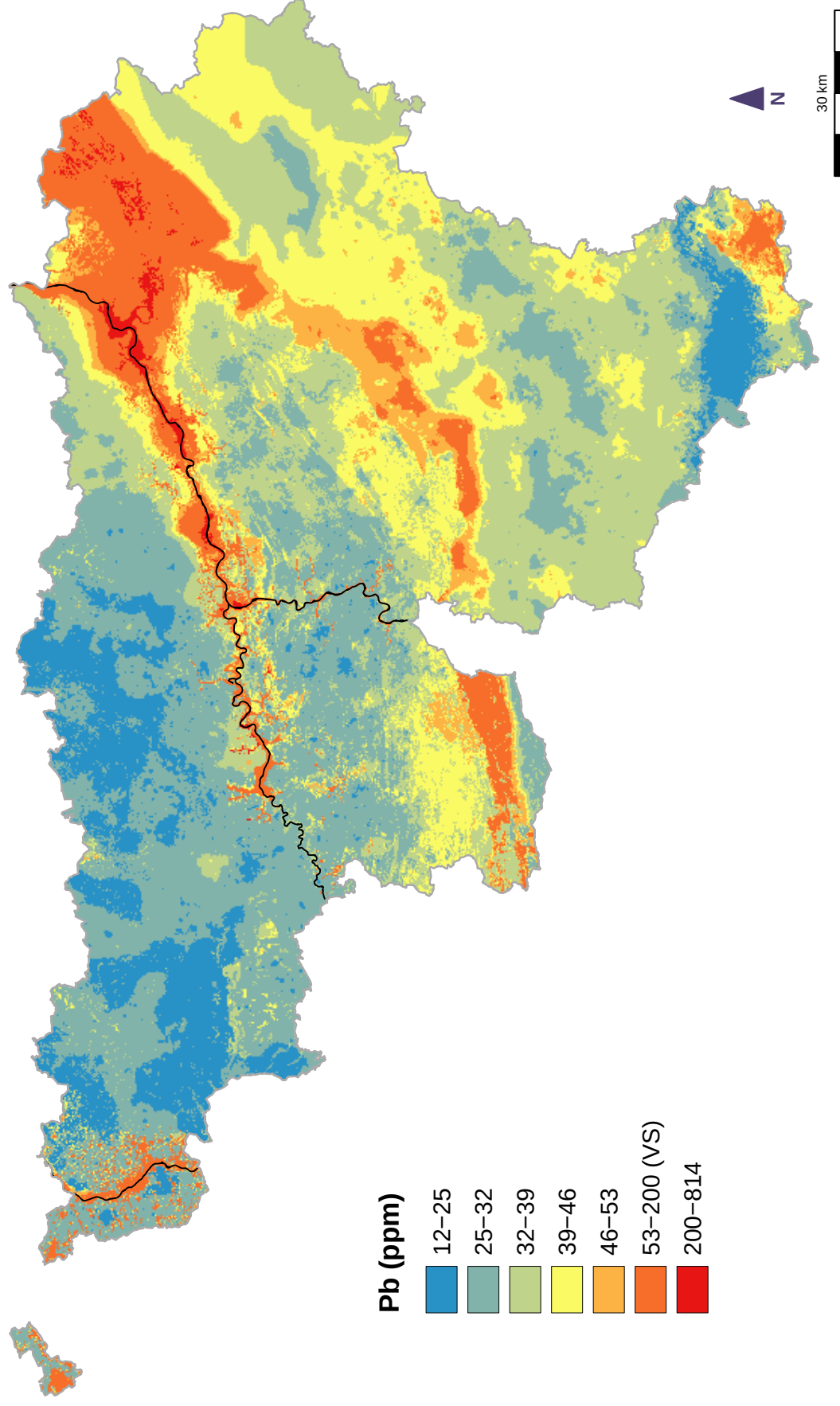
Teneur qui a moins d'1 chance sur 10 d'être dépassée dans un sol agricole



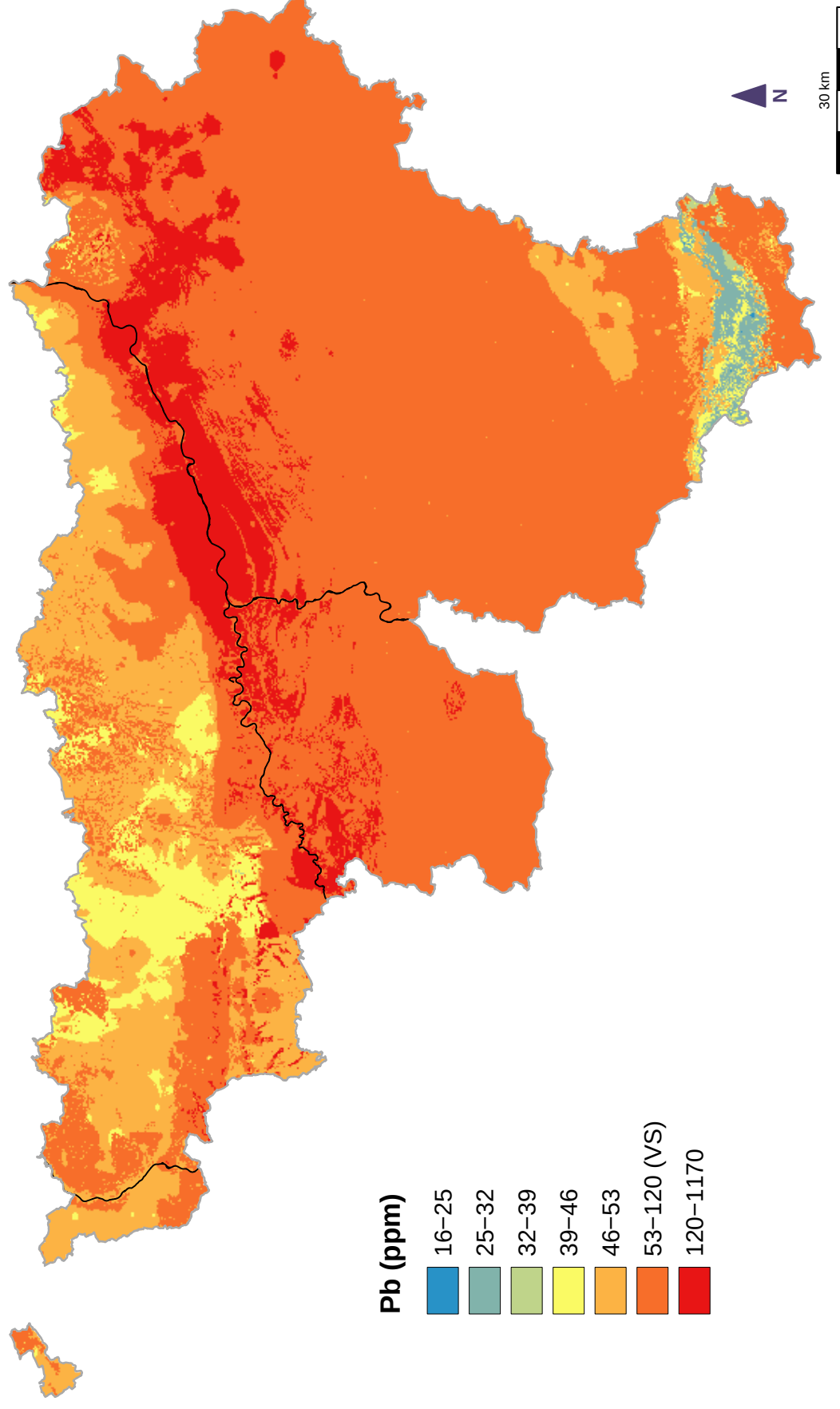
Teneur qui a moins d'1 chance sur 10 d'être dépassée dans un sol forestier



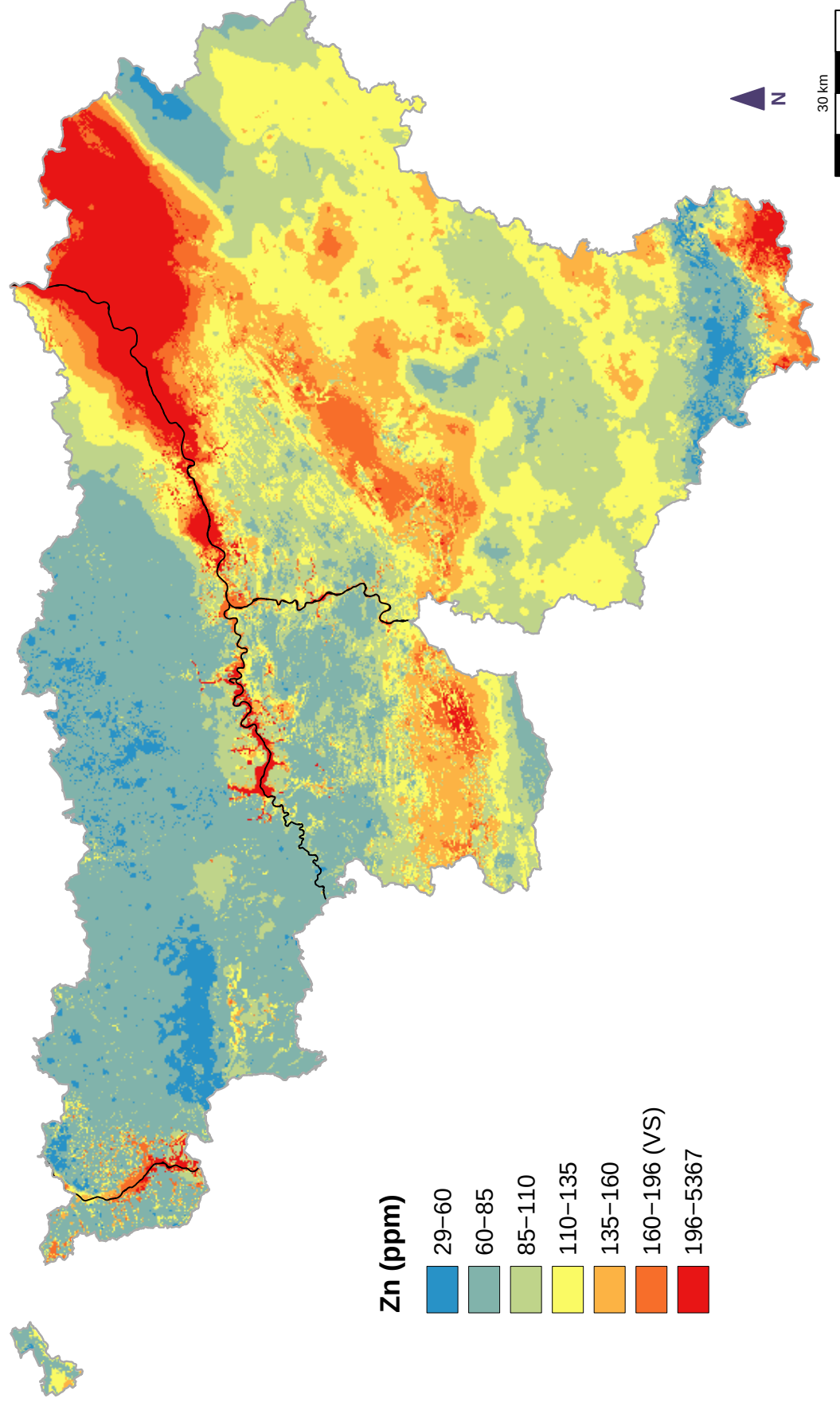
Teneur qui a moins d'1 chance sur 10 d'être dépassée dans un sol agricole



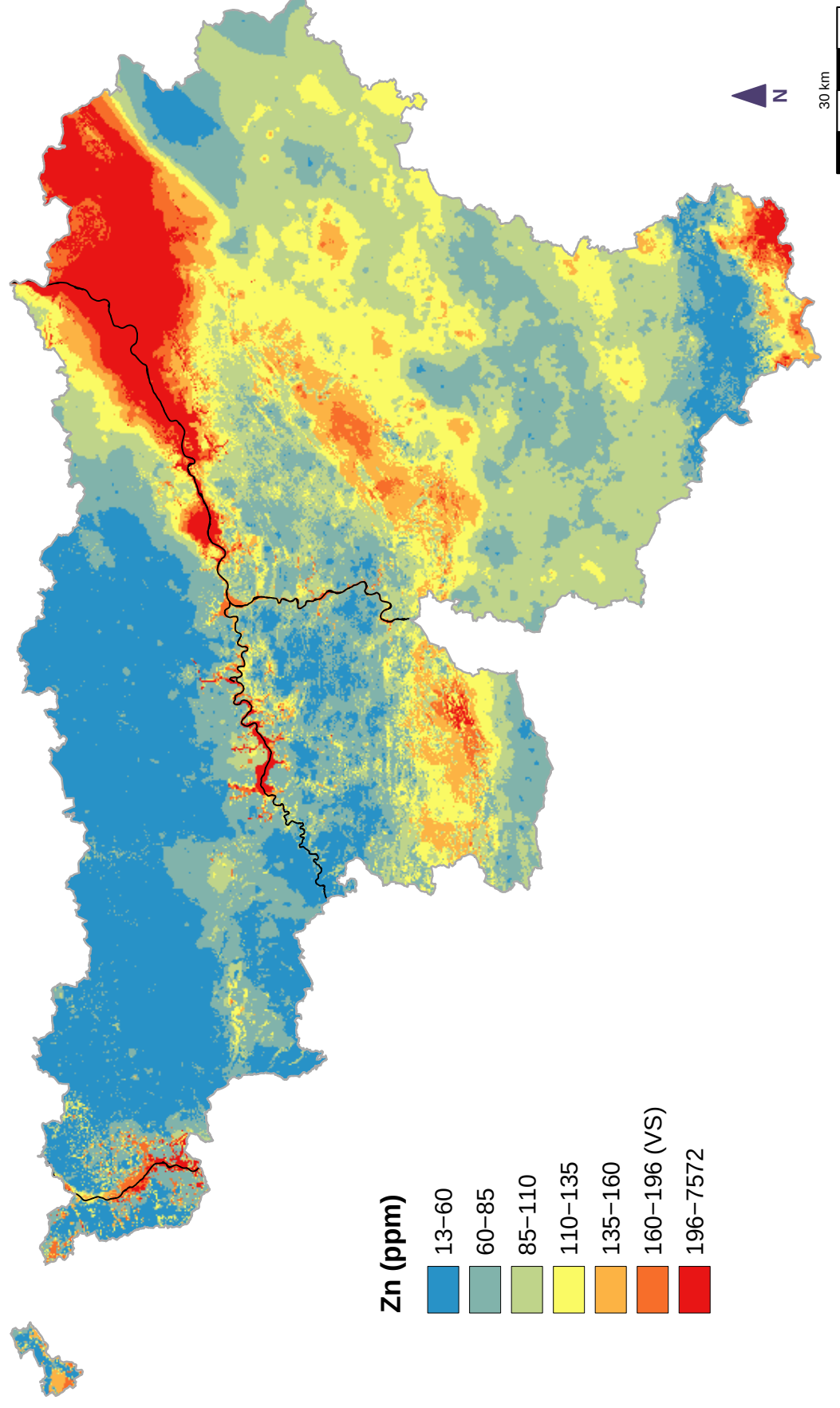
Teneur qui a moins d'1 chance sur 10 d'être dépassée dans un sol forestier



