



Université de Lille

Ecole doctorale Science de la Matière, du Rayonnement et de l'Environnement

Laboratoire de Génie Civil et Géo-Environnement LGCgE

Année universitaire 2022-2023

Thèse de DOCTORAT

Description multi-dimensionnelle de l'environnement à l'échelle des territoires :
contribution pour la recherche de déterminants environnementaux dans
l'étiologie des maladies chroniques.

Thèse préparée et soutenue publiquement
par Martin PAUMELLE
le 7 novembre 2023

En vue de l'obtention du grade de
Docteur de l'Université de Lille

Discipline : Science agronomiques et écologiques
Spécialité : Biologie de l'environnement, des organismes, des populations, écologie

Constitution du Jury :

Thèse dirigée par :	Annabelle Deram, PU Florent Occelli, MCU	Université de Lille Université de Lille
Président du jury :	Damien Cuny, PU	Université de Lille
Rapporteurs :	Anne-Peggy Hellequin, PU Guillaume Savoye, PU-PH	Université Paris Nanterre CHU Rouen
Examineurs :	Corinne Gower, MCU-PH Yohan Fayet, CR	CHU Reims Centre Léon Bérard

Remerciements

Je tiens d'abord à remercier la fondation I-Site, l'université de Lille et la métropole de Lille pour leur confiance et le financement de ces recherches. Je mesure la chance que j'ai eu de pouvoir travailler sur des sujets de recherche aussi passionnants pendant ces trois années. Merci aussi à l'Ed SMRE pour leur disponibilité et leur implication auprès des doctorants.

Bien sûr, un très grand merci au Pr. Annabelle Deram et Dr. Florent Ocelli pour leur soutien sans faille et leur accompagnement du début à la fin de ce périple. Merci pour vos conseils, vos encouragements et surtout votre bienveillance. En partageant mon expérience avec d'autres doctorant croisés ici ou là, je mesure la chance que j'ai eu d'avoir été encadré par vos soins. J'ai énormément appris et évolué pendant ces trois années. Vous m'avez donné confiance et permis de dépasser certains traits de mon caractère. Une pensée particulière pour Florent, qui m'a guidé et accompagné dans la préparation de mes premiers cours à l'ILIS.

Je souhaite adresser mes vifs remerciements aux Pr. Anne-Peggy Hellequin et Pr. Guillaume Savoye de me faire l'honneur de relire ce manuscrit et d'en être rapporteurs. Un grand merci également au Dr. Corinne Gower qui a suivi l'avancement de mes travaux au cours des comités de suivi, et qui a accepté de faire partie du jury de cette thèse. A ce titre, je tiens également à remercier chaleureusement le Dr. Yohan Fayet et le Pr. Damien Cuny.

Pendant ces trois années, j'ai aussi eu la chance d'être impliqué dans le projet ANR CROPS qui rassemble une grande diversité de personnes et de disciplines. Ce fut extrêmement enrichissant de pouvoir travailler ensemble sur un sujet commun aussi passionnant. Ce projet a notamment permis la conduite d'une campagne d'échantillonnage dans les clusters de la maladie Crohn. J'ai en tête de nombreuses anecdotes de toutes ces journées où on sillonnait le Nord. Pas toujours facile de prélever du lichen et des sols (ou de l'eau dans les cimetières). On a même eu affaire à la police mais on a montré patte blanche.

Je suis également très reconnaissant de l'accompagnement que m'ont offerts le Dr. Michaël Genin et le Dr. Camille Ternynck. Votre aide et votre expertise en statistiques spatiales a donné un poids supplémentaire à mes recherches.

J'ai une pensée aussi pour mes enseignants Dr. Juliette Faburé et Pr. Valérie Camel, qui m'ont encouragé dans mon souhait de continuer dans la recherche. Grosse pensée également pour le Dr. Joan Artigas, mon maître de stage de fin d'étude à Clermont-Ferrand. J'en garde d'excellents souvenirs et il m'a donné l'envie et la motivation de poursuivre dans cette voie.

Je suis heureux d'avoir pu réaliser ma thèse dans un labo comme le LGCgE, et plus particulièrement au LSVF en faculté de pharmacie. Beaucoup de bons souvenirs et de franches rigolades avec la fine équipe : Nicole, Alexandre, Sébastien, Damien, Caroline, Stéphane, Sylvain, Magali Ludivine, Régis, Pierre-Arthur. J'apprécie la chance que j'ai de pouvoir rester encore une année dans cet environnement. Elle est bien loin l'époque du Covid où chacun mangeait devant son ordinateur. Promis, je ramènerai plus souvent des gâteaux. J'aurai un peu plus de temps pour mitonner. Une grosse pensée aussi pour tous mes compagnons doctorant.es : Lara-Maria, Laakri, Yolène. Grand merci à Delphine de m'avoir mis le pied à l'étrier dans ce monde des données ouvertes et de la santé-environnement. Je pense également à tous les stagiaires avec qui j'ai eu la chance de travailler ou simplement d'échanger pendant ces trois années : Noémie, Lila, Camille, Sarah, Tracy, Amelle, Marie, Hugo.

Merci à toute ma famille qui m'a accompagné et encouragé depuis toujours. Je pense très fort à ma mère et à ma sœur Lucie. Merci pour tout. Une pensée aussi pour mon père, le silence n'est pas un oubli. Et puis bien sûr à tous mes amis sans qui je ne serais pas le même. Dalou, Malf, Bena, Malice, Zazie. Tous mes compagnons de musique avec qui j'ai pu me mettre au vert pour décompresser (Marco, Paul, Nums, Nitram). Et puis tous les autres....

Résumé

Parmi les maladies chroniques, la maladie de Crohn (MC) et l'insuffisance rénale chronique terminale (IRCT) présentent une étiologie multifactorielle encore en partie indéterminée pour laquelle un lien avec l'environnement est fortement suspecté. La répartition spatiale de leur incidence a été cartographiée à l'échelle communale dans le Nord de la France, à partir de deux registres de santé (Epimad et Nephronor). Ces disparités spatiales d'incidence constituent le point d'entrée pour investiguer les déterminants environnementaux susceptibles d'être impliqués dans la survenue de ces maladies.

La caractérisation de l'environnement et son lien à la santé est souvent appréhendée de manière cloisonnée. Les travaux se focalisent sur une source d'émission, un polluant, un milieu d'exposition. Bien que ces approches soient nécessaires, elles peuvent s'avérer limitées pour appréhender la complexité du lien entre environnement et santé, surtout pour des maladies multifactorielles dont les facteurs de risque environnementaux sont encore inconnus. Dans ce cas, il semble judicieux de privilégier des stratégies territoriales et multidimensionnelles, avant éventuellement de cibler des facteurs de risque environnementaux spécifiques. Dans ce contexte, comment mobiliser de multiples données environnementales ouvertes pour identifier les déterminants territoriaux de maladies multifactorielles ?

L'objectif principal de cette thèse est de proposer une description intégrée de l'environnement à l'échelle des territoires pour renseigner l'étiologie des maladies étudiées. La stratégie a consisté à collecter et réutiliser des données environnementales ouvertes. Cette démarche a permis d'identifier 24 sources de données et de générer 113 indicateurs spatialisés à l'échelle communale pour quatre départements. Ces indicateurs permettent de caractériser le niveau de contamination des milieux (air, eau, sols), le niveau des émissions polluantes, la localisation des sources d'émissions, l'occupation des sols, les pratiques agricoles, la naturalité des territoires et le climat. Plusieurs méthodologies ont été utilisées pour exploiter ces indicateurs et caractériser l'environnement sous un prisme multidimensionnel.

Une première approche a consisté à développer des indices spatiaux composites. Ces indices permettent de synthétiser, en une mesure globale, l'information contenue dans un grand nombre d'indicateurs. D'abord, un indice de vulnérabilité et un indice de résilience ont été calculés. Ils permettent de caractériser l'inégale répartition spatiale des déterminants territoriaux favorables et défavorables à la santé. Ensuite, des indices composites de multi-contamination des milieux (air, eau, sols) ont été construits.

Une deuxième approche a été développée en utilisant des méthodes de classification multivariées pour créer des typologies territoriales et décrire les profils environnementaux des communes. Ces résultats apportent une vision plus complexe des territoires, et ont permis d'appréhender comment les pressions environnementales se répartissent dans l'espace et se superposent les unes avec les autres.

Enfin, les résultats de ces approches multi-dimensionnelles ont été associés aux variations spatiales d'incidence des maladies chroniques, suggérant des liens potentiels entre l'environnement et la survenue de ces pathologies. Pour l'IRCT, des associations ont été observées avec la pression urbaine et la pollution atmosphérique en particules fines, corroborant la littérature existante. Pour la MC, des liens ont été suggérés avec les pratiques agricoles, la naturalité des territoires et la pollution métallique des sols. D'autres approches épidémiologiques doivent maintenant être envisagées pour éprouver ces hypothèses et poursuivre les recherches.

Mots-clés : Données environnementales ouvertes ; Indicateurs spatialisés ; Indices composites ; Profils territoriaux ; Maladie de Crohn ; Insuffisance rénale chronique terminale.

Abstract

Among chronic diseases, Crohn's disease (CD) and end-stage renal disease (ESRD) have a multifactorial etiology that remains partly unknown, with a strong suspicion of an environmental link. The spatial distribution of their incidence has been mapped at the municipal level in Northern France, using two health registers (Epimad and Nephronor). These spatial disparities in incidence serve as the starting point to investigate potential environmental determinants that may be involved in the onset of these diseases.

The characterization of the environment and its link to health is often approached in a fragmented manner, focusing on a specific emission source, pollutant, or exposure medium. While these approaches are necessary, they may be limited in comprehending the complexity of the relationship between the environment and health, especially for multifactorial diseases with unknown environmental risk factors. In such cases, it is relevant to prioritize territorial and multidimensional strategies before potentially targeting specific environmental risk factors. In this context, how can multiple open environmental data sources be leveraged to identify territorial determinants of multifactorial diseases?

The main objective of this thesis is to offer an integrated description of the environment at the territorial level to inform the etiology of the studied diseases. The strategy involved collecting and reusing open environmental data. This approach identified 24 data sources and generated 113 spatial indicators at the municipal level for four departments. These indicators allow for the characterization of contamination levels in various media (air, water, soil), pollutant emissions, the location of emission sources, land use, agricultural practices, the natural features of territories, and climate. Several methodologies were used to exploit these indicators and characterize the environment from a multidimensional perspective.

A first approach involved developing composite spatial indices. These indices synthesize information from many indicators into a single global measure. Initially, vulnerability and resilience indices were calculated. They characterize the uneven spatial distribution of environmental determinants that have a beneficial or detrimental impact on health. Subsequently, composite indices of multi-media contamination (air, water, soil) were constructed.

A second approach was developed using multivariate classification methods to create territorial typologies and describe the environmental profiles of municipalities. These results provide a more complex view of territories and have allowed to understand how environmental pressures are distributed in space and overlap with each other.

Finally, the results of these multidimensional approaches were linked to spatial variations in the incidence of chronic diseases, suggesting potential connections between the environment and the occurrence of these pathologies. For ES-CKD, associations were observed with urban pressure and fine particulate air pollution, corroborating existing literature. For CD, links were suggested with agricultural practices, the natural characteristics of territories, and metallic soil pollution. Further epidemiological approaches are now needed to test these hypotheses and advance research in this area.

Key words : Open environmental data ; Spatial indicators ; Composite indices ; Territorial profiles ; Crohn's disease ; End-stage renal disease.

Table des matières

Remerciements	3
Résumé	5
Abstract	7
Table des matières	8
Table des figures	11
Liste des tableaux	14
Liste des abréviations	15
Introduction générale	16
CHAPITRE I – Potentiels de vulnérabilité et de résilience des territoires	23
Introduction	23
Matériel & Méthodes	25
1. Développement d'un jeu de données multi-dimensionnel : économie, environnement physique, politique, santé, services, social.	25
2. Construction des indices de vulnérabilité et de résilience.	28
3. Mise en relation avec les variations spatiales d'incidence de la MC.	33
Résultats	34
1. Une vision intégrée des territoires : économie, environnement physique, politique, santé, services, social.	34
2. Potentiels de vulnérabilité et de résilience des communes.	36
3. Association avec les variations spatiales d'incidence de la MC.	41
Discussion	46
1. Extension du jeu de données à la Seine Maritime.	46
2. Mise en parallèle des résultats par rapport au projet Cisterres.	47
3. Intérêt de cette approche pour renseigner l'étiologie de la MC.	48
4. Représentativité des indices obtenus.	49
Conclusion	50
CHAPITRE II – Mobiliser des données ouvertes pour décrire l'environnement physique au sens du milieu de vie extérieur	51
Introduction	51
Matériel & Méthodes	52

Résultats.....	53
1. Niveaux de contamination des milieux.....	55
2. Niveaux des émissions polluantes	56
3. Localisation des sources d'émissions.....	57
4. Occupation des sols	59
5. Pratiques agricoles.....	60
6. Naturalité.....	62
7. Climat.....	63
8. Cas particuliers de certaines bases de données.....	64
Discussion	71
1. Vue d'ensemble sur le jeu de données.	71
2. Temporalité des données environnementales	77
Conclusion.....	80
CHAPITRE III – Evaluer la multi-contamination des milieux environnementaux.	81
Introduction	81
PARTIE 1 : Comment caractériser la multi-contamination à l'échelle des territoires ?	83
1. Matériel & Méthodes	83
2. Résultats	87
3. Discussion	100
4. Conclusion	106
PARTIE 2 : Multi-contamination et variations spatiales d'incidence des pathologies étudiées	107
1. Matériel & Méthodes	107
2. Résultats	107
3. Discussion	111
4. Conclusion	113
CHAPITRE IV – Décrire le profil environnemental des territoires : une stratégie multi-dimensionnelle pour caractériser l'environnement physique.....	114
Introduction	114
Matériels & Méthodes.....	116
1. Sélection des données d'entrée.....	117
2. Préparation des données.....	117
3. Méthode de classification.....	117
4. Interprétation des profils des communes.....	118
Résultats.....	118

1. Sélection des données d'entrée.....	118
2. Jeu de données final et réduction dimensionnelle (ACP).....	120
3. Classification obtenue.....	122
4. Description et interprétation des profils de communes	124
Discussion	129
1. Une description multi-dimensionnelle des territoires	130
2. Perspectives de recherche en santé-environnement	130
Conclusion.....	131
CHAPITRE V – L'apport des profils environnementaux des territoires pour expliquer les variabilités d'incidence des maladies chroniques.	132
Association entre les profils environnementaux et l'incidence de la MC	133
1. Matériel et méthodes	133
2. Résultats	133
3. Discussion	137
Association entre les profils environnementaux et l'incidence de l'IRCT.....	140
1. Matériel et méthodes	140
2. Résultats	140
3. Discussion	143
Conclusion.....	144
Discussion générale	145
Références bibliographiques	155
ANNEXE 1 : Tableau de présentation des indicateurs retenus pour la construction des indices de vulnérabilité et de résilience.....	175
ANNEXE 2 : Interprétation donnée aux composantes principales retenues pour le calcul des indices	182
ANNEXE 3 : Nuages de points et liens de corrélation entre les différents indices composites de multi-contamination.....	185
ANNEXE 4 : Multi-contamination et variations spatiales d'incidence des pathologies étudiées	188
ANNEXE 5 : Comparaison entre la classification obtenue avant et après consolidation des k-means.	192
ANNEXE 6 : Résultats des régressions écologiques univariées menées sur les indicateurs d'entrée ayant permis la détermination des profils environnementaux.....	193

Table des figures

Figure 1 : Représentation cartographique des SIRs lissés de la MC ajustés sur l'âge et le sexe au cours de la période 1990-2014.	18
Figure 2: Clusters spatiaux détectés pour la MC par statistique de scan isotonique à partir des SIRs lissés de la MC ajustés sur l'âge et le sexe au cours de la période 1990-2014.	19
Figure 3 : Représentation cartographique des SIRs lissés de l'IRCT au cours de la période 2005-2018, ajustés sur l'âge, le sexe, l'éloignement au centre lourd de dialyse, le DFG et le FDep.	20
Figure 4 : Occupation des sols et caractéristiques démographiques de la zone d'étude (Corine Land Cover, 2018 ; INSEE, 2018)	26
Figure 5 : Gestion des données manquantes (médiane des communes avoisinantes).	27
Figure 6 : Schéma général de la méthode de construction des indices de vulnérabilité et de résilience (schéma repris de la publication : Brousmiche et al., 2020a)	28
Figure 7 : Structure du jeu de données et principaux phénomènes abordés.	34
Figure 8 : Indice de défaveur sociale FDep (2017)	36
Figure 9 : Représentation cartographique de l'indice de vulnérabilité calculé à l'échelle de la commune.....	37
Figure 10 : Distribution de l'indice de vulnérabilité par département.....	38
Figure 11 : Représentation cartographique de l'indice de résilience calculé à l'échelle de la commune	39
Figure 12 : Distribution de l'indice de résilience par département	40
Figure 13 : Distribution spatiale de l'analyse conjointe réalisée pour l'indice de vulnérabilité et de résilience	41
Figure 14: Distribution de l'indice de vulnérabilité au sein des clusters spatiaux de la MC (surincidence marquée en rouge, sous-incidence en bleu).....	42
Figure 15 : Distribution de l'indice de résilience au sein des clusters spatiaux de la MC (surincidence marquée en rouge, sous-incidence en bleu).....	43
Figure 16 : Distribution des SIRs lissés de la MC par classes croissantes de vulnérabilité (a) ou de résilience (b).	44
Figure 17 : Distribution des SIRs lissés de la MC au sein des profils de communes issus de l'analyse conjointe.....	45
Figure 18 : Comparaison des résultats obtenus à l'issue de l'analyse conjointe lors du projet Cisterres (a) et lors du présent travail de thèse (b).....	47
Figure 19: Principaux phénomènes décrits et répartition des indicateurs au sein des dimensions.....	53
Figure 20 : Stratégie employée pour gérer les limites de quantification hétérogènes entre départements (exemple avec le paramètre Turbidité).	65
Figure 21: Concentration moyenne en nitrates dans les eaux de distribution (valeur moyenne sur la période 2016-2020).....	66
Figure 22 : Paramètres disponibles pour la qualité de l'eau potable à l'échelle du registre Epimad... 66	66
Figure 23 : Paramètres disponibles pour la qualité de l'eau potable à l'échelle du registre Néphronor.	67
Figure 24 : Densité surfacique calculée à l'échelle communale : (a) installations ICPE ; (b) stations d'épuration.....	68
Figure 25 : Alternatives au calcul de la densité surfacique : (a) distance entre le centroïde de la commune et la station d'épuration la plus proche ; (b) pourcentage de la surface communale située à moins de 7,5 km d'un incinérateur de déchets.....	70
Figure 26 : Temporalité des données environnementales retenues et exploitées.	78
Figure 27 : Valeurs de référence sélectionnées pour le calcul du RIM.	85

Figure 28 : Résultats obtenus pour le RIM appliqué au compartiment atmosphérique (RIM air).	88
Figure 29 : Ratios d'imprégnation individuels pour chaque polluant inclus dans le calcul de RIM air. 88	
Figure 30 : Résultats obtenus pour le RIM appliqué au compartiment Sol (RIM sol).	89
Figure 31 : Ratios d'imprégnation individuels pour chaque polluant inclus dans le calcul de RIM sol. 90	
Figure 32 : Résultats obtenus pour le RIM appliqué au compartiment Eau (RIM eau).....	91
Figure 33 : Ratios d'imprégnation individuels pour chaque polluant inclus dans le calcul de RIM eau 92	
Figure 34 : Résultats obtenus pour le compartiment atmosphérique en construisant un indice composite à partir d'une ACP : (a) Cartographie ; (b) cercles de corrélation des composantes principales retenues.	93
Figure 35 : Résultats obtenus pour le compartiment Sol en construisant un indice composite à partir d'une ACP : (a) Cartographie ; (b) cercles de corrélation des composantes principales retenues.	94
Figure 36 : Résultats obtenus pour le compartiment Eau en construisant un indice composite à partir d'une ACP : (a) Cartographie ; (b) cercles de corrélation des composantes principales retenues.	95
Figure 37 : Résultats obtenus pour le compartiment Air en construisant un indice composite à partir de la méthode min-max (MinMax air).	96
Figure 38 : Résultats obtenus pour le compartiment Sol en construisant un indice composite à partir de la méthode min-max (MinMax sol)	97
Figure 39 : Résultats obtenus pour le compartiment Eau en construisant un indice composite à partir de la méthode min-max (MinMax eau).....	97
Figure 40 : Résultats obtenus pour le compartiment Air en construisant un indice composite à partir de la méthode des rangs (RANG air)	99
Figure 41 : Résultats obtenus pour le compartiment Sol en construisant un indice composite à partir de la méthode des rangs (RANG sol).....	99
Figure 42 : Résultats obtenus pour le compartiment Eau en construisant un indice composite à partir de la méthode des rangs (RANG eau)	100
Figure 43 : Cartographies dans chacun des milieux des quatre indices composites étudiés.	101
Figure 44 : Nuages de points et coefficients de corrélation (Pearson et Spearman) entre l'indice RIM et les trois autres indices composites dans chacun des milieux	102
Figure 45 : Cercles des corrélations obtenus à l'issue de l'ACP appliquée au compartiment Eau.....	103
Figure 46 : Valeurs prises par l'ensemble des communes de la zone d'étude pour les paramètres Turbidité et Teneur en nitrates.	104
Figure 47 : Nuages de points et coefficients de corrélation (Pearson et Spearman) entre les indices et deux paramètres d'entrée : Turbidité et Teneur en nitrates.	104
Figure 48 : Nuages de points et coefficients de corrélation de Spearman entre les indices de multi-contamination dans chaque milieu (RIM) et les SIRs de l'IRCT.	107
Figure 49 : (a) Graphe et Tableau de distribution des SIRs de l'IRCT par quartile croissant de multi-contamination de l'air (RIMair) ; (b) Représentation cartographique de l'indice de multi-contamination de l'air (RIMair) ; (c) Représentation cartographique de la répartition spatiale d'incidence de l'IRCT.	108
Figure 50 : Nuages de points et coefficients de corrélation de Spearman entre les indices de multi-contamination dans chaque milieu (RIM) et les SIRs de la MC.....	109
Figure 51 : (a) Graphe et Tableau de distribution des SIRs de la MC par quartile croissant de multi-contamination des sols (RIMsol) ; (b) Représentation cartographique de l'indice de multi-contamination des sols (RIMsol) ; (c) Représentation cartographique de la répartition spatiale d'incidence de la MC.	110
Figure 52 : Principales étapes méthodologiques pour développer une typologie territoriale. Les résultats qui concernent le cas d'étude sont représentés en rouge (Paumelle et al., 2023).	116

Figure 53 : (a) Représentation cartographique de la classification obtenue après CAH et consolidation des k-means ; (b) Dendrogramme issu de la CAH ; (c) Gains d'inertie inter-classe obtenus avec la CAH selon le nombre de classes retenues (c) (Paumelle et al., 2023).	123
Figure 54 : Boxplots représentant la distribution des z-scores pour les 39 indicateurs d'entrée au sein de chaque profil (Paumelle et al., 2023).	126
Figure 55 : (a) Représentation cartographique de la typologie environnementale des communes ; (b) Représentation cartographique de la répartition spatiale d'incidence de la MC.	134
Figure 56: Résultats issus de la régression écologique comparant les RR associés à chaque profil environnemental.	135
Figure 57 : (a) Représentation cartographique de la typologie environnementale des communes ; (b) Représentation cartographique de la répartition spatiale d'incidence de l'IRCT.	141
Figure 58 : Graphe et Tableau de distribution des SIRs de l'IRCT au sein des sept profils environnementaux identifiés.	142
Figure 59 : Répartition spatiale des taux de prévalence de la MC standardisés sur l'âge et le sexe (2007-2014).	151
Figure 60 : Plan d'échantillonnage utilisé pour les campagnes de prélèvements réalisés au sein des huit clusters spatiaux d'incidence de la MC.	153

Liste des tableaux

Tableau 1 : Pools d'indicateurs retenus pour le calcul des indices de résilience et de vulnérabilité....	31
Tableau 2: Définition de quatre profils de communes lors de l'analyse conjointe des indices.	33
Tableau 3: Critères d'inclusion et d'exclusion de nouvelles sources de données.....	52
Tableau 4 : Niveaux de contamination des milieux - Présentation détaillée des données sources et des indicateurs produits	55
Tableau 5 : Niveaux des émissions polluantes - Présentation détaillée des données sources et des indicateurs produits	56
Tableau 6 : Localisation des sources d'émissions - Présentation détaillée des données sources et des indicateurs produits	57
Tableau 7 : Occupation des sols - Présentation détaillée des données sources et des indicateurs produits	59
Tableau 8 : Pratiques agricoles - Présentation détaillée des données sources et des indicateurs produits	60
Tableau 9 : Naturalité - Présentation détaillée des données sources et des indicateurs produits.....	62
Tableau 10 : Climat - Présentation détaillée des données sources et des indicateurs produits.....	63
Tableau 11 : Jeu de données final retenu pour la détermination des profils environnementaux. (modifié à partir de Paumelle et al., 2023).	121
Tableau 12 : Valeur médiane et écart interquartile pour chaque indicateur d'entrée au sein des sept profils environnementaux identifiés (Paumelle et al., 2023).	125
Tableau 13 : Principaux résultats issus des régressions écologiques univariées menées sur les 39 indicateurs d'entrée.	136

Liste des abréviations

ACP : Analyse en composantes principales
BASIAS : Base de données des anciens sites industriels et activités de services
BASOL : BAse de données des sites et SOLs pollués
BNVD : Banque Nationale des Ventes de produits phytopharmaceutiques par les Distributeurs agréés
CAH : Classification ascendante hiérarchique
CISTERRES : Cartographie d'Indices de vulnérabilité et d'adaptabilité au Service des TERRitoires en Environnement et Santé.
CLC : Corine Land Cover
COV : Carbone organique volatil
CROPS : CROhn disease and Pollution of Soils
DFG : Débit de filtration glomérulaire
EIQ : Ecart interquartile
EPCI : Établissement public de coopération intercommunale
ETM : Eléments traces métalliques
FDep : French Deprivation Index
GD4H : Green data for health
HAP : Hydrocarbure aromatique polycyclique
HDH : Health data hub
HVN : Haute valeur naturelle
IC : Intervalle de confiance
ICPE : Installation classée pour la protection de l'environnement
IESS : Inégalités environnementales et sociales de santé
INERIS : Institut national de l'environnement industriel et des risques
IR : Indice de résilience
IRCT : Insuffisance rénale chronique terminale
IRIS : Ilots Regroupés pour l'Information Statistique
IV : Indice de vulnérabilité
LQ : Limite de quantification
MC : Maladie de Crohn
MICI : Maladies inflammatoires chroniques de l'intestin
MRC-PE : Maladie rénale chronique et profils environnementaux
NDVI : Normalized Difference Vegetation Index
OMS : Organisation mondiale de la santé
OR : Odds ratio
PCB : Polychlorobiphényles
PNSE : Plan national santé environnement
Ppb : Partie par billion
Ppm : Partie par million
REIN : Réseau, Epidémiologie, Information, Néphrologie
RIM : Ratio d'imprégnation moyen
RMQS : Réseau de mesure de la qualité des sols
RPG : Registre parcellaire graphique
RR : Risque relatif
SIR : Ratio d'incidence standardisé
USEPA : United-States environmental protection agency

Introduction générale

L'environnement constitue un déterminant de santé majeur. L'Organisation Mondiale de la Santé (OMS) estime que 23% des décès prématurés à l'échelle mondiale peuvent être imputés à des facteurs environnementaux et comportementaux (qualité de l'air, qualité de l'eau, alimentation, mode de vie, etc...) (Prüss-Üstün et al., 2016). Par ailleurs, la part des décès attribuables à des maladies non transmissibles est en augmentation à l'échelle mondiale. Elles seraient à l'origine de 41 millions de décès chaque année, soit 74 % de l'ensemble des décès dans le monde (OMS, 2022). Les principaux facteurs de risque environnementaux sont la pollution de l'air, de l'eau et des sols, l'exposition aux substances chimiques, ou le changement climatique. En 2019, la pollution au sens large aurait causé 9 millions de décès à l'échelle mondiale (Fuller et al., 2022). En France, en 2021, la pollution atmosphérique d'origine humaine était associée à environ 40 000 décès par an pour les PM_{2.5}, et près de 7000 pour le dioxyde d'azote (NO₂) (Santé Publique France, 2021). Cependant, il demeure souvent difficile de caractériser et de quantifier précisément la relation entre l'environnement et la santé. L'état de santé d'une population est influencé par un ensemble complexe de déterminants individuels, de facteurs sociaux et environnementaux.

Certaines maladies chroniques présentent une étiologie multifactorielle encore en partie indéterminée et pour laquelle un lien avec l'environnement est fortement suspecté. C'est le cas pour deux maladies chroniques en particulier : la maladie de Crohn (MC) et l'insuffisance rénale chronique terminale (IRCT).

La maladie de Crohn (MC) est une maladie inflammatoire chronique de l'intestin (MICI) caractérisée par des phases de rémission et de rechute, résultant d'une activation excessive et incontrôlée du système immunitaire intestinal. La MC n'est pas une maladie rare ; elle touche plus de 3 millions de personnes dans le monde (1,5 million en Amérique du Nord et 1,5 million en Europe, dont 250 000 en France), ce qui en fait une préoccupation de santé publique majeure. Son étiologie est encore mal connue et plusieurs arguments suggèrent un lien avec des facteurs environnementaux. D'abord son incidence est en augmentation à l'échelle mondiale, en particulier dans les sociétés occidentales, depuis une trentaine d'années (Lerner et al., 2016). Une répartition spatiale d'incidence très hétérogène est constatée dans de nombreux pays et à de nombreuses échelles (Genin et al., 2013; Green et al., 2006; Michel et al., 2010; Nerich et al., 2010; Ng et al., 2013). De plus, des changements sont observés au niveau de l'âge de déclaration de la maladie (Panés et al., 2007) avec notamment de plus en plus d'enfants et de jeunes adultes concernés (Ghione et al., 2018; Gower-Rousseau et al., 2013). Des facteurs génétiques ont bien été identifiés (Hugot et al., 2001; Jostins et al., 2012; Ogura et al., 2001), cependant ceux-ci sont insuffisants pour expliquer les causes de la maladie. Des études se sont notamment intéressées aux jumeaux monozygotes et ont montré une concordance d'environ 50 % seulement (Orholm et al., 1991; Tysk et al., 1988), ce qui suggère de nouveau une influence non négligeable des comportements individuels et/ou des facteurs environnementaux. Un autre argument est l'augmentation significative de la MC constatée chez les populations migrantes se déplaçant d'une région à faible incidence vers une région à forte incidence (Agrawal et al., 2020).

Récemment un état de l'art a été publié sur le lien entre la MC et les pollutions environnementales, plusieurs conclusions ont pu être émises (Tenailleau et al., 2020). D'abord, à l'heure actuelle aucun facteur de risque environnemental n'a pu être clairement identifié pour la MC. Des associations ont été constatées (notamment avec certains agents biologiques ou avec certains métaux) mais les résultats sont contradictoires et le niveau de preuve insuffisant pour faire consensus. L'article identifie également des pistes encore peu documentées qu'il serait pertinent de considérer. D'abord le lien avec certains polluants qui n'ont que très peu ou jamais été étudiés : notamment le lien avec certains

métaux ou avec des perturbateurs endocriniens (produits phytosanitaires, produits pharmaceutiques, etc...). Par ailleurs, la plupart des études se sont concentrées sur le compartiment aérien ou aquatique et de nombreuses hypothèses environnementales n'ont pas encore été explorées. Pour finir, la nécessité et la pertinence de caractériser l'environnement sous un prisme multi-dimensionnel ont été clairement soulignées.

L'Insuffisance rénale chronique terminale (IRCT) correspond au dernier stade de la maladie rénale chronique. La fonction rénale est alors tellement impactée qu'il devient nécessaire d'engager une dialyse à long terme ou une greffe de rein. L'étiologie de cette maladie est également mal établie. De nouveau, plusieurs arguments suggèrent une influence de l'environnement. D'abord au cours des dernières décennies, son incidence est en augmentation constante dans le monde, bien qu'une certaine stabilisation soit observée depuis les années 2000 dans les pays développés (Aghighi et al., 2009; Counil et al., 2008; Eggers, 2011; Wakai et al., 2004; Zoccali et al., 2010). Par ailleurs, une forte hétérogénéité spatiale d'incidence est constatée dans de nombreux pays (Collins et al., 2015; Hommel et al., 2010; Occelli et al., 2014; Roderick et al., 1999; Rosansky et al., 1990; Shaw et al., 2013; Tanner et al., 2013; Usami et al., 2000; Wimmer et al., 2003). Ces variations spatio-temporelles sont partiellement liées au vieillissement de la population et à la prévalence croissante d'autres pathologies associées : hypertension, diabète, obésité (Collister et al., 2016). De plus, une forte influence de facteurs socio-démographiques et de facteurs liés à l'accessibilité aux soins a également été clairement identifiée (Boyle et al., 1996; Couchoud et al., 2010; van de Luijngaarden et al., 2012). Pour finir, plusieurs études ont mis en évidence des associations avec certains facteurs environnementaux : contamination en métaux (Cheng et al., 2018; Hellström et al., 2001; Sabath and Robles-Osorio, 2012; Tsai et al., 2018), pollution de l'air (Bowe et al., 2018a; C.-D. Wu et al., 2020a), produits phytosanitaires (Almaguer et al., 2014; G and Osama, 2010; Lebov et al., 2016). Mais ces derniers résultats sont encore minces et souvent contradictoires.

Pour contribuer à renseigner l'étiologie de ces deux maladies chroniques, leur répartition spatiale d'incidence a été cartographiée à l'échelle communale dans le Nord de la France, à partir des données issues de deux registres de santé. La répartition spatiale d'incidence de ces maladies constitue le point d'entrée pour investiguer les déterminants environnementaux susceptibles d'être impliqués dans la survenue de ces maladies à l'échelle des territoires. Ces registres constituent des leviers, véritables valeurs ajoutées, dans la caractérisation des liens environnement – pathologies.

Les cartographies d'incidence de la MC et de l'IRCT ont été générées à l'échelle des communes sur le périmètre de chacun des registres de santé. L'ensemble de ces analyses ont été réalisées par l'équipe de recherche METRICS (ULR 2694 CHU Lille), spécialisée en biostatistique en santé publique. Le contour administratif des communes utilisé est celui de l'IGN datant de juin 2016.

Le registre EPIMAD est le seul registre français dédié aux MICI et il est également considéré comme le plus important au monde (Gower-Rousseau et al., 1994). Mis en place en 1988, il a été créé pour collecter et analyser les données des patients atteints de MICI dans le nord de la France. Ce registre couvre quatre départements (Nord, Pas-de-Calais, Somme, Seine Maritime), ce qui correspond à une surface de 24,862 km² et une population de 6 millions de personnes (soit 9,1% de la population française).

Jusqu'en 2014, 19 266 cas de MICI ont été enregistrés. Il s'agit du premier système d'enregistrement de cas de MICI en France. Au cours de la période de 1988 à 2014, une augmentation de l'incidence de la MC a été observée, passant de 5,3 à 7,6 cas pour 100 000 habitants (Fumery et al., 2018). Par ailleurs, une augmentation significative des cas chez les adolescents a été constatée, passant de 4,2 à 9,5 pour

100 000 habitants. Ce registre de santé été conçu, entres autres, pour permettre d'étudier les variations spatio-temporelles de la MC.

Sur la base de ce registre, la cartographie d'incidence de la MC à l'échelle communale a été calculée en considérant l'ensemble des cas incidents sur la période 1990-2014 (10 792 cas). Les ratios d'incidence standardisés (SIR) sur l'âge et le sexe (standardisation indirecte) ont été calculés. Puis la distribution spatiale des SIRs a été lissée grâce à un modèle de régression bayésien hiérarchique (Besag et al., 1991a) (Figure 1).

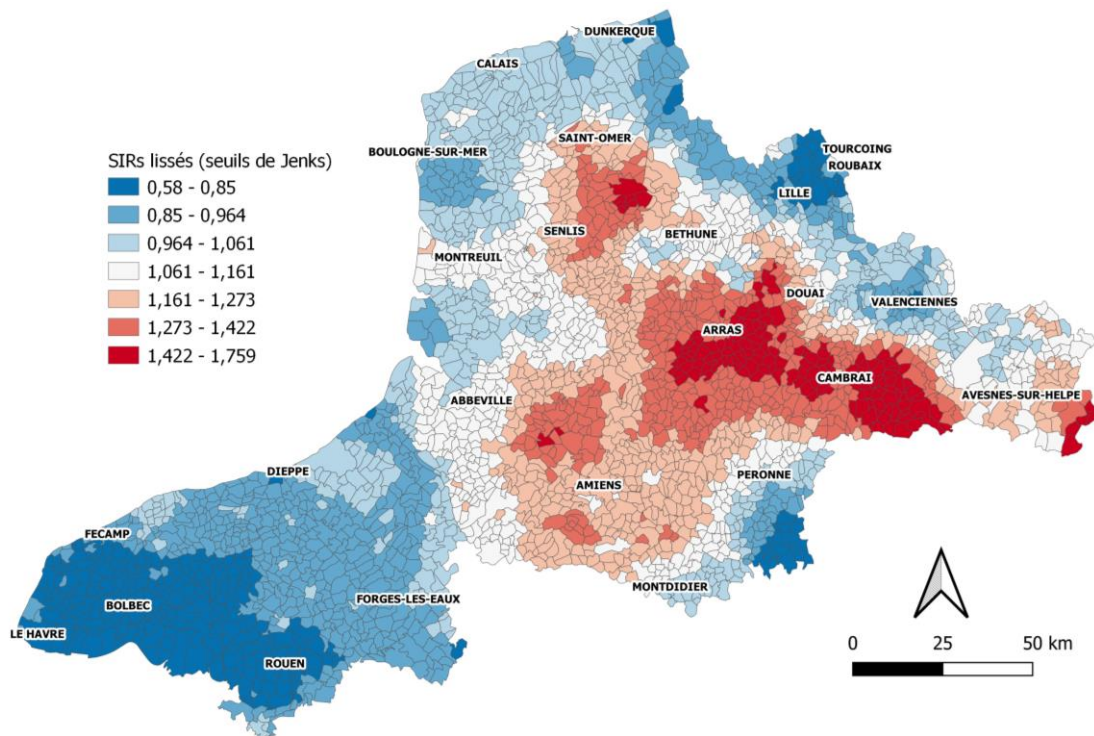


Figure 1 : Représentation cartographique des SIRs lissés de la MC ajustés sur l'âge et le sexe au cours de la période 1990-2014.

Un cluster spatial d'incidence est une zone géographique composée d'une ou plusieurs unités spatiales contiguës, caractérisées par un niveau d'incidence anormalement élevé (sur-incidence) ou faible (sous-incidence) par rapport à la zone d'étude dans son ensemble. Sur la base des SIRs lissés, des clusters spatiaux d'incidence ont été détectés sur la période 1990-2014 par statistique de scan isotonique (Kulldorff, 1997). Quatre clusters de sur-incidence et quatre clusters de sous-incidence ont été détectés. Dans sa version isotonique, la statistique de scan donne la variabilité du risque au sein même du cluster. Par commodité, les huit clusters ont été nommés à partir du nom d'une des communes présentes dans le cluster (Figure 2). Ainsi, les clusters de Cambrai, d'Arras, de Saint-Omer et d'Amiens sont des clusters de sur-incidence. Les clusters de Rouen, de Bolbec, de Lille et de Péronne sont des clusters de sous-incidence.

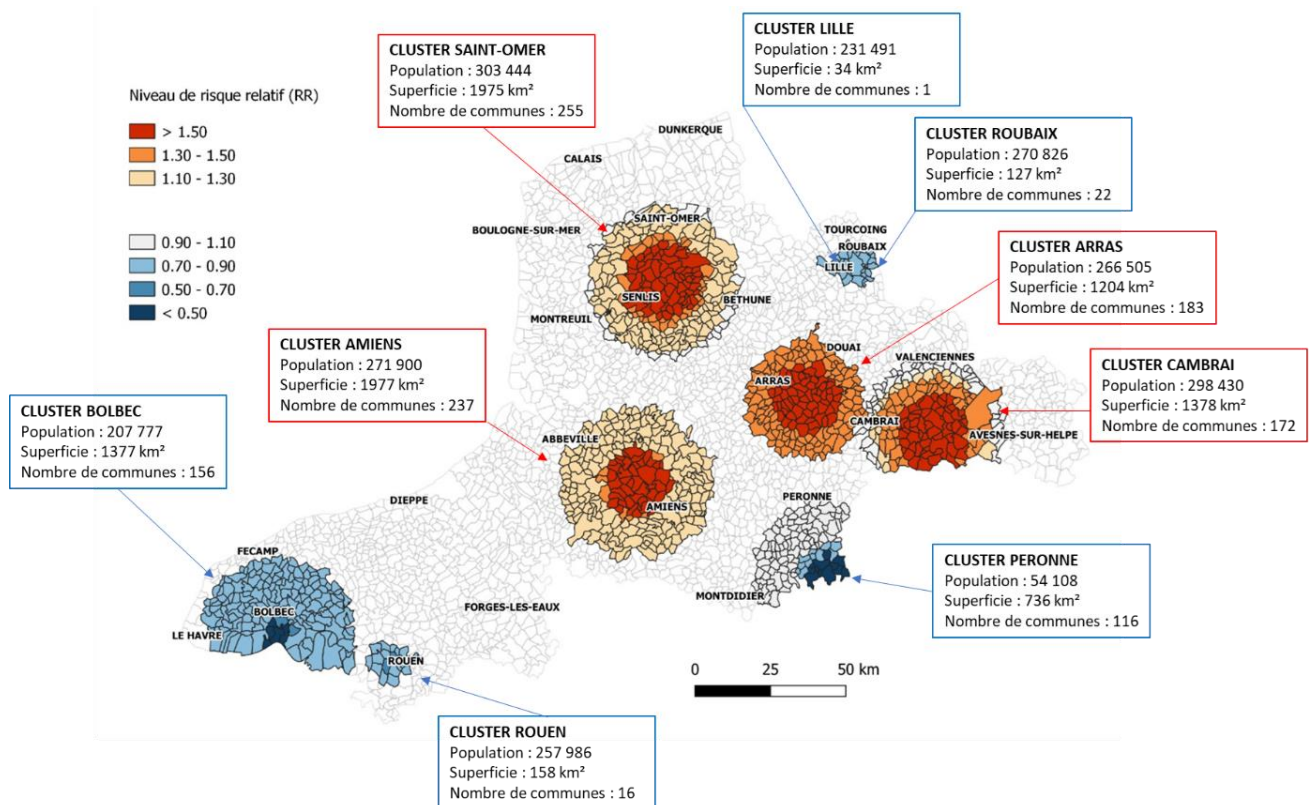


Figure 2: Clusters spatiaux détectés pour la MC par statistique de scan isotonique à partir des SIRs lissés de la MC ajustés sur l'âge et le sexe au cours de la période 1990-2014.

Le registre Néphronor est la déclinaison régionale du Réseau, Epidémiologie, Information, Néphrologie (REIN) (Couchoud et al., 2006). Il existe depuis 2005 et recense l'ensemble des patients atteints d'IRCT dans les départements du Nord et du Pas-de-Calais.

Sur la période 2005-2018, 9 339 patients âgés de plus de 18 ans ont été inclus dans le registre à la suite d'un diagnostic d'IRCT. Leur commune de résidence au moment du diagnostic a été enregistrée dans la base de données. L'incidence de l'IRCT a ensuite été calculée sur cette même période, à l'échelle des communes et sur tout le périmètre du registre. Les données d'incidence ont d'abord été ajustées sur l'âge et le sexe. Puis les données ont été ajustées sur d'autres facteurs de confusion pour prendre en compte l'accessibilité au système de soin, les pratiques de soin et la défaveur socio-économique. L'accessibilité aux soins a été étudiée en considérant le temps de trajet par la route entre le centroïde des communes et le centre lourd de dialyse le plus proche. Selon les centres de dialyse, des différences dans les pratiques de soin peuvent être observées sur l'initiation à la dialyse. Le débit de filtration glomérulaire (DFG) est un paramètre qui mesure le volume de plasma filtré par les reins par unité de temps, il permet de définir la gravité de l'insuffisance rénale. Selon le centre, les équipes de néphrologues initient plus ou moins tôt la dialyse, le DFG peut être dès lors plus ou moins élevé. Ce paramètre a donc été utilisé pour ajuster l'incidence et prendre en compte ce différentiel dans les pratiques de soin. Enfin, pour ajuster sur la défaveur socio-économique, le FDep a été utilisé (Rey et al., 2009). Ces travaux d'ajustement des données d'incidence ont été présentés en détail dans la thèse de recherche de Delphine Brousmiche (Brousmiche, 2022) (Figure 3).

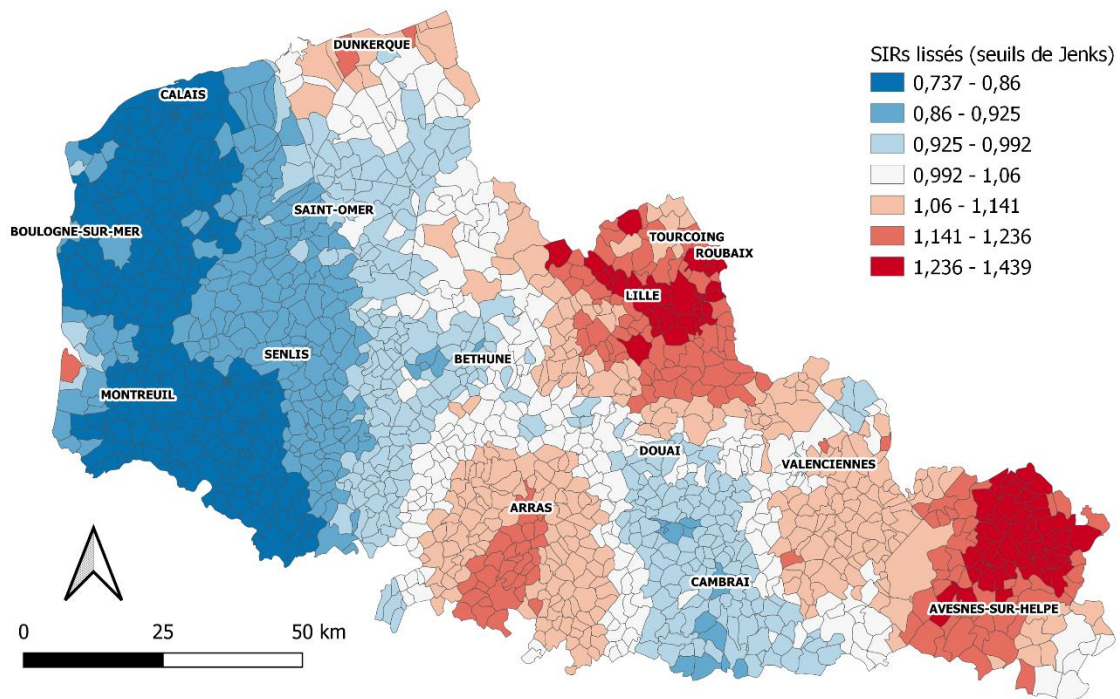


Figure 3 : Représentation cartographique des SIRs lissés de l'IRCT au cours de la période 2005-2018, ajustés sur l'âge, le sexe, l'éloignement au centre lourd de dialyse, le DFG et le FDep.

La caractérisation de l'environnement et son lien à la santé est souvent appréhendée de manière cloisonnée dans les études en population. La plupart des travaux se focalisent sur une source, un polluant, un milieu d'exposition ou un phénomène en particulier. Les études analysent par exemple l'association entre l'incidence ou la prévalence d'un événement de santé et la pollution de l'air (Ocelli et al., 2020a; Bove et al., 2018b; C.-D. Wu et al., 2020b), la qualité de l'eau potable (Cheng et al., 2017; Wasana et al., 2016), ou encore l'accès aux espaces verts (Lim et al., 2017; Richardson and Mitchell, 2010). Bien que ces approches soient nécessaires, elles sont parfois inappropriées pour appréhender la complexité du lien entre environnement et santé. En particulier, elles présentent des limites importantes lorsqu'il s'agit d'étudier des maladies multifactorielles. Cela est d'autant plus le cas dès lors que les facteurs de risque environnementaux sont encore inconnus, puisqu'il faut préalablement sélectionner des pistes environnementales. Dans ce cas, il semble plus judicieux de procéder par étapes, en commençant par des stratégies territoriales et multidimensionnelles, avant éventuellement de cibler un ou des facteurs de risque environnementaux spécifiques. Dans ce contexte, comment mobiliser de multiples données environnementales ouvertes pour identifier les déterminants territoriaux de maladies multifactorielles ?

Dans ce contexte, l'objectif principal de ce travail de thèse est précisément de proposer une description intégrée de l'environnement à l'échelle des territoires. Pour y parvenir, la stratégie retenue a consisté à collecter et réutiliser des données environnementales ouvertes. Ces données ont été exploitées pour générer des indicateurs spatialisés à l'échelle des communes et couvrir le périmètre des registres de santé. Tout l'enjeu était de parvenir à caractériser l'environnement le plus exhaustivement possible.

Une fois le jeu de données spatial constitué, plusieurs méthodologies multidimensionnelles ont été mobilisées pour caractériser l'environnement. Dans l'ensemble de la thèse, deux approches principales peuvent être distinguées. La première implique le développement d'indices spatiaux composites, tandis que la deuxième consiste à construire des typologies territoriales. Les indices spatiaux composites sont utilisés pour synthétiser, en une mesure globale, l'information contenue dans un grand nombre d'indicateurs spatialisés. Cette approche moyenne les facteurs influents qu'ils soient protecteurs, inductifs ou amplifiant et intègre les potentiels effets cocktail. Les typologies territoriales consistent à regrouper des unités spatiales en fonction de leurs caractéristiques environnementales communes, créant ainsi des catégories ou classes de territoires. De cette manière, il est possible d'appréhender les combinaisons de facteurs qui définissent ces territoires sur le plan environnemental.

Une fois caractérisées, les résultats de ces approches multidimensionnelles ont dans un second temps été associés aux variations spatiales d'incidence des maladies chroniques par le biais d'approches écologiques menées à l'échelle territoriale. Dans le cadre de cette thèse, ce travail a été réalisé pour la MC et l'IRCT. Mais ces recherches ont été menées de telle sorte que le même procédé puisse facilement être décliné par la suite en lien avec d'autres maladies ou événements de santé. Les approches mobilisées sont des approches géographiques et écologiques menées à l'échelle du territoire, il est donc important de rappeler que l'enjeu n'est pas de démontrer des liens de causalité, mais plutôt d'identifier des pistes étiologiques qui pourraient par la suite être approfondies, notamment à l'échelle individuelle.

Ce mémoire de thèse s'articule en quatre chapitres, suivant la chronologie des travaux de recherche entrepris. Chaque chapitre contribue à la compréhension du sujet, et leur séquence reflète l'évolution du raisonnement. L'ensemble étant itératif, nous avons fait le choix d'intégrer le matériel et méthode dans chacun des chapitres. Un chapitre servant d'appui méthodologique au suivant.

Dans le premier chapitre, une méthode préalablement développée au sein du laboratoire a été utilisée pour mesurer le cumul de déterminants environnementaux favorables ou défavorables à la santé à l'échelle des territoires. Un jeu de données socio-environnemental a été conçu pour décrire l'environnement à l'échelle des communes, en ciblant les dimensions suivantes : économie, social, services, politique, santé et environnement physique. Sur la base de ces données, deux indices spatiaux composites ont été calculés pour évaluer le potentiel de vulnérabilité et de résilience des communes. L'indice de vulnérabilité permet de rendre compte du cumul de déterminants environnementaux potentiellement défavorables à la santé. L'indice de résilience rend compte du cumul de déterminants environnementaux favorables à la santé, en d'autres termes sa capacité à faire face aux inégalités de santé. Ces indices composites ont ensuite été utilisés pour mesurer l'influence globale de l'environnement sur la répartition d'incidence de ces maladies.

Pour compléter l'analyse, l'environnement a été conceptualisé de manière différente dans la suite de la thèse. Les travaux se sont concentrés sur la description de l'environnement physique au sens du milieu de vie extérieur des populations. Le deuxième chapitre expose donc les recherches qui ont été menées pour décrire la diversité des pressions environnementales qui s'exercent sur un territoire. Partant de moins de 10 au début de la thèse, plus de 100 indicateurs spatialisés ont été développés. Le jeu de données ainsi constitué permet de caractériser spatialement le niveau de contamination de l'air, de l'eau et des sols, le niveau des émissions polluantes, la proximité aux sources d'émissions, l'occupation des sols, les pratiques agricoles, la naturalité des territoires et le climat.

Dans le troisième chapitre, les données de contamination ont été mobilisées pour caractériser la multi-contamination dans chacun des milieux (air, eau, sols). Pour ce faire, plusieurs méthodes de calcul ont

été testées à l'échelle communale pour être comparées. L'indice composite le plus approprié pour étudier le phénomène a finalement été retenu. Enfin, les associations entre la multi-contamination des milieux et les variations d'incidence de la MC et de l'IRCT ont été évaluées.

Dans le quatrième chapitre, un cadre méthodologique a été conçu pour développer une typologie environnementale des territoires. Cette approche a été appliquée à l'échelle des communes en considérant l'ensemble des données spatiales décrivant l'environnement physique. Les résultats permettent d'avoir une vision complexe des territoires et d'appréhender les combinaisons de facteurs environnementaux qui les caractérisent. Des profils environnementaux de communes ont été identifiés et décrits sur l'ensemble de la zone d'étude.

Dans le cinquième et dernier chapitre, les répartitions spatiales d'incidence de la MC et de l'IRCT ont été mises en lien avec la typologie environnementale des communes. L'objectif était de déterminer si certains profils environnementaux étaient associés à une incidence anormalement élevée ou faible pour ces maladies.

CHAPITRE I – Potentiels de vulnérabilité et de résilience des territoires.

Introduction

Pour aborder la complexité du lien entre environnement et santé à l'échelle du territoire, l'environnement doit être décrit de manière intégrée et multi-dimensionnelle. En particulier, il faut être capable de caractériser l'ensemble des déterminants territoriaux susceptibles de générer des inégalités spatiales de santé. Dans ce contexte, la notion d'environnement est à appréhender au sens large du terme. Il s'agit d'étudier l'environnement physique mais également l'environnement social, économique, ou politique.

Un autre enjeu important est de concevoir l'environnement comme générateur d'effets délétères mais également d'effets bénéfiques sur la santé des populations. Le plus souvent dans la littérature, l'environnement est considéré pour son impact négatif en ciblant des phénomènes comme la contamination des milieux (Occelli et al., 2020a; Bowe et al., 2018b; C.-D. Wu et al., 2020b), la proximité aux sources d'émission (Brender et al., 2011; Domingo et al., 2020) ou encore la défaveur socio-économique (Braveman et al., 2010; Morton et al., 2016; Reiss, 2013). Cependant, plus récemment, certains facteurs environnementaux ont été mis en évidence pour leur impact positif sur la santé des populations. Ce lien est bien établi pour l'accès aux espaces verts en milieu urbain qui joue un rôle protecteur dans la survenue de nombreux événements de santé : obésité (Jia et al., 2021), maladies cardiovasculaires (De la Fuente et al., 2021), santé mentale (McCormick, 2017). En dehors du cadre urbain, la notion d'espaces verts est moins pertinente et le concept d'aménités environnementales semble plus adapté. Il désigne la qualité des ressources naturelles au sein d'un territoire. Plusieurs publications suggèrent par exemple un effet bénéfique de la biodiversité sur la santé humaine, en particulier sur la régulation du système immunitaire (Aerts et al., 2018; Hanski et al., 2012; Rook, 2013a). Une interaction entre aménités environnementales et défaveur sociale est même évoquée : l'accessibilité aux espaces naturels pourrait atténuer l'impact de la défaveur sociale sur la santé des populations (Mitchell and Popham, 2008).

Pour se rapprocher de cette vision intégrée de l'environnement et des territoires, la stratégie retenue consiste à utiliser des données ouvertes pour générer des indicateurs spatialisés et constituer un jeu de données caractérisant les territoires. Ce travail de récolte et de structuration des données a été initié en amont du présent travail de thèse. En particulier, un cadre méthodologique a été défini lors des travaux de thèse de Delphine Brousmiche (Brousmiche, 2022) et dans le cadre du projet de recherche CISTERRES (Cartographie d'Indices de vulnérabilité et de résilience au Service des Territoires en Environnement et Santé ; soutenu par le Conseil Régional Hauts-de-France, l'ARS Hauts-de-France et le CPER CLIMIBIO). Ces recherches ont été menées sur le périmètre des Hauts-de-France et à l'échelle des communes. L'objectif était de caractériser les inégalités environnementales et sociales de santé (IESS) en développant deux indices spatiaux composites : un indice de vulnérabilité (IV) et un indice de résilience (IR) des communes. La vulnérabilité a été définie comme l'accumulation de déterminants de santé défavorables dans une zone géographique donnée et la résilience comme la capacité d'un territoire à faire face à ces inégalités de santé. Dans ce cadre, l'hypothèse est que l'environnement est à considérer comme générateur d'effets positifs ou négatifs inégalement répartis, et ayant des effets sur la santé des populations. La démarche qui avait été engagée se compose de trois grandes étapes, chacune ayant fait l'objet d'une publication scientifique : (i) une revue de la littérature portant sur le développement d'indices spatiaux composites en santé-environnement (Brousmiche et al., 2020b) ; (ii) un data article qui présente le jeu de données environnemental construit pour le calcul des indices de vulnérabilité et de résilience (Brousmiche et

al., 2021) ; (iii) un article de recherche sur la méthode de calcul et les principaux résultats obtenus (Brousmiche et al., 2020a).

Menée en amont, la revue de la littérature avait permis de dresser un état de l'art sur les indices spatiaux composites utilisés en santé-environnement (Brousmiche et al., 2020b). En particulier, elle avait permis de décrire les différentes manières de percevoir l'environnement et la nature des données qui sont utilisées pour le caractériser. Dans cet article, 23 indices composites existants avaient été sélectionnés et analysés. De manière générale, une grande hétérogénéité dans les pratiques avait été constatée, tant au niveau du nombre que du type d'indicateurs intégrés dans les indices composites. Les 23 indices composites sélectionnés étaient calculés sur la base de 4 à 76 indicateurs répartis au sein de 2 à 6 dimensions. Ces choix variaient principalement selon l'objectif recherché, le contexte géographique ou l'unité spatiale considérée. Sur la base de cet état de l'art, pour harmoniser les pratiques et favoriser la transposabilité des indices, un cadre méthodologique avait été proposé. Une des préconisations était de structurer l'environnement et les données associées en six dimensions distinctes : économie, environnement physique, politique, santé, services, social. Ce choix s'appuyait notamment sur des travaux de l'OMS sur les déterminants de santé (Dahlgren and Whitehead, 1991) et sur des travaux de la Commission pour la Mesure des Performances Economiques et du Progrès Social (CMPEPS) sur la qualité de vie (Stiglitz et al., 2009). Cette conception de l'environnement en six dimensions a ensuite été reprise et utilisée pour structurer le jeu de données environnemental. A partir de données ouvertes, 50 indicateurs spatialisés ont été générés. Chacun de ces indicateurs a été calculé à l'échelle des communes sur le périmètre des Hauts-de-France. L'ensemble de ces indicateurs ont été présentés au sein du data article (Brousmiche et al., 2021). Ensuite, à partir de ce jeu de données, un indice de vulnérabilité et un indice de résilience ont été conçus et calculés sur les 3 817 communes des Hauts-de-France (Brousmiche et al., 2020a). Enfin, une interprétation croisée des deux indices a été menée pour décrire à la fois le niveau de vulnérabilité des communes, mais également les ressources dont elles disposent pour y faire face.

Dans ce premier chapitre de thèse, le même cadre méthodologique a été repris pour tenter de mieux comprendre l'influence de l'environnement, au sens large, sur l'étiologie de la MC. La principale question qui se pose est la suivante : les potentiels de vulnérabilité et de résilience des territoires contribuent-ils à expliquer la variabilité spatiale d'incidence de la MC ?

La première étape a consisté à étendre le jeu de données environnemental pour couvrir le périmètre du registre Epimad. En l'occurrence, les données ont été étendues aux communes de la Seine Maritime, en plus de celles des Hauts-de-France. De plus, un travail d'enrichissement des données a été réalisé en parallèle, plusieurs nouveaux indicateurs spatialisés récemment générés par le laboratoire ont été ajoutés pour mieux caractériser la dimension « Environnement physique ». A partir du jeu de données ainsi constitué, les indices de vulnérabilité et de résilience ont ensuite été recalculés avant d'être mis en lien avec les données d'incidence de la MC.

Enfin, ce chapitre a également permis de mettre à l'épreuve la transposabilité de la méthode de calcul des indices de vulnérabilité et de résilience. Après avoir élargi la zone d'étude à un département supplémentaire, avoir ajouté quelques indicateurs dans le jeu de données : les résultats obtenus à l'échelle des Hauts-de-France en sont-ils fortement impactés ?

A noter que les indices de vulnérabilité et de résilience ont également été mis en lien avec l'hétérogénéité spatiale d'incidence de l'IRCT. Mais ce travail a été réalisé au début de la thèse en utilisant des données d'incidence qui sont aujourd'hui considérées comme obsolètes. Elles n'étaient ajustées que sur l'âge et le sexe, sans prise en compte des facteurs confondants étudiés par la suite (accessibilité aux centres de dialyse, pratiques de soin et défaveur sociale). Ainsi, ces résultats n'ont

pas été présentés dans ce chapitre. De plus, toujours en lien avec l'IRCT, un travail approfondi a été mené par la suite dans les travaux de thèse Delphine Brousmiche. L'indice de résilience et de vulnérabilité ont été recalculés en considérant cette fois des indicateurs spécifiquement sélectionnés en lien avec l'IRCT. Ces analyses ont été menées à l'échelle communale avec les données d'incidence ajustés sur tous les facteurs confondants précédemment cités. Les résultats obtenus démontrent une influence significative de la résilience des territoires sur la répartition spatiale d'incidence de l'IRCT (résultats en cours de publication).

Matériel & Méthodes

La méthodologie employée dans ce chapitre s'articule en trois étapes principales : (i) constitution du jeu de données environnemental, (ii) construction des indices de vulnérabilité et de résilience, (iii) mise en relation des indices avec les variations spatiales d'incidence de la MC.

1. Développement d'un jeu de données multi-dimensionnel : économie, environnement physique, politique, santé, services, social.

- a. Zone d'étude et unité spatiale de référence

Pour constituer le jeu de données, la zone d'étude retenue couvre six départements : Aisne (02), Nord (59), Oise (60), Pas-de-Calais (62), Somme (80) et Seine Maritime (76). En lien avec les données de santé à disposition, l'unité spatiale de référence considérée est la commune et le contour administratif utilisé est celui de l'IGN datant de juin 2016. Elle s'étend sur 38 367 km² et comprend 4 535 communes pour 7 259 991 habitants (INSEE, 2018). Les communes présentent une surface médiane de 6.96 [4.77 ; 10.35] km² (médiane [écart interquartile (EIQ)]), et une population médiane de 445 [220 ; 1016] habitants. L'occupation des sols y est répartie de manière hétérogène et tous les types d'occupation sont représentés (Figure 4). Des territoires artificialisés se répartissent autour des grandes agglomérations urbaines telles que Lille, Dunkerque Amiens, Rouen, Le Havre, ; dans l'ancien bassin minier entre Béthune et Valenciennes ; ou encore au Sud de l'Oise (à la frontière de l'Île-de-France). Par ailleurs, les territoires agricoles sont omniprésents et couvrent l'écrasante majorité de la zone d'étude. Enfin, les espaces naturels sont répartis de manière irrégulière sur le territoire. En particulier, les départements de l'Oise, de l'Aisne et de la Seine-Maritime présentent une proportion plus importante de forêts et d'espaces semi-naturels.

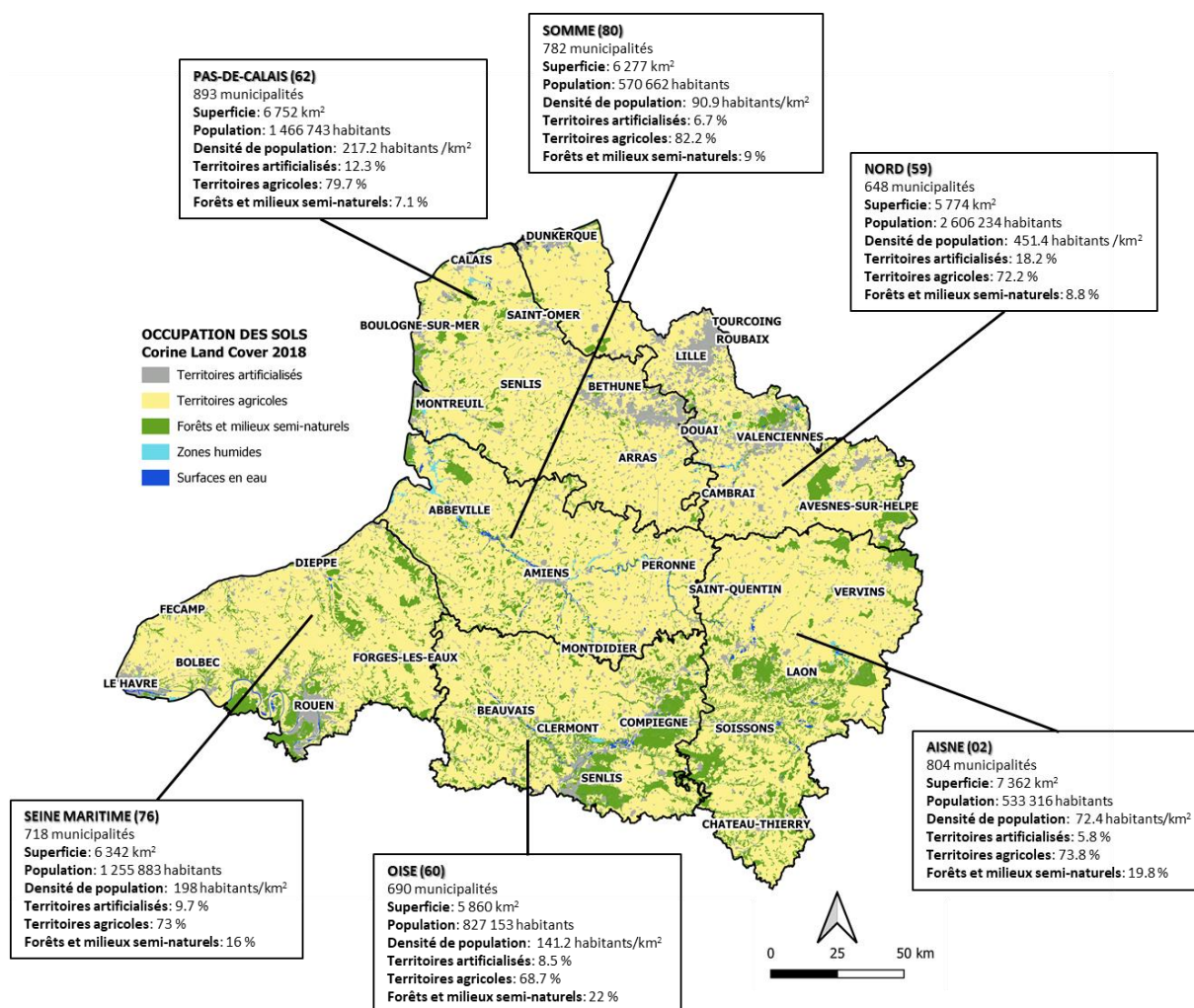


Figure 4 : Occupation des sols et caractéristiques démographiques de la zone d'étude (Corine Land Cover, 2018 ; INSEE, 2018)

b. Outils d'analyse mobilisés pour générer des indicateurs spatialisés

Les données brutes issues des bases de données environnementales ouvertes se présentent sous une diversité de formats. Certaines sont disponibles sous la forme de tables (bien souvent au format excel). D'autres sont mises à disposition sous la forme de données spatiales. Dans ce cas, on distingue deux formats de données principaux : « raster » et « shapefile » (ou « fichier de forme »). Le format raster représente l'information spatiale sous la forme d'une grille régulière de pixels. Chaque pixel contient une valeur qui peut représenter une caractéristique spécifique de l'environnement, comme la concentration en un polluant. Cette représentation est particulièrement adaptée pour caractériser des phénomènes continus. Le format « shapefile » ou « fichier de forme » est un format de stockage de données vectorielles (ponctuelles, linéaires ou surfaciques). Il permet d'archiver l'emplacement, la forme et les attributs de ces entités géographiques.

Quel que soit le format d'origine, ces données brutes ont été traitées pour générer des indicateurs spatialisés à l'échelle des communes et ainsi constituer un jeu de données environnemental. Selon la base de données considérée, ce traitement a été plus ou moins complexe. Tous les détails sont présentés dans un chapitre dédié à la constitution de ce jeu de données environnemental (chapitre 2). Mais de manière générale, on distingue :

- les indicateurs qui n'ont nécessité aucune action particulière (les données brutes ont pu être incluses telles quelles dans la base de données) ;
- les indicateurs qui ont nécessité des transformations élémentaires. Par exemple, dans de nombreux cas, les données brutes ont été rapportées à la surface ou à la population communale. Cette étape est impérative pour être en mesure par la suite de comparer les communes entre elles ;
- les indicateurs pour lesquels un traitement plus complexe a été nécessaire. Selon le cas de figure, les manipulations ont été menées sous QGIS (QGIS Development Team, 2021) ou à l'aide du logiciel R (R Core Team, 2021).

c. Gestion des données manquantes

Certains indicateurs intégrés dans le jeu de données présentent des données manquantes. Il est indispensable de mettre en place un processus spécifique pour y remédier car conserver ces lacunes peut entraîner une perte de précision et des biais dans la construction des indices. Ainsi, pour combler une donnée manquante dans une commune, la médiane des communes avoisinantes a été appliquée (Figure 5).

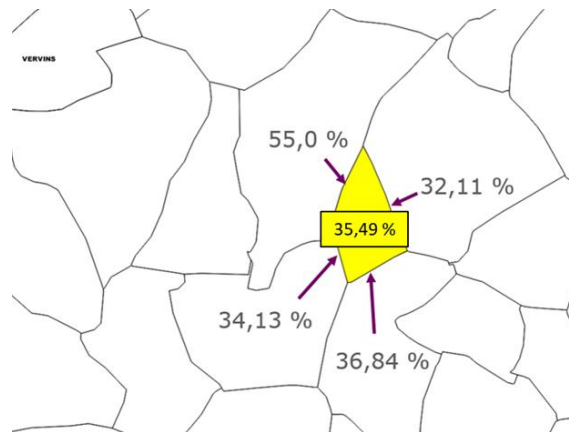


Figure 5 : Gestion des données manquantes (médiane des communes avoisinantes).

d. Critères d'inclusion et d'exclusion de nouveaux indicateurs dans le jeu de données

La première étape consiste à valider l'inclusion de chaque indicateur dans le jeu de données. Pour ce faire, s'appuyer sur des critères objectifs est indispensable. L'analyse de la littérature a permis de mettre en évidence une grande diversité de critères. Les critères d'inclusion finalement retenus sont les suivants :

- une donnée librement accessible, exhaustive et disponible à l'échelle nationale ;
- une donnée disponible ou calculable à l'échelle spatiale étudiée (commune) ;
- une donnée porteuse d'une information indépendante et non redondante par rapport au set déjà intégré ;
- une donnée suivie par le producteur ;
- une donnée adaptée au contexte territorial.