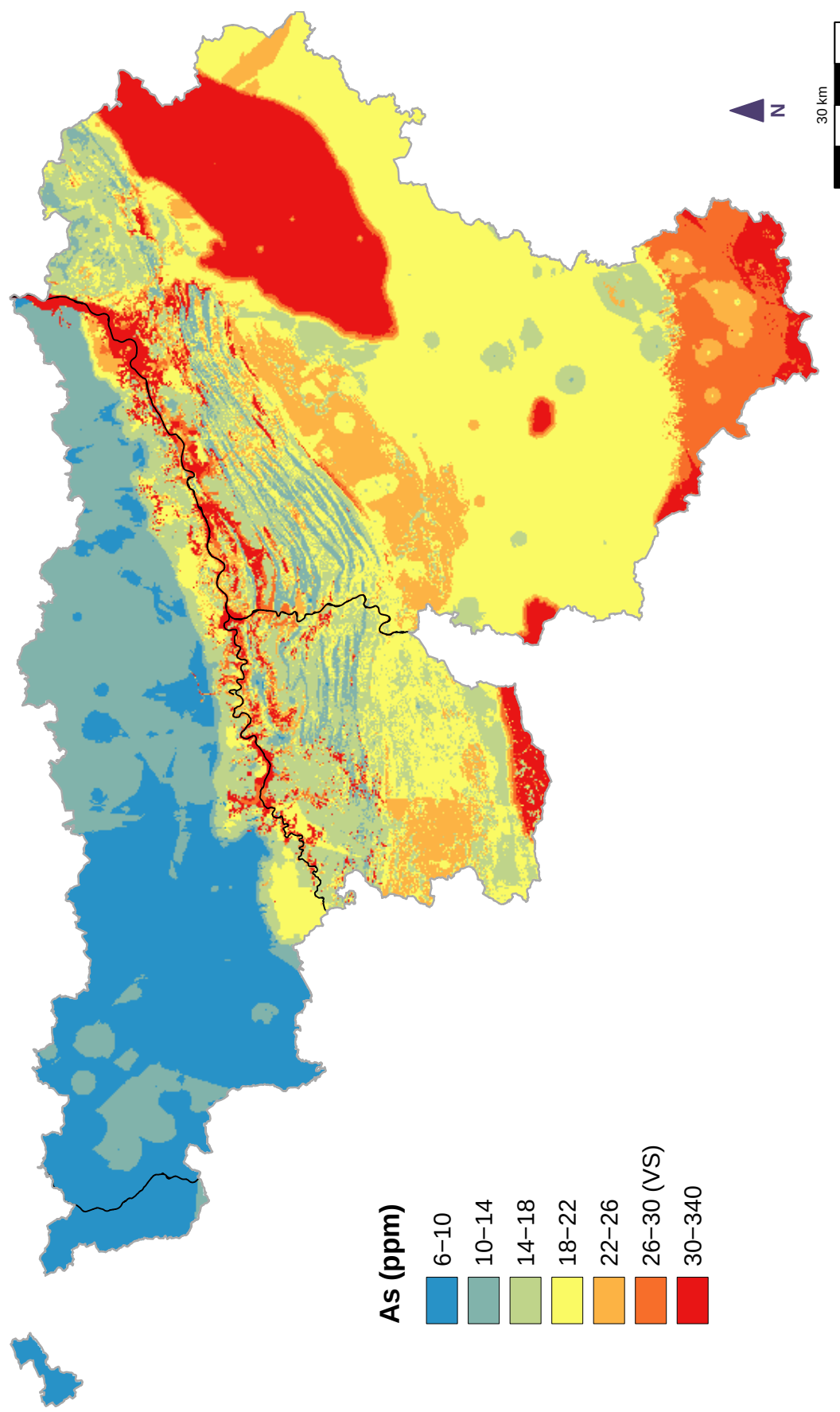
Teneur qui a moins d'1 chance sur 10 d'être dépassée dans un sol agricole



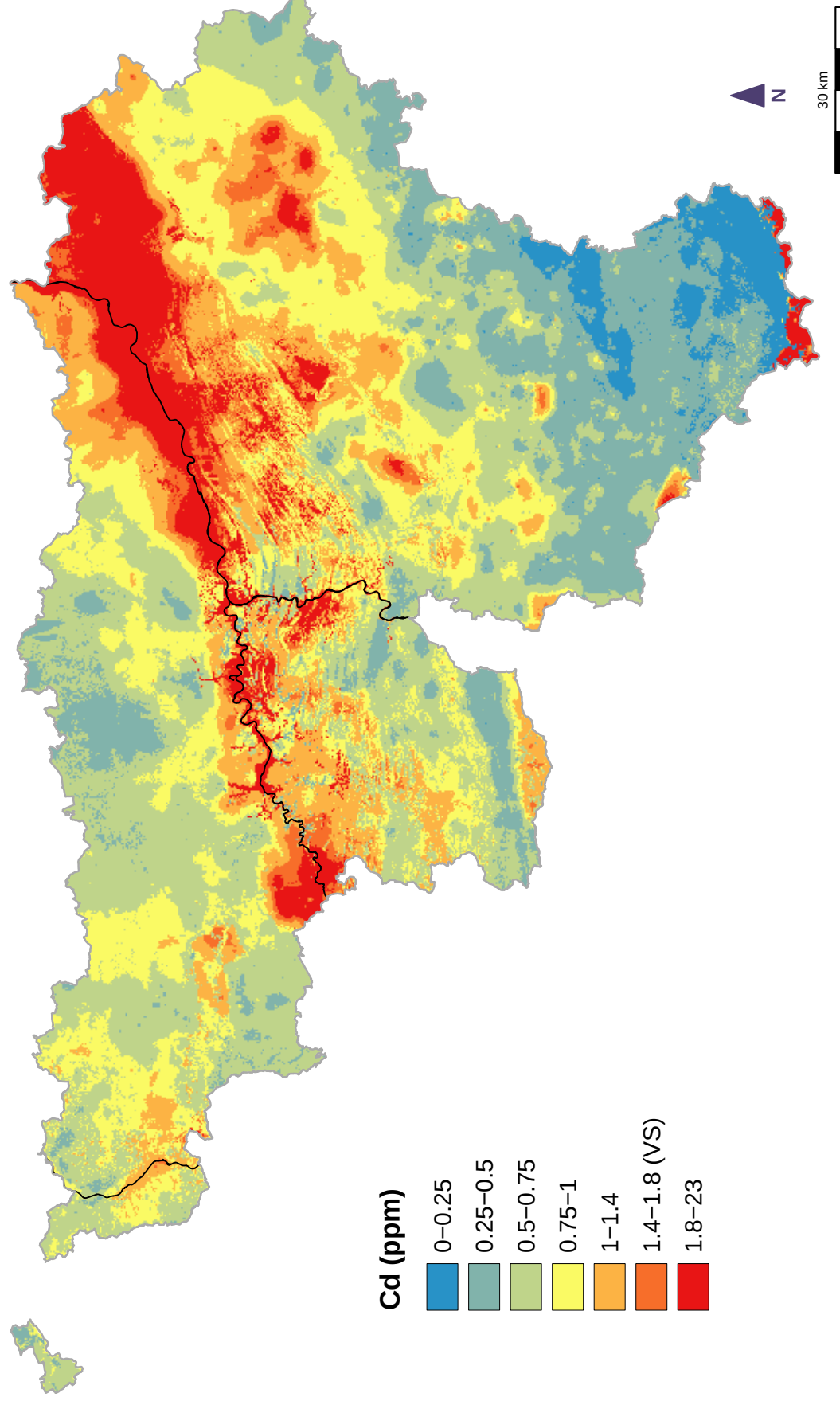
Teneur qui a moins d'1 chance sur 10 d'être dépassée dans un sol forestier



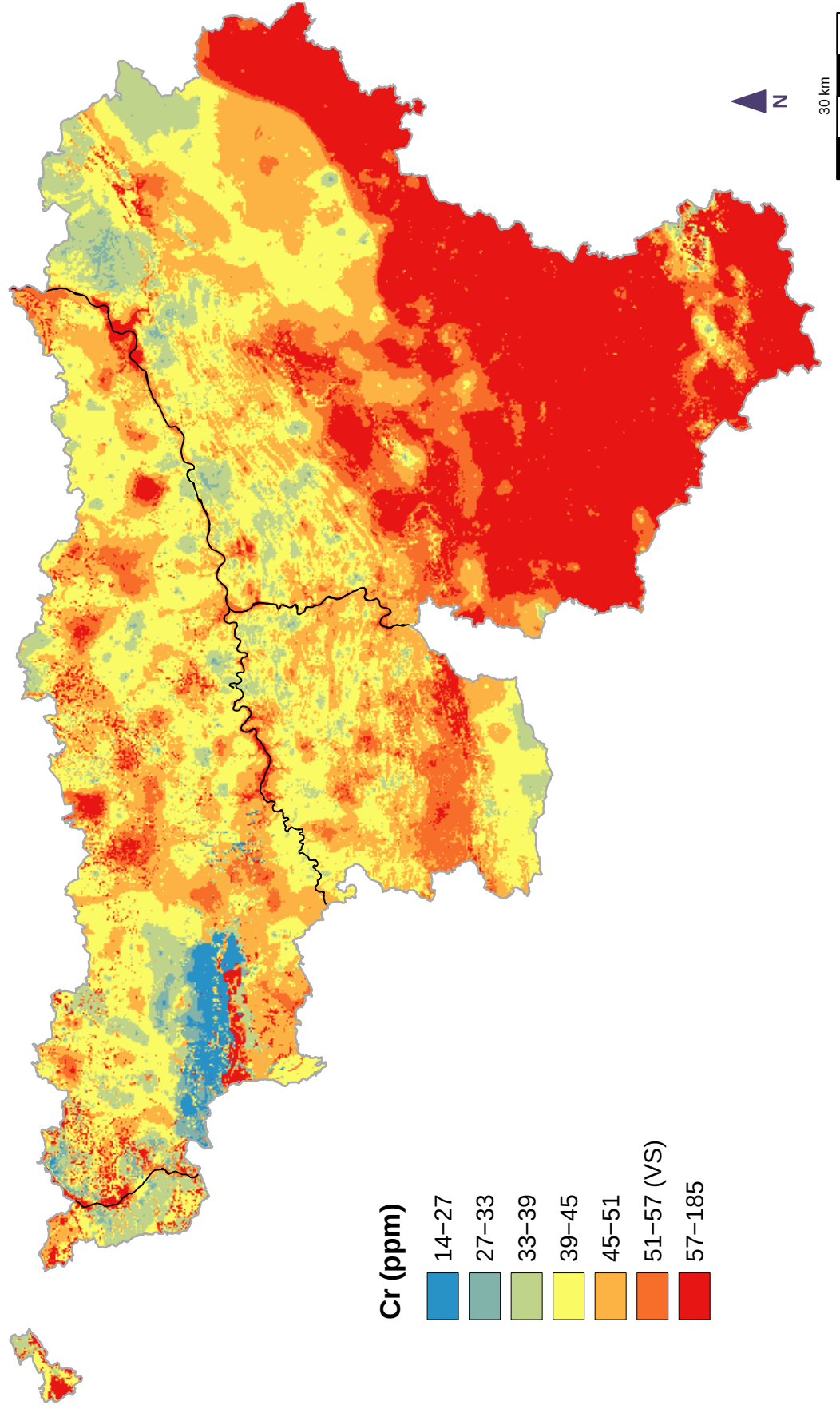
Teneur qui a moins d'1 chance sur 20 d'être dépassée dans un sol agricole ou forestier



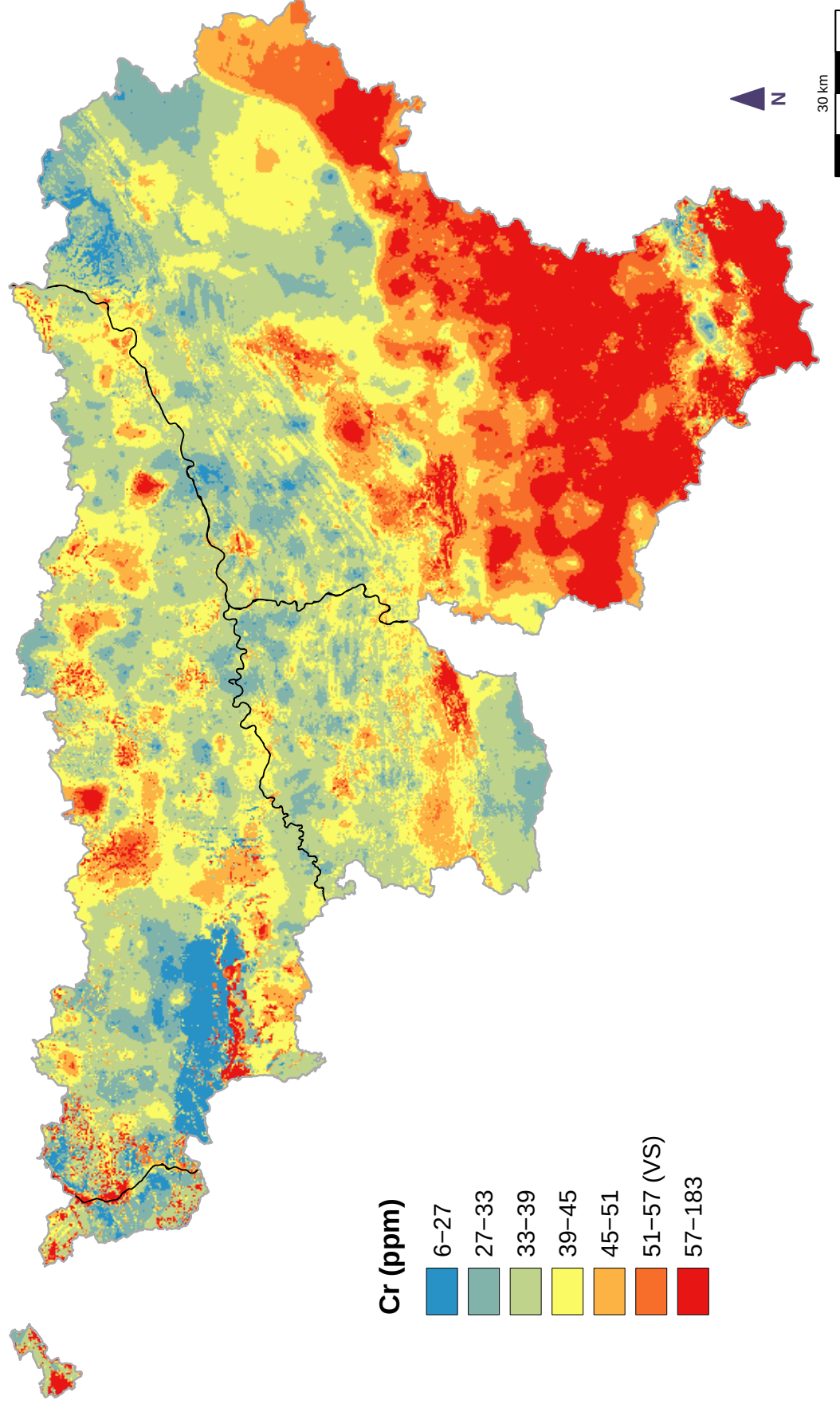
Teneur qui a moins d'1 chance sur 20 d'être dépassée dans un sol agricole ou forestier



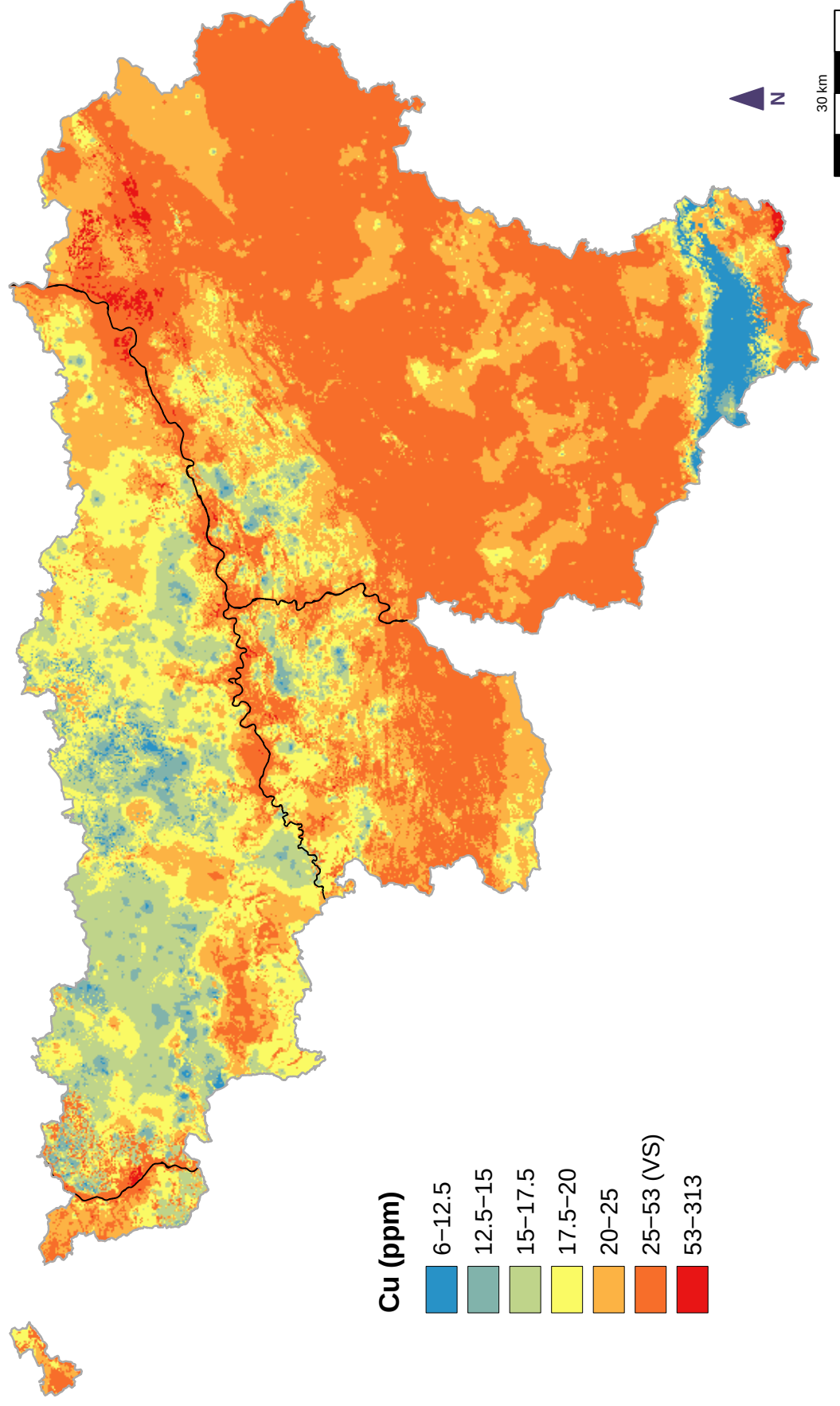
Teneur qui a moins d'1 chance sur 20 d'être dépassée dans un sol agricole



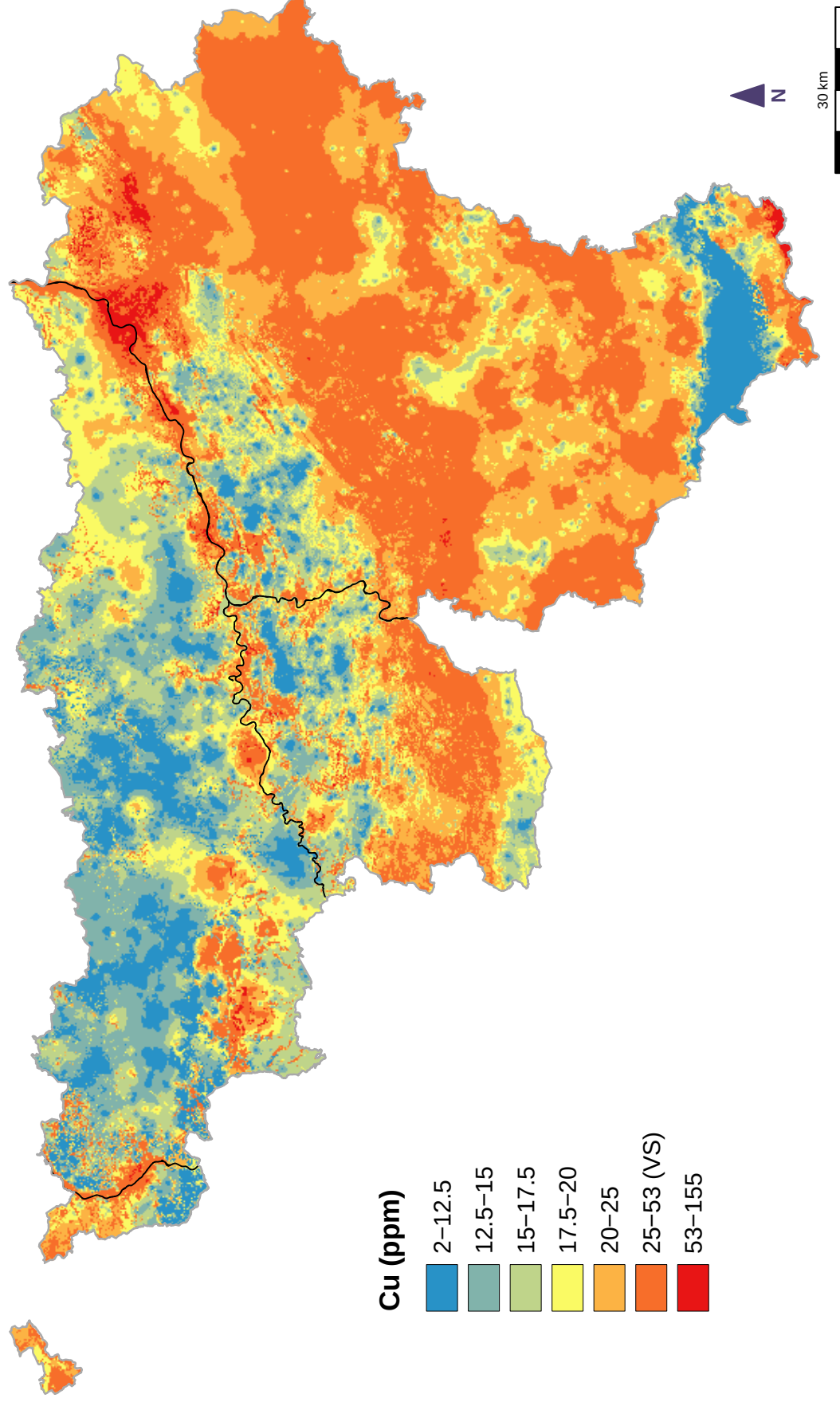
Teneur qui a moins d'1 chance sur 20 d'être dépassée dans un sol forestier



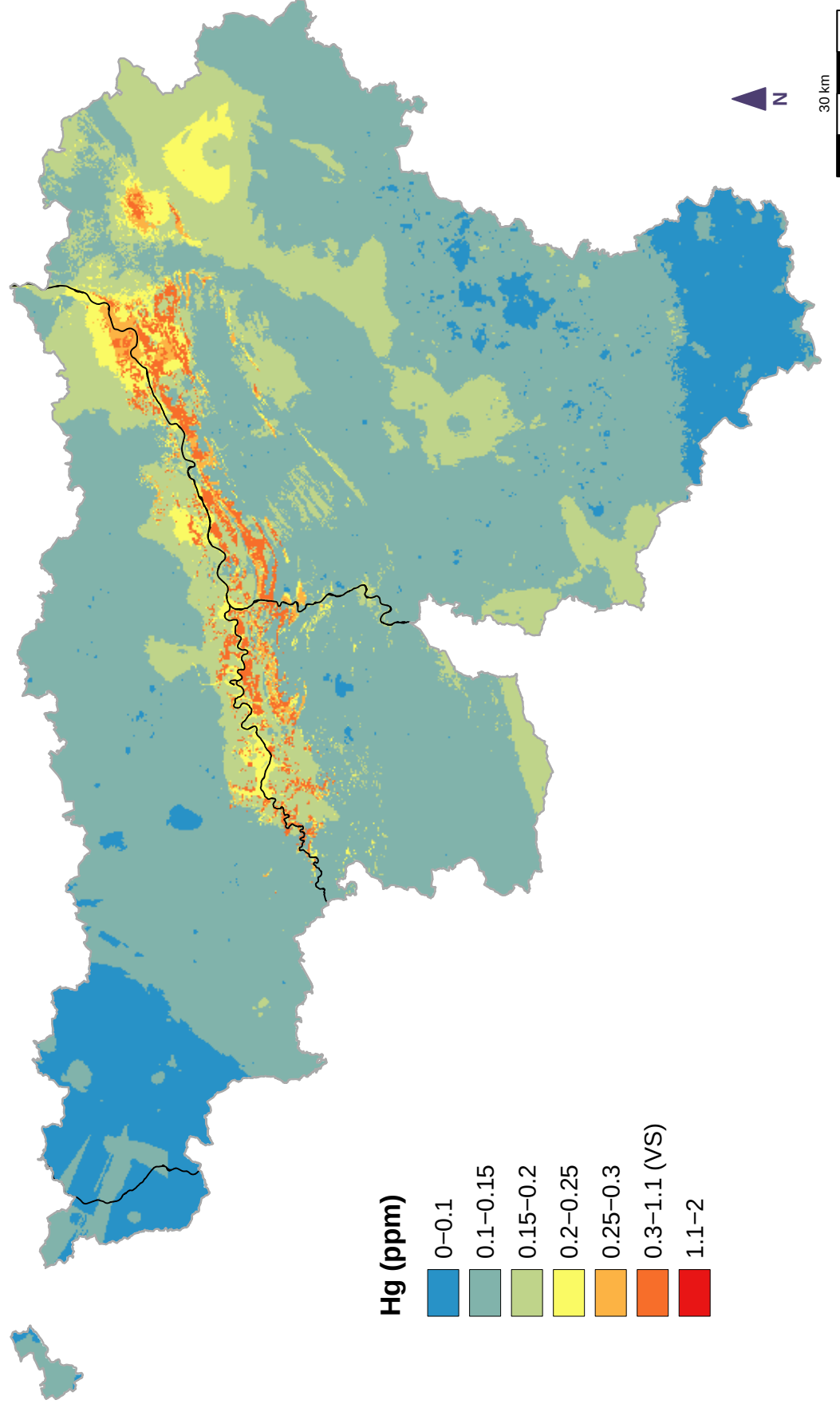
Teneur qui a moins d'1 chance sur 20 d'être dépassée dans un sol agricole



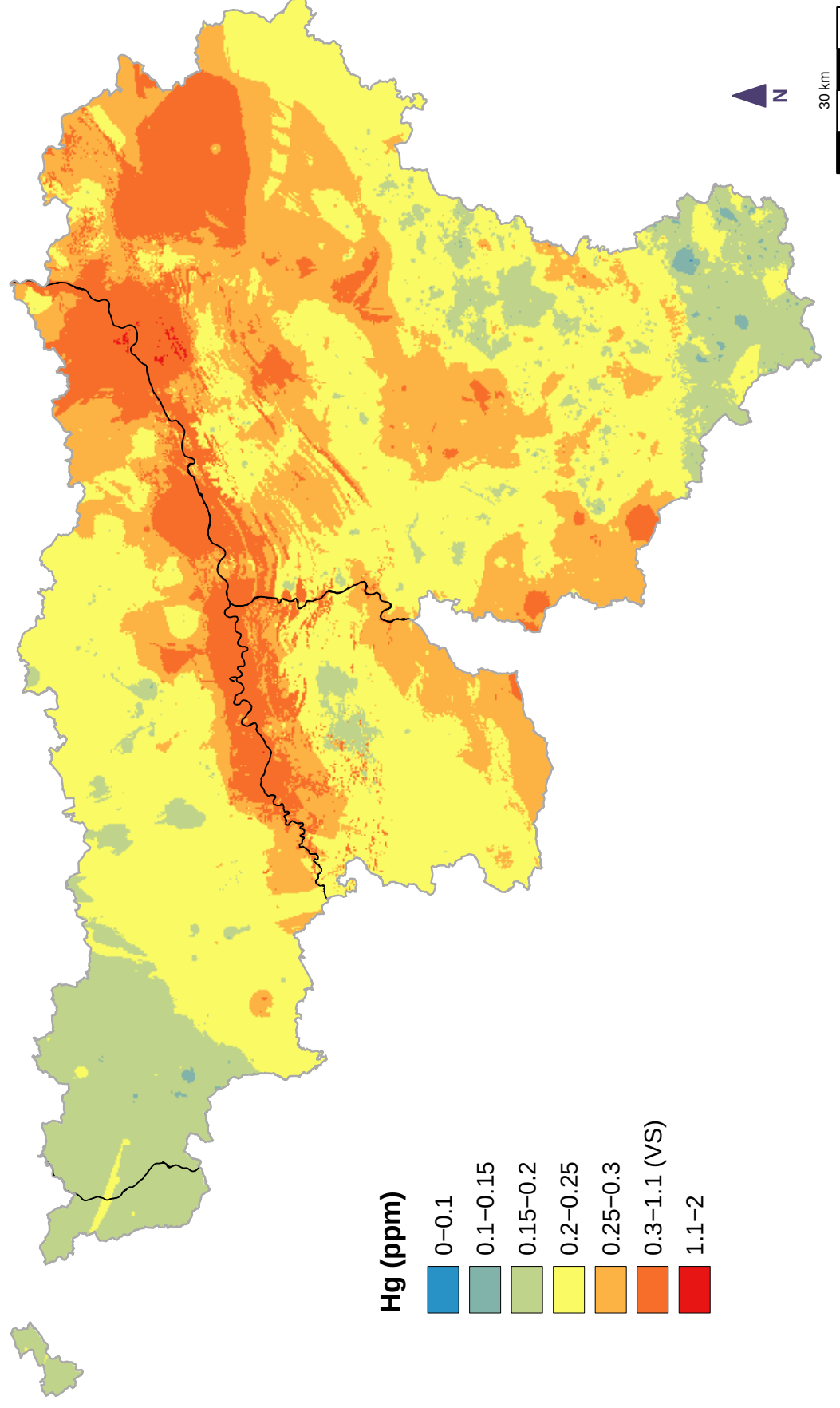
Teneur qui a moins d'1 chance sur 20 d'être dépassée dans un sol forestier