Les critères d'exclusion sont les suivants :

- une donnée décrivant des comportements individuels ;
- une donnée sans lien avec les problématiques de santé-environnement.

2. Construction des indices de vulnérabilité et de résilience.

La démarche générale de calcul des indices repose sur une analyse en composantes principales (ACP). En effet, deux ACP distinctes sont mises en œuvre sur deux pools d'indicateurs : un premier décrivant la vulnérabilité des communes et un second décrivant leur résilience. Plus largement, la construction des indices est composée de trois étapes : (i) préparation des données d'entrée, (ii) calcul des indices, (iii) discrétisation (Figure 6)

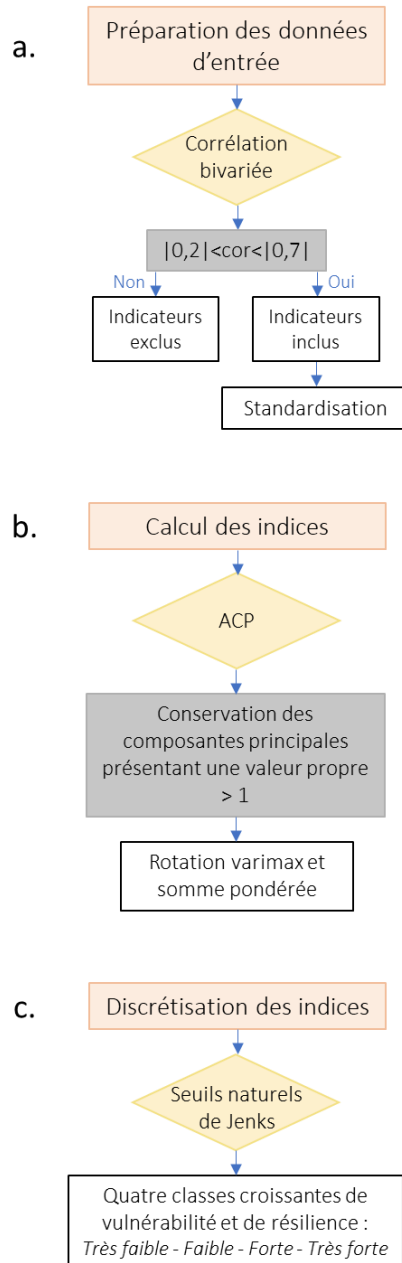


Figure 6 : Schéma général de la méthode de construction des indices de vulnérabilité et de résilience (schéma repris de la publication : Brousmiche et al., 2020a)

a. Préparation des données d'entrée.

i. *Réduction du jeu de données : structures de corrélation entre indicateurs.*

Une première étape consiste à faire une sélection dans le jeu de données initial en examinant les structures de corrélations entre indicateurs. L'objectif est d'éliminer les indicateurs trop fortement corrélés entre eux afin de limiter d'une part la redondance d'information et d'autre part l'effet miroir (corrélation inverse entre IR et IV). Un seuil de corrélation maximal (coefficient de corrélation de Pearson) est fixé à 0.7 (Dormann et al., 2013a; U.S. Environmental Protection Agency, 2020). Ensuite, une seconde étape de sélection est réalisée afin d'éliminer les indicateurs trop peu corrélés au reste du jeu de données. Pour cela, les indicateurs qui ne présentent aucune corrélation supérieure à 0,20 sont éliminés du jeu de données. Cette étape permet de réduire le « bruit de fond » et évite ainsi de générer des composantes principales corrélées à un seul et unique indicateur.

Ainsi, au cours de cette sélection basée sur les liens de corrélation, des choix ont donc dû être effectués entre indicateurs. Ces décisions ont été réalisées en tenant compte de la pertinence des indicateurs vis-à-vis de la problématique, de la fiabilité de la donnée source et de la représentativité globale du jeu de données.

Chacune de ces décisions a été détaillée et justifiée ci-dessous :

- i. La première étape consiste à trier les indicateurs corrélés un à un et très proches en termes de signification. Autrement dit, cela revient à faire des choix qui n'ont que peu de conséquences sur la représentativité globale du jeu de données. Ce cas de figure s'est présenté pour les couples d'indicateurs suivants (l'indicateur représenté en gras a été conservé dans le jeu de données) :
 - **APL_MG_2013** (accessibilité potentielle localisée aux médecins généralistes) et APL_PHAR_2013 (accessibilité potentielle localisée aux pharmaciens).
 - **P_FI_HAB** (potentiel financier par habitant) et POT_FISC_H (potentiel fiscal par habitant).
 - **SUBV_HAB_14** (montant des subventions versées par la commune par habitant) et P_SUBV_CH_FONC (part des subventions versées par la commune parmi les charges de fonctionnement).
- ii. Ensuite, étant donné le consensus scientifique sur l'utilisation du « French Deprivation index » (FDep - Rey et al., 2009), en particulier à l'échelle communale et dans le contexte français, cet indicateur a été conservé. En revanche, les indicateurs TX_CHOM (taux de chômage parmi les 15-64 ans) et PEU_DIPLO_1564 (part des peu diplômés parmi les 15-64 ans) ont été éliminés.
- iii. Les indicateurs portant sur la thématique du logement étaient fortement corrélés les uns aux autres : P_PROP (part des propriétaires dans les résidences principales), P_LOCA (part des locataires dans les résidences principales), P_LOCA_HLM (part des locataires en HLM dans les résidences principales) et HLM (nombre de résidences principales HLM pour 1000 habitants). Le choix s'est porté sur P_LOCA qui recouvre en grande partie l'information contenue dans P_LOCA_HLM et HLM.
- iv. Les trois indicateurs de pollution atmosphérique (NO₂, PM₁₀ et PM_{2,5}) étaient mutuellement corrélés. Au regard des connaissances scientifiques actuelles, il semblait plus pertinent d'étudier PM_{2,5} car les sources d'émission sont diverses (chauffage urbain, trafic routier, industrie, etc...). De plus, l'impact sanitaire des PM_{2,5} est particulièrement préoccupant, du fait de leur composition chimique (hydrocarbures, métaux, etc..) et de leur pénétration dans

l'organisme (pénétration profonde dans l'arbre respiratoire, dans le système digestif par déglutition, et passage de la barrière biologique) (Nan et al., 2023). Enfin d'un point de vue purement mathématique, $PM_{2,5}$ est celui des trois qui était le plus fortement corrélé aux deux autres.

- v. L'indicateur P_ARTIF (proportion de territoires artificialisés) était corrélé à Densite_pop (densité de population) et INTERSEC (densité d'intersections routières). Or la densité de population était décrite dans le jeu de données via un autre indicateur GRIL_COM_DENS (grille communale de densité). Ensuite, INTERSEC semblait assez pertinent car les intersections routières présentent un potentiel d'émission important dû aux ralentissements et aux accélérations des véhicules (Soulhac et al., 2009). Cependant, la pollution atmosphérique était déjà décrite via l'indicateur $PM_{2,5}$ qui est un marqueur plus direct du phénomène. Ainsi afin de conserver un maximum d'information, l'indicateur P_ARTIF a été conservé au détriment de Densite_pop et INTERSEC.
- vi. Enfin, les indicateurs trop faiblement corrélés au reste du jeu de données ont également été exclus. Il s'agit des cinq indicateurs suivants : P_NAISS_DOM_HAB_14 (part des naissances domiciliées au domicile de la mère par habitant), LNG_RT_H (longueur des routes primaires par habitant), EV_EMPL_PR (évolution de l'emploi pour les activités présentiels), TX_CR_ENT (taux de création d'entreprises), P_HUMID (proportion de zones humides).

A l'issue de cette sélection bivariée, 17 indicateurs ont été exclus et le set de données final comporte 44 indicateurs.

ii. Répartition des indicateurs entre vulnérabilité et résilience.

Avant d'engager le calcul des deux indices composites, les 44 indicateurs ont été répartis entre vulnérabilité (17 indicateurs) et résilience (27 indicateurs) pour constituer deux ensembles distincts (Tableau 1). Si une augmentation de l'indicateur était associée à impact bénéfique sur la santé, l'indicateur a été placé en résilience. Si une augmentation de l'indicateur pouvait être associée à un effet délétère ou négatif sur la santé, l'indicateur est placé en vulnérabilité.

Tableau 1 : Pools d'indicateurs retenus pour le calcul des indices de résilience et de vulnérabilité.

Indice de résilience (27 variables)			Indice de vulnérabilité (17 variables)		
SANTÉ	APL_MG_2013	Accessibilité potentielle localisée aux médecins généralistes	SANTÉ	P_DECES_DOM_HAB_14	Part des décès domiciliés par habitant
	P_SANTÉ_1000H	Nombre de professionnels de santé pour 1000 habitants, toute spécialité confondue	POLITIQUE	ANN_DET_H	Annuité de la dette de la commune, par habitant
	SHAN_MED	Indice de Shannon-Weaver pour les professions médicales et paramédicales	SOCIAL	P_PERS_plus65	Part de la population de plus de 65 ans dans la population totale
POLITIQUE	AGENDA21	Part de la population couverte par un agenda 21 communal ou infra-départemental		P_IMM_15	Part des immigrés dans la population totale
	R_ENS_HAB	Résultat d'ensemble comptable de la commune, par habitant		GRIL_COM_DENS	Grille communale de densité
	DEP_EQUIP_HAB_14	Dépenses pour les équipements de la commune, par habitant		P_MONOP	Part des familles monoparentales parmi les ménages
	P_EQUIP_INVEST	Part des dépenses pour les équipements parmi le total des dépenses d'investissements		P_COUV_AL	Part de la population couverte par une prestation de la branche famille
	SUBV_HAB_14	Montant des subventions versées par la commune, par habitant		FDEP_17	Indice de défaveur sociale Fdép
	PROD_TAX_HAB	Produit de la taxe d'habitation rapporté au nombre de résidences principales et de résidences secondaires		SERVICES	P_LOCA
	P_M_IMPOT	Part des ménages fiscaux imposés	ECONOMIE	T_PREC	Taux de précarité
P_FI_HAB	Potentiel financier par habitant	ENVIRONNEMENT	Emis_GES	Emission des gaz à effet de serre hors puits (PRG)	
AFR	Zones éligibles au zonage d'aide à finalité régionale		BASIAS	Densité de sites BASIAS par commune	
SOCIAL	TXSCO_1824		Taux de poursuite d'études des 18-24 ans	BASOL	Densité de sites BASOL par commune
	EMP_ED_FORM_14		Emploi dans l'éducation et la formation pour 1000 habitants	ICPE	Densité de sites ICPE par commune
	NB_EC_PUB_1000H		Nombre d'écoles publiques pour 1000 habitants	P_ARTIF	Proportion de territoires artificialisés
SERVICES	EQUIP_15		Part de la population ayant accès en moyenne aux 12 équipements de la gamme intermédiaire en 15 minutes ou moins	PM2,5_MEAN	Moyenne interannuelle (2011-2017) de la pollution atmosphérique en PM2,5
	COUV_4G		Part de surface couverte en 4G par l'opérateur mieux disant	ARTIF_0919	Pourcentage de la surface communale convertie en surface artificialisée entre 2009 et 2019
ECONOMIE	V_EMP_0813		Evolution de la population active de 15 à 64 ans		
	EV_POP_ACT_0914		Evolution de la part des actifs occupés de 15 ans et plus parmi la population totale		
	TX_EMP		Taux d'emploi des 15-64 ans		
	REV_MED_UC_12	Revenu médian par unité de consommation			
	P_AC_OCC_MMCOM	Part des actifs occupés de 15 ans et plus vivant et travaillant dans la même commune			
	SHANNON_CSP	Indice de Shannon-Weaver de représentation des catégories socio-professionnelles			
ENVIRONNEMENT	AC_EV	Part de la population communale ayant accès à un espace vert dans un rayon de 400 mètres			
	H_ESV	Monétarisation par habitant des services écosystémiques			
	NAT_HAB	m ² d'espaces verts ou espaces naturels par habitant			
	P_FOR/NAT	Proportion de forêts et milieux semi-naturels			

iii. Standardisation des indicateurs.

La question de l'interopérabilité des données s'est posée puisque les indicateurs sont exprimés dans des unités distinctes et prennent des gammes de valeurs très différentes. Pour les intégrer au sein d'un même indice, il est nécessaire de passer au préalable par une étape de standardisation (OECD et al., 2008b). Sans standardisation, l'importance donnée à chacun des indicateurs est extrêmement variable et la qualité de l'indice final est impactée. Par exemple, si un premier indicateur est exprimé en m² et un second en km², le premier est susceptible de présenter une variance beaucoup plus importante et donc de peser davantage dans le calcul de l'indice final. Cette opération permet également de garantir l'interopérabilité entre les indicateurs exprimés dans des dimensions différentes (ppm et pourcentage par exemple).

La méthodologie retenue est celle du z-score qui permet de centrer et réduire les indicateurs. Après transformation, ils présentent une moyenne de 0 et un écart type de 1. La formule est la suivante :

$$z_i = \frac{x_i - \bar{x}}{\sigma}$$

z_i : valeur centrée réduite de la donnée i

x_i : valeur initiale de la donnée i

\bar{x} : moyenne de l'indicateur

σ : écart-type de l'indicateur

Une fois standardisés, les indicateurs sont comparables et l'étape d'agrégation peut être entreprise.

b. Méthode de pondération et de calcul des indices.

L'objectif est de synthétiser dans une analyse multifactorielle l'ensemble des données en faisant ressortir les éventuels liens entre indicateurs. Une ACP est réalisée pour l'indice de vulnérabilité, une autre pour l'indice de résilience. Les composantes principales de l'ACP sont conservées lorsque leur valeur propre est supérieure à 1. Enfin, une rotation des axes factoriels est réalisée par le biais d'une rotation VARIMAX pour accroître la lisibilité et faciliter l'interprétation (Kaiser, 1958).

Les deux indices peuvent ensuite être calculés :

Indice (résilience ou vulnérabilité)

$$= \sum_1^n (\text{proportion de variance expliquée par une composante principale} * \text{score de cette composante})$$

Avec n le nombre de composantes présentant une valeur propre supérieure à 1.

Pour finir les deux indices obtenus sont standardisés grâce à la méthode min-max qui permet de placer toutes les valeurs sur un intervalle compris entre 0 et 1, selon la formule suivante :

$$z_i = \frac{x_i - \max(x)}{\max(x) - \min(x)}$$

z_i : valeur standardisée de la donnée i

x_i : valeur initiale de la donnée i

$\min(x)$: valeur minimale pour l'indicateur x

$\max(x)$: valeur maximale pour l'indicateur x

c. Discrétisation des indices et représentation cartographique.

L'objectif est de mettre en évidence les disparités spatiales de vulnérabilité et de résilience des communes. L'étape de cartographie doit donc conserver au mieux l'information contenue dans les indices, tout en la rendant accessible. L'agrégation des données en classes (discrétisation) est alors nécessaire pour améliorer la lisibilité des cartes. La méthodologie retenue est celle de Jenks, cette méthode permet de minimiser la variance intra-classe tout en maximisant la variance inter-classe (Jenks, 1967). Ainsi, quatre classes croissantes de résilience et quatre classes croissantes de vulnérabilité ont été retenues pour optimiser la clarté de représentation des indices. Elles ont été

intitulées de la manière suivante : très faible résilience/vulnérabilité, faible résilience/vulnérabilité, forte résilience/vulnérabilité, très forte résilience/vulnérabilité. Il est important de préciser que cette qualification est relative. Par exemple, une commune est considérée comme fortement vulnérable par rapport à l'ensemble des communes de la zone étudiée. Les cartographies ont été réalisées à l'aide du logiciel libre QGIS (QGIS Development Team, 2021).

d. Analyse conjointe des indices.

L'analyse conjointe de la vulnérabilité et de la résilience a permis de définir quatre profils de communes : « Résilient », « A surveiller », « Fragile avec des ressources », « Fragile ». Ces profils ont été définis à partir des seuils de Jenks utilisés précédemment pour discrétiser les indices (Tableau 2).

Tableau 2: Définition de quatre profils de communes lors de l'analyse conjointe des indices.

		Résilience (IR)			
		Très faible	Faible	Forte	Très forte
(IV) Vulnérabilité	Très faible	A surveiller		Résilient	
	Faible				
	Forte	Fragile		Fragile avec des ressources	
	Très forte				

3. Mise en relation avec les variations spatiales d'incidence de la MC.

a. Comparaisons entre clusters spatiaux d'incidence de la MC.

Les niveaux de vulnérabilité et de résilience ont été comparés entre les clusters spatiaux d'incidence de la MC. Préalablement, la distribution normale des indices de vulnérabilité et de résilience a été évaluée à l'aide du test de Lilliefors (Lilliefors, 1967). L'hypothèse de normalité ne s'est vérifiée dans aucun des cas examinés. Par conséquent, des tests de Kruskal-Wallis ont été réalisés afin de déterminer s'il existait des différences significatives entre les clusters d'incidence ($p < 0,05$) (McKight and Najab, 2010). En cas de significativité du test de Kruskal-Wallis, des tests post-hoc de Dunn avec ajustement de Bonferroni ont été effectués pour examiner les différences entre chaque paire de clusters (Dinno, 2015).

b. Association avec les SIRs de la MC.

En plus de la comparaison entre clusters spatiaux, les indices de vulnérabilité et de résilience ont également été mis en relation avec les SIRs lissés de la MC. Tout d'abord, des tests de corrélation de Spearman ont été utilisés pour évaluer la dépendance entre ces variables. Ensuite, la distribution des

SIRs a été comparée entre les quatre classes croissantes de vulnérabilité/résilience et entre les quatre profils de communes identifiés lors de l'analyse conjointe. L'hypothèse de normalité a été exclue dans chacun des cas examinés suite à la mise en œuvre de tests de Lilliefors. Ainsi, de nouveau, les comparaisons ont été menées à l'aide du test de Kruskal-Wallis, suivi de tests post-hoc de Dunn avec ajustement de Bonferroni.

Résultats

1. Une vision intégrée des territoires : économie, environnement physique, politique, santé, services, social.

a. Vue d'ensemble sur le jeu de données.

Au total, 61 indicateurs spatialisés ont été générés à partir de 17 sources de données distinctes. Ces indicateurs ont été répartis comme suit au sein des six dimensions décrites : Economie (10 indicateurs), Environnement physique (15 indicateurs), Social (11 indicateurs), Santé (6 indicateurs), Services (7 indicateurs), Politique (12 indicateurs). Les principaux phénomènes abordés dans chaque dimension ont été détaillés en Figure 7. A noter que parmi ces 61 indicateurs, 17 ont été exclus lors de l'analyse bivariée et 44 ont été retenus pour le calcul des indices de vulnérabilité et résilience.

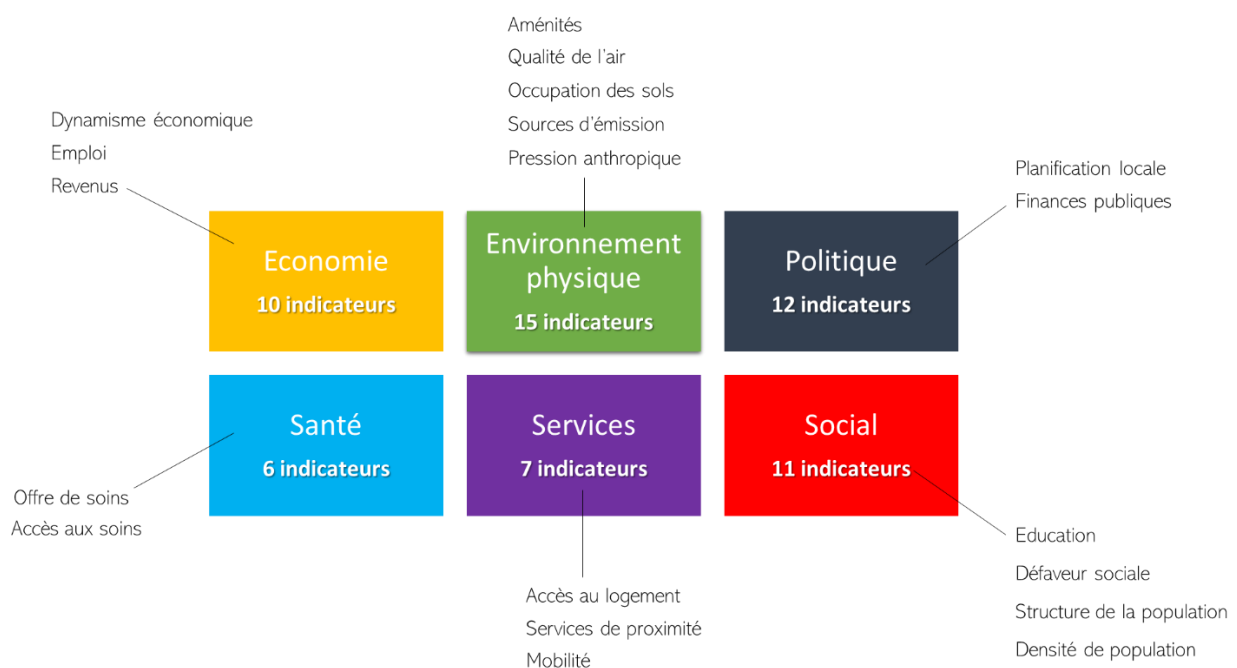


Figure 7 : Structure du jeu de données et principaux phénomènes abordés.

Au sein du jeu de données, on distingue :

- 19 indicateurs utilisés *in extenso* sans manipulation particulière des données brutes.
- 18 indicateurs pour lesquels une simple étape de normalisation a été réalisée. Dans ce cas, la grandeur a été rapportée à la population ou à la surface communale. Cette étape est primordiale pour être en mesure de comparer les communes les unes aux autres. Normaliser

sur la population communale permet d'exprimer une grandeur en pourcentage (par exemple : part de la population âgée de plus de 65 ans). Elle peut aussi permettre de caractériser l'offre de la commune sur le plan éducatif par exemple (nombre d'écoles publiques pour 1000 habitants). La normalisation sur la surface communale n'a été utilisée que pour des indicateurs qui appartiennent à la dimension Environnement physique. Soit pour exprimer une grandeur en pourcentage (par exemple : part de la surface communale couverte par des forêts et milieux semi-naturels). Soit pour caractériser la densité de sources d'émissions potentielles (par exemple : nombre d'intersections routières par unité de surface communale).

- 24 indicateurs pour lesquels un traitement spécifique des données a été nécessaire.

La description détaillée de chacun de ces indicateurs est présentée en annexe (ANNEXE 1). En plus de ces informations descriptives, la médiane, l'écart interquartile et le pourcentage de données manquantes sont également présentés pour chaque indicateur. Au maximum, le pourcentage de données manquantes s'élevait à 1.85% (soit 84 communes manquantes sur 4 535). Une cartothèque spécifique a été créée pour présenter ces 61 indicateurs spatialisés. Celle-ci est disponible sur demande mais elle n'a pas été intégrée directement au manuscrit.

Par rapport au jeu de données initial développé dans les Hauts-de-France, la dimension 'Environnement physique' compte 10 indicateurs de plus. Dans les cinq autres dimensions, un seul indicateur a été ajouté. Il s'agit d'un indice de défaveur sociale, il est présenté en détail ci-dessous.

b. Focus sur l'indice de défaveur sociale : FDep.

Pour évaluer le niveau de défaveur sociale, un indicateur a été calculé. Il s'agit du « French social Deprivation index » (FDep) décrit par Rey et al. (2009). Cet indicateur a été conçu spécifiquement pour le contexte français et il est adapté aux échelles locales (iris ou commune). De plus, il est considéré comme étant moins urbano-centré que d'autres indicateurs comme l'indice de Townsend (Townsend, 1987). Le FDep est un indice composite calculé à partir de quatre sous-indicateurs :

- Taux de chômage dans la population active des 15-64 ans (2017, Observatoire des territoires).
- Taux d'ouvriers dans la population active de 15 à 64 ans (2017, Insee base Emploi-Population active).
- Revenu médian par unité de consommation (2017, Observatoire des territoires)
- Taux de bacheliers dans la population de plus de 15 ans sortie du système scolaire (2017, Insee base Diplômes-Formation)

Ces sous-indicateurs ont d'abord été normalisés (centrés réduits) et une ACP a été appliquée. Le FDep correspond aux coordonnées des communes sur la 1^{ère} composante de l'ACP. Plus la valeur du FDep est grande, plus le niveau de défaveur sociale est élevé. Pour l'année 2017 et sur le périmètre de la zone d'étude, cette 1^{ère} composante explique 63,4% de la variabilité globale. En comparaison, elle expliquait 68% de la variabilité totale dans la publication originale (Rey et al., 2009). Le résultat est représenté cartographiquement en Figure 8. A noter que le score a été centré réduit comme il est d'usage de le faire.

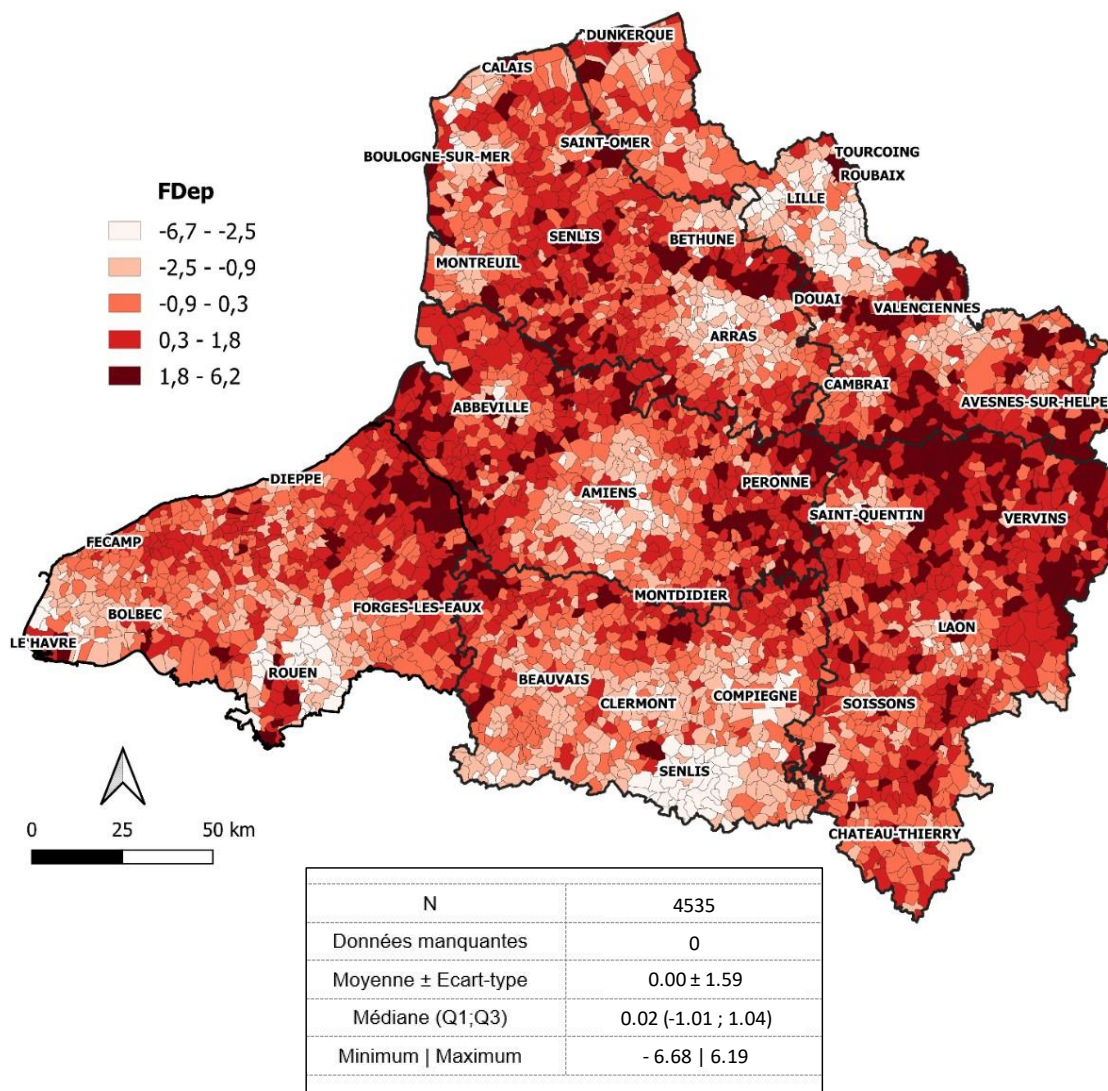


Figure 8 : Indice de défaveur sociale FDep (2017)

Des disparités spatiales importantes sont observées dans la zone d'étude. Le niveau de défaveur sociale est le plus faible en périphérie des principaux centres urbains. Cela concerne notamment l'agglomération lilloise, arrageoise, amiénoise, rouennaise et dans une moindre mesure havraise. Dans le sud de l'Oise, notamment en raison de la proximité avec l'Île de France, la défaveur sociale est également faible. Au contraire, le bassin minier est facilement identifiable car marqué par une défaveur sociale importante (communes situées entre Béthune et Valenciennes). Des zones de forte défaveur sociale sont également localisées dans des espaces ruraux moins densément peuplés : en particulier dans le Nord de l'Aisne, autour de Péronne ou encore de part et d'autre de la frontière entre les départements de la Somme et de la Seine-Maritime.

2. Potentiels de vulnérabilité et de résilience des communes.

a. Potentiel de vulnérabilité.

Les valeurs prises par l'indice de vulnérabilité varient entre 0 et 1 sur la zone d'étude. Plus l'indice tend vers 1, plus la commune en question est considérée comme vulnérable. La moyenne sur l'ensemble de

la zone s'élève à 0,138 avec un écart interquartile de [0,087 ; 0,155]. Cela témoigne de fortes inégalités sur la zone d'étude avec des valeurs extrêmes très élevées (marquées en rouge foncé sur la carte), et les trois quarts des communes qui prennent des valeurs inférieures à 0.155. Les communes sont subdivisées en quatre classes de vulnérabilité croissante obtenues à partir des seuils de Jenks (Figure 9).

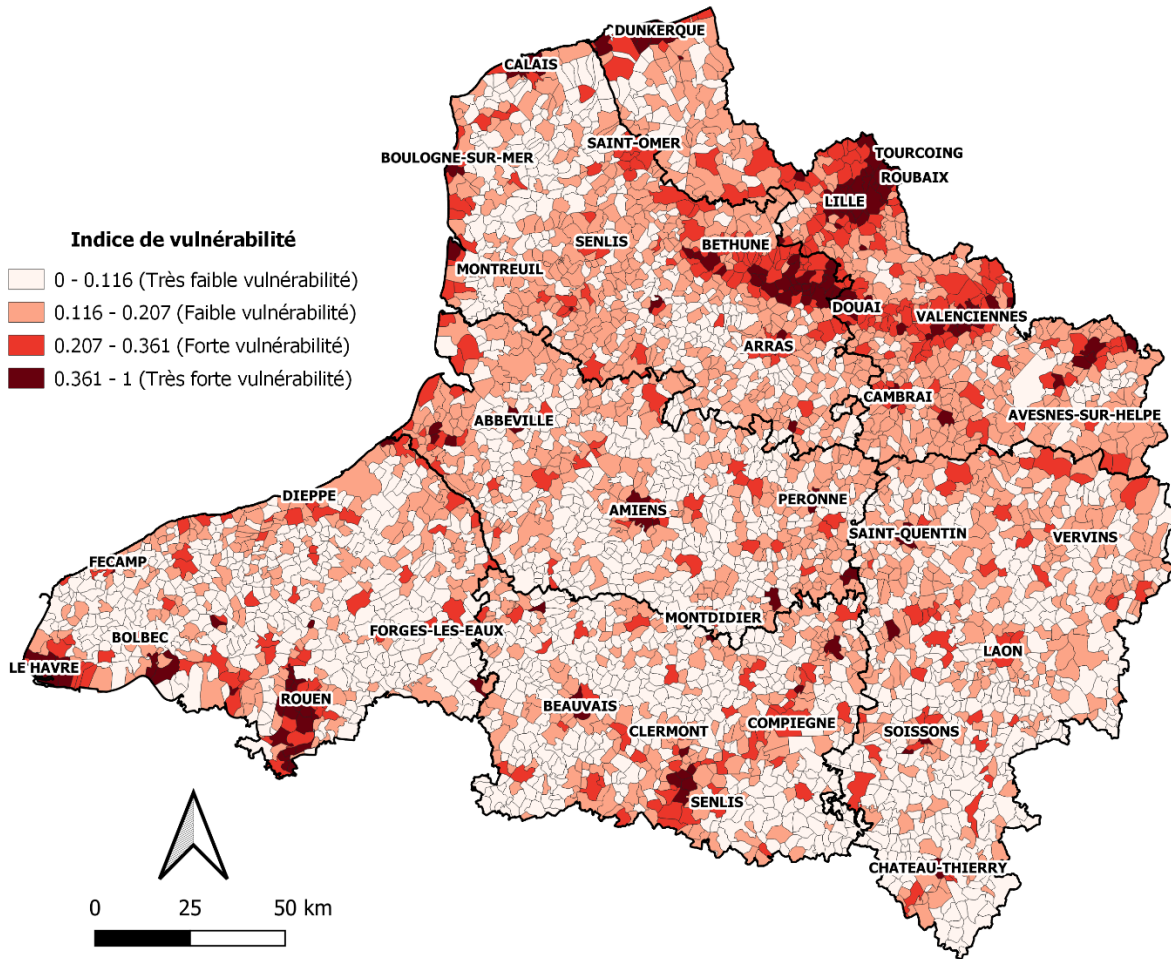


Figure 9 : Représentation cartographique de l'indice de vulnérabilité calculé à l'échelle de la commune

Ces classes ont été intitulées comme suit : très faible vulnérabilité (2 297 communes soit 50,7 %), faible vulnérabilité (1 646 communes soit 36,3 %), forte vulnérabilité (428 communes soit 9,4 %) et très forte vulnérabilité (164 communes soit 3,6 %). En premier lieu se dégage un groupe de communes particulièrement vulnérables qui se détachent nettement du reste de la zone étudiée. Ces communes forment une bande qui traverse les départements du Pas-de-Calais (62) et du Nord (59) d'ouest en est avec pour villes principales : Béthune, Lens, Douai ou encore Valenciennes. Ces communes appartiennent à l'ancien bassin minier dont la vulnérabilité importante est connue (Marot, 2019). Par ailleurs, les autres communes les plus vulnérables se répartissent principalement dans les grands centres urbains et leur périphérie (Le Havre, Rouen, Amiens, Arras, Lille, Dunkerque notamment). Au contraire, les communes très faiblement vulnérables se concentrent majoritairement à l'écart des centres urbains. La carte suggère même une différence nette entre départements : les communes très faiblement vulnérables semblent sous-représentées dans le Nord (59) et le Pas-de-Calais (62) et sur-représentées dans la Somme (80), la Seine-Maritime (76), l'Oise (60) et l'Aisne (02).

Ces différences se confirment de manière plus évidente sur le graphique de comparaison entre départements (Figure 10).