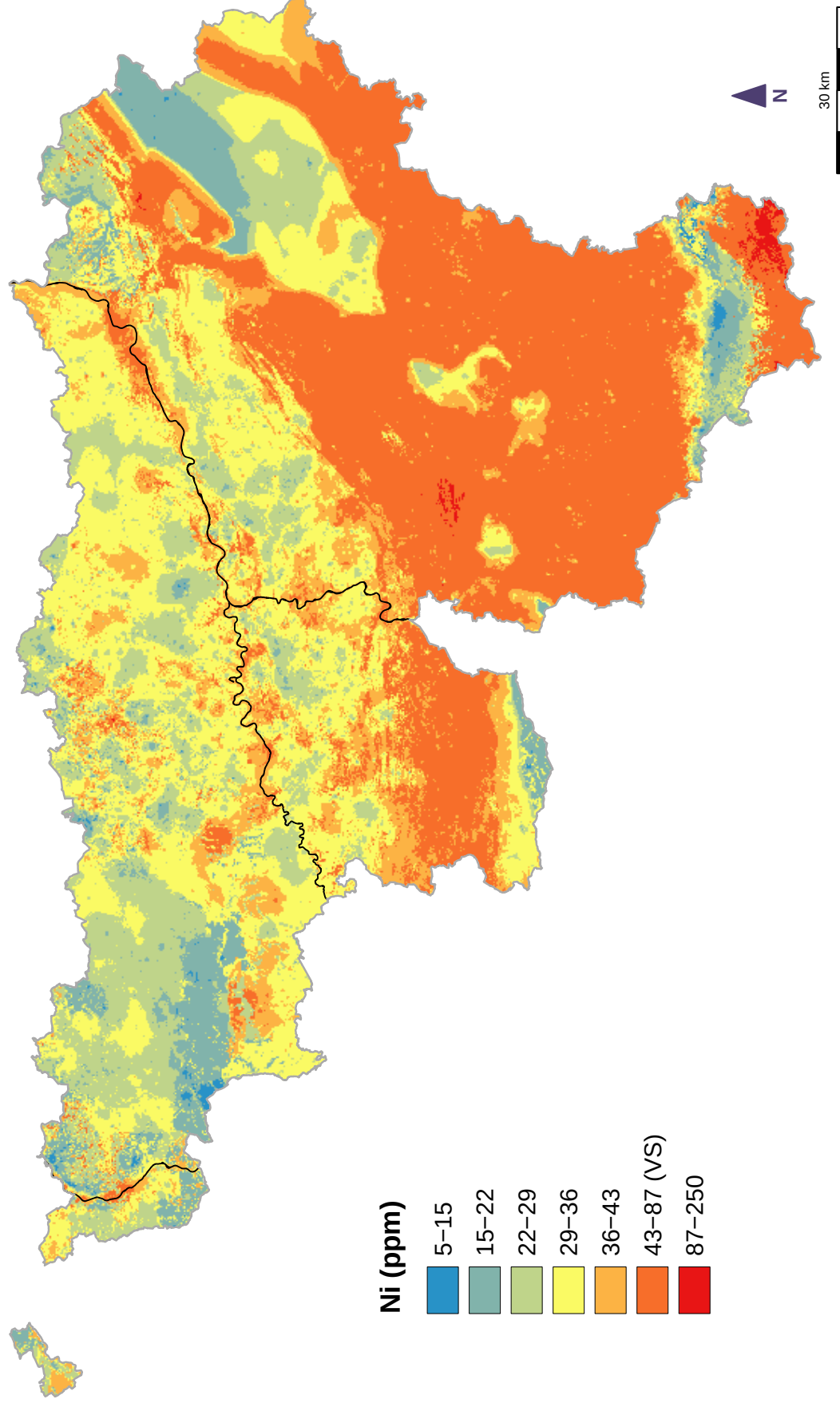
Teneur qui a moins d'1 chance sur 20 d'être dépassée dans un sol agricole



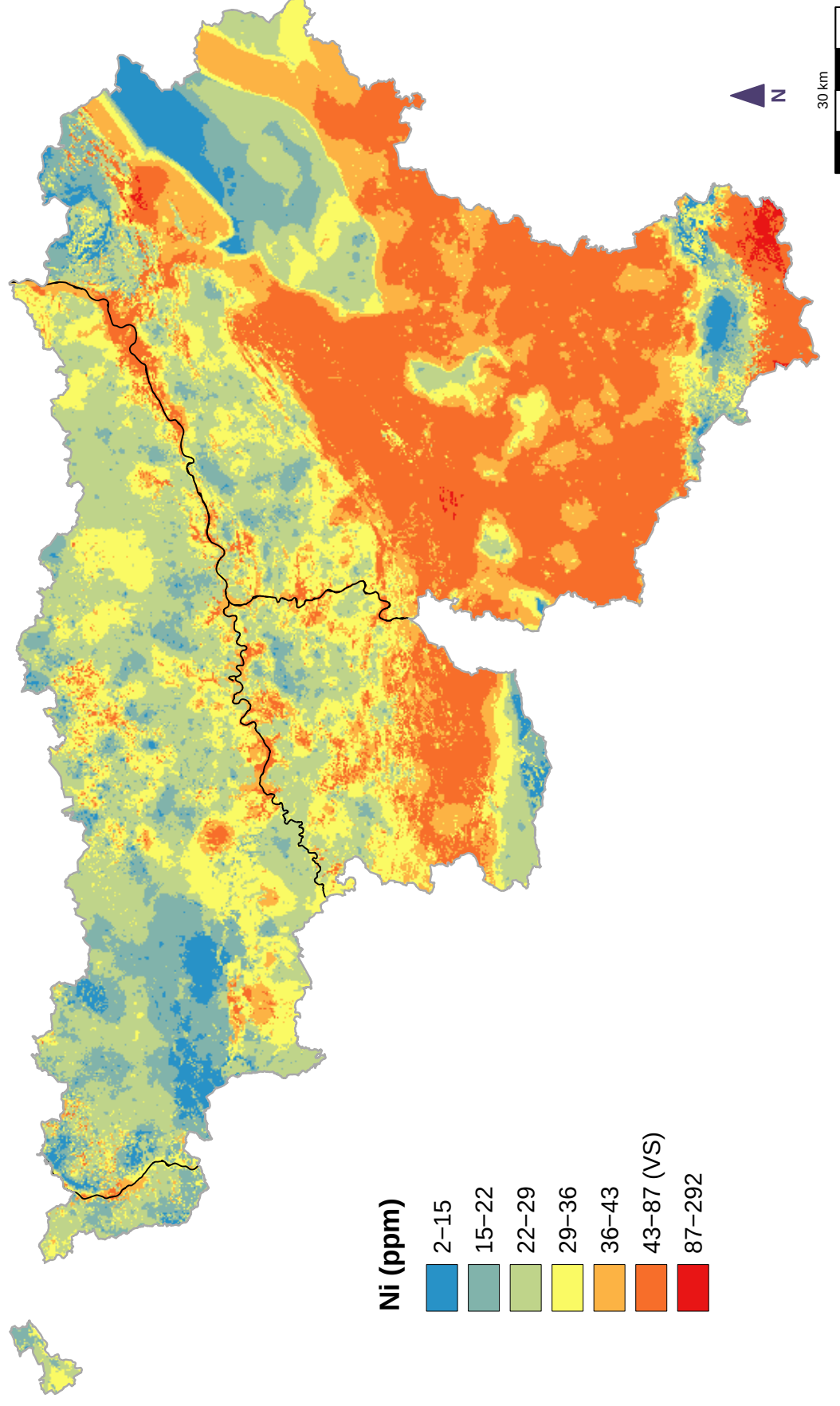
Teneur qui a moins d'1 chance sur 20 d'être dépassée dans un sol forestier



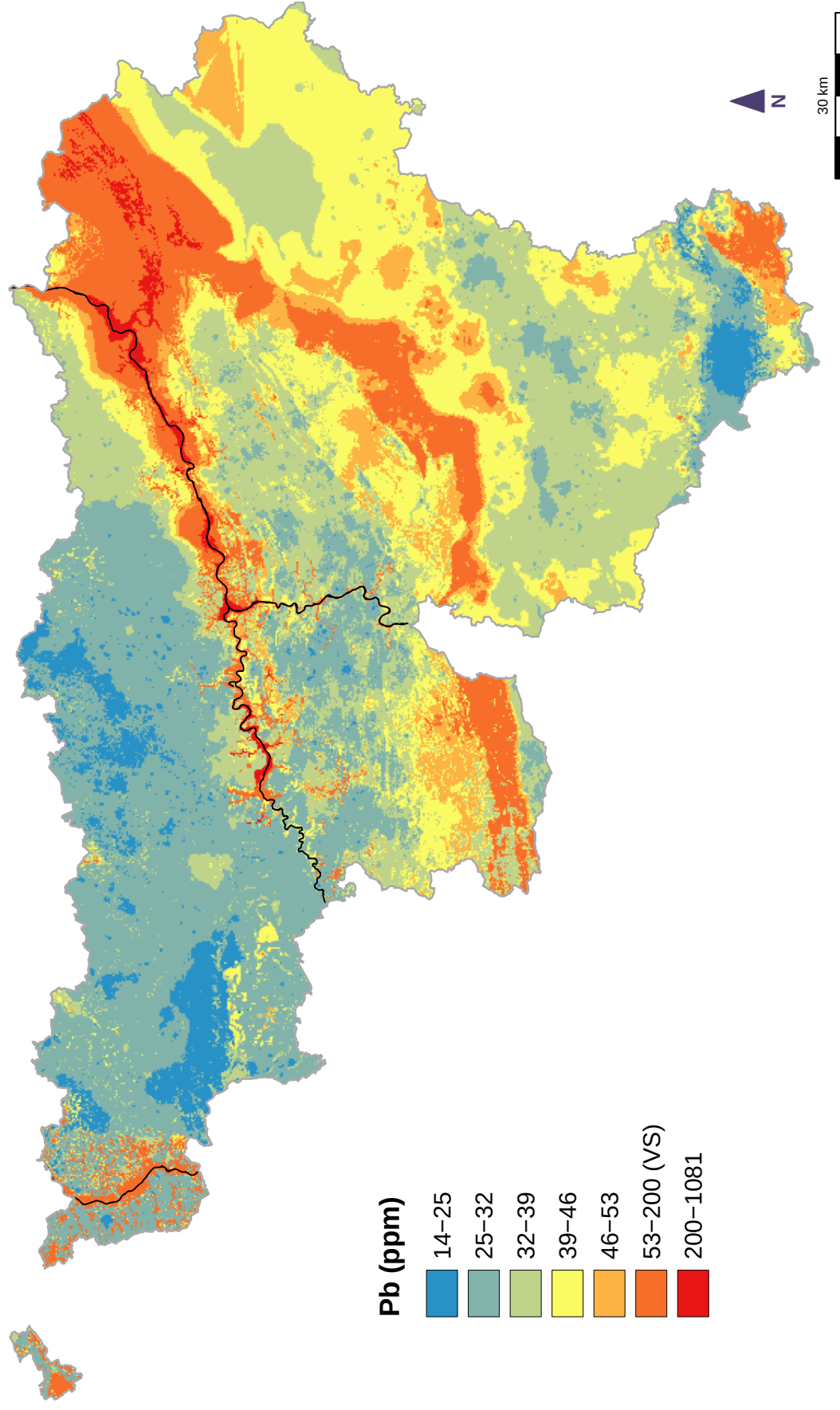
Teneur qui a moins d'1 chance sur 20 d'être dépassée dans un sol agricole



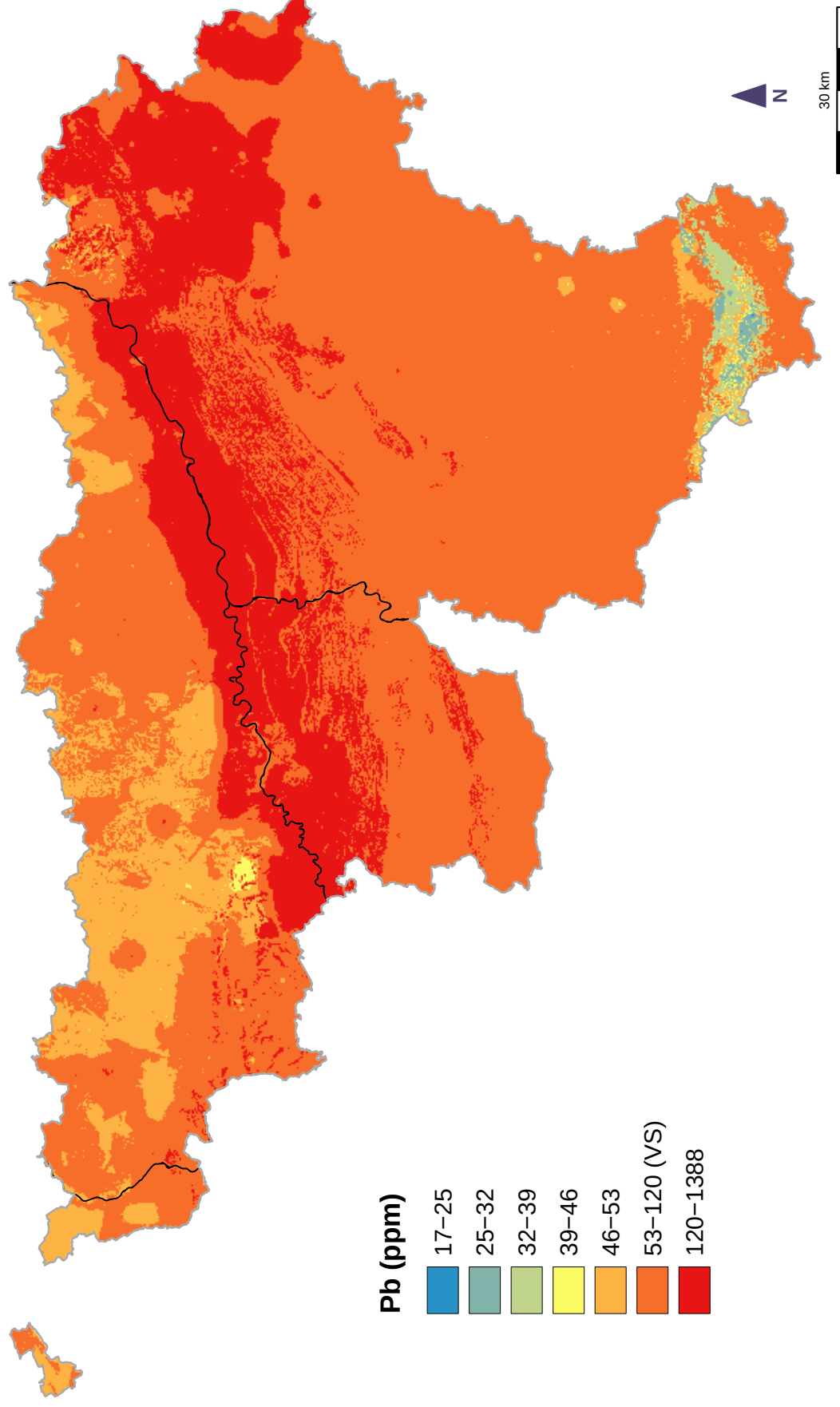
Teneur qui a moins d'1 chance sur 20 d'être dépassée dans un sol forestier



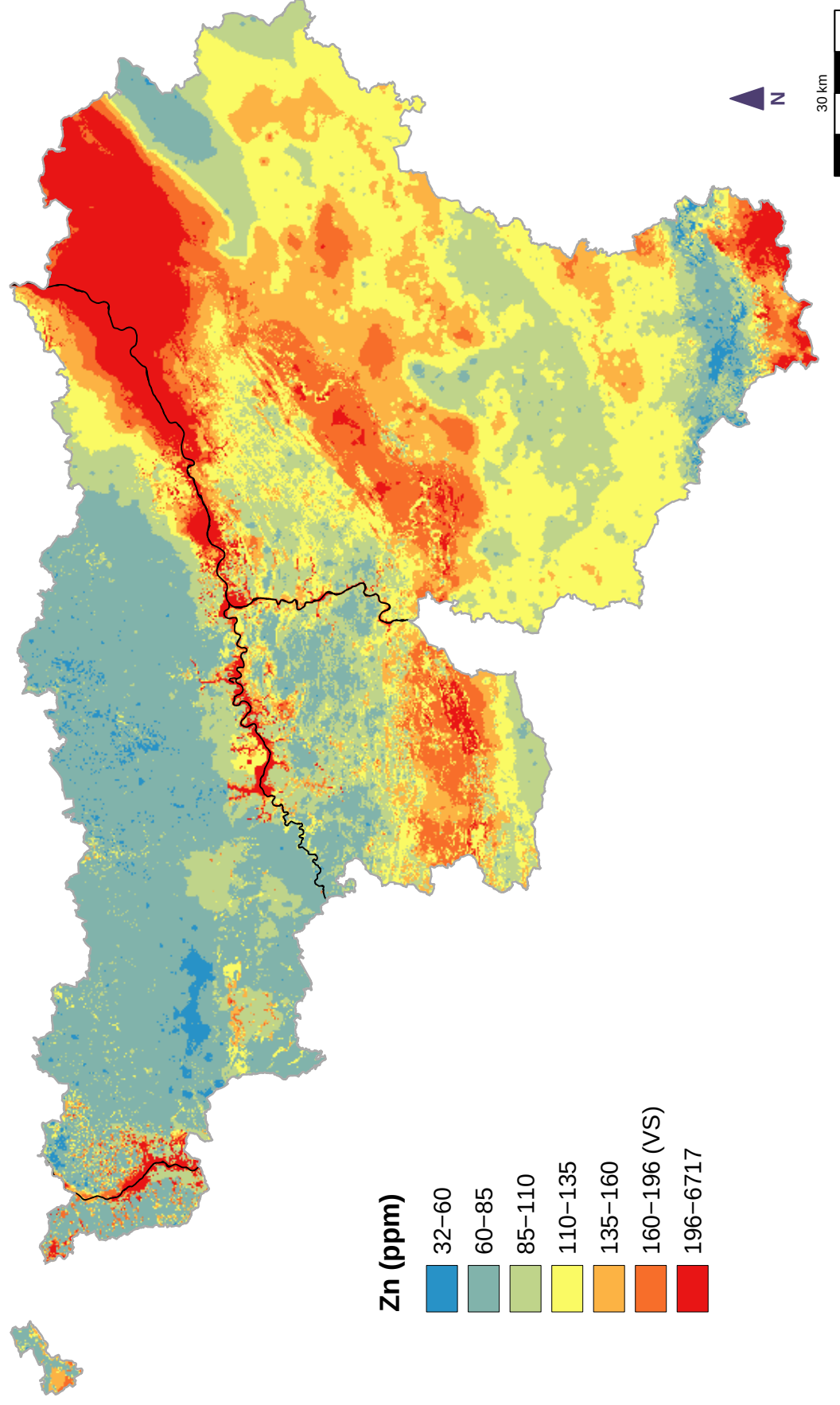
Teneur qui a moins d'1 chance sur 20 d'être dépassée dans un sol agricole



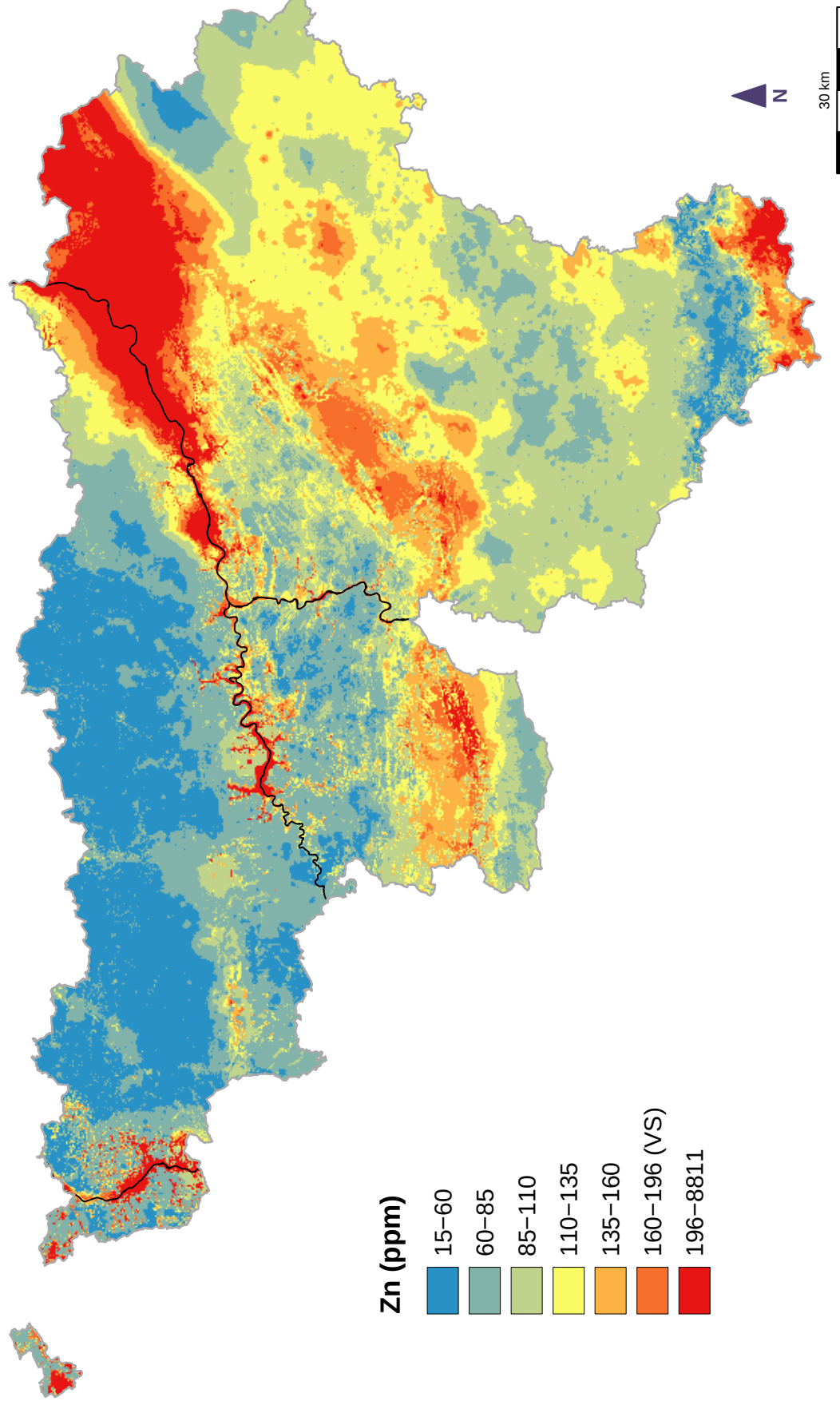
Teneur qui a moins d'1 chance sur 20 d'être dépassée dans un sol forestier



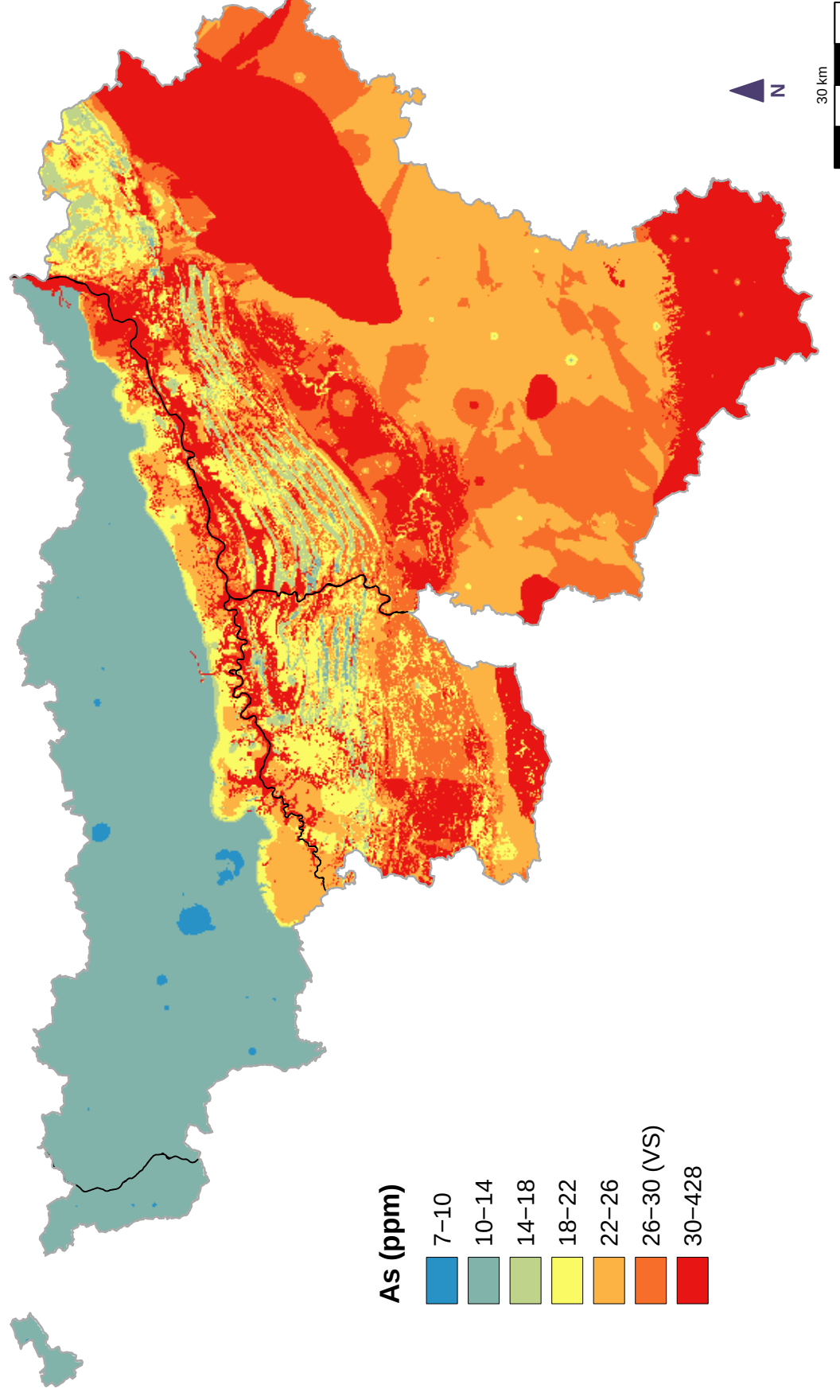
Teneur qui a moins d'1 chance sur 20 d'être dépassée dans un sol agricole