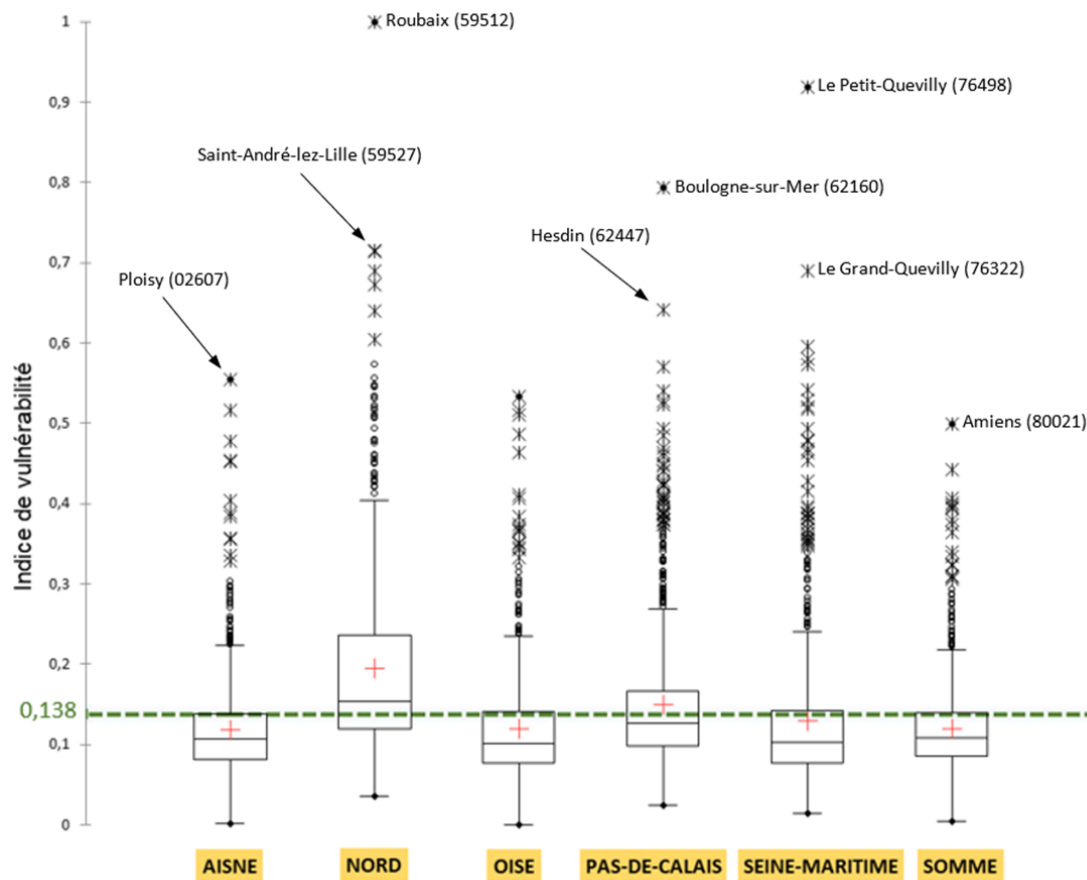


Figure 10 : Distribution de l'indice de vulnérabilité par département

Le département du Nord (59) et dans une moindre mesure celui du Pas-de-Calais (62) présentent une vulnérabilité accrue par rapport aux autres. Ce résultat est à mettre en lien avec les conclusions précédentes sur l'ancien bassin minier et les nombreuses aires urbaines comme Lille, Dunkerque, Calais, Boulogne-sur-Mer (ces deux départements sont d'ailleurs en moyenne les plus densément peuplés de la zone d'étude). La commune de Roubaix située dans l'agglomération lilloise est celle qui présente l'indice de vulnérabilité le plus élevé de toute la zone d'étude. Le département de Seine Maritime (76) ne présente pas de valeur moyenne aussi élevée que les deux départements précédents. Cependant il contient des communes particulièrement vulnérables notamment en périphérie de Rouen (Le Petit-Quevilly et Le Grand-Quevilly).

b. Potentiel de résilience.

Les valeurs prises par l'indice de résilience varient entre 0 et 1 sur la zone d'étude (Figure 11). Plus l'indice tend vers 1, plus la commune en question est considérée comme résiliente. La moyenne sur l'ensemble de la zone s'élève à 0,336 avec un écart interquartile de [0,292 ; 0,377]. L'écart interquartile est de nouveau réduit, suggérant qu'un certain nombre de communes présentent des valeurs extrêmes. Les communes sont subdivisées en quatre classes de résilience croissante obtenues à partir des seuils de Jenks.

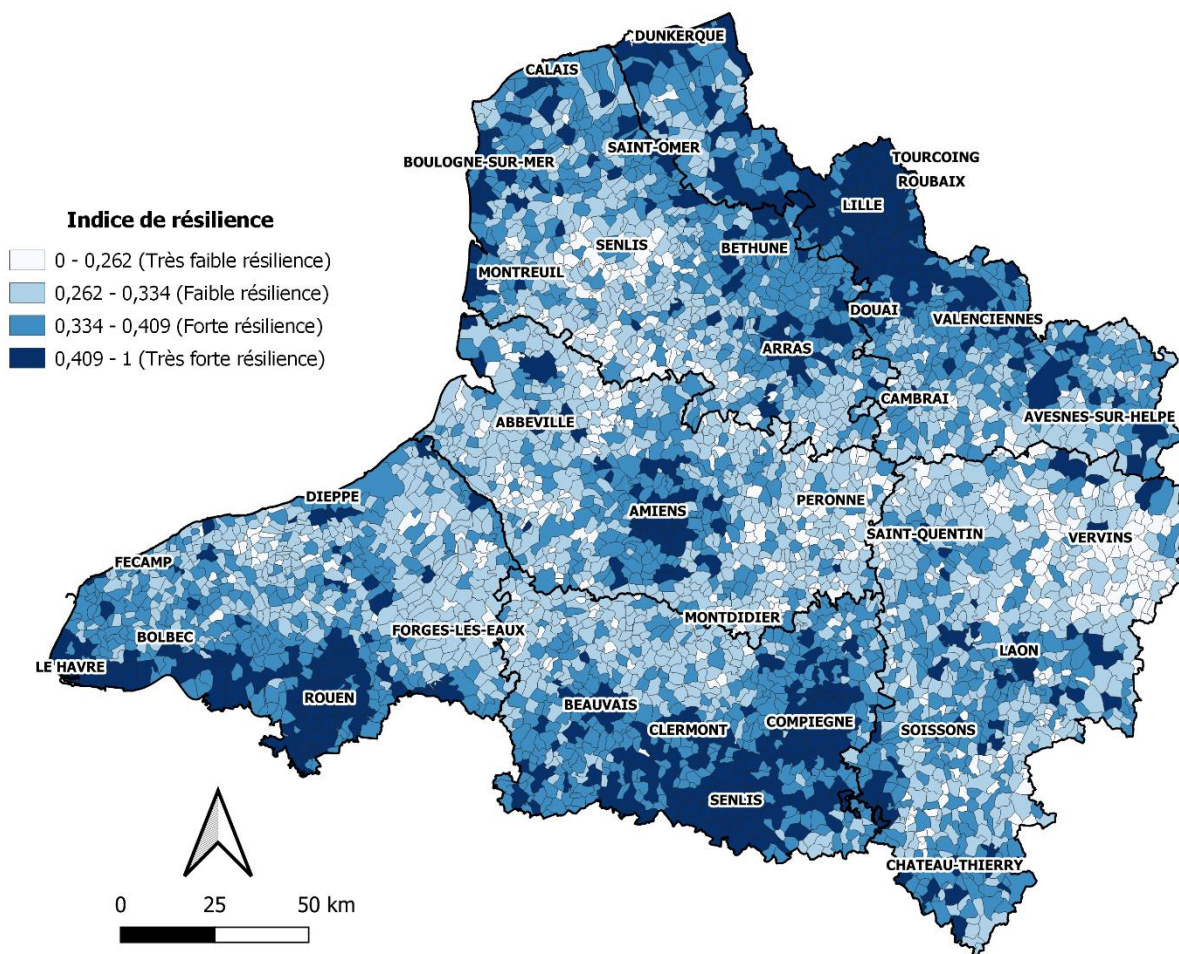


Figure 11 : Représentation cartographique de l'indice de résilience calculé à l'échelle de la commune

Les quatre classes de résilience ont été nommées de la manière suivante : très faible résilience (555 communes soit 12,2 %), faible résilience (1 771 communes soit 39,1 %), forte résilience (1 579 communes soit 34,8 %) et très forte résilience (630 communes soit 13,9 %). Les communes les plus résilientes semblent se répartir principalement autour des aires urbaines (Le Havre, Rouen, Senlis, Compiègne, Amiens, Arras, Lille, Dunkerque notamment). Au contraire, les zones moins densément peuplées semblent caractérisées par des indices de résilience plus faibles. En particulier, les départements de la Somme et de l'Aisne qui sont les moins densément peuplés, présentent globalement une résilience assez réduite.

Ces différences sont plus clairement illustrées sur le graphique de comparaison entre départements (Figure 12).

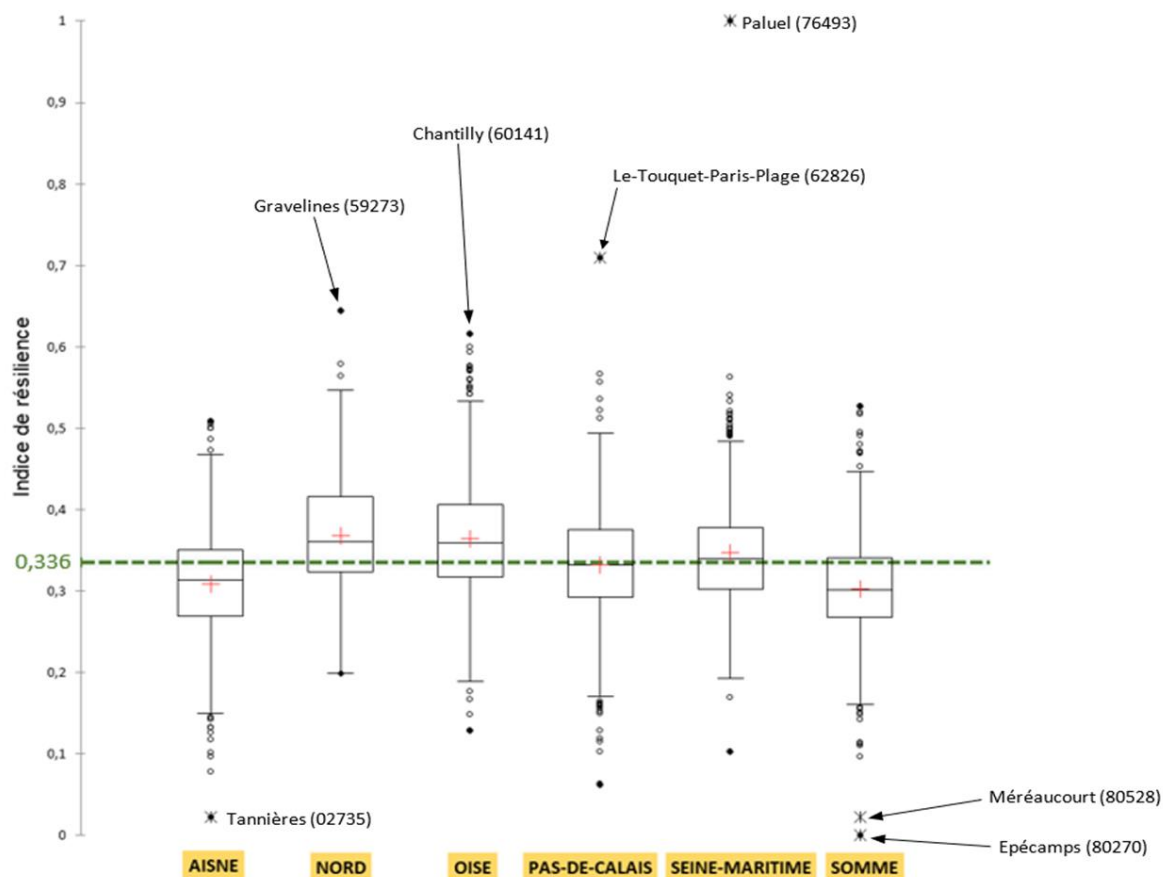


Figure 12 : Distribution de l'indice de résilience par département

Les départements du Nord (59) et de l'Oise (62) sont en moyenne plus résilients à l'inverse des départements de la Somme (80) et de l'Aisne (02). Les communes de Paluel (indice de résilience le plus élevé de la zone d'étude) et de Gravelines sont très résilientes et présentent toutes deux une particularité intéressante puisqu'elles abritent chacune une centrale nucléaire. Les communes de Chantilly et du Touquet-Paris-Plage sont respectivement les communes les plus résilientes du département de l'Oise (60) et du Pas-de-Calais (62) et sont reconnues comme abritant des populations aisées. A contrario, les communes apparemment les moins résilientes de la zone d'étude (par exemple Tannières, Méréaucourt et Epécamps) sont de très petites communes rurales qui ne présentent pas plus de vingt habitants.

c. Analyse conjointe des deux indices.

Les communes de la zone d'étude se répartissent de la manière suivante au sein des quatre profils : « Résilient » (1 690 communes soit 37,3 %), « A surveiller » (2 253 communes soit 49,7 %), « Fragile avec des ressources » (519 communes soit 11,4 %), « Fragile » (73 communes soit 1,6 %) (Figure 13).

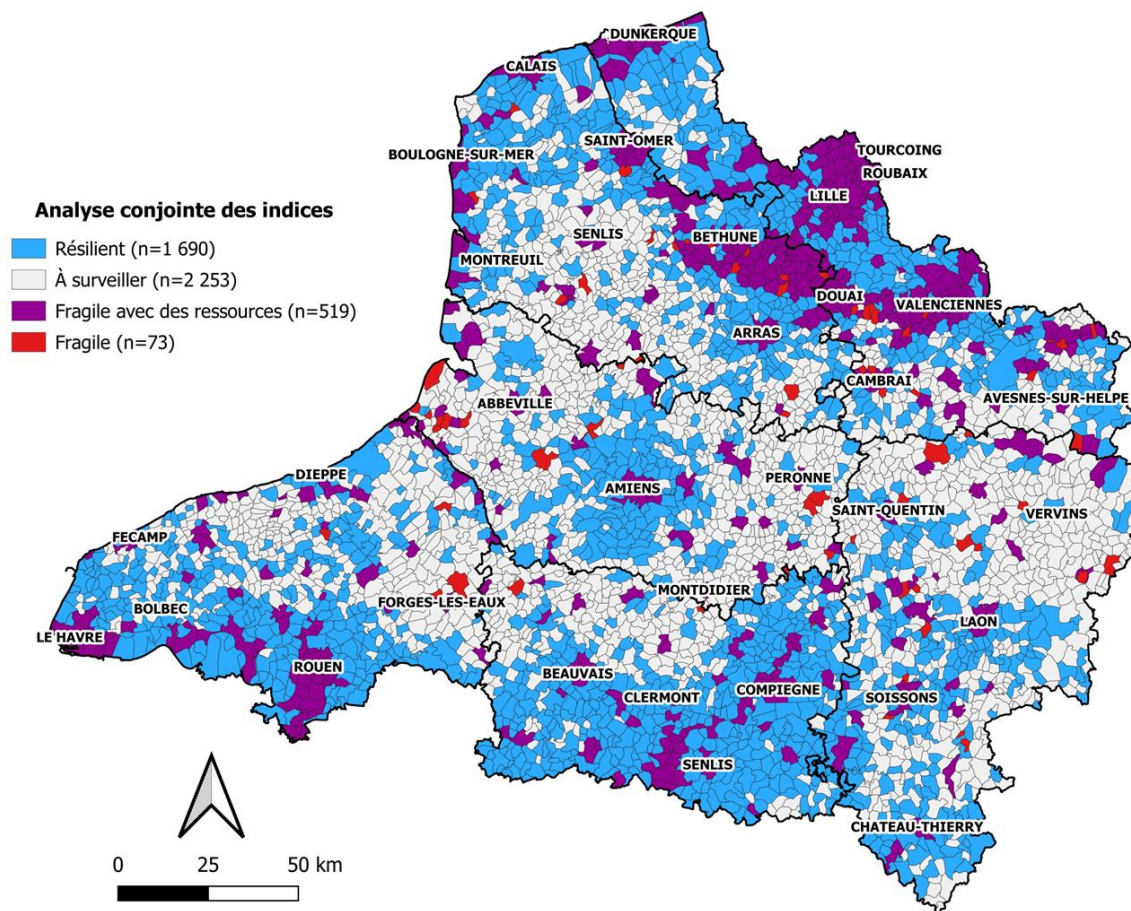


Figure 13 : Distribution spatiale de l'analyse conjointe réalisée pour l'indice de vulnérabilité et de résilience

Le profil « Fragile avec des ressources » se concentre quasi exclusivement dans les centres urbains et leur périphérie. Ces communes sont fortement agrégées spatialement et forment des noyaux qui correspondent principalement aux agglomérations urbaines de la zone d'étude. Par ailleurs, ces noyaux sont quasi systématiquement insérés dans une couronne de communes appartenant au profil « Résilient ». Ce profil « Résilient » se concentre donc essentiellement en grande périphérie des centres urbains. En outre, les communes rurales les plus en marge des centres urbains sont majoritairement classées dans le profil « A surveiller ». Pour finir, les communes « Fragiles » sont très peu agrégées spatialement (hormis quelques-unes dans le bassin minier). On distingue certaines communes sous influence urbaine, comme dans le bassin minier et autour de Cambrai. Les autres sont plutôt de profil rural et se répartissent de manière relativement homogène dans la zone d'étude.

3. Association avec les variations spatiales d'incidence de la MC.

a. Comparaison entre clusters spatiaux.

La distribution de l'indice de vulnérabilité a été comparée entre les huit clusters spatiaux d'incidence de la MC (Figure 14).

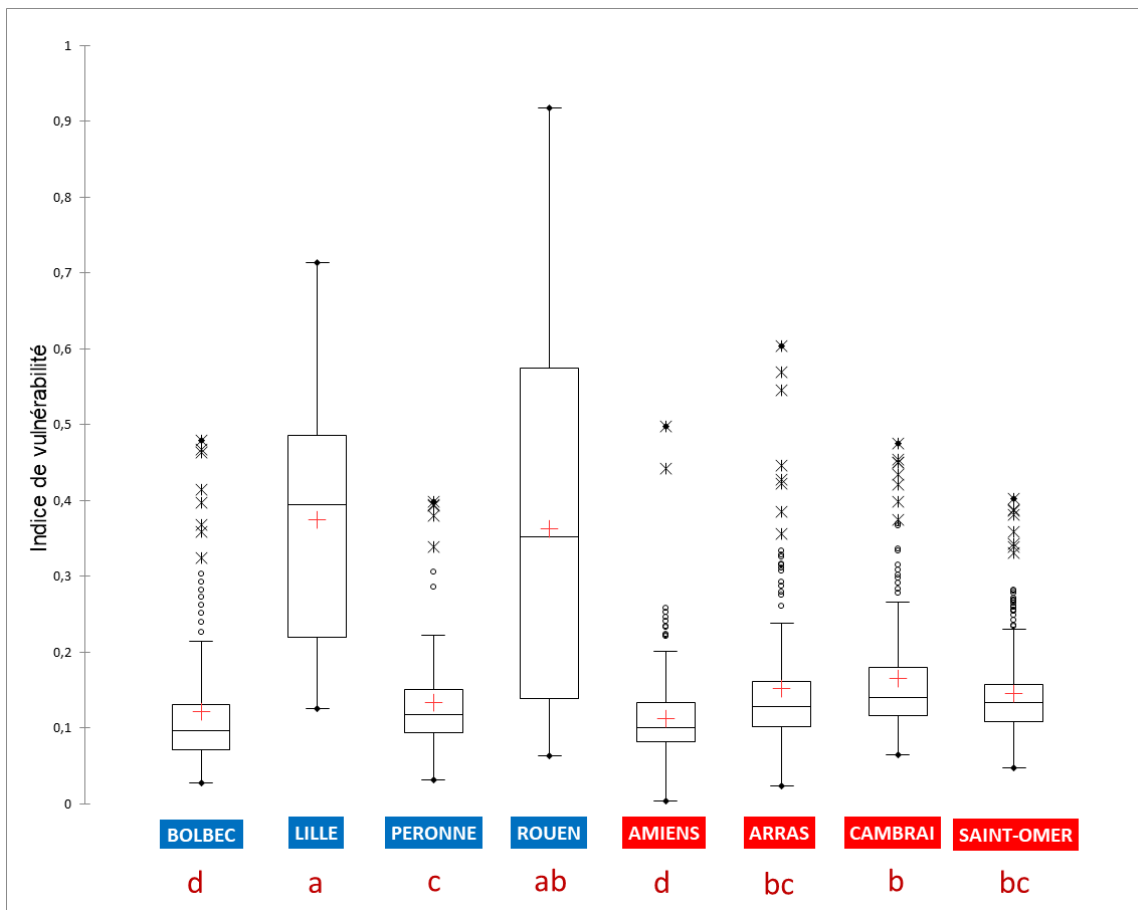


Figure 14: Distribution de l'indice de vulnérabilité au sein des clusters spatiaux de la MC (surincidence marquée en rouge, sous-incidence en bleu)

Si deux clusters partagent une lettre en commun alors ils ne sont pas significativement différents, au contraire si deux clusters ne partagent aucune lettre en commun alors ils sont significativement différents (Test de Kruskal-Wallis avec comparaisons multiples de Dunn)

Les clusters de Lille et de Rouen se distinguent très nettement des autres en termes de vulnérabilité. C'est d'autant plus vrai pour le cluster de Lille qui présente un niveau de vulnérabilité significativement plus élevé que tous les autres à l'exception de Rouen. Le cluster de Rouen est lui moins atypique puisqu'il est significativement différent des clusters d'Amiens, de Bolbec et de Péronne. Pour ces deux clusters, la vulnérabilité est globalement plus élevée mais la variabilité est également beaucoup plus importante. L'agglomération lilloise et rouennaise sont les deux plus grandes aires urbaines de toute la zone d'étude. Ainsi une des hypothèses possibles pour expliquer leurs spécificités en termes de vulnérabilité est la suivante : dans la plupart des grandes agglomérations urbaines, d'une commune à l'autre il peut y avoir des grandes disparités en termes de précarité, de niveau socio-économique, d'environnement physique. C'est ce qui pourrait expliquer cette variabilité très marquée. A noter par ailleurs que les clusters d'Amiens (sur-incidence) et de Bolbec (sous-incidence) présentent un niveau de vulnérabilité significativement inférieur à tous les autres clusters.

La distribution de l'IR a été comparée entre les huit clusters spatiaux d'incidence de la MC (Figure 15).

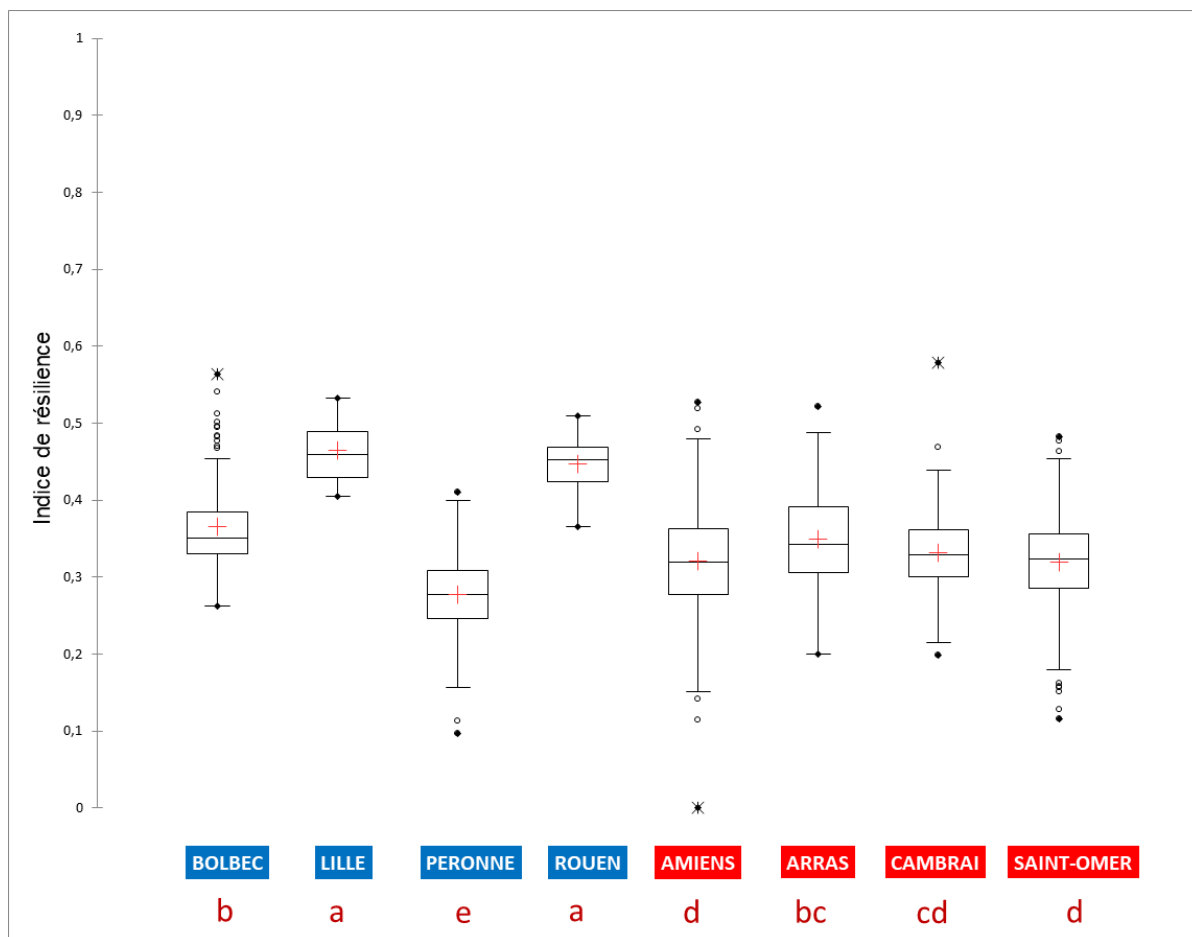


Figure 15 : Distribution de l'indice de résilience au sein des clusters spatiaux de la MC (surincidence marquée en rouge, sous-incidence en bleu)

Si deux clusters partagent une lettre en commun alors ils ne sont pas significativement différents, au contraire si deux clusters ne partagent aucune lettre en commun alors ils sont significativement différents (Test de Kruskal-Wallis avec comparaisons multiples de Dunn)

La comparaison de l'indice de résilience met en évidence des profils de résilience très différents selon les clusters. Les deux clusters de sous-incidence de Lille et de Rouen sont significativement différents de tous les autres clusters, ils présentent des niveaux de résilience plus élevés. Le cluster de sous-incidence de Péronne est lui aussi significativement différent de tous les autres mais au contraire il présente le niveau de résilience le plus faible. Les autres clusters présentent des niveaux de résilience intermédiaires. On peut également noter que parmi les clusters de sur-incidence, le cluster d'Arras présente un niveau de résilience significativement plus élevé que les clusters d'Amiens et de Saint-Omer. Ainsi de la même manière que pour l'indice de vulnérabilité, les deux clusters de sous-incidence de Lille et de Rouen (qui correspondent aux deux plus grandes aires urbaines de la zone d'étude) sont très atypiques par rapport aux autres clusters. L'une des hypothèses que l'on peut émettre est que les communautés d'agglomération (métropoles) ont un poids important en termes de politiques publiques.

b. Mise en lien avec les données d'incidence.

La distribution des SIRs de la MC a été comparée au sein des quatre classes croissantes de vulnérabilité ou de résilience (Figure 16).

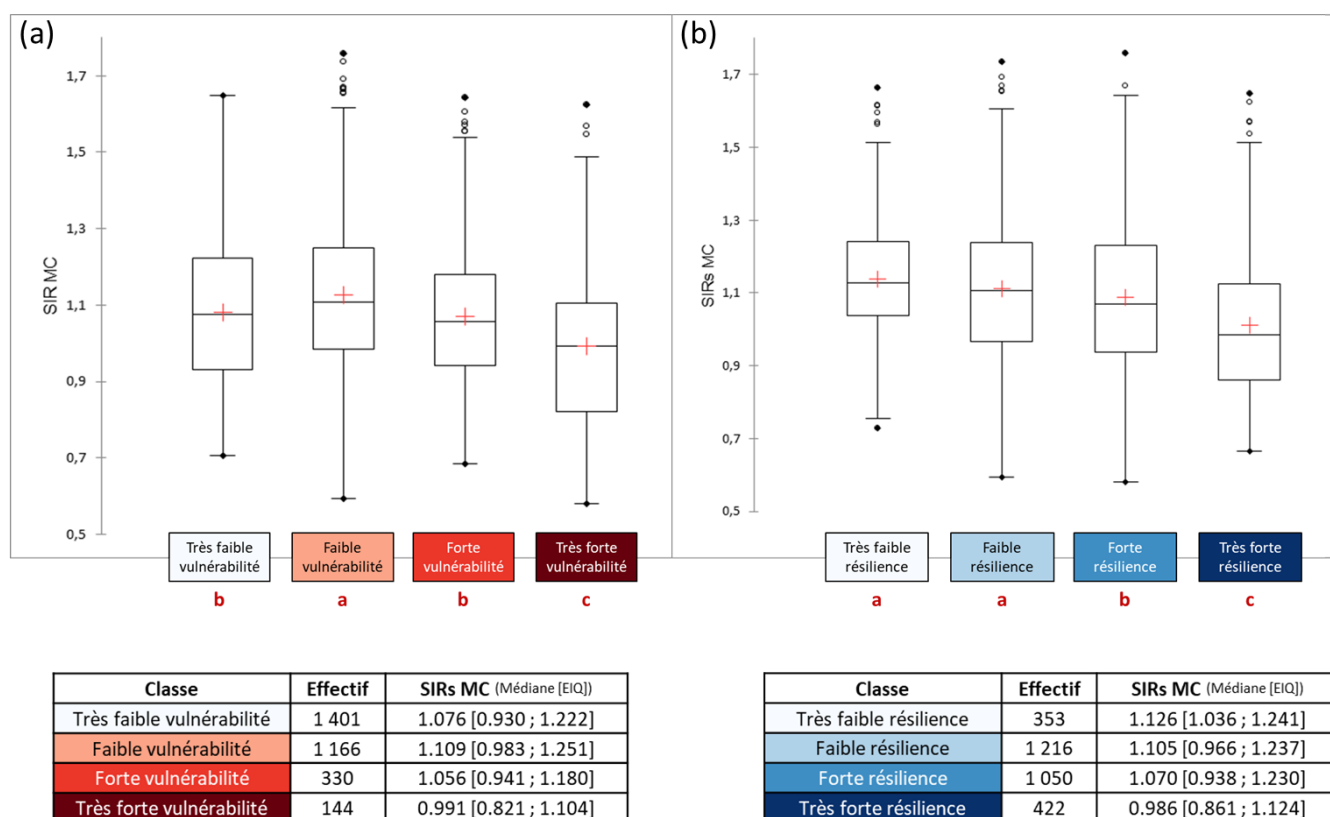


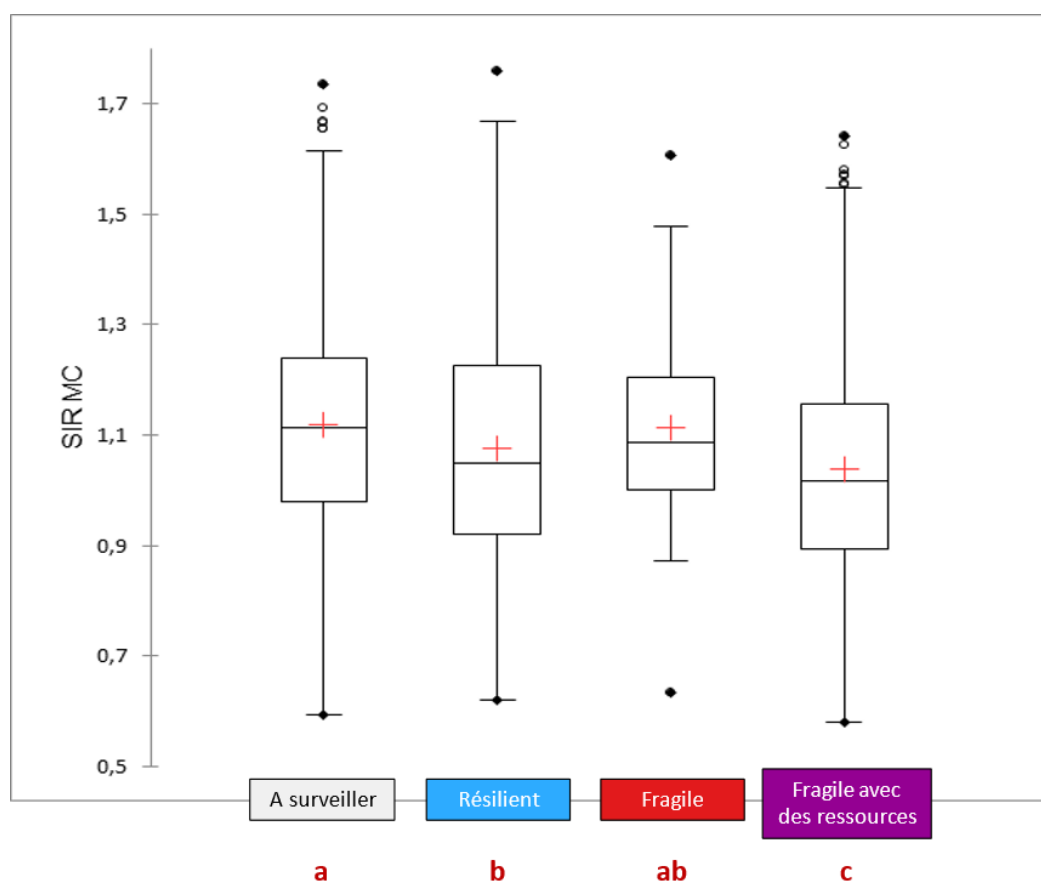
Figure 16 : Distribution des SIRs lissés de la MC par classes croissantes de vulnérabilité (a) ou de résilience (b).

Si deux classes partagent une lettre en commun alors elles ne sont pas significativement différentes, au contraire si deux classes ne partagent aucune lettre en commun alors elles sont significativement différentes (Test de Kruskal-Wallis avec comparaisons multiples de Dunn)

Pour la vulnérabilité, l'incidence de la MC est significativement plus élevée au sein des communes identifiées comme faiblement vulnérables. Au contraire, les communes les plus vulnérables (Très forte vulnérabilité) présentent le niveau d'incidence le plus faible. Les deux classes restantes (Très faible vulnérabilité & Forte vulnérabilité) présentent un niveau d'incidence intermédiaire. Ces résultats sont difficilement interprétables, il ne semble pas y avoir de relation particulière entre le niveau de vulnérabilité des territoires et la répartition spatiale d'incidence de la MC. De plus, le coefficient de corrélation de Spearman entre IV et SIRs s'élève à 0,005 (non significatif).

Pour la résilience, le niveau d'incidence diminue à mesure que la classe de résilience augmente. A l'exception des deux premières classes qui ne se distinguent pas, ces différences sont significatives. De plus, le coefficient de corrélation de Spearman entre IR et SIRs s'élève à -0,19. Ce lien de corrélation est faible mais il est significatif. De ce fait, ces résultats suggèrent une potentielle influence de la résilience des territoires sur la répartition spatiale d'incidence de la MC.