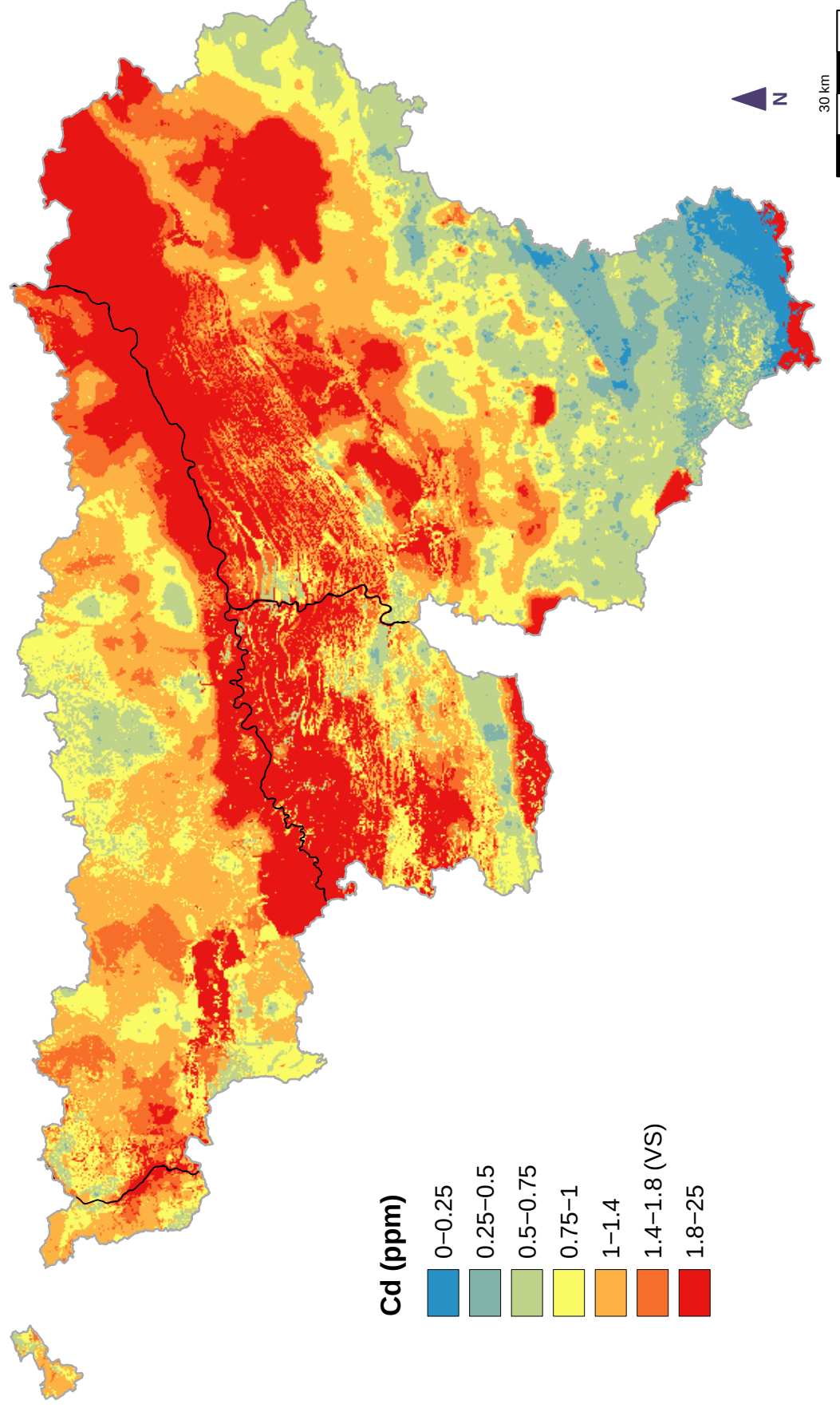
Teneur qui a moins d'1 chance sur 20 d'être dépassée dans un sol forestier



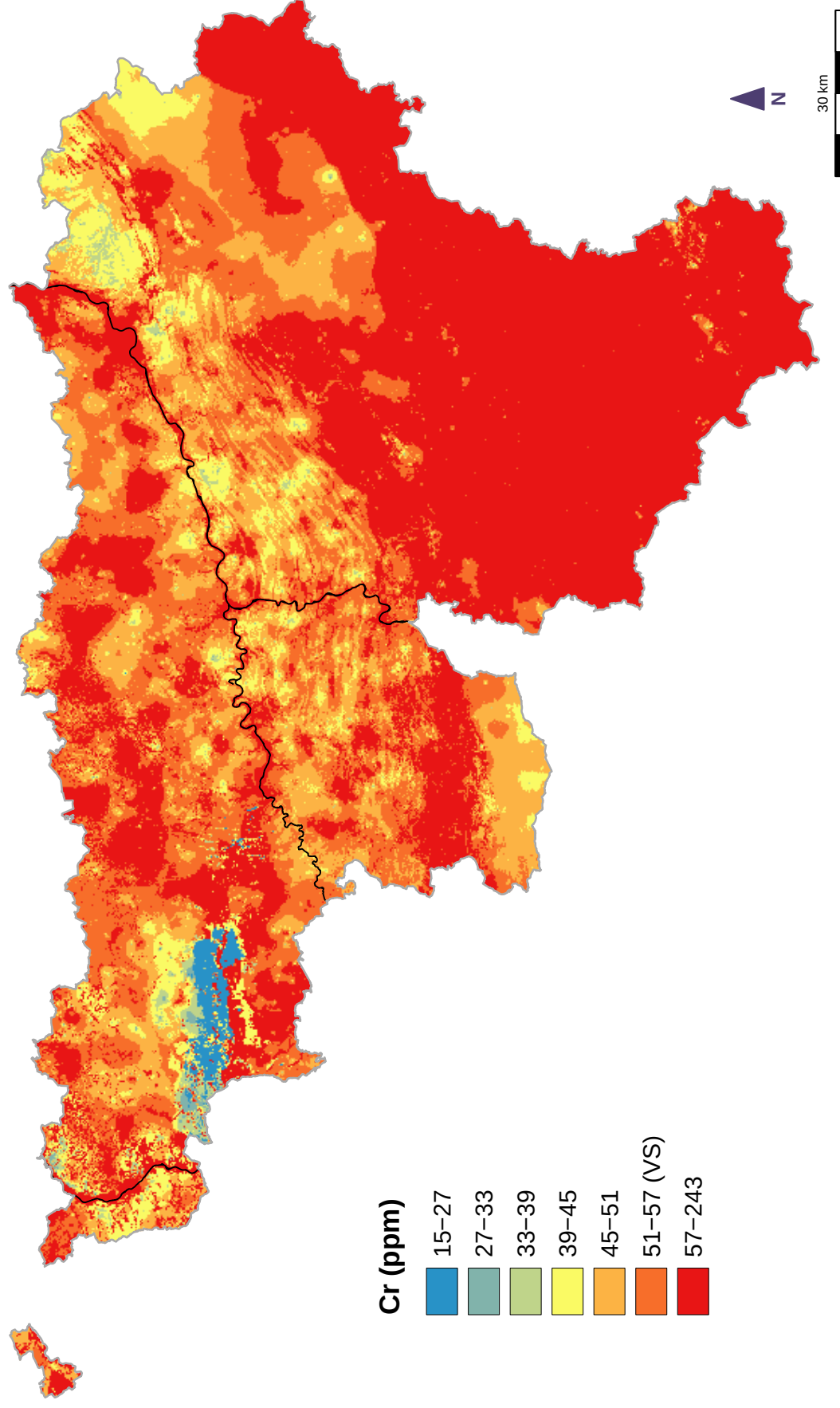
Teneur qui a moins d'1 chance sur 100 d'être dépassée dans un sol agricole ou forestier



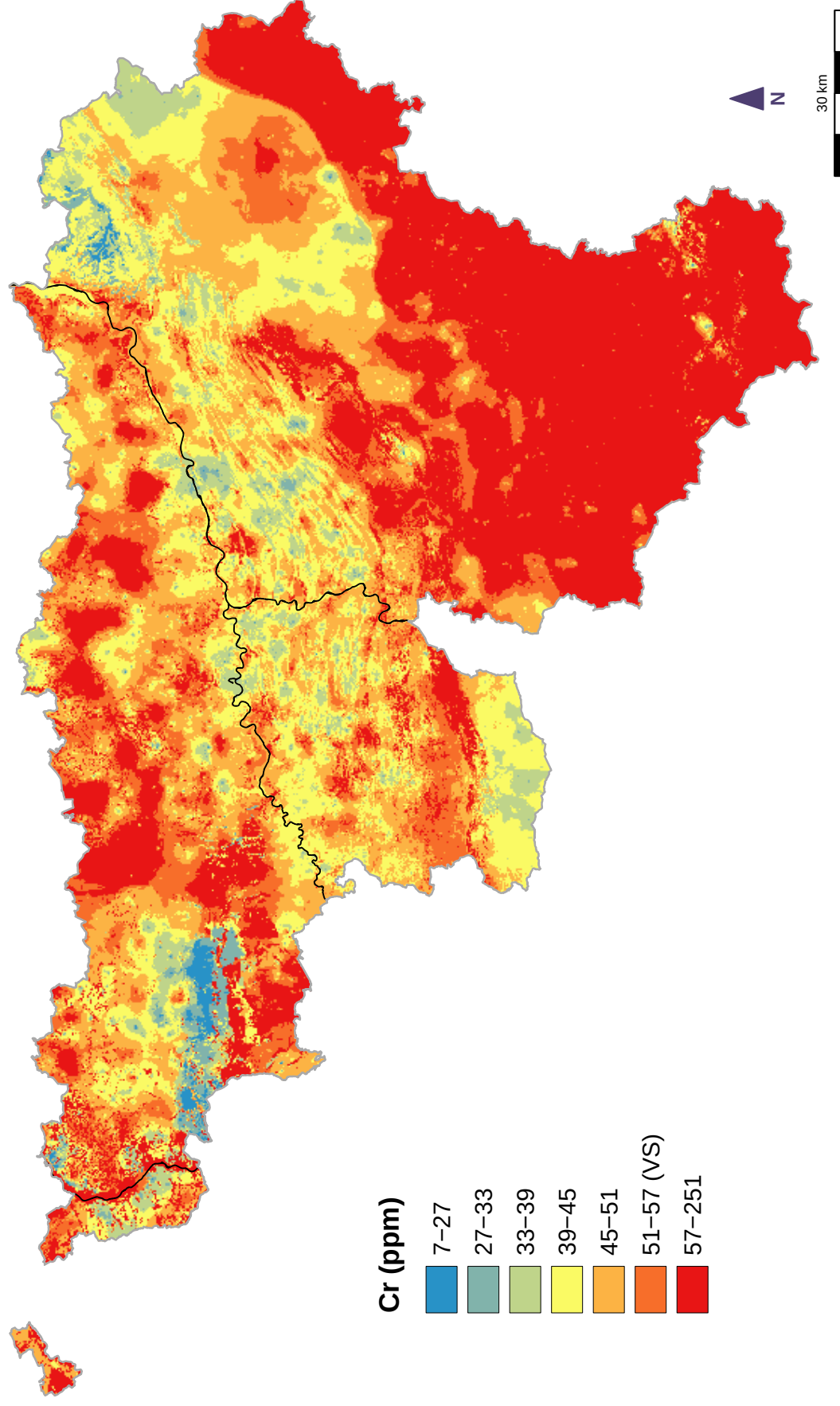
Teneur qui a moins d'1 chance sur 100 d'être dépassée dans un sol agricole ou forestier



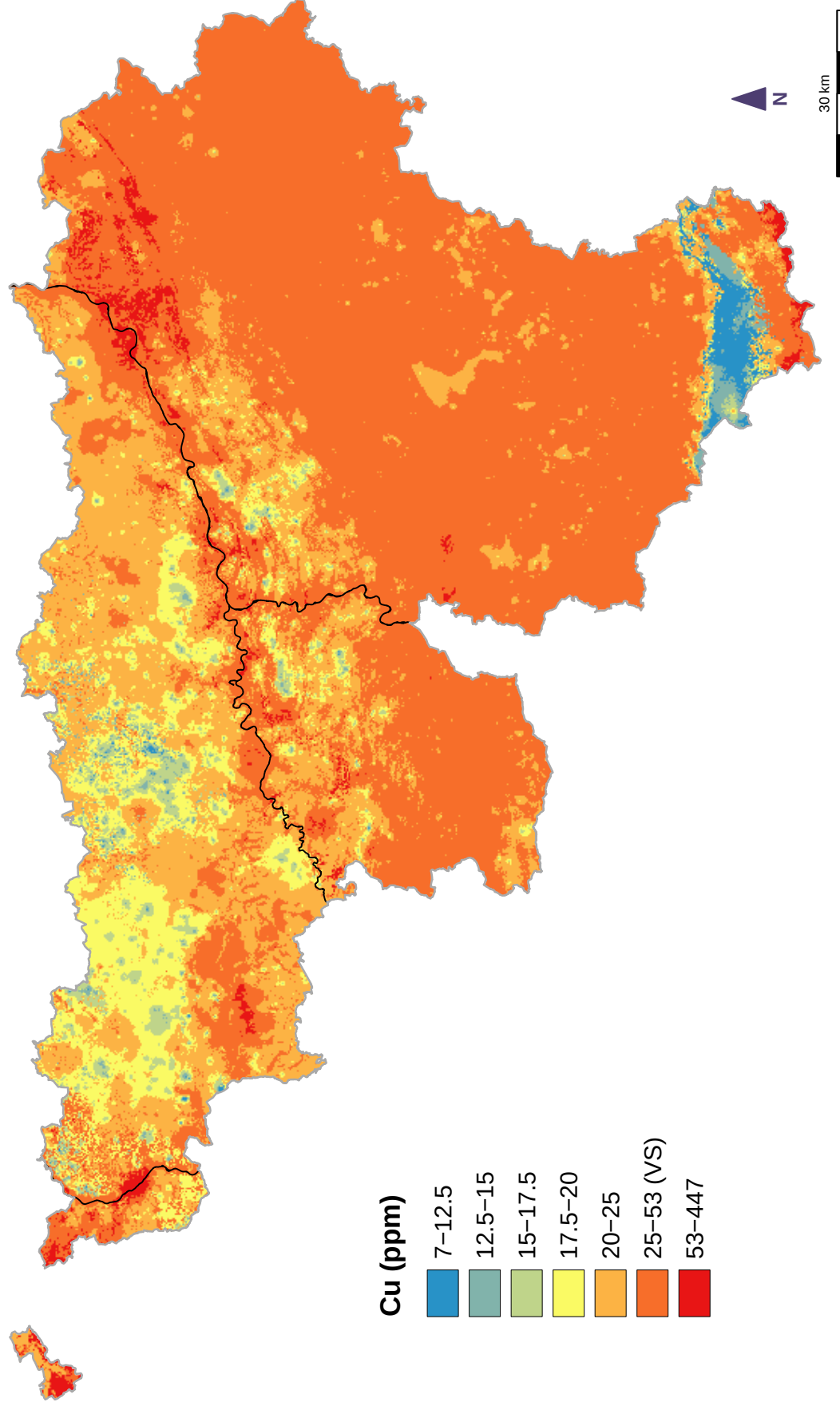
Teneur qui a moins d'1 chance sur 100 d'être dépassée dans un sol agricole



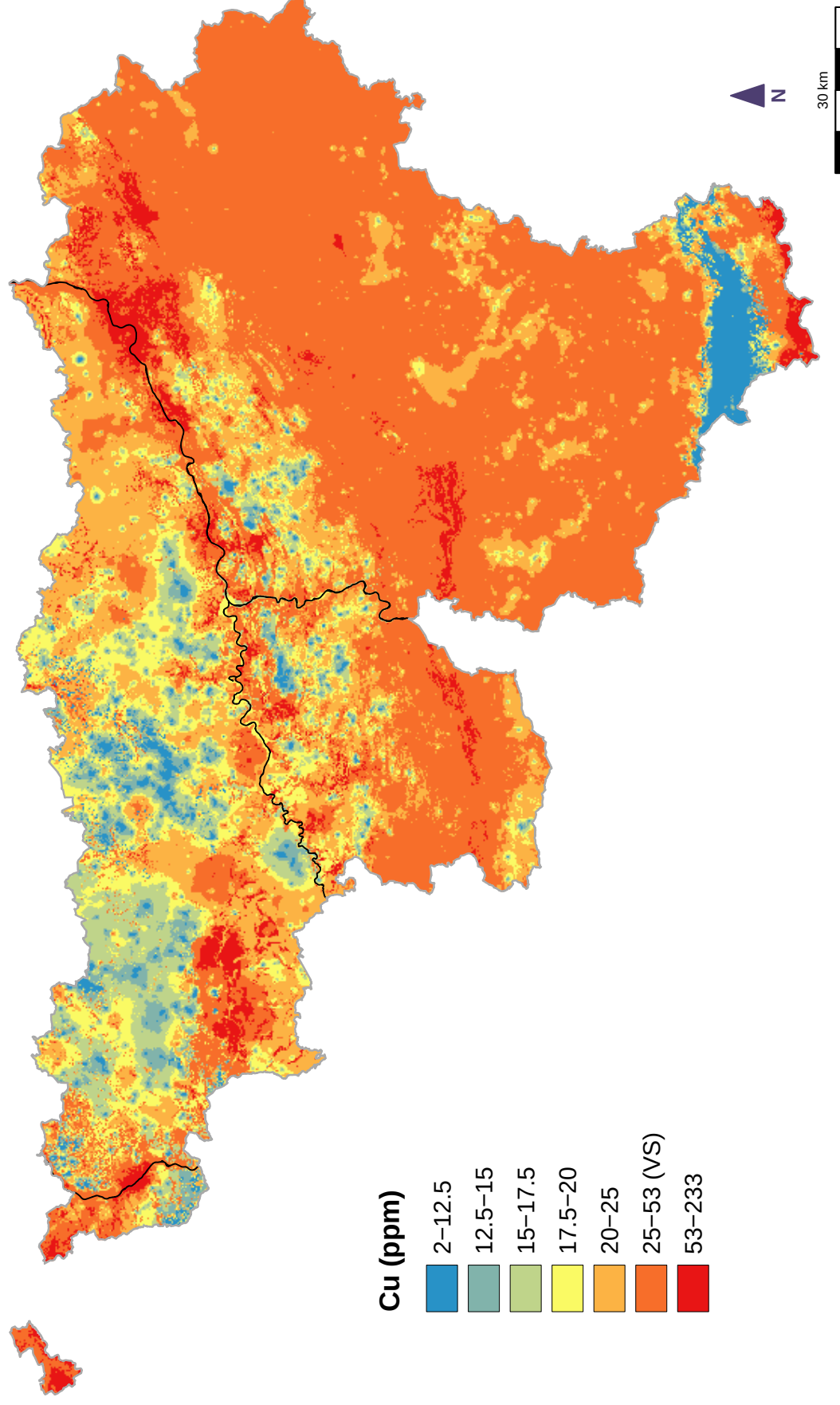
Teneur qui a moins d'1 chance sur 100 d'être dépassée dans un sol forestier



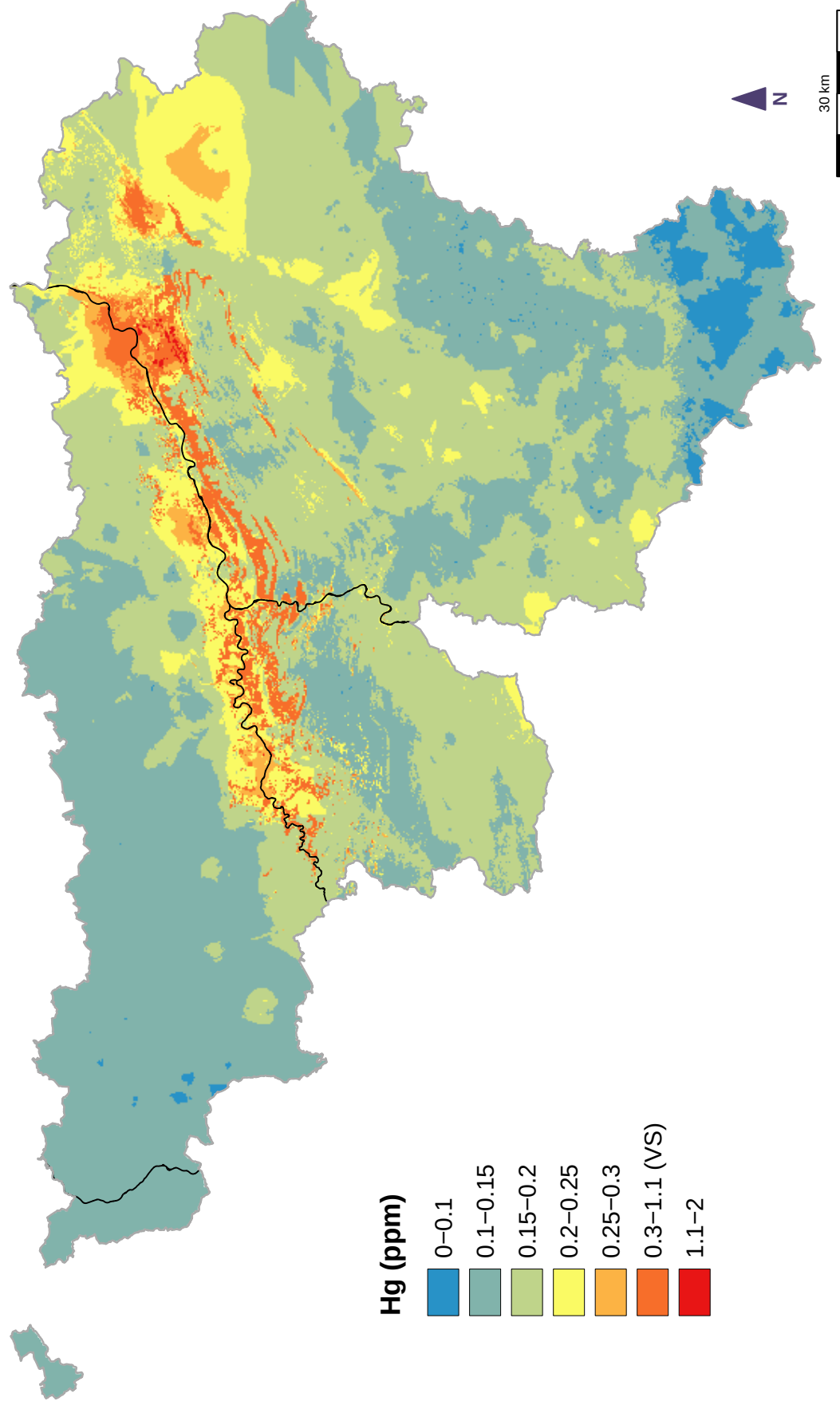
Teneur qui a moins d'1 chance sur 100 d'être dépassée dans un sol agricole



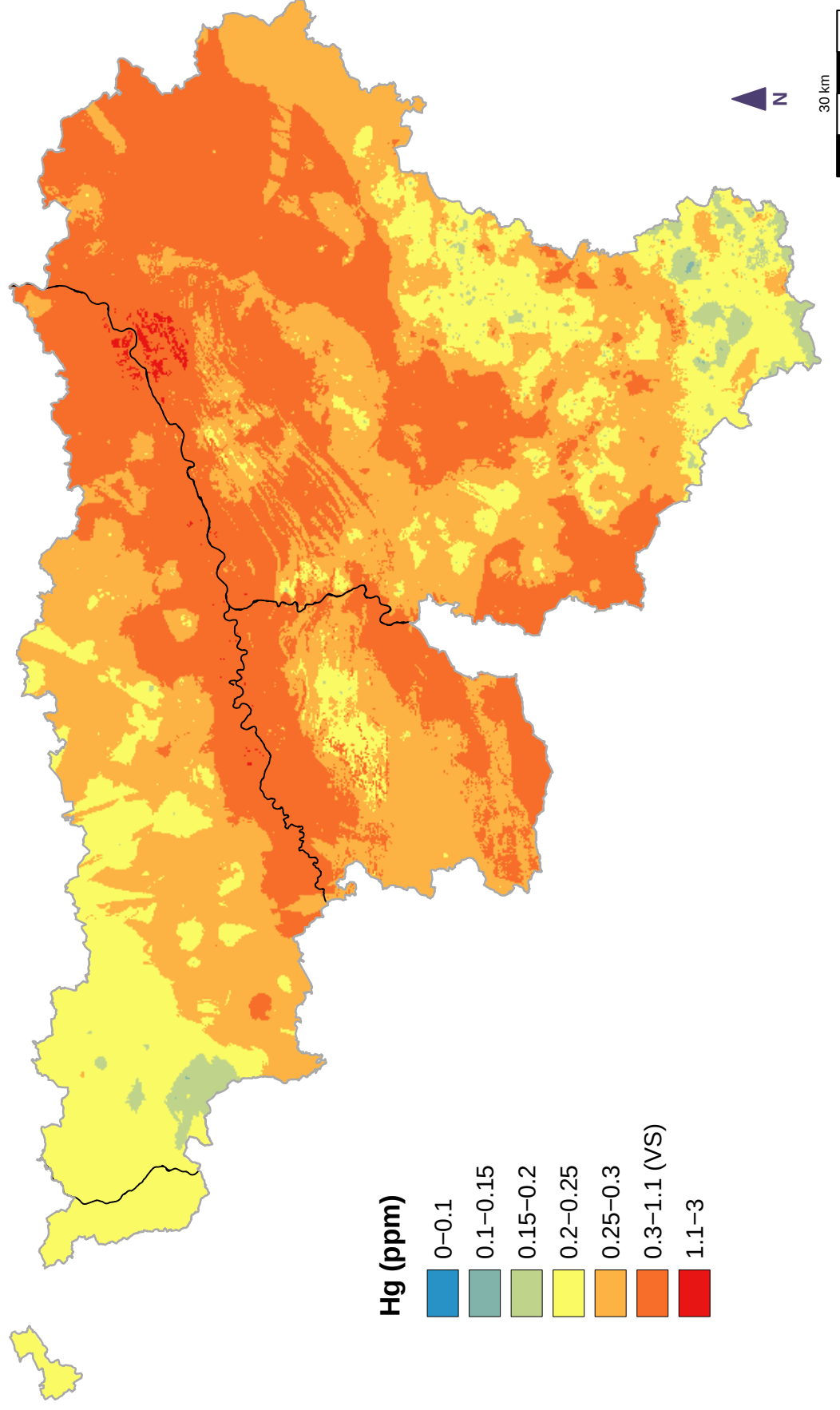
Teneur qui a moins d'1 chance sur 100 d'être dépassée dans un sol forestier



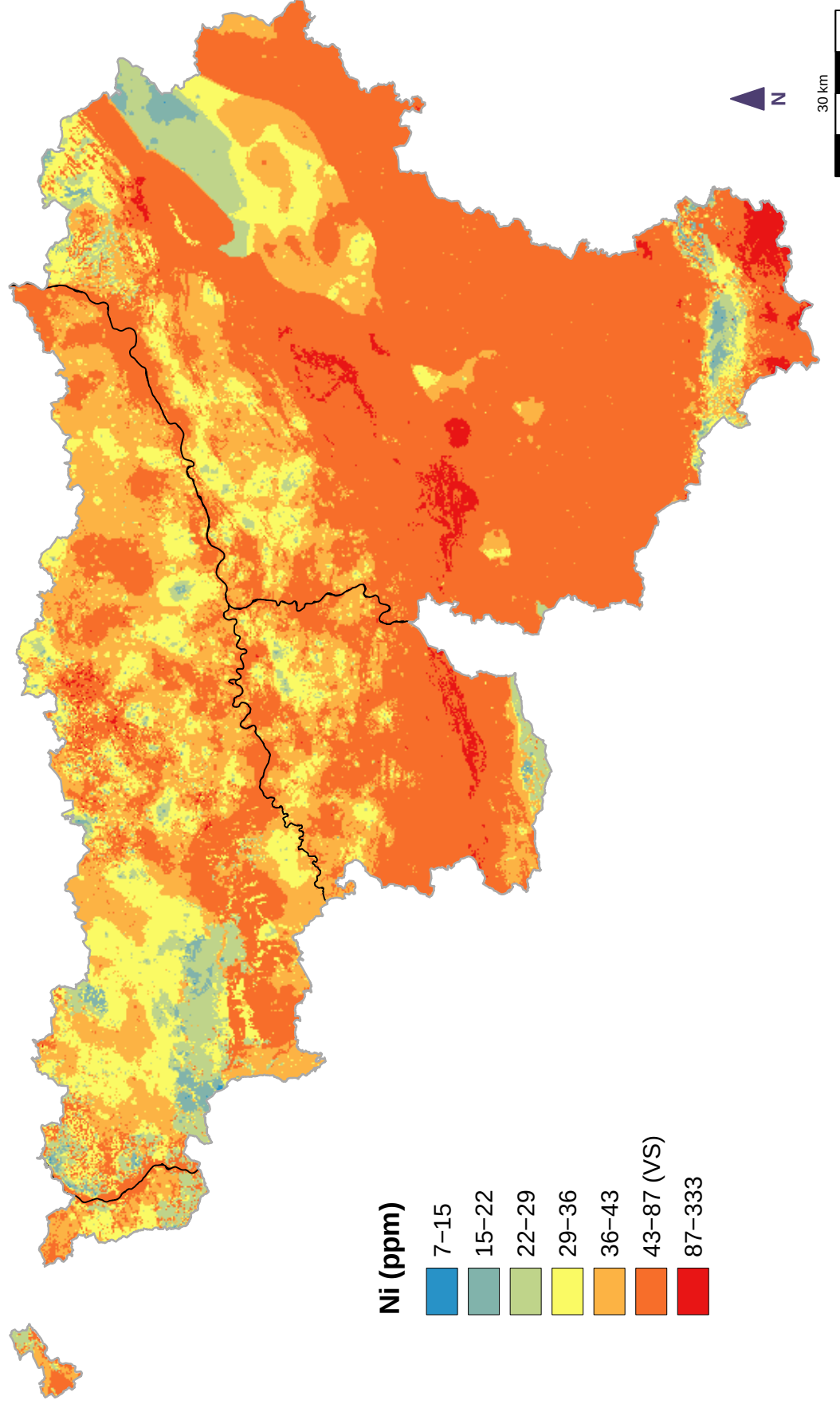
Teneur qui a moins d'1 chance sur 100 d'être dépassée dans un sol agricole