Enfin, la distribution des SIRs de la MC a été comparée au sein des quatre profils de communes identifiés lors de l'analyse conjointe (Figure 17).



Classe	Effectif	SIRs MC (Médiane [EIQ])
A surveiller	1 516	1.113 [0.979 ; 1.239]
Résilient	1 051	1.050 [0.920 ; 1.227]
Fragile	53	1.087 [1.001 ; 1.205]
Fragile avec des ressources	421	1.017 [0.895 ; 1.157]

Figure 17 : Distribution des SIRs lissés de la MC au sein des profils de communes issus de l'analyse conjointe

Si deux classes partagent une lettre en commun alors elles ne sont pas significativement différentes, au contraire si deux classes ne partagent aucune lettre en commun alors elles sont significativement différentes (Test de Kruskal-Wallis avec comparaisons multiples de Dunn)

D'abord, le profil « Résilient » présente un niveau d'incidence inférieur au profil « A surveiller ». De même, le profil « Fragile avec des ressources » présente un niveau d'incidence inférieur au profil « Fragile ». Ainsi dans les deux cas, pour un niveau de vulnérabilité équivalent, le profil avec le niveau de résilience le plus élevé présente l'incidence la plus faible. De nouveau, ces résultats suggèrent une potentielle influence de la résilience des territoires sur la répartition d'incidence de la MC.

Discussion

Dans ce premier chapitre, le principal objectif était d'évaluer la vulnérabilité et la résilience des territoires pour tenter d'expliquer la variabilité spatiale d'incidence de la MC. Pour y parvenir, la méthodologie développée dans le cadre du projet de recherche Cisterres a été reprise. Initialement appliquée sur la région Hauts-de-France, le calcul des indices a été étendu à la Seine Maritime pour couvrir le périmètre du registre Epimad. En parallèle, les résultats produits ont également permis d'évaluer la robustesse de la méthode de calcul des indices.

1. Extension du jeu de données à la Seine Maritime.

Au-delà d'avoir étendu les données à la Seine Maritime, de nouveaux indicateurs ont été ajoutés par rapport au jeu initial. Au sein de la dimension Social, un indicateur de défaveur sociale a été ajouté. Initialement, la dimension Environnement physique comportait cinq descripteurs permettant de caractériser : l'accès aux espaces verts, l'accès aux espaces naturels, la monétarisation des services écosystémiques, les émissions de gaz à effet de serre et la localisation des sites et sols pollués. Dix indicateurs supplémentaires ont été intégrés. Ces dix nouveaux descripteurs permettent de caractériser : la pollution atmosphérique, la densité du réseau routier, la localisation des activités industrielles, l'occupation des sols et la dynamique temporelle d'artificialisation des sols.

Par rapport au jeu initial, des changements ont également dû être opérés sur le plan temporel. Par exemple, certains producteurs écrasent régulièrement leurs données pour les remplacer par des plus récentes. Ainsi, l'accès à des millésimes plus anciens est limité. C'est le cas pour la base de données des anciens sites industriels et activités de services (BASIAS) ou pour certains indicateurs développés par l'Observatoire des territoires. Sur la base de ce retour d'expérience, plusieurs recommandations peuvent être émises. D'abord, un téléchargement et un archivage régulier des données est recommandé. Il existe d'ailleurs des outils d'extraction de données qui permettent d'automatiser ce processus. Quelle que soit la méthode employée, cet archivage permet de constituer un historique et d'être en mesure, à terme, d'intégrer une dimension temporelle aux phénomènes étudiés. Cependant, il arrive que certains producteurs fassent évoluer leurs données et leurs méthodes de calcul. Ainsi, la comparaison historique peut devenir caduque. Il est donc important de surveiller les évolutions éventuelles pour s'assurer que les données soient bien comparables dans le temps. Par ailleurs, dans la mesure du possible, il est recommandé de réaliser ce travail d'archivage sur tout le périmètre national. Dans ce travail de thèse par exemple, pour étendre le jeu de données à un seul département supplémentaire, il a fallu télécharger à nouveau la quasi-totalité des données brutes et recalculer les indicateurs.

Cependant, au-delà des quelques difficultés rencontrées, l'extension du jeu de données à un département supplémentaire n'a pas posé de problème majeur, mes méthodes mises en place dans le projet Cisterres ont pu être aisément appliquées dans ce nouveau contexte. Les données brutes ont la plupart du temps pu être téléchargées à l'échelle nationale, laissant la possibilité par la suite d'étendre la base de données à d'autres départements (voire à la France entière). De nouveaux indicateurs ont même été ajoutés. Au sein de la dimension Social, un indicateur de défaveur socio-économique a été intégré. Pour la dimension Environnement physique, initialement elle comportait cinq descripteurs pour caractériser : l'accès aux espaces verts, l'accès aux espaces naturels, la monétarisation des services écosystémiques, les émissions de gaz à effet de serre et la localisation des sites et sols pollués. Dix indicateurs supplémentaires ont été intégrés. Ces dix nouveaux descripteurs

permettent de caractériser : la pollution atmosphérique, la densité du réseau routier, la localisation des activités industrielles, l'occupation des sols et la dynamique temporelle d'artificialisation des sols.

2. Mise en parallèle des résultats par rapport au projet Cisterres.

Dans le cadre du projet CISTERRES, IV et IR avaient été calculés. De fait, il est possible de mettre en parallèle les résultats précédemment obtenus avec ceux produits dans le cadre de notre analyse. Toutefois, une véritable analyse de sensibilité, ainsi que des comparaisons statistiques sont à proscrire car les indicateurs et les millésimes ne sont pas strictement identiques. L'objectif ici, est donc de vérifier si, en dépit de ces différences, les résultats restent en phase avec ceux précédemment obtenus. En d'autres termes, une commune en situation de fragilité dans Cisterres peut-elle avoir un statut de commune résiliente dans la thèse ? Les résultats cartographiques issus de l'analyse conjointe ont donc été comparés sur les Hauts-de-France (Figure 18).

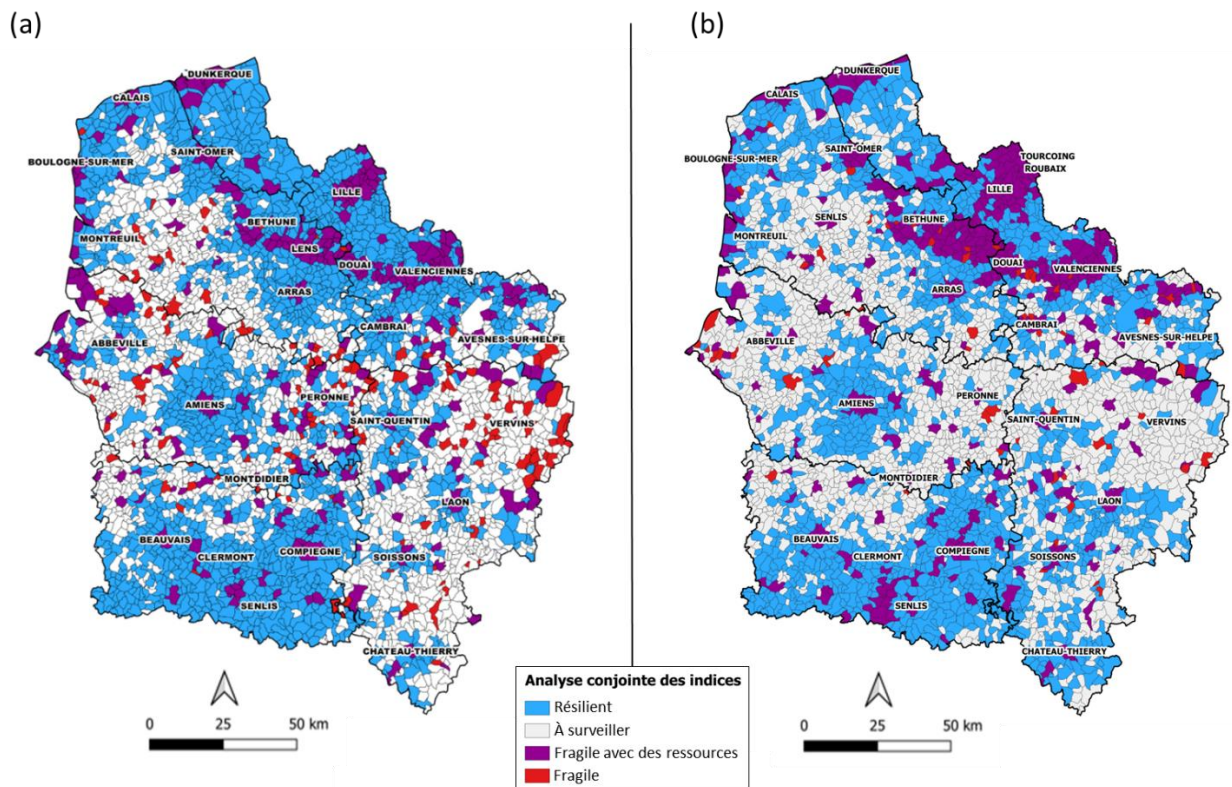


Figure 18 : Comparaison des résultats obtenus à l'issue de l'analyse conjointe lors du projet Cisterres (a) et lors du présent travail de thèse (b)

Si l'on s'intéresse aux indices composites IV et IR, les liens de corrélation sont très élevés. L'indice IR présente un coefficient de corrélation de Pearson de 0,81 avec celui calculé lors du projet Cisterres. Ce coefficient s'élève à 0.70 pour IV. Les indicateurs en eux-mêmes sont donc relativement stables. Certaines communes voient malgré tout leur classe (discrétisation Jenks) de vulnérabilité ou de résilience croître ($n=1\ 567$ et 1470 respectivement). Finalement, à l'issue de l'analyse conjointe, sur les $3\ 817$ communes de la région Hauts-de-France, $1\ 107$ (29%) ne sont plus classées de la même manière par rapport aux résultats obtenus lors du projet Cisterres.

Parmi les limites inhérentes à la méthode, la dépendance des résultats vis-à-vis de la zone d'étude avait déjà été clairement évoquée dans l'article (Brousmiche et al., 2020a). Les comparaisons menées ici permettent de l'illustrer concrètement. Cette évaluation de la vulnérabilité et de la résilience des

territoires est strictement dépendante de la zone d'étude considérée. En particulier, l'étape de discrétisation des indices semble particulièrement sensible et les résultats qui en sont issus doivent être considérés en tenant compte de la zone d'étude, puisqu'il s'agit surtout d'un classement relatif. Une commune classée comme très fortement vulnérable ne l'est en réalité que par rapport à la zone d'étude considérée. Ainsi, l'objectif de la démarche est avant tout de discriminer les unités spatiales les unes des autres au sein d'une zone d'étude clairement définie. Mener des comparaisons entre deux zones étudiées séparément serait une erreur.

3. Intérêt de cette approche pour renseigner l'étiologie de la MC.

Parmi les objectifs, l'un était de mesurer, à l'échelle de la commune et sur le périmètre du registre Epimad, l'accumulation de facteurs environnementaux favorables et défavorables. Il s'agissait ensuite d'examiner si ce cumul pouvait expliquer, au moins partiellement, la répartition spatiale d'incidence de la MC.

Le résultat le plus notable concerne l'association inverse qui est suggérée entre résilience des territoires et incidence de la MC. Pour confirmer ces observations, une méthode de régression spatiale spécifique devrait être mise en œuvre. Toutefois, en l'état, ce résultat renforce la thèse selon laquelle l'environnement et les déterminants territoriaux jouent certainement un rôle dans la répartition spatiale d'incidence de la MC. Pour connaître les indicateurs influents, il est nécessaire de disséquer les indices composite (synthèse de 27 indicateurs). Pour ce faire, une analyse a été menée sur les composantes principales qui ont permis le calcul de l'indice de résilience (lors de l'ACP). En analysant les liens de corrélation entre ces composantes principales et les indicateurs d'entrée, une interprétation a pu être donnée à chaque composante (ANNEXE 2). Ainsi, certaines composantes principales décrivent plutôt l'accès aux espaces verts, le dynamisme économique ou encore l'accès aux soins. Cependant, aucune tendance particulière ne s'est dégagée lors de l'examen des liens de corrélation entre les scores de ces composantes principales et les SIRs de la MC. C'est la raison pour laquelle ces résultats n'ont pas été présentés ici.

Plusieurs publications suggèrent un rôle bénéfique de l'accès aux espaces verts sur la MC (Elten et al., 2021; Zhang et al., 2022), ou plus largement de l'exposition à la biodiversité (da Luz Moreira et al., 2022). En ce qui concerne le niveau socio-économique des populations, la défaveur sociale semble avoir une influence négative sur la gravité des symptômes chez les personnes déjà atteintes de la MC (Walker et al., 2018). En revanche, les résultats sont très contradictoires sur le rôle de la défaveur sociale dans le déclenchement de la maladie. Certaines études ne trouvent aucune association (Declercq et al., 2010a), d'autres tendent à montrer que la MC touche davantage les populations défavorisées (Genin et al., 2020; Li et al., 2019; Nerich et al., 2010), tandis que d'autres encore suggèrent une sur-représentation de la MC au sein des populations les plus aisées (Armitage et al., 2004; Hu et al., 2014). Une revue de la littérature relativement récente n'a pas permis de trancher sur la question (Wardle et al., 2017).

Au-delà de la mise en lien avec les SIRs, des comparaisons ont également été menées entre les clusters spatiaux d'incidence. Les résultats de ces comparaisons se révèlent assez complexes à interpréter. D'abord, aucune tendance globale n'a pu être mise en évidence entre clusters de sur-incidence et clusters de sous-incidence. Certains clusters se distinguent cependant assez nettement des autres. En particulier, les clusters de sous-incidence de Lille et de Rouen présentent un niveau de résilience élevé et une très forte variabilité pour l'indice de vulnérabilité. Cette spécificité peut probablement être attribuée à leur nature atypique, ils correspondent aux deux plus grandes agglomérations urbaines de

la zone d'étude. Le cluster de sous-incidence de Péronne présente le niveau de résilience le plus faible. Les clusters d'Amiens (sur-incidence) et de Bolbec (sous-incidence) se distinguent par un faible niveau de vulnérabilité. Il convient de noter que certains clusters sont très étendus spatialement et qu'il est donc difficile de les caractériser sur le plan géographique et territorial. Malgré tout, ces résultats renforcent plutôt l'hypothèse selon laquelle chaque cluster pourrait être lié à une histoire et à des facteurs environnementaux spécifiques.

4. Représentativité des indices obtenus.

Dans ce travail, nous avons pu constater que la distinction nette entre facteurs de vulnérabilité et facteurs de résilience n'est pas toujours évidente à gérer. Certains facteurs environnementaux sont ambigus et il est difficile de les considérer sous ce prisme. De ce fait, certains descripteurs qui ne rentrent pas dans cette classification doivent être exclus de l'analyse. Par exemple, comment appréhender l'occupation des sols agricoles ? La présence de cultures céréalières dans une commune contribue-elle à sa vulnérabilité ou à sa résilience ?

De plus, pour étudier des pathologies dont l'étiologie est en grande partie inconnue, il semble important de prendre en compte l'environnement de manière exhaustive, sans préjugés particuliers, afin de capturer l'ensemble des influences potentielles. En effet, il est tout à fait envisageable qu'un facteur de risque pour une pathologie donnée puisse constituer un facteur de protection pour une autre, soulignant ainsi la complexité des relations entre environnement et santé.

Par ailleurs, la répartition spatiale des indices de vulnérabilité et de résilience semble fortement influencée par le continuum urbain-rural. Les centres urbains et leur périphérie sont très souvent caractérisés à la fois par une forte vulnérabilité et une forte résilience. Cette tendance se confirme lorsque l'on examine les liens de corrélation avec des marqueurs urbains. En particulier, le coefficient de corrélation de Pearson est de 0,85 entre l'indice de vulnérabilité et P_ARTIF (part de la surface communale occupée par des territoires artificialisés). Il s'élève à 0,74 avec la densité de population. Pour l'indice de résilience, le lien de corrélation le plus élevé concerne l'indicateur SHAN_CSP (indice de Shannon-Weaver pour mesurer la diversité des catégories socio-professionnelles représentées), il s'élève à 0,62. Ce constat pose question, il montre à quel point ces indices sont influencés par le gradient urbain.

Au-delà des méthodes mobilisées, ce résultat s'explique principalement par les données d'entrée sélectionnées. Plusieurs lacunes peuvent être identifiées dans le jeu de données en particulier pour la dimension Environnement physique. La pollution atmosphérique est décrite grâce à des données modélisées. En revanche, aucune donnée de contamination n'a pour l'instant été intégrée pour l'eau et les sols. Seules des variables indirectes qui décrivent notamment les sources de pollution étaient considérées. Par ailleurs, ces liens de corrélation démontrent clairement que les pressions urbaines sont sur-représentées dans le jeu de données. Il n'y a pour l'instant quasiment aucun descripteur lié aux pressions rurales C'est probablement en grande partie ces lacunes qui expliquent les limites dans les résultats obtenus. Il est donc crucial de parvenir à une description beaucoup plus exhaustive des pressions environnementales qui s'exercent sur un territoire.

Conclusion

Initialement, les indices de vulnérabilité et de résilience ont été développés pour caractériser les inégalités environnementales et sociales de santé. Au départ, l'ambition était notamment de proposer une méthode transposable pour identifier les points noirs environnementaux et ainsi prioriser l'action publique. Cette méthodologie a été reprise ici pour tenter de renseigner l'étiologie de la MC. Des résultats intéressants ont été obtenus en lien avec la résilience des territoires. Cependant, ces résultats n'ont pas réellement pu être approfondis. Par nature, ces indices mesurent un cumul de facteurs environnementaux (favorables ou défavorables). En revanche, ils sont moins pertinents pour décomposer l'analyse et tester des hypothèses étiologiques plus ciblées. Pour y parvenir, d'autres stratégies doivent donc être déployées.

Ce premier chapitre a également permis de mettre en évidence un point crucial. Le jeu de données environnemental qui a été construit présente des lacunes très importantes. En particulier, pour caractériser l'environnement physique, les pressions urbaines sont largement sur-représentées. Ainsi, avant d'envisager de nouvelles stratégies, il faut avant tout s'attacher à enrichir ces données environnementales. L'enjeu est de parvenir à décrire la diversité des pressions qui peuvent s'exercer sur un territoire. Par ailleurs, pour l'instant seule la contamination de l'air a été traitée, les autres compartiments (Eau et Sols) doivent absolument être décrits également. Ces points seront abordés dans le Chapitre 2.

CHAPITRE II – Mobiliser des données ouvertes pour décrire l’environnement physique au sens du milieu de vie extérieur.

Introduction

Les indices de vulnérabilité et de résilience ont été élaborés à partir d’une première base de données qui décrivait l’environnement au sens large, en cherchant à englober tous les déterminants territoriaux susceptibles de contribuer aux inégalités de santé. Ce jeu de données était structuré en six dimensions : Environnement physique, Economie, Social, Santé, Services, Politique. Les résultats obtenus ont mis en évidence une certaine dépendance des indices de vulnérabilité et de résilience vis-à-vis du milieu urbain. Cette tendance a été attribuée à un déséquilibre dans les données d’entrée. En particulier, les indicateurs intégrés dans la dimension « Environnement physique » traitaient essentiellement des pressions environnementales urbaines. Par exemple, la qualité de l’air a été décrite mais un effort de recherche doit être mené pour caractériser la qualité de l’eau et des sols. De la même manière, l’étude de la proximité aux sources d’émissions s’est pour l’instant limitée aux routes, aux sites et sols pollués, et aux des installations classées pour la protection de l’environnement. Dans ce contexte, il faut parvenir à une description beaucoup plus exhaustive de l’environnement physique. La stratégie retenue reste la même : développer des indicateurs spatialisés à l’échelle des communes à partir de données environnementales ouvertes. Ce second chapitre expose les recherches qui ont été menées pour développer de nouveaux indicateurs et construire un jeu de données permettant de décrire la diversité des pressions environnementales qui s’exercent sur un territoire. Tout au long de ce chapitre et dans la suite de la thèse, l’analyse a donc été recentrée sur l’environnement physique au sens du milieu de vie extérieur.

Il n’existe pas à ce jour de base de données permettant de décrire l’environnement physique de façon holistique à l’échelle des territoires. En France, quelques initiatives ont cependant été menées pour inventorier les bases de données existantes. Par exemple, l’Institut national de l’environnement industriel et des risques (Ineris) a produit un rapport en 2016 pour identifier et décrire l’ensemble des bases de données spatiales disponibles (INERIS, 2016). Ce premier rapport a été mis à jour en 2018 (INERIS, 2018). En 2022, pendant le présent travail de thèse, l’initiative Green Data for health (GD4H) a été lancée. Il s’agit d’une plateforme portée par le Commissariat Général au Développement Durable (CGDD) permettant d’inventorier les bases de données environnementales existantes pour faciliter leur mobilisation par les acteurs de la recherche en santé-environnement. Cette initiative s’inscrit dans le cadre du 4^{ème} Plan National Santé Environnement (PNSE4, 2021), plus précisément dans le cadre de l’Axe 4 qui vise à « mieux connaître les expositions et les effets de l’environnement sur la santé des populations et sur les écosystèmes ».

Malgré ces initiatives, la structuration et la bonne réutilisation des données environnementales reste un véritable défi pour la recherche en santé-environnement. À travers ce travail, nous apportons des éléments de réflexion concernant la réutilisation des données spatialisées. L’objectif est de contribuer à répondre aux questions suivantes : Quelles dimensions environnementales considérer ? Quelles sont les données disponibles ? Lesquelles sont pertinentes pour la construction d’indicateurs spatialisés à l’échelle communale ? Comment construire et calculer ces indicateurs ? Quelles sont les limites de ces données ?

Matériel & Méthodes

Deux zones d'études différentes ont été considérées. Dans les deux cas, l'unité spatiale de référence est la commune. La première comporte 3 041 communes et couvre le périmètre du registre Epimad (Nord, Pas-de-Calais, Somme, Seine-Maritime). La seconde est incluse dans la première, elle couvre le périmètre du registre Néphronor (Nord, Pas-de-Calais) et comporte 1 541 communes. Nous verrons plus bas que cette distinction a son importance, car la disponibilité et la qualité des données peut être variable selon le territoire d'étude.

Dans un premier temps, un travail a été mené pour inventorier les sources de données environnementales existantes. Cet inventaire a permis d'identifier 106 bases de données potentielles. Chaque source d'information a été confrontée à des critères d'inclusion et d'exclusion (Tableau 3).

Tableau 3: Critères d'inclusion et d'exclusion de nouvelles sources de données.

Critères d'inclusion
Donnée exhaustive et disponible à l'échelle nationale
Donnée adaptée au contexte territorial
Donnée disponible ou calculable à l'échelle spatiale étudiée (commune)
Critères d'exclusion
Donnée ne décrivant pas l'environnement physique au sens du milieu de vie extérieur
Donnée sans lien avec les problématiques de santé-environnement

La plupart du temps, les données sources identifiées ont été exclues en raison d'une couverture spatiale insuffisante. Pour ne citer qu'un exemple, la base de données Phyt'Atmo qui s'intéresse aux concentrations de pesticides dans l'air a été écartée pour cette raison. Une autre raison majeure d'exclusion était tout simplement la non-pertinence des données par rapport à la problématique environnement-santé. Par exemple, la base de données du BRGM qui identifie les zones de sismicité a été écartée de l'analyse. Finalement, 24 bases de données sources ont été retenues.

Comme dans le chapitre I, le jeu de données final comporte des indicateurs utilisés *in extenso* sans manipulation particulière des données brutes, des indicateurs pour lesquels une simple étape de normalisation a été opérée, et enfin des indicateurs dont la construction a nécessité un traitement plus complexe.

Pour la gestion des données manquantes, la même stratégie que celle utilisée dans le chapitre 1 a été reprise. En ce qui concerne la temporalité des indicateurs, à chaque fois que cela était possible, plusieurs millésimes ont été pris en compte et la moyenne a été calculée pour être incluse dans le jeu de données.

Les étapes du data-management des bases de données sources ainsi que les protocoles mis en place pour générer l'ensemble des indicateurs spatialisés sont décrits en détail par la suite dans la section Résultats. En effet, pour une meilleure lisibilité et une lecture plus agréable, tous les éléments ont été regroupés dans de grands tableaux.

En parallèle, des cartographies de chaque indicateur ont été générées. la méthode des seuils de Jenks a été retenue pour discrétiser les indicateurs (Jenks, 1967). De plus, des paramètres statistiques descriptifs ont également été fournis : moyenne, écart-type, médiane, quartiles, minimum, maximum.

Enfin, il est important de noter que la phase de collecte des données environnementales s’est arrêtée en décembre 2021. Il est donc possible que d’autres jeux de données aient été mis en ligne par la suite, notamment dans le cadre du projet GD4H, et qu’ils ne soient par conséquent pas repris dans ce travail.

Résultats

A l’échelle du registre Epimad, l’exploitation des 24 bases de données sources a permis de générer 113 indicateurs spatialisés. Les indicateurs ont été répartis en sept dimensions thématiques identifiées pour caractériser l’environnement physique : (i) niveaux de contamination des milieux, (ii) niveaux des émissions polluantes, (iii) localisation des sources d’émissions, (iv) occupation des sols, (v) pratiques agricoles, (vi) naturalité, (vii) climat. Les principaux phénomènes décrits dans le jeu de données et la répartition des indicateurs au sein des sept dimensions thématiques sont présentés en Figure 19.

Sur le périmètre du registre Néphronor, 20 indicateurs supplémentaires ont pu être calculés. Ils sont tous issus de la même base de données source et décrivent les niveaux de contamination dans les réseaux de distribution d’eau potable.

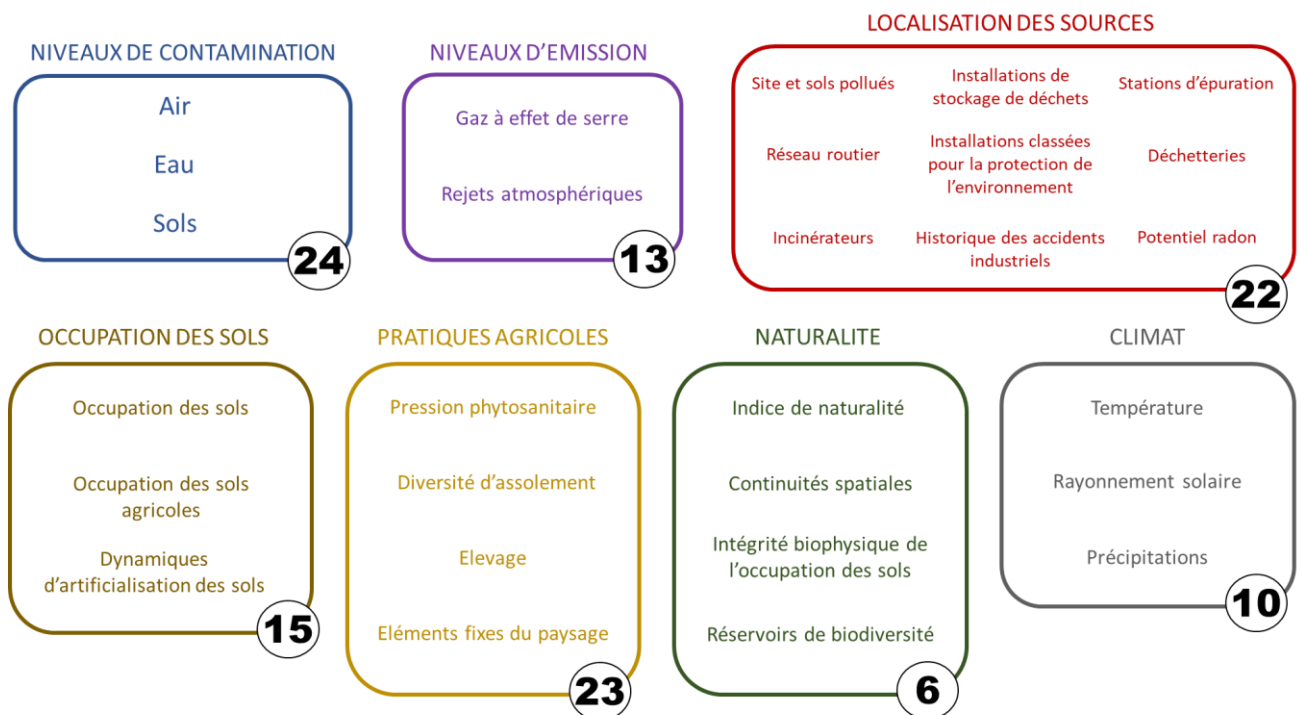


Figure 19: Principaux phénomènes décrits et répartition des indicateurs au sein des dimensions
Les chiffres correspondent au nombre d'indicateurs intégrés dans chaque sous-dimension

Cette figure permet d’avoir une vision d’ensemble sur le jeu de données et son contenu. Le détail sur les données sources utilisées et sur les indicateurs développés est présenté ci-après sous forme de tableaux. Pour plus de clarté, chaque tableau correspond à une dimension thématique (Tableaux 4, 5, 6, 7, 8, 9, 10). Ces tableaux fournissent une description approfondie de chaque source de données utilisée : nom, producteur, description détaillée de son contenu, format des données brutes, temporalité des données disponibles. Une description précise des indicateurs spatialisés produits est

également fournie : méthodes de calcul employées, nombre d'indicateurs produits, nature de ces indicateurs, principales limites identifiées.

En plus des tableaux, pour deux thématiques spécifiques, une présentation plus approfondie est proposée. Des choix méthodologiques importants ont été pris ou des difficultés particulières ont été rencontrées lors de l'exploitation de ces données sources.

En parallèle, une cartothèque a été produite, elle constitue un livrable de la thèse mais n'a pas été intégrée en annexe du manuscrit (elle fait plus de 150 pages). Elle donne accès à la représentation cartographique de tous les indicateurs spatialisés développés. En plus de ces cartographies, des statistiques descriptives sont également fournies pour chaque indicateur.

1. Niveaux de contamination des milieux

Tableau 4 : Niveaux de contamination des milieux - Présentation détaillée des données sources et des indicateurs produits

SOUS-DIMENSION	SOURCES DE DONNÉES					CONSTRUCTION DES INDICATEURS SPATIALISÉS			
	Nom	Producteur	Description	Temporalité des données disponibles	Format des données brutes	Traitement des données	Nombre d'indicateurs produits	Liste des indicateurs produits	Limites d'interprétation de ces indicateurs
Niveaux de contamination des milieux	IAQD	Agence Européenne de l'Environnement (AEE)	Ces données sont issues des réseaux de surveillance de la qualité de l'air à l'échelle européenne. Des cartographies sont proposées pour un certain nombre de polluants atmosphériques. Elles sont mises à jour annuellement. Les concentrations s'expriment en $\mu\text{g.m}^{-3}$. Pour chaque polluant, les données brutes se présentent sous la forme d'une couche raster par millésime.	2011, 2012, 2013, 2014, 2015, 2016, 2017	Rasters	Pour chaque polluant, les données brutes se présentent sous la forme de sept couches rasters (une couche par millésime). Ces couches ont été découpées selon le contour des communes de la zone d'étude. Les données ont ensuite été agrégées en calculant la valeur moyenne dans chaque commune. Enfin, la moyenne entre les sept millésimes a été calculée et l'indicateur final s'exprime en $\mu\text{g.m}^{-3}$.	3	NO2_AIR : Moyenne interannuelle de la pollution atmosphérique en NO_2 PM10_AIR : Moyenne interannuelle de la pollution atmosphérique en PM_{10} PM2.5_AIR : Moyenne interannuelle de la pollution atmosphérique en $\text{PM}_{2.5}$	Equivalait à une pollution de fond, ne tient pas compte de la pollution de proximité. Note : depuis 2022, l'INERIS met à disposition les données modélisées à partir du modèle CHIMERE pour les mêmes polluants sur la période 2000-2022. En plus de la concentration totale, la concentration pondérée par la population est proposée pour chaque commune.
	Réseau de Mesure de la Qualité des Sols (RMQS)	Groupement d'intérêt scientifique Sol & Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE)	A notre connaissance, il s'agit de la seule base de données ouverte sur la qualité des sols qui couvre le territoire national. Le réseau RMQS repose sur le suivi de 2240 sites de prélèvement répartis uniformément sur le territoire national selon une maille carrée de 16 km de côté. Les données utilisées sont issues de la première campagne de mesure conduite par l'INRAE sur la période 2000-2009. La teneur des sols en neuf éléments traces métalliques a été mesurée. Ces données sont exprimées en mg de contaminant par kg de sol (poids sec). Par ailleurs, à partir de ces données de terrain, les producteurs ont conduit des interpolations spatiales pour fournir des estimations à une résolution de 1000 m.	Données issues d'une campagne de prélèvements menée sur la période 2000-2009	Rasters	Les données brutes se présentent sous la forme de couches raster. Ces couches ont été découpées selon le contour des communes de la zone d'étude. Les données ont ensuite été agrégées en calculant la valeur moyenne dans chaque commune. Les indicateurs sont exprimés en milligrammes de contaminant par kilogramme de sol (poids sec).	15	AS_SOL : Teneur du sol en arsenic total CD_SOL : Teneur du sol en cadmium total CD_EXTR_SOL : Teneur du sol en cadmium extractible CR_SOL : Teneur du sol en chrome total CR_EXTR_SOL : Teneur du sol en chrome extractible CU_SOL : Teneur du sol en cuivre total CU_EXTR_SOL : Teneur du sol en cuivre extractible HG_SOL : Teneur du sol en mercure total NI_SOL : Teneur du sol en nickel total NI_EXTR_SOL : Teneur du sol en nickel extractible PB_SOL : Teneur du sol en plomb total PB_EXTR_SOL : Teneur du sol en plomb extractible ZN_SOL : Teneur du sol en zinc total ZN_EXTR_SOL : Teneur du sol en zinc extractible TL_SOL : Teneur du sol en thallium total	Les mailles utilisées pour le plan d'échantillonnage sont des mailles carrées de 16km de côté. Il faut donc bien garder à l'esprit que l'interpolation spatiale qui a été réalisée peut être imprécise (en particulier en milieu urbain).
	SISE-Eaux	Direction générale de la santé (DGS)	En France, l'eau du robinet est étroitement surveillée pour garantir sa sécurité sanitaire conformément à la Directive européenne 98/83/CE et au Code de la santé publique. Les Agences Régionales de Santé (ARS) assurent cette surveillance en effectuant des contrôles microbiologiques, physico-chimiques et radiologiques. Les laboratoires agréés par le Ministère de la Santé effectuent plus de 300 000 prélèvements d'eau chaque année, ce qui représente de l'ordre de 12 millions d'analyses. Toutes les données de surveillance sanitaire des eaux de distribution sont rassemblées dans la base de données SISE-Eaux. Chaque résultat d'analyse est associé à une unité de distribution (UDI). Une UDI se définit comme « un ensemble de canalisations de distribution de l'eau potable au sein desquelles la qualité de l'eau est considérée comme homogène ». Tous les abonnés raccordés au réseau public d'eau potable sont associés à une UDI. La France compte au total plus de 25 300 unités de distribution.	2016, 2017, 2018, 2019, 2020	Tables	L'objectif étant d'être le plus représentatif possible de la qualité de l'eau consommée par les populations, les mesures réalisées au niveau des points de captage ou des installations de traitement ont été retirées de l'analyse. Seuls les prélèvements réalisés dans les UDI ont été conservés. Ensuite pour chaque UDI, tous les prélèvements réalisés sur la période d'étude ont été pris en compte. Pour chaque paramètre, la valeur moyenne a été calculée à l'échelle des UDI en considérant tous les résultats d'analyse. Enfin, grâce à des tables d'équivalence, la correspondance entre UDI et commune a pu être gérée. Pour les communes alimentées par plusieurs UDI, la moyenne des UDI a été prise comme valeur pour la commune. Après avoir exploité cette base de données, certains paramètres ne présentaient aucune mesure quantifiée au cours de la période sur toute la zone d'étude. Dans ce cas, les paramètres ont été exclus. De même, un grand nombre de données manquantes a été constaté pour certains paramètres qui ont dû être exclus.	6 (à l'échelle du registre Epimad) 26 (à l'échelle du registre Néphronor)	A l'échelle du registre Epimad (Nord, Pas-de-Calais, Somme, Seine-Maritime), de nombreux paramètres n'ont pas pu être inclus pour cause de trop nombreuses données manquantes. Voici les 6 indicateurs disponibles : ENTERO : Dénombrement Entérocoques PH : pH moyen de l'eau CONDUCT : Conductivité moyenne de l'eau TURBNFU : Turbidité moyenne de l'eau CL2LIB : Concentration moyenne en chlore libre dans l'eau CL2TOT : Concentration moyenne en chlore total dans l'eau A l'échelle du registre Néphronor (Nord, Pas-de-Calais), le nombre de paramètres disponibles s'avère bien plus important. En plus des 6 indicateurs précédents, 20 indicateurs supplémentaires ont été calculés. D'abord 4 indicateurs de contamination microbiologique : ECOLI : Dénombrement Escherichia coli COLIF : Dénombrement Coliformes GT22 : Dénombrement micro-organismes revivifiables à 22°C GT36 : Dénombrement micro-organismes revivifiables à 36°C Et 16 indicateurs de contamination chimique : Nitrates (NO3-), nitrites (NO2-), ammonium (NH4+), plomb (PB_EAU), nickel (NI_EAU), fer (FE_EAU), cuivre (CU_EAU), chrome (CR_EAU), antimoine (SB_EAU), chloroforme (CLF), bromoforme (BRF), bromodichlorométhane (DCLMBR), chlorodibromométhane (DBRMCL), trihalométhanes (THM4), chlorure de vinyle (CLVYL) et perchlorate (PCLAT).	Une forte hétérogénéité dans la temporalité des mesures effectuées au sein des UDI et des phénomènes de pics de pollutions amènent au questionnement de la représentativité des indicateurs vis-à-vis d'une pollution chronique de l'eau.