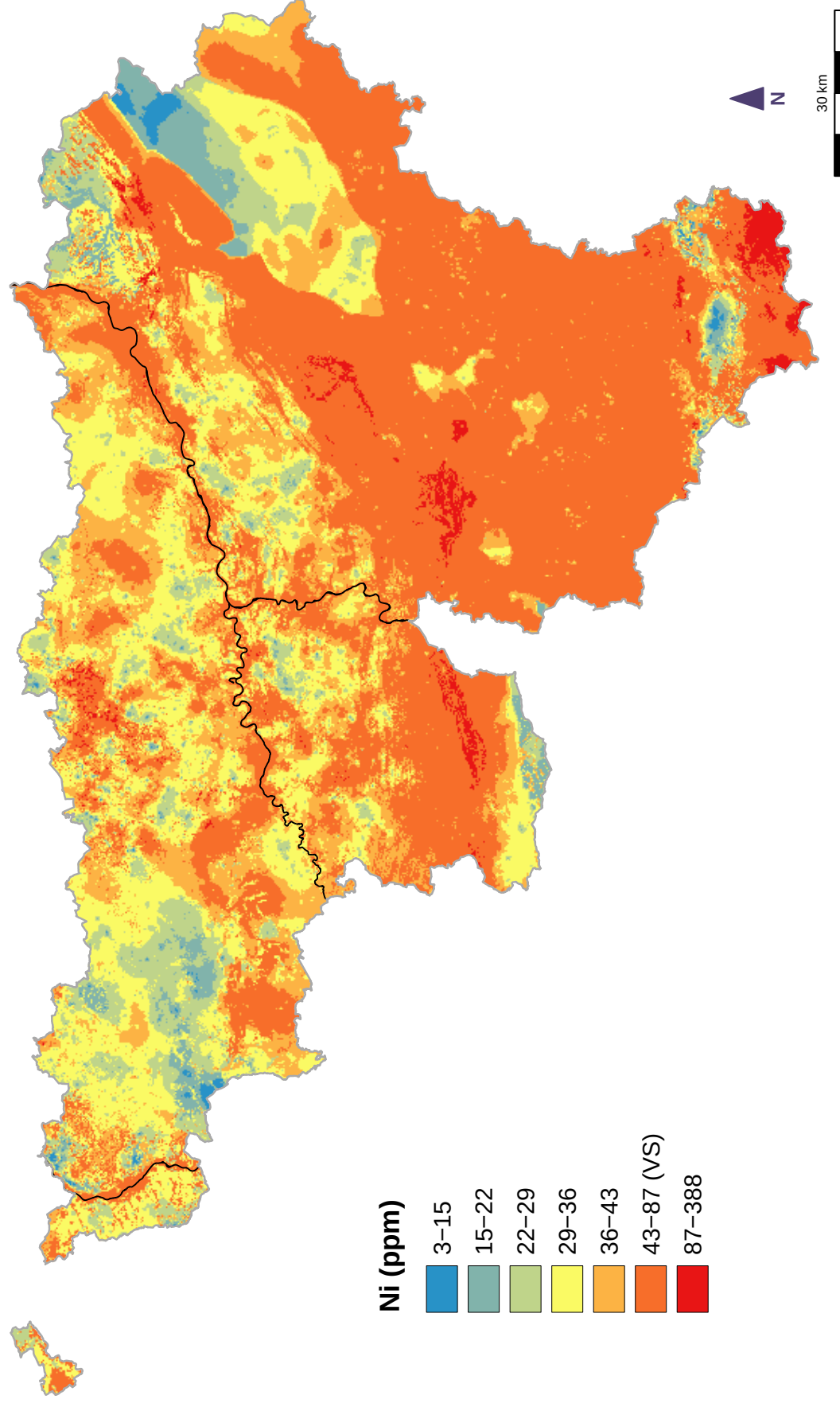
Teneur qui a moins d'1 chance sur 100 d'être dépassée dans un sol forestier



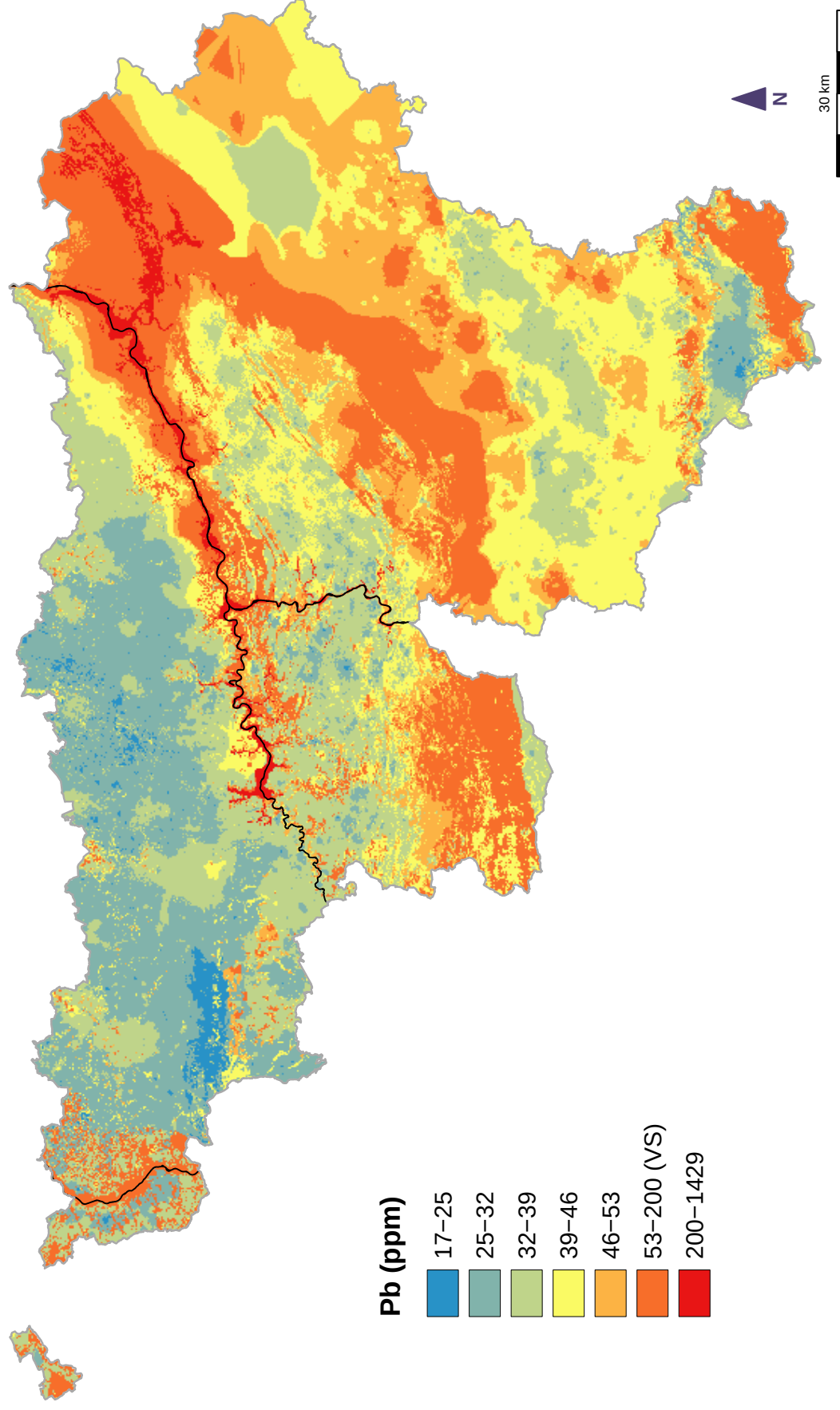
Teneur qui a moins d'1 chance sur 100 d'être dépassée dans un sol agricole



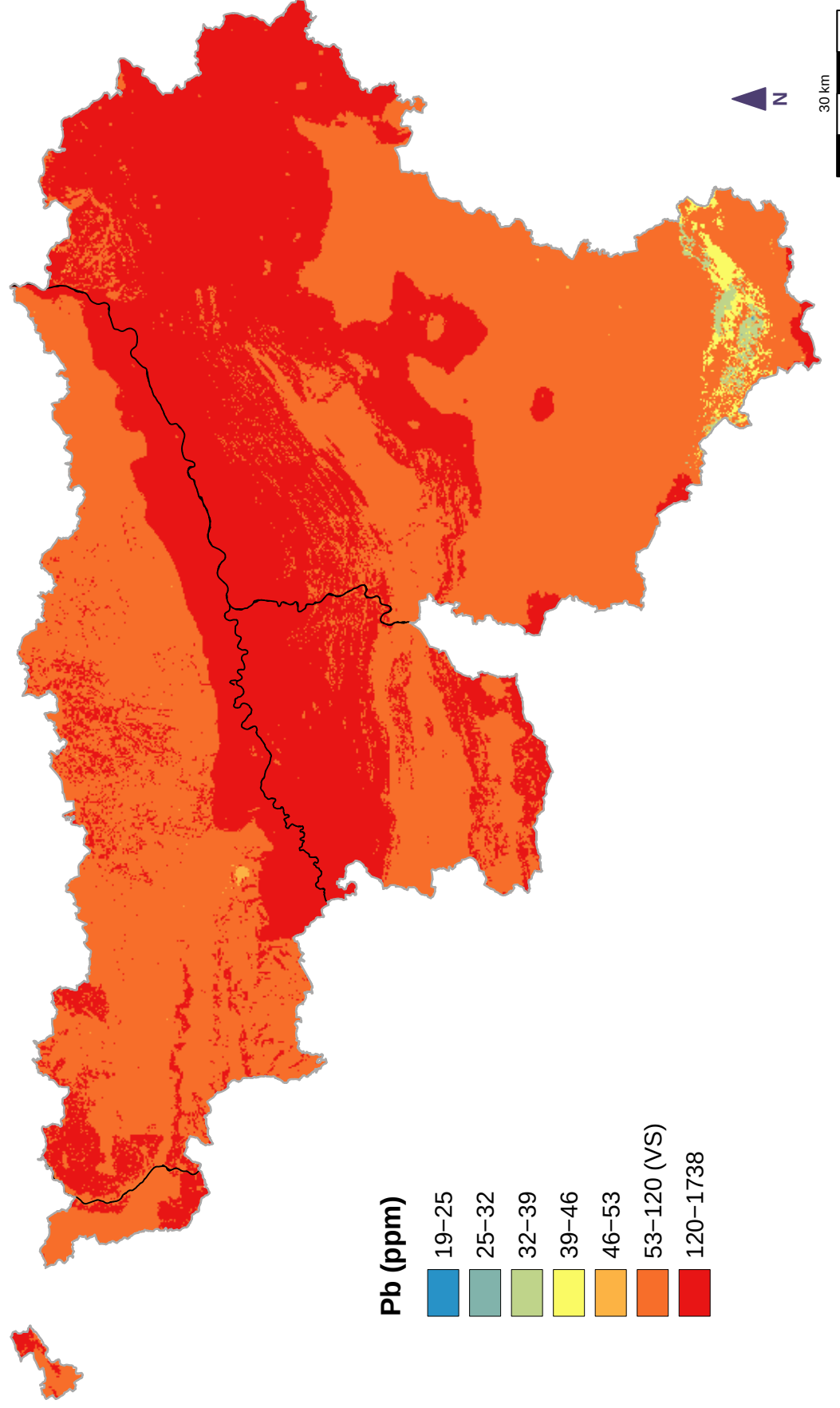
Teneur qui a moins d'1 chance sur 100 d'être dépassée dans un sol forestier



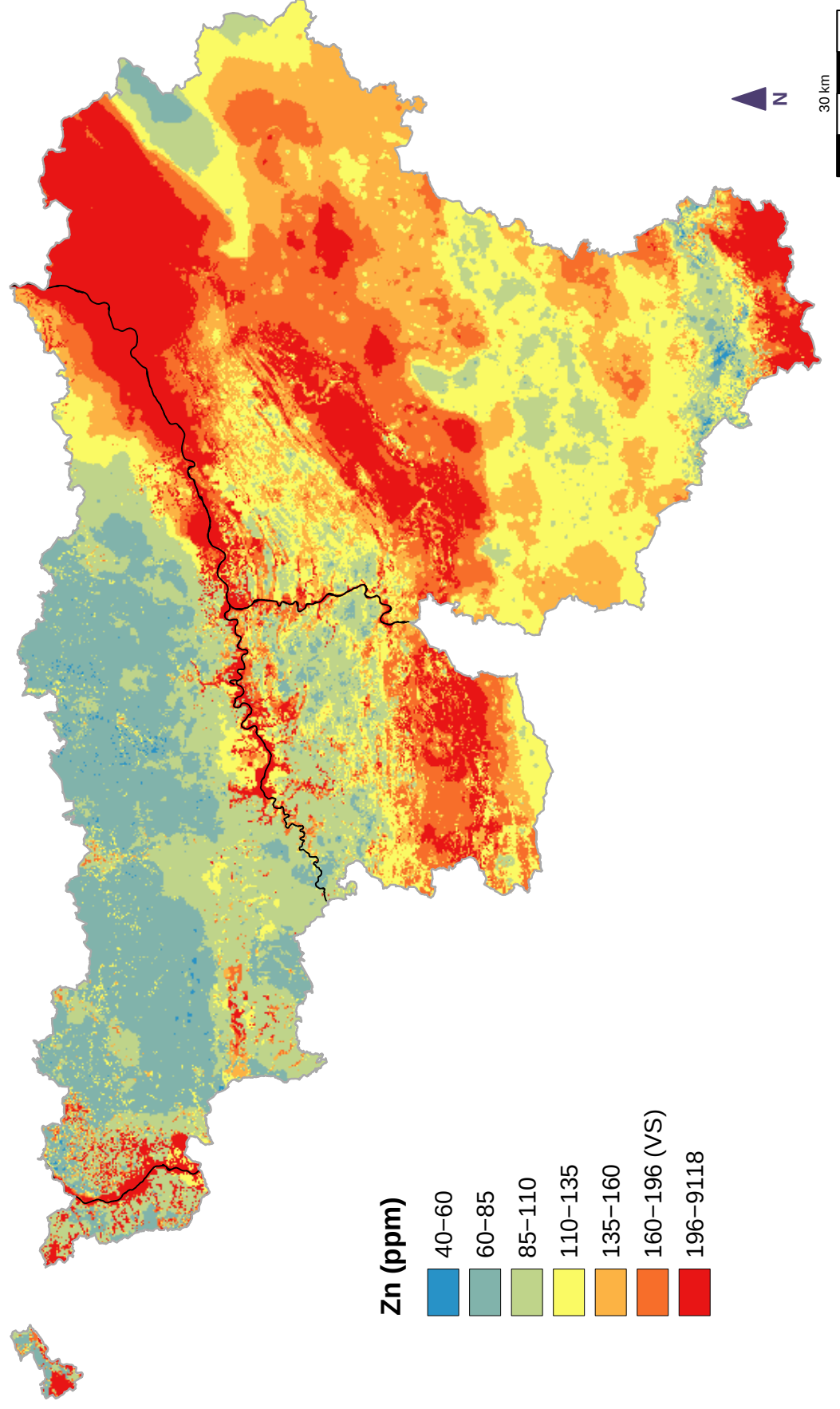
Teneur qui a moins d'1 chance sur 100 d'être dépassée dans un sol agricole



Teneur qui a moins d'1 chance sur 100 d'être dépassée dans un sol forestier



Teneur qui a moins d'1 chance sur 100 d'être dépassée dans un sol agricole



Teneur qui a moins d'1 chance sur 100 d'être dépassée dans un sol forestier