2. Niveaux des émissions polluantes

Tableau 5 : Niveaux des émissions polluantes - Présentation détaillée des données sources et des indicateurs produits

SOUS-DIMENSION	SOURCES DE DONNÉES				CONSTRUCTION DES INDICATEURS SPATIALISÉS				
	Nom	Producteur	Description	Temporalité des données disponibles	Format des données brutes	Traitement des données	Nombre d'indicateurs produits	Liste des indicateurs produits	Limites d'interprétation de ces indicateurs
Niveaux des émissions polluantes	Emissions atmosphériques tous secteurs confondus	Centre technique de référence en matière de pollution atmosphérique et de changement climatique (Citepa)	Le Citepa met à disposition des estimations des émissions annuelles totales des principaux polluants atmosphériques au niveau des établissements publics de coopération intercommunale (EPCI). Ces estimations sont calculées conformément au guide méthodologique pour l'élaboration des inventaires territoriaux des émissions atmosphériques (Bréhon et al., 2018) et conformément au référentiel français OMINEA (Andre et al., 2018). Les données brutes pour chaque polluant sont exprimées en kg d'émissions par an.	2012, 2015	Table	Pour chacun des polluants étudiés, la valeur moyenne des deux millésimes a été calculée après avoir rapporté les niveaux d'émission à la surface des EPCI. Enfin, grâce à des tables d'équivalence, un changement d'unité spatiale a été mené pour passer à l'échelle communale. Les indicateurs s'expriment en kg d'émissions par hectare et par an.	12	NOX_EM : Emissions atmosphériques d'oxydes d'azote Nox SO2_EM : Emissions atmosphériques de dioxyde de soufre SO ₂ PM10_EM : Emissions atmosphériques de particules en suspension PM ₁₀ PM2.5_EM : Emissions atmosphériques de particules fines PM _{2.5} CO_EM : Emissions atmosphériques de monoxyde de carbone CO COVNM_EM : Emissions atmosphériques de Composés Organiques Volatils Non Méthaniques NH3_EM : Emissions atmosphériques d'ammoniac NH ₃ NI_EM : Emissions atmosphériques de nickel Ni PB_EM : Emissions atmosphériques de plomb Pb C6H6_EM : Emissions atmosphériques de benzène C ₆ H ₆ CD_EM : Emissions atmosphériques de cadmium Cd AS_EM : Emissions atmosphériques d'arsenic As	<p>Les émissions sont rapportées à l'échelle des EPCI.</p> <p>Les données ne sont disponibles que pour deux années.</p>
	Emissions de gaz à effet de serre tous secteurs confondus	Observatoire des territoires	L'Observatoire des territoires fournit des données au niveau communal sur les émissions totales de gaz à effet de serre, mesurées en milliers de tonnes équivalent CO ₂ . Ces émissions évaluent l'effet cumulatif des substances qui contribuent à l'effet de serre, :le dioxyde de carbone (CO ₂), le méthane (CH ₄), le protoxyde d'azote (N ₂ O), les hydrofluorocarbures (HFC), les perfluorocarbures (PFC) et l'hexafluorure de soufre (SF ₆). Les données ne prennent pas en compte les puits de carbone, tels que l'absorption de CO ₂ par la végétation (photosynthèse) et le méthane (CH ₄) par les sols forestiers. De plus, les émissions des sources biotiques (forêts non gérées, prairies naturelles, zones humides, rivières, feux de forêt) ne sont pas prises en compte. Seules les émissions liées aux activités humaines ou gérées par l'homme sont incluses.	2012	Table	Donnée brute incluse sans traitement préalable. L'indicateur s'exprime en milliers de tonnes équivalent CO ₂	1	EMIS_GES : Emission des gaz à effet de serre hors puits	<p>Seules les émissions liées aux activités humaines ou gérées par l'homme sont incluses. De plus, les données ne prennent pas en compte les puits de carbone.</p> <p>Les données ne sont disponibles que pour l'année 2012.</p>

3. Localisation des sources d'émissions

Tableau 6 : Localisation des sources d'émissions - Présentation détaillée des données sources et des indicateurs produits (1/2)

SOUS-DIMENSION	SOURCES DE DONNÉES				CONSTRUCTION DES INDICATEURS SPATIALISÉS				
	Nom	Producteur	Description	Temporalité des données disponibles	Format des données brutes	Traitement des données	Nombre d'indicateurs produits	Liste des indicateurs produits	Limites d'interprétation de ces indicateurs
Localisation des sources d'émissions	Base de données des anciens sites industriels et activités de services (BASIAS)	Bureau de Recherches Géologiques et Minières (BRGM)	BASIAS regroupe les sites pollués ou potentiellement pollués sur lesquels étaient établis des activités de service ou des activités industrielles. Cette base permet de recenser les sites susceptibles de générer une pollution de l'environnement. Dans la base de données BASIAS, pour chaque site est associée la commune concernée, le nom de l'activité, la localisation du site et les activités réalisées avec le code NAF (Nomenclature d'Activité Française). Les données BASIAS utilisées dans ce travail ont été téléchargées en 2020.	2020	Shapefile (entités ponctuelles)	A partir des données ponctuelles, le nombre de sites BASIAS par commune a été comptabilisé. Puis la densité a été calculée en rapportant le nombre de sites à la superficie communale (ha). L'indicateur s'exprime donc en nombre de sites par hectare de surface communale.	1	BASIAS : Densité de sites BASIAS	Les sites BASIAS sont considérés de façon homogène quelle que soit leur taille ou leur historique.
	BASOL	Ministère de la transition écologique	BASOL est une base de données qui recense les sites et sols pollués ou potentiellement pollués nécessitant une intervention des pouvoirs publics, que ce soit à des fins préventives ou curatives. Ces sites peuvent contenir des polluants susceptibles de causer des nuisances pour la population ou l'environnement. Pour chaque site, la base de données fournit sa localisation, son identifiant unique BASOL et son code d'activité ICPE (Installations Classées pour la Protection de l'Environnement). Ces données ont été téléchargées en 2020.	2020	Shapefile (entités ponctuelles)	A partir des données ponctuelles, le nombre de sites BASOL par commune a été comptabilisé. Puis la densité a été calculée en rapportant le nombre de sites à la superficie communale (ha). L'indicateur s'exprime donc en nombre de sites par hectare de surface communale.	1	BASOL : Densité de sites BASOL	Les sites BASOL sont considérés de façon homogène quelle que soit leur taille ou leur historique.
	Base de données nationale des Installations Classées (S3IC)	Ministère de la transition écologique	La base de données S3IC regroupe les installations industrielles ou agricoles soumises à autorisation ou à enregistrement, en fonctionnement, en construction ou en cessation d'activité. Une installation classée pour la protection de l'environnement (ICPE) correspond à une exploitation « susceptible de créer des risques ou de provoquer des pollutions ou nuisances, notamment pour la sécurité et la santé des riverains ». La base de données fournit des informations telles que le nom de l'ICPE, sa localisation, son activité principale et son régime de classement. Les données utilisées ont été téléchargées en 2020.	2020	Shapefile (entités ponctuelles)	A partir des données ponctuelles, le nombre d'installations ICPE par commune a été comptabilisé. Puis la densité a été calculée en rapportant le nombre d'installations à la superficie communale (ha). L'indicateur s'exprime donc en nombre d'installations par hectare de surface communale.	1	ICPE : Densité de sites ICPE	Les installations ICPE sont considérées de façon homogène quelle que soit leur taille ou leur activité.
	Portail d'information sur l'assainissement communal	ADEME	Le portail d'information sur l'assainissement communal met à disposition des données sur l'assainissement collectif en France métropolitaine. Elles permettent de géolocaliser les stations d'épuration et de connaître leur niveau de conformité au regard de la directive « Eaux Résiduaires Urbaines ». Les données utilisées concernent l'année 2015.	2015	Shapefile (entités ponctuelles)	A partir des données ponctuelles, deux types d'indicateurs ont été développés : la densité surfacique et la distance à l'installation la plus proche. Pour la densité surfacique, le nombre d'installations par commune a été comptabilisé et rapporté à la superficie (ha). Dans ce cas, l'indicateur s'exprime en nombre d'installations par hectare. Pour le second type d'indicateur, la distance entre le centroïde des communes et l'installation la plus proche a été mesurée. Dans ce cas, l'indicateur s'exprime en km.	4	STEPS_dst : Densité de stations d'épuration des eaux usées STEPS_dst_nc : Densité de stations d'épuration des eaux usées non conformes STEPS_prox : Distance entre le centroïde de la commune et la station d'épuration la plus proche. STEPS_prox_nc : Distance entre le centroïde de la commune et la station d'épuration non conforme la plus proche.	Les stations d'épuration sont considérées de façon homogène quelle que soit leur taille.
	BD TOPO	Institut national de l'information géographique et forestière (IGN)	La base de données topographique (BD TOPO) contient des informations sur les segments de route, tels que leur nature, leur numéro, leur importance, leur classement administratif, le nombre de voies et la commune à laquelle ils appartiennent. La version utilisée dans ce travail date de l'année 2020.	2020	Shapefile (entités linéaires)	D'abord, un point a été créé pour chacune des intersections présentes à la jonction des tronçons de routes. Après avoir vérifié que ces points étaient uniques, le nombre de points par commune a été déterminé. Puis la densité d'intersection a été calculée en divisant le nombre de points dans une commune par sa superficie. L'indicateur s'exprime donc en nombre d'intersections de routes par hectare de surface communale.	1	INTERSEC : Densité d'intersections de routes	Toutes les intersections ont été considérées de la même manière quelle que soit l'importance des routes.

Tableau 6 : Localisation des sources d'émissions - Présentation détaillée des données sources et des indicateurs produits (2/2)

SOUS-DIMENSION	SOURCES DE DONNÉES					CONSTRUCTION DES INDICATEURS SPATIALISÉS			
	Nom	Producteur	Description	Temporalité des données disponibles	Format des données brutes	Traitement des données	Nombre d'indicateurs produits	Liste des indicateurs produits	Limites d'interprétation de ces indicateurs
Localisation des sources d'émissions	Sinoe	Ministère de la transition écologique	La base de données SINOE localise l'ensemble des installations de gestion et de traitement des déchets. Les données utilisées ont été téléchargées au cours de l'année 2021.	2021	Shapefile (entités ponctuelles)	Trois types d'indicateurs ont été développés : la densité surfacique, la distance à l'installation la plus proche, et le pourcentage de surface communale située dans l'aire d'influence d'une installation. Pour la densité surfacique, le nombre d'installations a été comptabilisé dans chaque commune et rapporté à sa superficie. Dans ce cas, l'indicateur s'exprime en nombre d'installations par hectare. D'autre part, la distance entre le centroïde des communes et l'installation la plus proche a été mesurée. Dans ce cas, l'indicateur s'exprime en km. Enfin, pour les incinérateurs et les installations de stockage, en se basant sur la littérature, l'aire d'influence de ces installations a été prise en compte. Pour les incinérateurs, des tampons de 10km de rayon ont été considérés autour de chaque installation (avec fusion des tampons en cas de chevauchement). Dans chaque commune, le pourcentage de surface communale située à moins de 10km d'un incinérateur a été calculé. Pour les installations de stockage, un rayon de 2km a été pris.	12	<p>DECHET_dst : Densité d'installations de traitement ou de gestion des déchets.</p> <p>DECHET_prox : Distance entre le centroïde de la commune et l'installation de traitement ou de gestion des déchets la plus proche.</p> <p>TRI : Densité de centres de tri</p> <p>INCIN : Densité d'incinérateurs ou co-incinérateurs (tous types de déchets)</p> <p>INCIN_BUFF : Pourcentage de la surface communale située à moins de 7,5 km d'un incinérateur de déchets (tous types de déchets)</p> <p>DECHETERIE : Densité de déchèteries</p> <p>METHAN : Densité d'unités de méthanisation (tous types de méthaniseurs)</p> <p>COMPOST : Densité de plateformes de compostage</p> <p>STOCKAGE : Densité d'installations de stockage des déchets (tous types de déchets)</p> <p>STOCKAGE_BUFF : Pourcentage de la surface communale située à moins de 2 km d'une installation de stockage des déchets (tous types de déchets)</p> <p>PLASTIC : Densité de sites de négoce et de valorisation des plastiques</p> <p>DEE : Densité de sites de traitement des déchets électriques et électroniques</p>	Les installations de traitement et de gestion des déchets sont considérées de façon homogène quelle que soit leur taille.
	Analyse, Recherche et Information sur les Accidents (ARIA)	Bureau d'Analyse des Risques et Pollutions Industriels (BARPI)	La base de données ARIA répertorie les incidents, accidents ou presque accidents qui ont porté, ou auraient pu porter atteinte à la santé, à la sécurité publique ou à l'environnement. Les données utilisées ont été téléchargées au cours de l'année 2020.	Accidents recensés sur la période 1960-2020	Shapefile (entités ponctuelles)	A partir des données ponctuelles, le nombre d'accidents recensés par commune a été déterminé. Puis la densité a été calculée en rapportant le nombre d'accidents à la superficie communale (ha). L'indicateur s'exprime donc en nombre d'accidents par hectare de surface communale.	1	ACCIDENT : Densité d'accidents industriels recensés depuis 1960 dans la base de données ARIA	Tous les accidents industriels sont considérés de façon homogène quel que soit leur type et leur importance.
	Potentiel radon	Institut de radioprotection et de sûreté nucléaire (IRSN)	La cartographie du potentiel radon réalisée par l'IRSN classe les communes en trois catégories en fonction des formations géologiques et de leur teneur en uranium. Catégorie 1 : Formations géologiques avec les plus faibles teneurs en uranium. Les concentrations en radon dans la plupart des bâtiments sont faibles. Catégorie 2 : Formations géologiques avec des teneurs en uranium faibles, mais des facteurs géologiques spécifiques (failles, mines souterraines) qui peuvent faciliter le transfert du radon vers les bâtiments. Catégorie 3 : Formations géologiques avec des teneurs en uranium plus élevées. Dans ces communes, une proportion plus importante de bâtiments présente des concentrations élevées en radon.	2019	Table	Donnée brute incluse sans traitement préalable.	1	RADON : Potentiel radon	Note : l'IRSN a depuis publié un indicateur d'exposition au radon ($\mu\text{Sv}/\text{an}$) équivalent à la période 2024-2019.

4. Occupation des sols

Tableau 7 : Occupation des sols - Présentation détaillée des données sources et des indicateurs produits

SOUS-DIMENSION	SOURCES DE DONNÉES					CONSTRUCTION DES INDICATEURS SPATIALISÉS			
	Nom	Producteur	Description	Temporalité des données disponibles	Format des données brutes	Traitement des données	Nombre d'indicateurs produits	Liste des indicateurs produits	Limites d'interprétation de ces indicateurs
Occupation des sols	Corine Land Cover (CLC)	Ministère de la transition écologique	CLC est une base de données géographique qui décrit l'état biophysique des sols dans 39 pays européens et sur plusieurs millésimes (1990, 2000, 2006, 2012, 2018). L'occupation des sols est catégorisée en cinq classes : « Territoires artificialisés », « Territoires agricoles », « Forêts et milieux semi-naturels », « Zones humides », « Surfaces en eau ».	1990, 2000, 2006, 2012	Shapefile (entités polygones)	Un indicateur a été créé pour chacune des cinq classes en prenant en compte les millésimes 1990, 2000, 2006 et 2012. Prenons l'exemple des territoires artificialisés. Pour chaque millésime, la couche d'occupation du sol a été intersectée avec le contour des communes. Une fois l'intersection réalisée, dans chaque commune, la surface des territoires artificialisés a été rapportée à la surface communale. Enfin l'indicateur final P_ARTIF a été généré en calculant la moyenne sur les quatre millésimes (1990, 2000, 2006, 2012). Il s'exprime en pourcentage de surface communale occupée par des territoires artificialisés.	5	<p>P_ARTIF : Pourcentage de surface communale occupée par des territoires artificialisés</p> <p>P_NAT : Pourcentage de surface communale occupée par des forêts et milieux semi-naturels</p> <p>P_AGR : Pourcentage de surface communale occupée par des territoires agricoles</p> <p>P_HUMID : Pourcentage de surface communale occupée par des zones humides</p> <p>P_EAU : Pourcentage de surface communale occupée par des surfaces en eau</p>	La base de données CLC permet l'identification de zones présentant une surface minimale de 25 ha. Ainsi les données manquent de précision pour étudier certains phénomènes. Par exemple, elles ne permettent pas de rendre compte du mitage urbain.
	Registre Parcellaire Graphique (RPG)	Institut national de l'information géographique et forestière (IGN)	Le RPG est une base de données géographique qui a été instaurée dans le cadre de la politique agricole commune (PAC). Mises à jour chaque année à partir des déclarations des exploitants pour les aides de la PAC, ces données fournissent des informations sur l'usage agricole des sols. Il s'agit d'un système d'information géographique qui représente à l'échelle 1/5000ème, la culture principale au sein des îlots culturaux (ensemble de parcelles contiguës appartenant à une même exploitation). Dans le cadre de ce travail, trois millésimes successifs ont été collectés et utilisés : 2010, 2011, 2012.	2010, 2011, 2012	Shapefile (entités polygones)	Pour chaque millésime, la couche d'occupation des sols agricoles a été intersectée avec le contour des communes. Une fois l'intersection réalisée, dans chaque commune, la surface occupée par les différents groupes culturaux a été rapportée à la surface communale. Enfin les indicateurs finaux ont été générés en calculant la moyenne sur les trois millésimes (2010, 2011, 2012). Ils s'expriment en pourcentage de surface communale occupée par le groupe cultural en question.	9	<p>ARABLE : Pourcentage de surface communale occupée par des terres arables</p> <p>PRAIRIE : Pourcentage de surface communale occupée par des prairies permanentes ou temporaires</p> <p>PERENNE : Pourcentage de surface communale occupée par des cultures pérennes</p> <p>SAU : Pourcentage de la surface communale occupée par la SAU totale</p> <p>CEREALES : Pourcentage de la surface communale occupée par des cultures de céréales</p> <p>OLEAGINEUX : Pourcentage de la surface communale occupée par des cultures d'oléagineux</p> <p>PROTEAGINEUX : Pourcentage de la surface communale occupée par des cultures de protéagineux</p> <p>LEGUMES_F : Pourcentage de la surface communale occupée par des cultures de légumes ou fleurs (dont pommes de terre)</p> <p>CULT_IND : Pourcentage de la surface communale occupée par d'autres cultures industrielles (betterave, etc...)</p>	Ces données sont issues des déclarations des agriculteurs dans le cadre de la PAC. Une partie des surfaces agricoles n'est donc pas représentée. Mais cela ne représente qu'une faible part. En effet, Preux et al. (2014) ont comparé les résultats du RPG avec les résultats du recensement agricole en Basse Normandie. En 2010, 23 900 exploitations étaient recensées en Basse Normandie contre 20 800 enregistrées dans le RPG. En termes de surfaces agricoles, cela représente un différentiel de 7,5 %.
	Portail de l'artificialisation des sols	Cerema	L'Observatoire National de l'Artificialisation propose des données sur les dynamiques temporelles d'artificialisation des sols. Ces données sont calculées à partir des Fichiers fonciers, une base de données nationale d'origine fiscale mise à jour annuellement pour le traitement de la taxe foncière. Cette base de données a été détournée de son usage initial pour évaluer le changement d'usage des sols. Plus spécifiquement, les données fournissent à l'échelle communale le flux d'artificialisation entre 2009 et 2019. Le flux d'artificialisation correspond au nombre d'hectares qui sont passés du statut "Naturel", "Agricole" ou "Forestier" à "Artificialisé".	Evolution entre 2009 et 2019	Table	Dans chaque commune, le flux d'artificialisation a simplement été rapportée à la surface communale. L'indicateur final s'exprime donc en pourcentage de surface communale convertie en surface artificialisée entre 2009 et 2019.	1	ARTIF_0919 : Pourcentage de la surface communale convertie en surface artificialisée entre 2009 et 2019	Les Fichiers fonciers ne traitent que les parcelles cadastrées. Par exemple, les routes et les cours d'eau ne sont pas représentés. Cependant, ces surfaces ne représentent qu'environ 4 % du territoire métropolitain (Bocquet et al., 2022).

5. Pratiques agricoles

Tableau 8 : Pratiques agricoles - Présentation détaillée des données sources et des indicateurs produits (1/2)

SOUS-DIMENSION	SOURCES DE DONNÉES					CONSTRUCTION DES INDICATEURS SPATIALISÉS			
	Nom	Producteur	Description	Temporalité des données disponibles	Format des données brutes	Traitement des données	Nombre d'indicateurs produits	Liste des indicateurs produits	Limites d'interprétation de ces indicateurs
Pratiques agricoles	Base Nationale des Ventes Distributeurs (BNVD)	Office Français de la Biodiversité (OFB)	La BNVD a été créée en 2009 à des fins fiscales dans le cadre de la redevance pour pollutions diffuses. Depuis cette date, les distributeurs agréés de produits phytosanitaires sont contraints de déclarer annuellement le bilan de leurs ventes. Pour chaque vente est renseigné le code postal du distributeur concerné ainsi que la quantité de produit ou de substance active vendue (en kg). Et depuis 2013, les distributeurs sont également contraints de rendre public le registre des achats. Ainsi désormais, à chaque vente est également associée le code postal de l'acheteur. Pour les deux premiers millésimes (2013, 2014), ces données d'achat sont considérées comme incomplètes. Dans ce travail, cinq millésimes consécutifs ont été pris en compte (2015 à 2019).	2015, 2016, 2017, 2018, 2019	Tableur	Les données d'achat sont disponibles par produit commercial ou par substance active. Le choix de considérer les substances actives a été retenu. Ensuite, les indices de pression phytosanitaire ont été calculés en considérant l'ensemble des substances actives ou en considérant les substances par fonction (herbicides, insecticides, fongicides, facteurs de croissance). Des indices ont également été calculés individuellement pour quelques substances actives en particulier. Dans tous les cas, les indices sont calculés en rapportant la quantité de substance(s) active(s) achetée(s) à la superficie du code postal. La dernière étape a consisté à changer d'unité spatiale. Des tables d'équivalence ont permis d'attribuer à chaque commune la valeur du code postal qui lui est associé. Le cas particulier des communes présentant plusieurs codes postaux a été géré en attribuant à chaque commune la valeur moyenne des codes postaux qui la composent. Tous les indicateurs obtenus s'expriment donc en kg de substance(s) active(s) achetée par hectare et par an.	15	<p>TOTPEST : Quantité totale de substances actives achetée par hectare (ha) de surface communale et par an (toutes substances confondues)</p> <p>4OPEST : Quantité totale de substances actives achetée par ha de surface communale et par an (en ne considérant que les 40 premières substances)</p> <p>FONGICID : Quantité totale de substances fongicides achetées par ha de surface communale et par an</p> <p>HERBICID : Quantité totale de substances herbicides achetée par ha de surface communale et par an</p> <p>REG_CROISS : Quantité totale de substances régulatrices de croissance achetée par ha de surface communale et par an</p> <p>INSECTICID : Quantité totale de substances insecticides achetée par ha de surface communale et par an</p> <p>PROSULF : Quantité de prosulfocarbe achetée par ha de surface communale et par an</p> <p>GLYPHO : Quantité de glyphosate achetée par ha de surface communale et par an</p> <p>MCOZEB : Quantité de mancozebe achetée par ha de surface communale et par an</p> <p>VASELINE : Quantité d'huile de vaseline achetée par ha de surface communale et par an</p> <p>CYPERMET : Quantité de cyperméthrine achetée par ha de surface communale et par an</p> <p>IMIDACLO : Quantité d'imidaclopride achetée par ha de surface communale et par an</p> <p>PHOSMET : Quantité de phosmet achetée par ha de surface communale et par an</p> <p>CLPYR-M : Quantité de chlorpyrifos méthyl achetée par ha de surface communale et par an</p> <p>CLPYR-E : Quantité de chlorpyrifos éthyl achetée par ha de surface communale et par an</p>	<p>Les achats de produits phytosanitaires au cours d'une année ne reflètent pas nécessairement leur usage réel.</p> <p>Le code postal de l'acheteur correspond à la localisation de son siège social. Un biais est donc introduit, notamment pour les exploitations dont les parcelles sont réparties sur plusieurs codes postaux.</p> <p>Afin de garantir l'anonymat, pour les codes postaux rassemblant moins de cinq acheteurs, la donnée est rendue indisponible.</p>
	S3IC	Ministère de la transition écologique	La base de données S3IC regroupe les installations industrielles ou agricoles soumises à autorisation ou à enregistrement, en fonctionnement, en construction ou en cessation d'activité. Une installation classée pour la protection de l'environnement (ICPE) correspond à une exploitation « susceptible de créer des risques ou de provoquer des pollutions ou nuisances, notamment pour la sécurité et la santé des riverains ». La base de données fournit des informations telles que le nom de l'ICPE, sa localisation, son activité principale et son régime de classement. Les données utilisées ont été téléchargées en 2020.	2020	Shapefile (entités ponctuelles)	Les données S3IC ont d'abord été filtrées pour ne retenir que les élevages. A partir de ces données ponctuelles, le nombre d'élevages classés ICPE a été comptabilisé et rapporté à la surface de la commune. L'indicateur final s'exprime donc en nombre d'élevages classés ICPE par hectare de surface communale.	1	ELEVAGE : Densité d'élevages (classés ICPE)	Les élevages sont soumis à enregistrement ou autorisation à partir du moment où les effectifs d'animaux dépassent un certain seuil. Par exemple, un élevage porcin est soumis à enregistrement à partir de 400 têtes et soumis à autorisation à partir de 800 têtes. De ce fait, les données sont loin d'être exhaustives et seuls les plus gros élevages sont recensés dans S3IC. En 2017, la base de données recensait 19 742 élevages, soit moins de 10 % des élevages français (Parisse et al., 2018). De plus, l'indicateur final ne donne pas d'information sur le nombre d'animaux.
	Recensement agricole	Agreste	Les données Agreste issues du recensement agricole recensent à l'échelle de la commune le nombre d'unités de gros bétail (UGB). L'UGB est une unité de référence permettant d'agrèger le bétail de différentes espèces et de différents âges en utilisant des coefficients spécifiques basés sur leurs besoins nutritionnels ou alimentaires. Par exemple un bovin de moins d'un an compte pour 0,4 UGB tandis qu'une poule pondeuse compte pour 0,014 UGB. Ainsi cette donnée permet à la fois de prendre en compte le nombre d'animaux et leur besoin alimentaire qui peut être considéré comme un proxy indirect de la quantité de déchets produits. Les données utilisées dans ce travail sont issues du recensement agricole de 2010.	2010	Tableur	Trois indicateurs distincts ont été développés. Dans chaque commune, le nombre d'UGB a été rapporté à : la surface communale, la surface agricole, ou la surface en prairies. L'indicateur final s'exprime donc en UGB par hectare de surface communale, par hectare de surface agricole, ou par hectare de surface en prairies.	3	<p>UGB/ha : Nombre d'unités de gros bétail (UGB) par hectare de surface communale</p> <p>UGB/ha SAU : Nombre d'unités de gros bétail (UGB) par hectare de surface agricole utile</p> <p>UGB/ha Pr : Nombre d'unités de gros bétail (UGB) par hectare de prairie</p>	

Tableau 8 : Pratiques agricoles - Présentation détaillée des données sources et des indicateurs produits (2/2)

SOUS-DIMENSION	SOURCES DE DONNÉES					CONSTRUCTION DES INDICATEURS SPATIALISÉS			
	Nom	Producteur	Description	Temporalité des données disponibles	Format des données brutes	Traitement des données	Nombre d'indicateurs produits	Liste des indicateurs produits	Limites d'interprétation de ces indicateurs
Pratiques agricoles	Haute valeur naturelle (HVN)	Solagro	<p>Il s'agit des seules données utilisées qui ne sont pas librement accessibles. Elles ont été produites par une organisation appelée Solagro qui détient le statut d'entreprise associative. Après simple sollicitation, les données nous ont été fournies à titre gracieux.</p> <p>Le concept de Haute Valeur Naturelle est apparu au début des années 2000, notamment avec l'adoption de la résolution de Kiev au sein de l'Union Européenne (2003). Cette résolution avait pour objectif de protéger et soutenir les zones agricoles à HVN. Chaque État membre a été chargé de mettre en place une méthode adaptée pour identifier et qualifier ces zones. Dans ce contexte, Solagro a travaillé avec le Centre commun de recherche de la Commission européenne pour développer une méthode d'évaluation (Pointereau et al., 2010; Poux and Pointereau, 2014). Ils ont développé un indicateur appelé score HVN qui correspond à la somme de trois sous-indices : (i) la diversité d'assolement, qui tient compte de la variété des cultures présentes sur les exploitations ; (ii) l'extensivité des pratiques, qui évalue le niveau d'intrants et la charge animale ; (iii) la densité des infrastructures agroécologiques, qui identifie la présence d'éléments fixes du paysage ou d'habitats semi-naturels.</p>	2000, 2010	Table	Donnée brute incluse sans traitement préalable.	4	<p>HVN : Score HVN ROTATIONS : Diversité d'assolement PRATIQUES : Extensivité des pratiques PAYSAGES : Pourcentage d'éléments paysagers dans la surface agricole</p>	

6. Naturalité

Tableau 9 : Naturalité - Présentation détaillée des données sources et des indicateurs produits

SOUS-DIMENSION	SOURCES DE DONNÉES				CONSTRUCTION DES INDICATEURS SPATIALISÉS				
	Nom	Producteur	Description	Temporalité des données disponibles	Format des données brutes	Traitement des données	Nombre d'indicateurs produits	Liste des indicateurs produits	Limites d'interprétation de ces indicateurs
Naturalité	Indice de naturalité	Union Internationale pour la Conservation de la Nature (UICN)	<p>Dans le cadre du projet CartNat, l'UICN a développé à l'échelle nationale un indicateur appelé gradient de naturalité potentielle (Guetté et al., 2021). La méthodologie développée s'appuie sur le concept de naturalité défini préalablement (Guetté et al., 2018). Cette définition distingue trois critères pour qualifier la naturalité du paysage :</p> <p>(i) l'intégrité biophysique de l'occupation du sol, qui traduit un état d'éloignement d'un paysage à ce qu'il serait sans transformation anthropique ; (ii) la spontanéité des processus, qui évalue le degré d'intervention humaine sur les dynamiques naturelles ; (iii) les continuités spatiales, qui correspondent au degré de connectivité des paysages. Pour chacun de ces trois critères, un sous-indice a été développé par l'UICN.</p> <p>L'intégrité biophysique de l'occupation des sols a été mesurée en comparant l'occupation des sols actuelle avec des données historiques datant du XVIIIe et XIXe siècle. La spontanéité des processus a été évaluée en utilisant deux proxys d'influence humaine : la distance aux routes et la densité de bâti. Les continuités spatiales ont été évaluées en utilisant un modèle mathématique appelé « Omniscape » (Landau et al., 2021). Enfin l'indice final, appelé gradient de naturalité potentielle, a été calculé en sommant sans pondération les trois sous-indices précédents.</p>	2021	Table	Donnée brute incluse sans traitement préalable.	4	<p>NATURALITE : Indice de naturalité potentielle BIOPHY : Intégrité biophysique de l'occupation du sol SPONTANE : Spontanéité des processus CONTINU : Continuités spatiales</p>	Le calcul du sous-indice de spontanéité des processus se base uniquement sur la distance aux routes et la densité de bâti.
	Trame verte et bleue (TVB)	Inventaire National du Patrimoine Naturel (INPN)	<p>La TVB est un outil conçu nationalement pour améliorer la continuité écologique et mieux intégrer la préservation de la biodiversité dans l'aménagement des territoires. La TVB désigne un réseau écologique constitué de corridors écologiques et de réservoirs de biodiversité. Les réservoirs de biodiversité sont définis comme des espaces où la biodiversité est la plus riche et la mieux représentée. Les corridors écologiques constituent des voies de déplacements entre réservoirs. La cartographie nationale de la TVB est mise à disposition par l'INPN.</p>	2014	Shapefile (entités polygones)	La couche "Trame verte et bleue" a d'abord été intersectée avec le contour des communes de la zone d'étude. Ensuite, dans chaque commune, la surface de réservoirs de biodiversité a été évaluée et rapportée à la surface communale. L'indicateur final s'exprime donc en pourcentage de surface communale occupée par des réservoirs de biodiversité.	1	RESERVOIR : Pourcentage de la surface communale occupée par des réservoirs de biodiversité	Initialement, ce travail a été réalisé à l'échelle de chaque région avec des méthodologies parfois différentes. Le muséum national d'histoire naturel (MNHN) a travaillé en collaboration avec l'INPN pour standardiser ces données et ainsi faciliter la comparaison à l'échelle nationale. Malgré ce travail, des biais entre régions peuvent persister.
	Open Street Map (OSM)	Open Street Map	<p>OSM est une base de données géographique libre et collaborative, créée par une communauté mondiale de contributeurs. Elle vise à fournir des données cartographiques détaillées et actualisées sur tous les aspects de notre environnement. Contrairement aux bases de données traditionnelles, OSM est construit à partir de contributions volontaires. Cette approche participative permet d'obtenir une couverture mondiale étendue et une diversité de données géographiques, allant des routes et des bâtiments aux points d'intérêt tels que les parcs, les restaurants, les commerces, etc... La couche utilisée dans ce travail est la couche "Land use".</p>	2020	Shapefile (entités polygones)	<p>Dans la couche « Land Use », seules les catégories suivantes ont été retenues pour représenter les espaces verts : « forest », « park » et « recreation ground ». Dans un premier temps, les espaces verts de moins de 0,5 ha ont été exclus de l'analyse. Ensuite des tampons de 400 m autour des espaces verts retenus ont été appliqués (avec fusion des tampons en cas de chevauchement). Ces données spatiales sur les espaces verts ont ensuite été recoupées avec la population carroyée à 200 m de l'année 2015 (après transformation des carreaux en centroïdes). L'objectif était d'estimer dans chaque commune la population située à moins de 400 m d'un espace vert. Enfin le rapport entre la population située à moins de 400 m d'un espace vert et la population totale dans la commune a été calculé. L'indicateur final correspond donc au pourcentage de la population communale qui réside à moins de 400 m d'un espace vert.</p>	1	AC_EV : Part de la population communale ayant accès à un espace vert dans un rayon de 400 mètres	

7. Climat

Tableau 10 : Climat - Présentation détaillée des données sources et des indicateurs produits

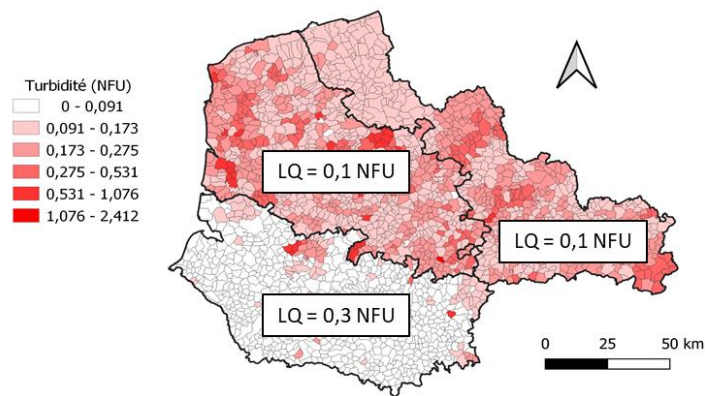
SOUS-DIMENSION	SOURCES DE DONNÉES					CONSTRUCTION DES INDICATEURS SPATIALISÉS			
	Nom	Producteur	Description	Temporalité des données disponibles	Format des données brutes	Traitement des données	Nombre d'indicateurs produits	Liste des indicateurs produits	Limites d'interprétation de ces indicateurs
Climat	E-obs dataset	Agence Européenne de l'Environnement (AEE)	E-OBS est une base de données ouverte qui rassemble des données climatiques à l'échelle européenne depuis 1950 jusqu'à aujourd'hui. Les données sont disponibles sur un pas de temps quotidien depuis 1950 et la résolution spatiale est une maille carrée de 11,1 km de côté. Les données sont donc très volumineuses et se présentent sous la forme de rasters spatio-temporels. Dans le cadre de ce travail, la base de données E-OBS a été téléchargée sur la période 1950-2019.	Données climatiques quotidiennes sur la période 1950-2019	Rasters	<p>Trois grands types d'indicateurs ont été générés.</p> <p>D'abord, la valeur moyenne de certains paramètres a été évaluée à l'échelle communale sur toute la période 1950-2019. Dans ce cas, l'indicateur s'exprime dans la même unité que le paramètre en question.</p> <p>Ensuite, certains indicateurs correspondent au nombre de jours où un paramètre a dépassé une certaine valeur (au cours de la période 1950-2019). Dans ce cas, l'indicateur s'exprime en nombre de jours de dépassement.</p> <p>Enfin, d'autres indicateurs ont été calculés pour mesurer l'évolution d'un paramètre entre deux périodes temporelles. Dans ce cas, l'indicateur est un pourcentage d'évolution, ou un nombre de jours d'écart.</p>	10	<p>Tmoy : Température moyenne journalière (valeur moyenne sur la période 1950-2019)</p> <p>Rmoy : Rayonnement global journalier (valeur moyenne sur la période 1950-2019)</p> <p>Pmoy : Niveau des précipitations journalières (valeur moyenne sur la période 1950-2019)</p> <p>DTmoy : Taux de variation de la température moyenne journalière entre les périodes 1950-1959 et 2010-2019</p> <p>DRmoy : Taux de variation du rayonnement global journalier entre les périodes 1950-1959 et 2010-2019</p> <p>DPmoy : Taux de variation du niveau des précipitations journalières entre les périodes 1950-1959 et 2010-2019</p> <p>Nbj_TSUp30 : Nombre de jours où la température maximale journalière a dépassé 30°C sur la période 1950-2019</p> <p>Nbj_PSup20 : Nombre de jours où le niveau des précipitations journalières a dépassé 20 mm sur la période 1950-2019</p> <p>DNbj_Tsup30 : Différence entre le nombre de jours où la température maximale journalière a dépassé 30°C entre la période 1985-2020 et la période 1950-1984</p> <p>DNbj_Psup20 : Différence entre le nombre de jours où le niveau des précipitations journalières a dépassé 20 mm entre la période 1985-2020 et la période 1950-1984</p>	

8. Cas particuliers de certaines bases de données.

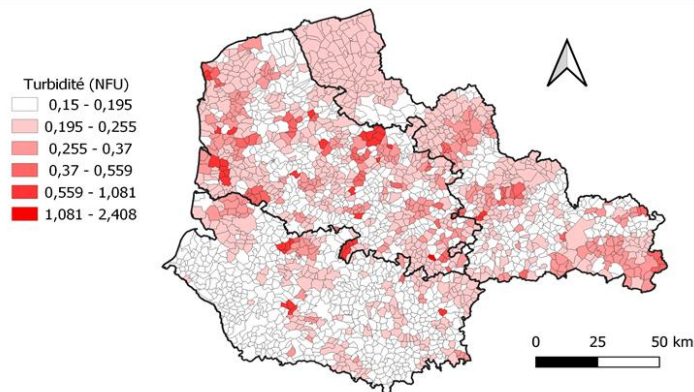
a. Niveaux de contamination dans les eaux de distribution (Sise-Eaux).

Pour exploiter la base de données Sise-Eaux, toutes les mesures réalisées au niveau des points de captage ou des installations de traitement ont été exclues. Seuls les prélèvements réalisés au niveau des UDI ont été pris en compte. En conservant tous les résultats sans distinction, le nombre de données manquantes aurait été bien inférieur pour de nombreux paramètres. Cependant, les indicateurs n'auraient pas été représentatifs de la qualité de l'eau réellement consommée par les populations. De plus, la comparabilité des indicateurs entre communes aurait également posé question. En effet, il aurait été problématique de comparer des résultats d'analyse réalisés au niveau des UDI (après traitement) à des résultats d'analyse réalisés au niveau des points de captage (avant traitement).

Par ailleurs, l'exploitation des données brutes a mis en évidence un biais important dans la qualité de mêmes données en fonction des laboratoires d'analyses. En effet selon le département considéré, les échantillons d'eau sont analysés par des laboratoires différents. Pour certains paramètres, cela se traduit au niveau des limites de quantification qui sont hétérogènes d'un département à un autre (Figure 20). Ainsi pour chaque paramètre concerné, la limite de quantification la plus élevée (LQ_{\max}) a été identifiée sur la zone d'étude. Puis toutes les mesures inférieures ou égales à cette LQ_{\max} ont été écrasées et remplacées par la valeur $\frac{LQ_{\max}}{2}$. Cette opération induit une perte d'information, mais elle était impérative pour pouvoir comparer les résultats d'un département à un autre.



Avant gestion des LQ



Après gestion des LQ

Figure 20 : Stratégie employée pour gérer les limites de quantification hétérogènes entre départements (exemple avec le paramètre Turbidité).

Après avoir exploité cette base de données, il s'est avéré que certains contaminants présentaient des concentrations très faibles et avec aucune mesure quantifiée sur toute la zone d'étude au cours de la période 2016-2020 (toutes les mesures étaient inférieures à la limite de quantification). Dans ce cas, les paramètres n'ont pas été inclus dans la base de données car absolument pas discriminants. Les huit paramètres concernés sont les suivants : benzo(a)pyrène, benzo(b)fluoranthène, benzo(k)fluoranthène, benzo(g,h,i)pérylène, indéno(1,2,3-cd)pyrène, cadmium, epichlorohydrine et acrylamide.

Par ailleurs, un grand nombre de données manquantes a été constaté pour un certain nombre de paramètres (en particulier dans les départements de la Somme et de la Seine Maritime). Un exemple est donné en Figure 21 avec la concentration en nitrates. Dans un tel cas de figure, les données manquantes sont trop nombreuses et trop agrégées spatialement pour envisager une imputation spatiale. A l'échelle du registre Epimad, le paramètre a donc dû être exclu de l'analyse.

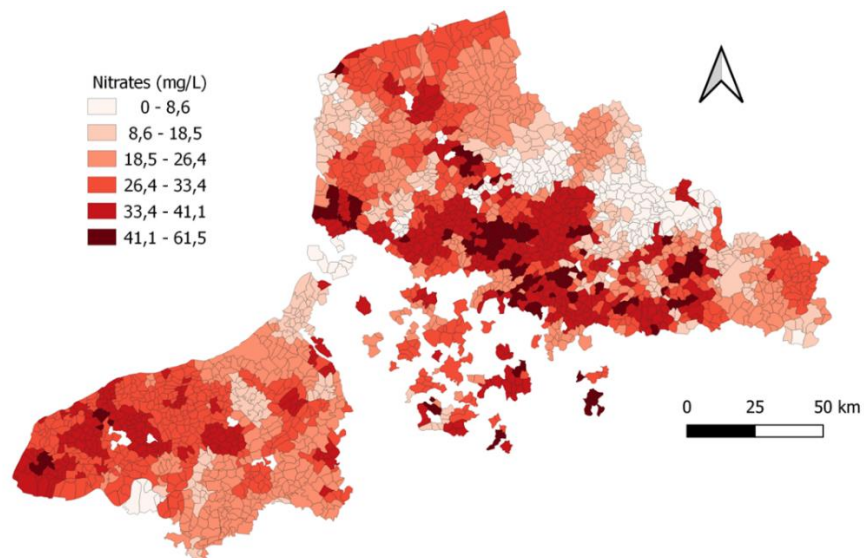


Figure 21: Concentration moyenne en nitrates dans les eaux de distribution (valeur moyenne sur la période 2016-2020).

Ainsi à l'échelle du registre Epimad, de nombreux paramètres n'ont pas pu être inclus dans le jeu de données. Seuls six paramètres étaient suffisamment exhaustifs pour être inclus (Figure 22).

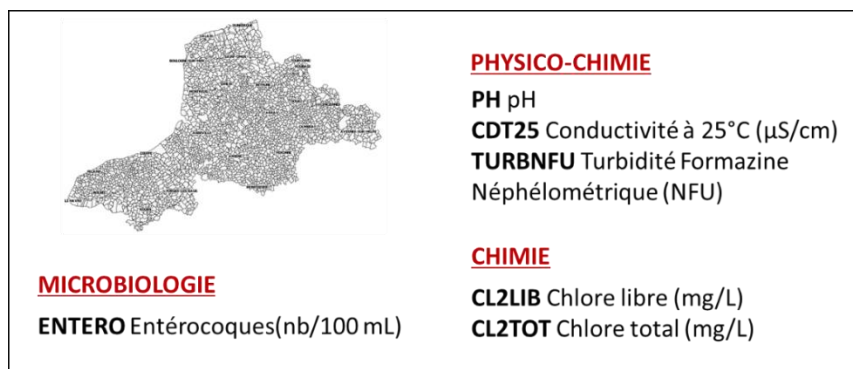


Figure 22 : Paramètres disponibles pour la qualité de l'eau potable à l'échelle du registre Epimad.

En revanche, en se plaçant à l'échelle du registre Néphronor (départements du Nord et du Pas-de-Calais), le nombre de paramètres disponibles s'avère bien plus important (voir Figure 23).


<p>PARAMETRES PHYSICO-CHIMIQUES</p> <p>PH pH CDT25 Conductivité à 25°C (µS/cm) TURBNFU Turbidité Formazine Néphélométrie (NFU)</p> <p>TRIHALOMETHANES</p> <p>CL2LIB Chlore libre (mg/L) CL2TOT Chlore total (mg/L) BRF Bromoforme (µg/L) CLF Chloroforme (µg/L) DBRMCL Chlorodibromométhane (µg/L) DCLMBR Bromodichlorométhane (µg/L) THM4 Somme des 4 trihalométhanes (µg/L)</p> <p>AUTRES</p> <p>CLVYL Chlorure de vinyle (µg/L) PCLAT Perchlorate (µg/L)</p>	<p>MICROBIO</p> <p>ECOLI (nb/100 mL) ENTERO (nb/100 mL) COLIF (nb/100 mL) GT22 Dénombrement micro-organismes revivifiables à 22°C (nb/100 mL). GT36 Dénombrement micro-organismes revivifiables à 36°C (nb/100 mL).</p> <p>METAUX</p> <p>PB Plomb (µg/L) NI Nickel (µg/L) FET Fer (µg/L) CU Cuivre (mg/L) CR Chrome (µg/L) SB Antimoine (µg/L)</p> <p>AZOTE</p> <p>NH4+ (mg/L) NO3- (mg/L) NO2- (mg/L)</p>	
--	--	--

Figure 23 : Paramètres disponibles pour la qualité de l'eau potable à l'échelle du registre Néphronor.

b. Localisation des sources d'émissions

Les données concernant la localisation des sources d'émissions se matérialisent par un échantillon de points géolocalisés. La construction d'indicateurs agrégés à la commune à partir de ces données est une problématique complexe. Trois stratégies ont été mises en œuvre et peuvent être comparées. La finalité est de parvenir à transformer ces données ponctuelles en un indicateur spatialisé disponible dans chaque commune de la zone d'étude.

La première stratégie employée consiste à calculer une densité surfacique. Le nombre de sites présents dans une commune est rapporté à la surface de la commune. Cette stratégie est simple, peu sujette à l'erreur, et bien adaptée pour les sources d'émissions qui présentent un effectif important sur la zone d'étude : BASIAS (42 338 sites), ICPE (4379 sites), accidents industriels (4756 sites), intersections de routes (331 608 intersections de routes). En revanche, la pertinence de cet indicateur est plus discutable lorsque la source d'émission étudiée présente un faible effectif sur la zone d'étude : stations d'épuration (829 sites), installations de gestion et de traitement des déchets (614 sites). Dans ce cas de figure, il est préférable de considérer la densité surfacique comme étant une variable catégorielle, indiquant la présence ou l'absence de la source étudiée (figure 24).

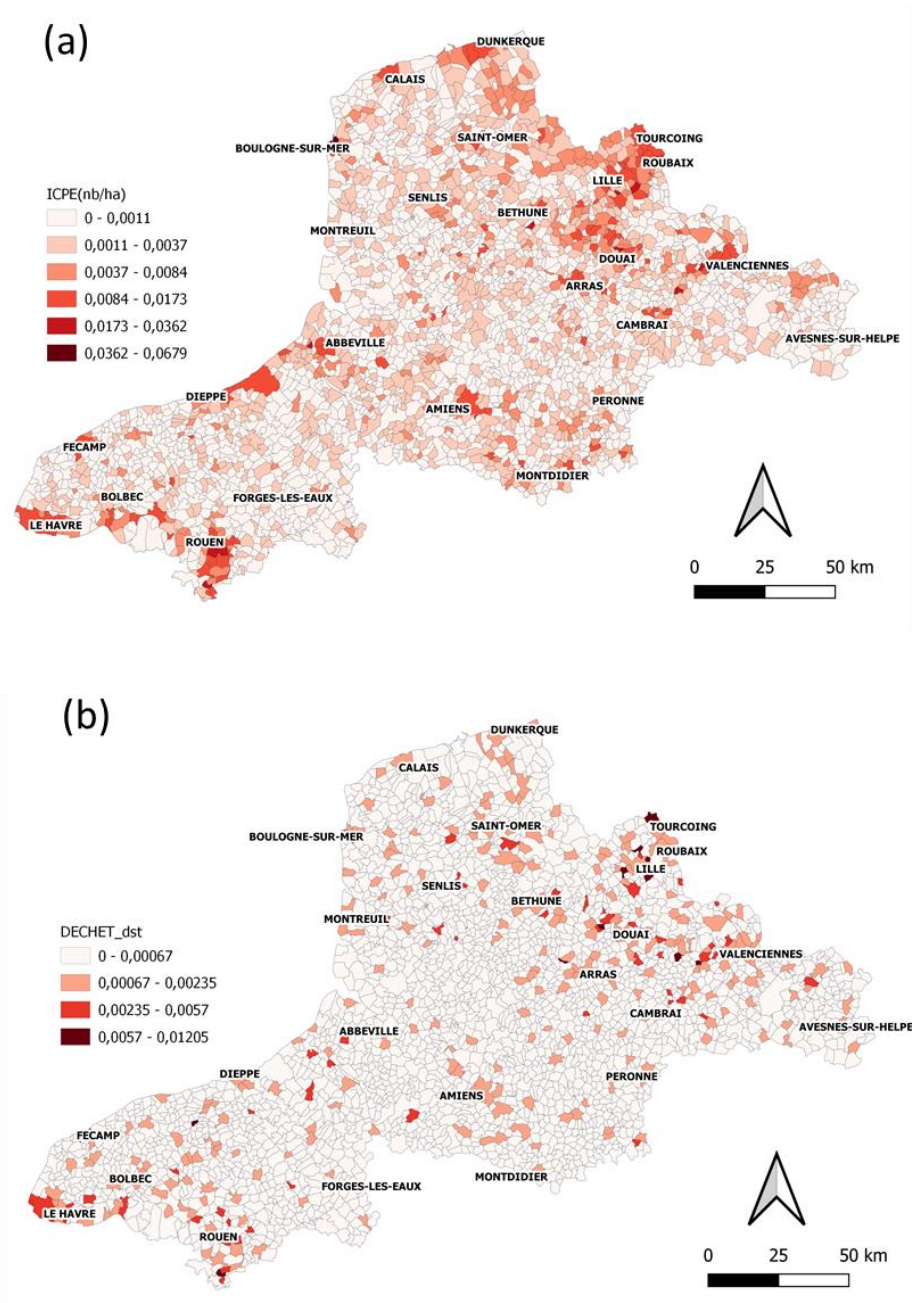


Figure 24 : Densité surfacique calculée à l'échelle communale : (a) installations ICPE ; (b) stations d'épuration.

Deux autres stratégies ont donc été déployées pour étudier les sources d'émissions qui présentent un trop faible effectif dans la zone d'étude. En particulier, ce travail a été effectué pour les sources d'émissions dont l'effectif était inférieur au nombre total de communes dans la zone d'étude (3 041 communes).

Une première option consiste à calculer dans chaque commune la distance entre le centroïde de la commune et le site le plus proche. De cette manière, on génère un indicateur de proximité qui s'exprime en km (Figure 25a). Plus l'indicateur prend des valeurs élevées, plus la commune peut être considérée comme éloignée de la source d'émission étudiée. Cette méthode de calcul a été appliquée

aux stations d'épuration et aux installations de traitement et de gestion des déchets (tous types d'installations confondus).

Une seconde option a été employée pour prendre en compte l'aire d'influence de ces sources d'émission dans le calcul de l'indicateur. Cette démarche n'a été appliquée que sur les incinérateurs et les installations de stockage de déchets. En effet, pour ces deux types d'installations, de nombreuses études ont été menées pour évaluer leur impact sanitaire sur les populations résidant à proximité. Pour les incinérateurs, la plupart des articles considèrent une zone tampon de 3 ou 4 km (Candela et al., 2015; Gatti et al., 2017, 2017; Ji et al., 2022; Xu et al., 2019). Cependant, une publication fait autorité. Il s'agit d'une étude menée au Royaume-Uni où l'incidence du cancer a été suivie chez plus de 14 millions de personnes résidant à proximité de 72 incinérateurs (Elliott et al., 1996). En considérant une zone tampon de 7,5 km de rayon autour des incinérateurs, ils ont observé une diminution significative du risque relatif à mesure que l'on s'éloigne du centre des buffers. Une zone d'influence de 7,5 km de rayon a donc été retenue comme référence pour les incinérateurs. Ainsi, dans chaque commune, la proportion de surface communale située à moins de 7,5 km d'un incinérateur a été calculée (Fig 25b). Enfin, pour les installations de stockage, les publications considèrent également des zones d'influence très diverses (Davoli et al., 2010; Higgs and Langford, 2009; Knox, 2000). De nouveau, une publication en particulier a été prise comme référence (Elliott et al., 2001). Il s'agit d'une étude menée au Royaume-Uni sur près de 10 000 sites d'enfouissement de déchets et sur la base de plus de 8 millions de naissances. Un excès de risque relatif a été observé pour certaines anomalies congénitales chez les populations résidant à moins de 3 km d'une installation de stockage. Une aire d'influence de 3 km a donc été prise pour le calcul de l'indicateur.

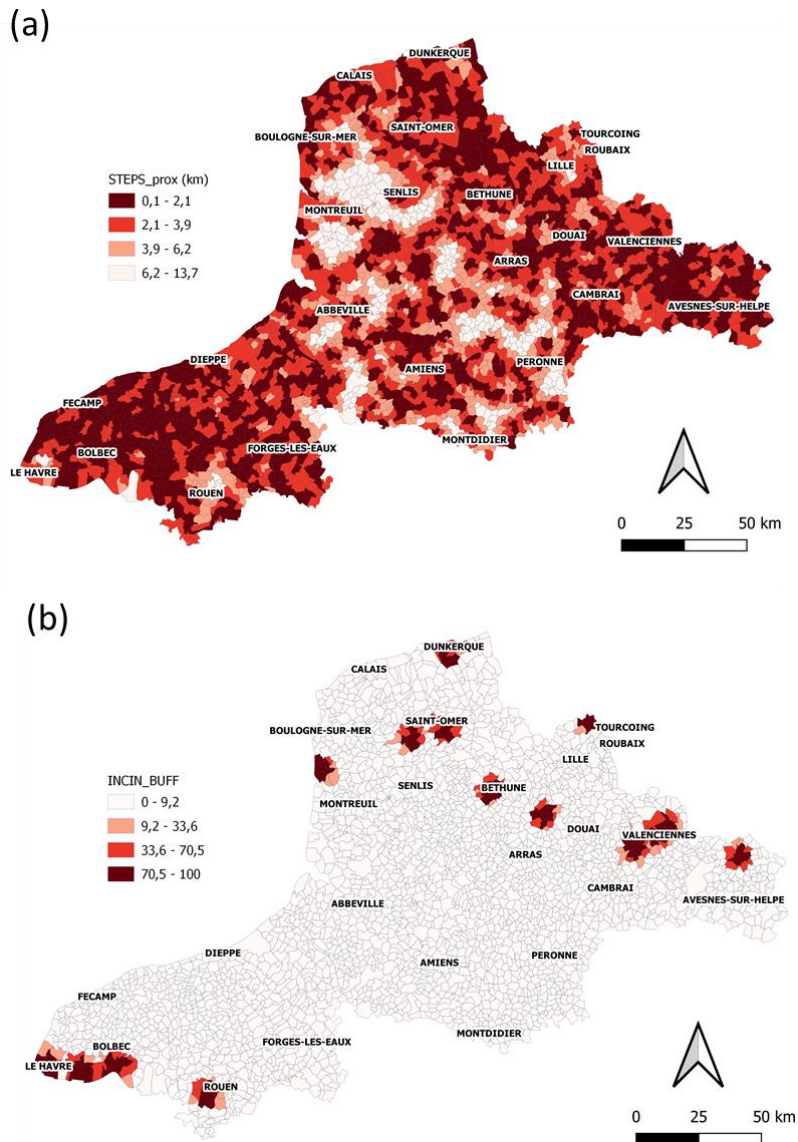


Figure 25 : Alternatives au calcul de la densité surfacique : (a) distance entre le centroïde de la commune et la station d'épuration la plus proche ; (b) pourcentage de la surface communale située à moins de 7,5 km d'un incinérateur de déchets.

Pour résumer, la densité surfacique permet de prendre en compte la présence de plusieurs installations dans une même commune. En revanche, l'inconvénient majeur est que seule la commune d'implantation de l'installation est concernée au moment du calcul. Or les nuisances associées à une source d'émissions ne s'arrêtent pas aux limites administratives d'une commune. Les deux autres alternatives présentent justement l'avantage de prendre en compte la distance à la source d'émission (pas uniquement sa présence ou son absence au sein d'une commune). En revanche, contrairement à la densité surfacique, l'aspect cumulatif et la présence de plusieurs installations n'est pas pris en compte.

Discussion

1. Vue d'ensemble sur le jeu de données.

L'objectif était d'identifier, de collecter et de traiter des bases de données ouvertes pour générer des indicateurs spatialisés et caractériser l'environnement physique. Dans cette optique, une avancée significative a été réalisée puisque 24 sources de données ont été identifiées, exploitées et 113 indicateurs ont été générés à l'échelle du registre Epimad (20 indicateurs supplémentaires disponibles pour Néphronor). Le jeu de données élaboré s'étend sur une vaste zone géographique, englobant quatre départements, avec une résolution spatiale fine au niveau des communes. Pour assurer une couverture spatiale exhaustive, les indicateurs présentant un trop grand nombre de données manquantes ont été écartés de l'analyse. Chaque indicateur est donc disponible pour l'ensemble des 3 041 communes de la zone d'étude. Ils ont été répartis en sept sous-dimensions thématiques : (i) niveaux de contamination des milieux, (ii) niveaux des émissions polluantes, (iii) localisation des sources d'émissions, (iv) occupation des sols, (v) pratiques agricoles, (vi) naturalité, (vii) climat. Ces données permettent d'une part de caractériser les activités humaines : industries, agriculture, transport. D'autre part, elles permettent de décrire les pressions exercées par ces activités sur l'environnement : pollution des milieux (air, eau, sols), émissions, artificialisation des sols, changement climatique. Les aménités environnementales sont également appréhendées via des indicateurs de la naturalité des territoires. Les paragraphes suivants dressent un bilan pour chaque sous-dimension thématique et évoquent les perspectives de production et de structuration des données.

Niveaux de contamination

AIR

Trois indicateurs de contamination atmosphérique ont été retenus dans le jeu de données ($PM_{2.5}$, PM_{10} , NO_2), ils correspondent aux niveaux de contamination moyens sur la période 2011-2017. Les données brutes sont fournies sous la forme d'une grille (format raster) avec une résolution spatiale de 10km de côté. Les modèles de pollution de l'air sont bien documentés par l'Agence européenne de l'environnement (Horálek et al., 2021). Les cartographies sont générées principalement à partir des données mesurées issues des réseaux de surveillance. La méthode de modélisation combine ces données de surveillance avec les résultats d'un modèle de transport chimique, et des données complémentaires comme l'altitude ou la météorologie.

Plus récemment au cours de la thèse, des modélisations annuelles sur la période 2000-2022 ont été développées par l'Institut national de l'environnement industriel et des risques (Ineris). Ces nouvelles données sont issues du modèle chimie-transport CHIMERE développé en partenariat avec le Centre national de la recherche scientifique (CNRS). Contrairement au modèle européen, ce modèle n'est pas calibré à partir des données mesurées. Il utilise en entrée les flux d'émissions et les données météorologiques pour simuler la dispersion, la transformation et l'accumulation des polluants dans l'atmosphère. Les concentrations moyennes annuelles des polluants réglementés ont été produites. De plus, les concentrations moyennes annuelles ont également été pondérées par la population pour fournir un proxy qui se rapproche davantage d'une mesure de l'exposition.

En dehors des contaminants réglementés (particules, oxydes d'azote et ozone), d'autres polluants comme les métaux, les hydrocarbures ou les produits phytosanitaires ne sont pas décrits par des données de contamination couvrant le territoire national. Par ailleurs, un effort de recherche important est mené actuellement pour mieux caractériser les sources et la composition chimique des

particules fines. Cela implique notamment l'identification des composés organiques, inorganiques, métalliques et toxiques qui les composent et leurs implications respectives sur la santé humaine (Yang et al., 2019).

EAU

Les données sur la qualité de l'eau dans les réseaux de distribution d'eau potable ont été exploitées. Elles couvrent l'ensemble du territoire (base de données SISE-Eaux ; Ministère des Solidarités et de la Santé, 2023). Six indicateurs ont pu être générés à l'échelle du registre Epimad (Nord, Pas-de-Calais, Somme et Seine-Maritime), contre 26 sur le périmètre du registre Néphronor (Nord et Pas-de-Calais). Les données ne sont mises à disposition qu'à partir de 2016, bien que la base SISE-Eaux ait été créée en 1994. Des difficultés ont été rencontrées lors de leur exploitation. La méthodologie mise en œuvre pour résoudre ces problèmes est présentée en détail dans les résultats. Selon le département français considéré, les analyses sont confiées à des laboratoires différents, ce qui entraîne des limites de quantification très hétérogènes et, par conséquent, des contraintes d'homogénéisation des valeurs mesurées. Par ailleurs, une forte proportion de données manquantes a été constatée pour de nombreux paramètres dans certaines zones du territoire. Bien que cette base de données soit d'envergure nationale et qu'il s'agisse d'un milieu sous surveillance depuis de nombreuses années, la spatialisation des données à l'échelle nationale serait un réel défi.

Des données relatives à la qualité des eaux continentales existent également, résultant d'une surveillance effectuée sur les masses d'eau de surface et les masses d'eau souterraines par les Agences de l'eau. À ce jour, ces données reflètent principalement la conformité ou la non-conformité aux normes environnementales émanant de la Directive Cadre sur l'Eau. Ce sont des données qualitatives qui décrivent l'état chimique et écologique des masses d'eau (bon ou mauvais). Elles n'ont cependant pas été considérées dans ce travail à cause de leur caractère qualitatif. De plus, leur exploitation n'aurait pas permis d'obtenir une mesure dans chaque commune puisque la répartition des masses d'eau est très inégale sur le territoire.

SOLS

Les données du programme RMQS ont été exploitées pour caractériser la contamination des sols en métaux : arsenic, cadmium, chrome, cuivre, mercure, nickel, plomb, zinc, thallium. Elles sont issues d'une campagne nationale qui a été menée sur la période 2000-2009 (Arrouays et al., 2003). Au total, 2 240 sites de prélèvements ont été répartis uniformément sur le territoire national. Les mailles utilisées pour le plan d'échantillonnage sont des mailles carrées de 16km de côté. A partir de ces données de terrain, les producteurs ont conduit des interpolations spatiales pour fournir des estimations à une résolution de 1000 m. Il faut donc bien garder à l'esprit que l'interpolation spatiale réalisée est relativement imprécise (en particulier en milieu urbain). Par ailleurs, il faut rappeler que les données RMQS sont principalement issues de prélèvements réalisés dans les sols agricoles et forestiers.

Une seconde campagne de prélèvements est actuellement en train d'être menée (2016-2027). Ces nouvelles données permettront d'actualiser les niveaux de contamination en éléments traces métalliques (ETM) dans les sols et d'évaluer l'évolution dans le temps. Pour cette nouvelle campagne, en plus des métaux, des tests de faisabilité sont en cours pour pouvoir suivre la biodiversité dans les sols et les teneurs en pesticides (Froger et al., 2023a). Ces nouvelles données permettraient d'enrichir considérablement la connaissance des sols sur le territoire français.

Enfin, des mesures de la qualité des sols sont effectuées en routine dans le cadre de projets de développement urbain (notamment via des bureaux d'étude). Ces données existent mais elles ne sont pour l'instant pas centralisées et partagées dans une base de données ouverte. Elles seraient pourtant très précieuses et permettraient d'obtenir une résolution spatiale beaucoup plus fine pour caractériser la contamination des sols urbains.

Niveaux des émissions polluantes

Les données d'émission utilisées proviennent du Centre technique de référence en matière de pollution atmosphérique et de changement climatique (Citepa). Les indicateurs ont pu être calculés uniquement à partir de deux années, 2012 et 2015, pour lesquelles des données étaient disponibles. Les polluants suivants ont été décrits : oxydes d'azote, PM_{2.5}, PM₁₀, dioxyde de soufre, monoxyde de carbone, composés organiques volatils non méthaniques, ammoniac, benzène, nickel, plomb, cadmium, arsenic. À l'heure actuelle, les données ouvertes sur les niveaux d'émissions sont agrégées au niveau des établissements publics de coopération intercommunale (EPCI). La résolution spatiale de ces données n'est donc pas très précise, mais elles demeurent pertinentes pour l'étude des polluants qui ne sont pas pris en compte par les indicateurs de contamination.

De plus, une autre source d'information utilisée est un indicateur élaboré par l'Observatoire des Territoires, qui fournit des données communales sur les émissions totales de gaz à effet de serre. Cet indicateur correspond au pouvoir de réchauffement global (PRG), exprimé en milliers de tonnes équivalent CO₂.

Proximité aux sources d'émissions

Plusieurs sources d'émissions potentielles ont été étudiées : sites et sols pollués, installations classées pour la protection de l'environnement (icpe), installations de traitement et de gestion des déchets, stations d'épuration, routes. Le contexte de l'étude impose d'évaluer la proximité aux sources d'émissions en construisant des indicateurs agrégés à l'échelle communale. Cette contrainte est assez difficile à gérer mais des stratégies ont été mises en œuvre pour y parvenir, elles ont été présentées en détail dans les résultats de ce chapitre. Trois méthodes différentes ont été appliquées selon la source étudiée : densité surfacique, calcul basé sur la distance entre le centroïde des communes et la source d'émission la plus proche, calcul prenant en compte l'aire d'influence des sources d'émissions.

Sur cette thématique, beaucoup de questions se posent sur la robustesse des indicateurs produits. Pour le moment, la taille des installations n'est pas prise en compte dans le calcul des indicateurs. Par exemple, tous les incinérateurs sont considérés de la même manière alors qu'ils présentent des tailles ou des capacités de traitement différentes. À l'heure actuelle, ces informations ne sont pas mises à disposition dans les bases de données ouvertes mobilisées. Cependant, si cela était amené à évoluer, une pondération selon la taille de l'installation pourrait être appliquée. De même, pour la densité d'intersections routières, il serait très pertinent d'intégrer dans le calcul de l'indicateur des informations sur la densité du trafic. Enfin, des biais ont été identifiés sur la précision géographique des données brutes. Par exemple, la base de données des anciens sites industriels et activités de services (BASIAS) souffre d'imprécisions au niveau de la localisation de certains sites.

En parallèle, un dernier indicateur a été intégré dans cette sous-dimension. La base de données ARIA (Analyse, Recherche et Information sur les Accidents) est publiée par le Bureau d'Analyse des Risques et Pollutions Industriels (BARPI). Elle répertorie les incidents, accidents ou presque accidents qui ont

porté, ou auraient pu porter atteinte à la santé, à la sécurité publique ou à l'environnement. Les données utilisées ont été téléchargées au cours de l'année 2020. Seuls les événements ayant le statut « Accidents » ont été conservés et la densité d'accidents par hectare de surface communale a été calculée.

Occupation des sols

L'occupation des sols agricoles a été étudiée à partir du registre parcellaire graphique (RPG). Le Registre Parcellaire Graphique (RPG) est une base de données géographique administrée par l'Agence de Services et de Paiement, elle a été instaurée dans le cadre de la politique agricole commune (PAC). Mises à jour chaque année à partir des déclarations des exploitants souhaitant prétendre à une aide de la PAC, ces données fournissent des informations sur l'usage agricole des sols. Il s'agit d'un système d'information géographique représentant chaque année au 1/5000ème les îlots culturaux et leur culture principale (ensemble de parcelles contigües appartenant à une même exploitation). A noter que seules les parcelles des agriculteurs déclarés font l'objet d'un référencement et d'une représentation graphique dans cette base de données. Avant 2015, selon le millésime considéré, la nomenclature de l'usage du sol comporte entre 24 et 28 grands groupes culturaux principaux. Dans le cadre de ce travail, trois millésimes successifs ont été collectés et utilisés : 2010, 2011, 2012.

Les postes d'occupation des sols plus traditionnels ont également été étudiés par l'intermédiaire de la base de données Corine Land Cover (CLC) qui est mise à jour tous les six ans. Cinq postes généraux ont été décrits : territoires artificialisés, territoires agricoles, forêts et milieux semi-naturels, zones humides, surfaces en eau. La moyenne de quatre millésimes a été calculée : 1990, 2000, 2012. Il faut préciser que le CLC propose également des niveaux de classification plus détaillés. Cependant, dans le cadre de cette thèse, il n'a pas été jugé pertinent d'aller jusqu'à ce niveau de détail. Toutefois, dans un contexte de recherche différent, ces données pourraient s'avérer pertinentes. De plus, la base de données CLC permet l'identification de zones présentant une surface minimale de 25 ha. Ainsi les données manquent de précision pour étudier certains phénomènes. Par exemple, elles ne permettent pas de rendre compte du mitage urbain.

L'Observatoire National de l'Artificialisation propose en libre accès des données sur les dynamiques temporelles d'artificialisation des sols. Ces données ont été calculées à partir des Fichiers fonciers, il s'agit d'une base de données nationale d'origine fiscale mise à jour annuellement pour le traitement de la taxe foncière. Cette base de données a été détournée de son usage initial pour évaluer le changement d'usage des sols. Ainsi la variable ARTIF_0919 correspond au pourcentage de surface communale convertie en surface artificialisée entre 2009 et 2019. Plus précisément, il s'agit du rapport entre le flux d'artificialisation sur la période 2009-2019 et la surface communale. Le flux d'artificialisation correspond au nombre d'hectares qui sont passés du statut "Naturel, Agricole ou Forestier" à "Artificialisé" dans la base des Fichiers fonciers. Le même type d'indicateurs aurait pu être calculé à partir du CLC, mais les résultats obtenus auraient été beaucoup moins fiables, du fait de la résolution spatiale plus grossière de la base de données.

Pratiques agricoles

La question des produits phytosanitaires n'a pas pu être traitée via les données de contamination dans l'eau, dans les sols ou dans l'air. La base de données Phyt'Atmo avait pourtant été identifiée, elle rassemble les mesures de pesticides dans l'air ambiant réalisées par les AASQA. Mais la couverture

spatiale était trop parcellaire pour envisager d'exploiter ces données à l'échelle communale. Dans les eaux de distribution d'eau potable, les données manquantes étaient trop nombreuses pour être comblées. Ainsi, pour traiter la question, d'autres alternatives ont été mobilisées. D'abord, les données d'occupation des sols agricoles constituent un premier proxy de l'usage agricole des produits phytosanitaires. En l'absence de données plus précises, des indicateurs décrivant l'occupation des sols sont souvent utilisés dans des études en santé-environnement (Cornelis et al., 2009; Coste et al., 2020). Pour mieux approcher la notion d'usage, les données d'occupation peuvent être couplées avec des données sur les pratiques culturales (Bukalasa et al., 2017; Ward et al., 2000). L'indice de fréquence de traitement (IFT) est souvent utilisé pour cela, il mesure le nombre de doses homologuées de produits phytosanitaires appliquées par hectare pendant une campagne culturale. L'intérêt de cet indice est de permettre d'agréger des substances très différentes et donc de mesurer une pression pesticide globale. En menant des enquêtes auprès des agriculteurs, ces IFT peuvent être estimées par type de culture et pour une région donnée. En France, l'Agreste met à disposition les IFT moyens par grand type de culture pour chaque région administrative (Agreste, 2020a). Très récemment, ce type d'approche a été utilisée par Solagro pour construire l'indicateur Adonis à l'échelle des communes et sur le périmètre national (Chayre et al., 2023; Chayre and Pointereau, 2023). Ils ont combiné des données d'occupation des sols agricoles avec les IFT moyens régionaux. De cette manière, ils sont parvenus à générer à l'échelle communale un indice de pression phytosanitaire. Une méthode très similaire avait préalablement été développée en Belgique pour cartographier l'utilisation des pesticides agricoles (Habran et al., 2022). La limite principale de ces méthodes est de considérer des IFT moyens régionaux, qui ne permettent pas de prendre en compte les disparités infrarégionales dans les pratiques agricoles.

Pour ce travail de thèse, une autre stratégie a été retenue avec l'utilisation de la BNVD qui a permis de construire des indicateurs de pression phytosanitaire. La BNVD a été créée en 2009 à des fins fiscales dans le cadre de la redevance pour pollutions diffuses. Depuis cette date, les distributeurs agréés de produits phytosanitaires sont contraints de déclarer annuellement le bilan de leurs ventes. Pour chaque vente est renseigné le code postal du distributeur concerné ainsi que la quantité de produit ou de substance active vendue (en kg). Et depuis 2013, les distributeurs sont également contraints de rendre public le registre des achats. Ainsi désormais, à chaque vente est également associée le code postal de l'acheteur. Des indices de pression phytosanitaire ont été calculés en considérant l'ensemble des substances actives ou en considérant les substances par fonction (herbicides, insecticides, fongicides, facteurs de croissance). Des indices ont également été calculés individuellement pour quelques substances actives spécifiques. Tous ces indicateurs s'expriment en kg de substance(s) active(s) achetée(s) par hectare de surface communale et par an.

Dans quelques rares pays, les agriculteurs consignent avec précision tous les traitements phytosanitaires qu'ils réalisent au cours de l'année. Ces données sont ensuite centralisées, ce qui permet d'évaluer de manière plus précise l'utilisation spatiale des pesticides. En particulier, depuis 1990, le programme « Pesticide Use Reporting » a été mis en place aux Etats-Unis dans l'Etat de Californie (Richardson et al., 2017). En Europe, à notre connaissance, le Danemark est le seul pays ayant mis en place ce système depuis 2010 (Kudsk et al., 2018). A terme, ces pratiques seront peut-être amenées à être mises en place en France également.

La base de données S3IC recense les installations classées pour la protection de l'environnement (ICPE) et classe ces installations selon leur domaine d'activité. De cette base de données ont été extraits tous les élevages soumis à enregistrement ou autorisation. Ensuite la densité d'élevages ICPE par hectare de surface communale a été calculée. Il est important de noter que les élevages sont soumis à

enregistrement ou autorisation à partir du moment où les effectifs d'animaux dépassent un certain seuil. Par exemple, un élevage porcin est soumis à enregistrement à partir de 400 têtes et soumis à autorisation à partir de 800 têtes. De ce fait, ces données sont loin d'être exhaustives et seuls les plus gros élevages sont recensés. En 2017, la base de données recensait 19 742 élevages à l'échelle nationale, soit moins de 10 % des élevages français (Parisse et al., 2018). De plus, l'indicateur final ne donne pas d'information sur le nombre d'animaux. Par ailleurs, la variable de densité d'élevages ICPE n'est pas informative en termes de nombre d'animaux.

Pour obtenir des données plus précises et qui ne s'appuient pas uniquement sur le nombre d'élevages mais également sur le nombre d'animaux, des données de l'Agreste issues du recensement agricole de 2010 ont été utilisées. En particulier, le nombre d'unités de gros bétail (UGB) par hectare de surface communale a été calculé. L'unité de gros bétail (UGB) est une unité de référence permettant d'agréger le bétail de différentes espèces et de différents âges en utilisant des coefficients spécifiques établis initialement sur la base des besoins nutritionnels ou alimentaires de chaque type d'animal. Par exemple un bovin de moins d'un an compte pour 0,4 UGB tandis qu'une poule pondeuse compte pour 0,014 UGB. Ainsi cette variable permet à la fois de prendre en compte le nombre d'animaux et leur besoin alimentaire qui peut être considéré comme un proxy indirect de la quantité de déchets produits.

Les données du dernier recensement agricole de 2020 ont été publiées récemment, au cours de l'été 2022. Dans le cadre de cette thèse, la phase de collecte des données était malheureusement close mais il serait intéressant pour la suite d'intégrer ces nouvelles données. Elles sont à priori très proches de celles proposées pour le recensement précédent, mais elles permettraient d'actualiser des données qui commencent à dater.

Les pratiques agricoles ont également été caractérisées sur le plan environnemental avec l'indicateur de haute valeur naturelle (HVN), développé par Solagro (Pointereau et al., 2010; Poux and Pointereau, 2014). Son calcul s'appuie sur trois sous-indices : (i) la diversité d'assolement, qui tient compte de la variété des cultures présentes sur les exploitations ; (ii) l'extensivité des pratiques, qui évalue le niveau d'intrants et la charge animale ; (iii) la densité des infrastructures agroécologiques, qui identifie la présence d'éléments fixes du paysage ou d'habitats semi-naturels. Ces indicateurs constituent une véritable valeur ajoutée et contribuent grandement à la caractérisation des pressions et aménités environnementales liées à l'agriculture.

Naturalité

Mesurer la biodiversité au niveau municipal sur l'ensemble du territoire national est un véritable défi qui nécessiterait des protocoles de prospection standardisés pour tous les taxons. C'est pourquoi la naturalité des territoires a été étudiée en utilisant des stratégies plus indirectes. Par exemple, un indice de naturalité fourni par l'UICN a été utilisé (Guetté et al., 2021). Son calcul repose sur trois sous-indicateurs : (i) l'intégrité biophysique de l'occupation du sol, qui mesure l'écart d'un paysage par rapport à son état en l'absence de transformations anthropiques ; (ii) la spontanéité des processus, qui évalue l'ampleur de l'intervention humaine dans les dynamiques naturelles ; (iii) la continuité spatiale, qui correspond au degré de connectivité des paysages. Par ailleurs, un autre indicateur a été développé pour évaluer la présence de réservoirs de biodiversité dans les communes. Les données utilisées pour cet indicateur sont issues du programme national « Trame verte et bleue », qui visait notamment à identifier et cartographier les réservoirs de biodiversité (Billon et al., 2017). Enfin, un indicateur d'accessibilité aux espaces verts a été construit. Il évalue la part de la population communale ayant accès à un espace vert dans un rayon de 400 m autour de son lieu d'habitation. Cependant, il convient

de rappeler que cette notion est assez urbano-centrée et que le calcul n'a probablement pas la même portée pour étudier le phénomène en milieu rural. A l'avenir, des réflexions pourront être menées pour mieux caractériser ce phénomène sur des territoires aussi contrastés.

Climat

Les données climatiques sont issues de la base de données européenne E-Obs qui rassemble des données climatiques à l'échelle européenne depuis 1950 jusqu'à aujourd'hui. Les données sont disponibles sur un pas de temps quotidien et la résolution spatiale est une maille carrée de 11,1 km de côté. Les données sont très volumineuses et se présentent sous la forme de rasters spatio-temporels. Trois grands types d'indicateurs ont été générés. D'abord, la valeur moyenne a été évaluée à l'échelle communale sur toute la période 1950-2019 pour les paramètres suivants : température, niveaux de précipitations, rayonnement global. Dans ce cas, l'indicateur s'exprime dans la même unité que le paramètre en question. Ensuite, certains indicateurs ont été construits pour quantifier le nombre de jours où un paramètre a dépassé une certaine valeur sur l'ensemble de la période 1950-2019. Ce mode de calcul a été utilisé pour étudier les épisodes de canicules et les fortes précipitations. Enfin, des indicateurs ont été calculés pour mesurer l'évolution d'un paramètre climatique entre deux périodes temporelles (en l'occurrence la période 1985-2019 a été comparée à la période 1950-1984).

Dans le cadre de cette thèse, ces indicateurs climatiques ont été agrégées sur de longues périodes. Mais la résolution quotidienne des données ouvre des perspectives diverses en santé-environnement. Certaines études s'intéressent par exemple à l'impact des épisodes caniculaires sur la santé des populations (Benmarhnia et al., 2017; Chien et al., 2016; Sheridan and Lin, 2014).

2. Temporalité des données environnementales

Au-delà des phénomènes décrits, il est également important de s'intéresser à la temporalité des données environnementales disponibles. Un schéma a été créé pour avoir une vision d'ensemble sur la temporalité de chacune des données sources exploitées (Figure 26).

Sous dimension	Phénomène étudié (Source)	1950	1960	1990	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
		Niveaux de contamination des milieux	Qualité de l'air (AirBase)																							
Qualité des sols (RMQS)																										
Qualité des eaux de distribution (Sise-Eaux)																										
Niveaux des émissions	Emissions atmosphériques (Citepa)																									
	Emissions de gaz à effet de serre (Observatoire des territoires)																									
Localisation des sources d'émissions	Sites et sols pollués (BASIAS)																									
	Sites et sols pollués (BASOL)																									
	Installations classées pour la protection de l'environnement (S3IC)																									
	Stations d'épuration (ADEME)																									
	Intersections routières (BD TOPO)																									
	Installations de gestion et de traitement des déchets (Sinoe)																									
	Accidents industriels (ARIA)																									
Occupation des sols	Potential radon (IRSN)																									
	Occupation des sols (CLC)																									
	Occupation des sols agricoles (RPG)																									
Pratiques agricoles	Dynamiques d'artificialisation des sols (Cerema)																									
	Pression phytosanitaire (BNVD)																									
	Elevage (S3IC)																									
	Elevage (Recensement agricole)																									
Naturalité	Haute Valeur Naturelle (Solagro)																									
	Naturalité potentielle (UICN)																									
	Réservoirs de biodiversité (Trame verte et bleue)																									
Climat	Accès aux espaces verts (Open Street Map)																									
	Climat (E-obs dataset)																									

Figure 26 : Temporalité des données environnementales retenues et exploitées.

La figure illustre une couverture temporelle sur plus de 50 ans pour certaines sources de données spécifiques : données climatiques et inventaire des accidents industriels. Les données d'occupation du sol permettent également de remonter assez loin dans le temps. En particulier, le premier millésime de la base de données Corine Land Cover date de 1990. Cependant, une réelle densification de la disponibilité des données est constatée depuis les années 2010.

Le mode de publication des données est très variable selon le phénomène décrit. Par exemple, les données sur la contamination métallique des sols s'étalent sur la période 2000-2009, ce qui correspond en fait à la période de campagne au cours de laquelle les sols français ont été prélevés. Pour la localisation des sources d'émission, les bases de données sont constamment mises à jour mais elles ne fournissent pas d'historique de ces modifications (BASOL, BASIAS, ICPE). L'année indiquée sur la figure 26 correspond donc à la date où ces données ont été téléchargées. Enfin, de nombreuses bases de données sont millésimées annuellement (e.g. occupation des sols agricoles, pression phytosanitaire, émissions polluantes, pollution atmosphérique, etc...).

La temporalité des données environnementales est une question importante dans une perspective de santé-environnement. La résolution temporelle et la disponibilité des données dans le temps sont très variables selon la source de données considérée. Il est donc important d'avoir une bonne connaissance des phénomènes décrits pour être en mesure d'évaluer la pérennité de ces données dans le temps. Par exemple, la contamination métallique dans les sols est relativement stable dans le temps. Wang et al. (2023) ont étudié la contamination métallique dans plus de 1 030 sols agricoles chinois entre 2000 et 2022. Ils ont observé une tendance marginale à la hausse pour certains métaux uniquement (cadmium et mercure). Dans ce cas, une seule mesure à un instant donné peut être considérée comme représentative de la contamination des sols sur plusieurs années. En revanche, l'appréciation sera différente s'il s'agit par exemple de la contamination microbiologique dans les eaux de distribution.

Pour le moment, lorsque les données le permettaient, les indicateurs ont été construits en agrégeant temporellement les données. La plupart des indicateurs ont été construits en appliquant la moyenne sur une période donnée à partir de plusieurs millésimes. Cette stratégie permet d'être plus représentatif d'une dynamique spatiale globale, en réduisant l'incertitude liée aux variations temporelles. Par exemple, pour l'occupation des sols agricoles ou la pression phytosanitaire, plusieurs années ont été considérées pour le calcul des indicateurs. Cela permet de prendre en compte la rotation des cultures, les conditions climatiques ou la pression parasitaire qui varient d'une année à l'autre. Pour la contamination des eaux de distribution, le fait d'avoir intégré cinq années de mesures dans le calcul paraît également essentiel. La représentativité temporelle du résultat sur une seule année poserait question. Par exemple, si une unité de distribution n'a fait l'objet que d'une seule mesure au cours de l'année, juste après un épisode orageux, le risque est de surestimer le niveau de contamination et de ne pas être représentatif de la qualité de l'eau sur le long terme. Alors qu'en prenant en compte l'ensemble des mesures réalisées sur 5 ans, comme cela a été fait, on réduit les incertitudes et on optimise la comparabilité entre unités spatiales.

Cette méthode d'agrégation temporelle présente cependant des inconvénients également. Elle entraîne une perte d'information et masque certains phénomènes ponctuels, comme les pics de pollution, qui peuvent avoir un impact significatif sur la santé. Dans un autre contexte, cette piste aurait pu être explorée. Mais dans le cadre de cette thèse, il fallait avant tout concentrer les efforts sur la caractérisation de la contamination des milieux sur le long terme. D'autant que les données de santé sont elles-mêmes agrégées temporellement sur de longues périodes, et qu'il s'agit de pathologies qui se développent sur le long terme. Ainsi, l'agrégation temporelle était la meilleure stratégie à adopter dans ce contexte, elle permet de réduire les incertitudes et d'optimiser la comparabilité entre unités spatiales.

Conclusion

Ce travail de collecte, de synthèse et d'analyse critique des données environnementales répond à plusieurs enjeux importants liés à recherche en santé-environnement. D'abord, de nombreuses sources de données environnementales librement accessibles ont été identifiées, décrites et exploitées. A partir de 24 bases de données sources soigneusement sélectionnés, 113 indicateurs spatialisés ont été développés et cartographiés à l'échelle communale sur le périmètre du registre Epimad (3 041 communes). Un regard critique a également été formulé sur les données sources utilisées et sur les indicateurs construits. Par ailleurs, l'utilisation exclusive de données librement accessibles et disponibles à l'échelle nationale offre la possibilité d'étendre ce travail sur l'ensemble du territoire national.

Par rapport au constat initial formulé à la fin du chapitre 1, l'environnement physique est désormais décrit de manière beaucoup plus exhaustive. Le jeu de données ainsi construit permet de décrire le niveau de contamination des milieux (air, eau, sols) ; le niveau des émissions polluantes ; la localisation des sources d'émissions ; l'occupation des sols ; les pratiques agricoles ; la naturalité des territoires ; le climat. L'enjeu principal est désormais de sélectionner les méthodes les plus appropriées pour exploiter ce jeu de données.

CHAPITRE III – Evaluer la multi-contamination des milieux environnementaux.

Introduction

Le jeu de données construit dans le chapitre précédent peut dans un premier temps être exploité en se focalisant sur les données de multi-contamination des milieux environnementaux. Au total, 38 indicateurs de contamination ont été générés à l'échelle communale : 3 polluants atmosphériques sur la période 2011-2017 (AirBase ; Horálek et al., 2021), 9 éléments traces métalliques dans les sols sur la période 2000-2009 (RMQS ; Arrouays et al., 2003), 26 indicateurs de qualité de l'eau sur la période 2016-2020 (SISE-Eaux ; Ministère des Solidarités et de la Santé, 2023). Ces données nous amènent à envisager la caractérisation de la multi-contamination des milieux et l'évaluation des associations avec la distribution spatiale d'incidence de l'IRCT et de la MC. Il nous faut donc nous interroger sur la méthode à utiliser ou à développer pour apprécier la multi-contamination de l'air, de l'eau et des sols à l'échelle communale à partir de ces données ouvertes. Il faut également considérer, tester et valider la pertinence de ces données pour mettre en évidence des pistes étiologiques en lien avec la MC et l'IRCT.

Les indices composites de multi-contamination constituent des outils pertinents pour évaluer la qualité globale des milieux environnementaux. Ces indices sont utilisés pour synthétiser plusieurs indicateurs de contamination environnementale en une mesure globale, permettant ainsi d'évaluer et de comparer les niveaux de contamination dans les différents milieux. Ils permettent d'appréhender l'impact cumulatif des contaminants sur l'environnement ou sur la santé humaine (Genin et al., 2022a; Hamroun et al., 2023a; Sun et al., 2019). Ce type d'approche a été largement utilisé pour évaluer la qualité de l'air (Kanchan et al., 2015), et dans une moindre mesure pour évaluer la qualité de l'eau (Han et al., 2022), et des sols (Kowalska et al., 2018). Dans la littérature, de nombreuses méthodologies différentes sont proposées, sans qu'il n'y ait consensus. Les méthodes de calcul se distinguent essentiellement par la nature et le nombre de polluants considérés, et par la méthode d'agrégation utilisée pour construire l'indice (Burgass et al., 2017; OECD et al., 2008a). Ces choix dépendent principalement des objectifs de l'étude, de la disponibilité des données et des connaissances sur les contaminants considérés. Quel que soit l'objectif et le milieu environnemental étudié, plusieurs stratégies peuvent être distinguées.

Une première manière de raisonner consiste à comparer les valeurs mesurées dans l'environnement à des valeurs de référence. Le type de référence choisi dépend essentiellement du contexte de l'étude. Pour des études de portée environnementale, les niveaux de fond sont souvent utilisés comme référence. Ces niveaux correspondent aux concentrations attendues ou habituellement mesurées dans l'environnement et sont souvent spécifiques d'une région donnée. Ce type de méthode est par exemple utilisée pour la biosurveillance lichénique avec le « Tomlinson Pollution Load Index » (Angulo, 1996) ou le « Ratio d'Imprégnation Moyen » (RIM) (Occelli et al., 2016). Pour la contamination des sols, on parlera généralement de fond géochimique (Barbieri, 2016). Pour des études s'intéressant plus spécifiquement à l'impact sur la santé humaine, des seuils sanitaires peuvent être utilisés comme valeurs de référence. Par exemple, un indice de qualité de l'air a été développé par l'USEPA, son calcul s'appuie sur des seuils de concentration qui correspondent aux normes nationales de qualité de l'air ambiant (USEPA, 1999). Quelle que soit la portée de cet indicateur (santé ou environnement), l'enjeu principal est de détecter des anomalies ou dépassements. Mais des variations existent au niveau de la méthode d'agrégation retenue. Par exemple, pour calculer l'indice ATMO largement utilisé en France

pour évaluer la qualité de l'air, un sous-indice est calculé pour chaque polluant et c'est le sous-indice maximal qui constitue l'indice final caractérisant la qualité de l'air globale (Garcia and Colosio, 2001). Pour le RIM, un sous-indice est calculé pour chaque polluant en appliquant le ratio entre la valeur mesurée et la valeur de référence sélectionnée (Occelli et al., 2016). Puis l'indice de multi-contamination est généré en calculant la moyenne des différents sous-indices.

D'autres méthodes reposent sur le calcul d'indices composites dont l'objectif principal est de discriminer les unités spatiales les unes des autres, sans volonté particulière de générer un indice qui s'interprète facilement. Dans ce cas, on parle généralement de score de multi-contamination. De nombreuses méthodes de calcul sont possibles. Très souvent dans la littérature, une étape préalable de standardisation des indicateurs est menée avant l'agrégation, elle permet d'attribuer un poids relativement similaire aux différents polluants. Par exemple, la méthode du min-max permet de ramener tous les indicateurs de contamination sur une même échelle comprise entre 0 et 1. Une autre alternative consiste à classer les unités spatiales par ordre croissant et à leur attribuer un rang (Lanier et al., 2019a; Occelli et al., 2020b). Dans ce cas, les écarts de valeurs ne sont pas conservés mais l'objectif est avant tout de hiérarchiser les unités spatiales les unes par rapport aux autres. Pour ce qui est de l'étape d'agrégation, la somme ou la moyenne est essentiellement utilisée. La question porte surtout sur l'inclusion d'une pondération experte des polluants, par exemple en fonction de leur dangerosité intrinsèque.

Enfin, une autre alternative consiste à utiliser des méthodes statistiques multivariées comme l'ACP (Bishoi et al., 2009; Genin et al., 2022a; Hamroun et al., 2023b; Sun et al., 2019). Dans ce cas, le mode de calcul est très similaire à celui utilisé pour les indices de vulnérabilité et de résilience. Dans la pratique, cela revient finalement à pondérer chacun des polluants dans le calcul de l'indice. Cependant, cette pondération est objectivée par un traitement statistique et se base sur l'analyse des données elles-mêmes.

Le premier objectif de ce chapitre est d'évaluer la multi-contamination de l'air, de l'eau et des sols à l'échelle des communes grâce à des indices composites que nous allons comparer. A partir des données disponibles en open data, et avec l'objectif d'établir des associations avec des données de santé, quelle est la méthodologie la plus appropriée pour construire ces indices et étudier le phénomène de multi-contamination ? Une fois la méthodologie sélectionnée, il s'agira d'étudier les associations entre les indices de multi-contamination dans chacun des milieux et la répartition spatiale d'incidence des pathologies étudiées.

PARTIE 1 : Comment caractériser la multi-contamination à l'échelle des territoires ?

Dans cette première partie, nous avons entrepris une évaluation comparative de quatre méthodes pour la construction d'indices composites de multi-contamination. L'objectif est d'analyser les résultats pour décrire les éventuelles différences entre ces approches. Si des différences importantes sont observées, une décision sera prise quant à la méthode à privilégier. Si les résultats sont sensiblement les mêmes, nous opterons pour la simplicité en choisissant l'indice le plus facile à interpréter.

1. Matériel & Méthodes

a. Données de multi-contamination environnementale

Les données utilisées décrivent la contamination des milieux (air, eau, et sol). Des indicateurs spatialisés ont été développés dans le chapitre II pour caractériser les niveaux de contamination en divers polluants à l'échelle des communes.

Trois polluants atmosphériques ont été décrits sur le périmètre du registre Epimad (n=3 041 communes). Les indicateurs spatialisés développés correspondent à la moyenne interannuelle de la contamination en PM_{2.5}, PM₁₀ et NO₂ sur la période 2011-2017 (AirBase ; Horálek et al., 2021).

Des indicateurs spatialisés de contamination des sols ont été développés à partir du réseau RMQS dont la campagne s'est échelonnée entre 2000 et 2009 (Arrouays et al., 2003). Neuf éléments traces métalliques ont été décrits à l'échelle communale sur le périmètre du registre Epimad (n=3 041 communes) : arsenic, cadmium, chrome, cuivre, mercure, nickel, plomb, thallium, zinc.

La qualité des eaux de distribution d'eau potable a été étudiée via la base de données SISE-Eaux (Ministère des Solidarités et de la Santé, 2023). Les indicateurs spatialisés développés à l'échelle des communes correspondent à la valeur moyenne de l'ensemble des mesures réalisées sur la période 2016-2020. A l'échelle du registre Néphronor (n= 1 541 communes), 26 indicateurs spatialisés ont été développés. Ils permettent d'étudier d'une part les 8 paramètres suivants : pH, conductivité, turbidité, dénombrement des entérocoques, dénombrement Escherichia coli, dénombrement des coliformes, dénombrement des microorganismes revivifiables à 22°C, dénombrement des microorganismes revivifiables à 36°C. D'autre part, les concentrations des 18 polluants suivants ont été étudiées : chlore libre, chlore total, nitrates, nitrites, ammonium, plomb, nickel, fer, cuivre, chrome, antimoine, chloroforme, bromoforme, bromodichlorométhane, chlorodibromométhane, trihalométhanes, chlorure de vinyle, perchlorate. En se plaçant à l'échelle du registre Epimad (n=3 041 communes), seuls six indicateurs spatialisés ont pu être calculés en raison de données manquantes trop nombreuses.

Ainsi, pour étudier la multi-contamination des milieux, la zone d'étude retenue correspond au périmètre du registre Néphronor (Nord et Pas-de-Calais). Les indices de multi-contamination ont donc été calculés séparément pour chacun des milieux environnementaux sur les 1 541 communes de la zone d'étude.

b. Méthodologies testées

La première méthode est basée sur un indice de multi-contamination appelé Ratio d'Imprégnation Moyen (RIM). Cet indice a été développé par Ocelli et al. (2016) en s'inspirant des travaux de Chen et

al. (2005). Initialement il a été développé pour la biosurveillance lichénique. La construction de cet indice s'appuie sur la comparaison entre des concentrations mesurées et des valeurs de référence.

La seconde méthode mobilisée repose sur une analyse en composantes principales (ACP). Le mode de calcul est très proche de celui utilisé pour construire les indices de vulnérabilité et de résilience.

La troisième méthode s'appuie sur une étape préalable de standardisation des indicateurs entre 0 et 1. Ces indicateurs sont ensuite sommés pour obtenir un score de multi-contamination.

Enfin, une dernière alternative a été testée. Pour chaque indicateur de contamination, les unités spatiales sont d'abord classées par ordre croissant et un rang leur est attribué. Ensuite, la somme de ces rangs est calculée pour obtenir le score.

Ratio d'Imprégnation Moyen (RIM)

Le ratio d'imprégnation moyen (RIM) est un indicateur initialement développé pour la biosurveillance lichénique et plus spécifiquement la multi-contamination métallique (Occelli et al., 2016). C'est un indice d'enrichissement du milieu qui représente le niveau d'imprégnation moyen par rapport aux concentrations naturellement retrouvées dans les lichens. Pour chaque polluant, un sous-indice est calculé en calculant le ratio entre la concentration mesurée et le niveau de fond. L'indice global (RIM) est obtenu en calculant la moyenne de ces sous-indices. Autrement dit, le RIM révèle le niveau de dépassement global par rapport aux niveaux de fond. L'indice se calcule de la manière suivante :

Ratio d'Imprégnation Moyen (RIM)

$$RIM = \frac{\sum Ci/Bi}{n}$$

C_i = concentration de l'élément
 B_i = niveaux de fonds
 n = nombre d'éléments

Cet indice s'interprète facilement. Lorsque le RIM prend une valeur supérieure à 1, cela signifie qu'en moyenne pour les polluants considérés, les concentrations mesurées sont supérieures aux valeurs de référence. Par ailleurs, il est également possible d'examiner individuellement le résultat des sous-indices, pour analyser les niveaux de dépassement pour chaque polluant.

Pour le sol, les valeurs de référence utilisées sont issues du référentiel pédo-géochimique du Nord-Pas-de-Calais (Sterckeman et al., 2002). Parmi les données RMQS sur la contamination des sols, on distingue les concentrations totales et les concentrations de la phase extractible. Or les données sur la phase extractible ne sont pas disponibles pour certains métaux (arsenic, mercure, thallium). Ainsi, ce sont les indicateurs mesurant les concentrations totales qui ont été utilisées dans le calcul de l'indice. Les polluants suivants ont été considérés : arsenic, cadmium, chrome, cuivre, mercure, nickel, plomb, thallium, zinc.

Pour l'air et l'eau, les valeurs de référence utilisées correspondent à des normes sanitaires (Figure 27). Pour le compartiment atmosphérique, les lignes directrices mondiales sur la qualité de l'air de l'OMS (valeurs moyennes annuelles mises à jour en 2021) ont été considérées (OMS, 2021). Trois polluants

atmosphériques ont été considérés dans le calcul de l'indice : PM₁₀, PM_{2,5} et NO₂. Pour le compartiment eau, les limites de qualité des eaux destinées à la consommation humaine (définies par le code de la santé publique) ont été utilisées. Les paramètres suivants ont donc été pris en compte dans le calcul de l'indice : chlorure de vinyle, chrome, cuivre, nitrites, nitrates, nickel, plomb, antimoine, trihalométhanes, turbidité de l'eau. Les paramètres qui n'étaient pas régis par une limite de qualité ont donc été exclus du calcul du RIM. Ainsi, pour les autres indices de multi-contamination testés, afin de garantir la comparabilité des résultats, les mêmes paramètres ont été pris en compte dans le calcul.

Lignes directrices mondiales de l'OMS sur la qualité de l'air (valeurs moyennes annuelles) :	Référentiel Pédo-Géochimique du Nord-Pas-de-Calais : (Sterckeman et al., 2007)	Limites de qualité des eaux destinées à la consommation humaine (code de la santé publique) :
PM10 : 15 µg/m ³ PM2.5 : 5 µg/m ³ NO2 : 10 µg/m ³	Arsenic : 10 mg/kg Cadmium : 0,44 mg/kg Chrome : 53,1 mg/kg Cuivre : 15,4 mg/kg Mercure : 0,079 mg/kg Nickel : 20,9 mg/kg Plomb : 35,8 mg/kg Zinc : 69,7 mg/kg Thallium : 0,42 mg/kg	Antimoine : 0,10 µg/L Chlorure de vinyle : 0,50 µg/L Chrome : 50 µg/L Cuivre : 2 µg/L Nickel : 20 µg/L Nitrates : 50 mg/L Nitrites : 0,50 mg/L Plomb : 10 µg/L Trihalométhanes : 100 µg/L Turbidité : 1 NFU

Figure 27 : Valeurs de référence sélectionnées pour le calcul du RIM.

Score ACP

La seconde méthode mobilisée consiste à construire un indice composite pour chaque milieu à partir d'une analyse en composantes principales (ACP). La méthode s'inspire du travail de construction des indices de vulnérabilité et de résilience (Brousliche et al., 2020a). Le mode de calcul est identique, simplement la première étape de sélection des données n'est pas nécessaire. En effet, pour construire un indice composite, il est théoriquement recommandé d'examiner les structures de corrélation entre indicateurs pour limiter la redondance d'information en éliminant certains indicateurs trop fortement corrélés entre eux. Cependant, dans ce contexte de multi-contamination, si deux contaminants sont fortement corrélés entre eux, cela ne constitue pas une redondance d'information. Cela signifie plutôt que leur répartition dans l'espace est similaire. Par conséquent, il n'y a pas de raison d'exclure certains contaminants, car l'objectif est précisément d'étudier la multi-contamination et d'approcher la notion de cumul. Pour résumer, le calcul de l'indice se décompose donc en quatre étapes :

- 1) Normalisation des variables (centrage et réduction)
- 2) ACP (seules les composantes principales présentant une valeur propre supérieure à 1 sont conservées)
- 3) Calcul de l'indice :

$$\sum_1^n$$
 (*proportion de variance expliquée par une composante * score de cette composante*)
n : nombre de composantes présentant une valeur propre supérieure à 1.
- 4) Standardisation de l'indice entre 0 et 1 pour une meilleure lisibilité (méthode du min-max)

La particularité de cette méthode est que chacun des polluants est pondéré dans le calcul de l'indice. Mais cette pondération est objectivée par un traitement statistique (ACP), elle s'appuie sur l'analyse des données d'entrée.

Score des Rangs

Cette méthode s'inspire d'un travail préalable qui visait à évaluer l'impact cumulatif des déterminants de santé (Lanier et al., 2019a). Pour chaque indicateur de contamination, les unités spatiales ont été classées par ordre croissant puis un rang leur a été attribué. Ce rang correspond à leur position dans la série statistique. Le rang 1 est attribué à la valeur la plus basse et à chaque valeur supérieure, une incrémentation de 1 est appliquée. En cas d'égalité entre deux communes, la continuité des rangs est maintenue. Concrètement, si deux communes présentent la même valeur et occupent la troisième position, les deux hériteront du rang 3, tandis que la valeur suivante se verra attribuer le rang 4. Ainsi, dans la pratique, les rangs ne s'échelonnent pas toujours entre 1 et 1 541 (nombre total de communes dans la zone d'étude). Enfin, le score de multi-contamination est obtenu en calculant la somme de ces rangs. Cette méthode se différencie des autres dans le sens où les indicateurs de contamination sont discrétisés et que les écarts de valeurs ne sont donc pas conservés pour le calcul de l'indice final.

Score Min-Max

Les indicateurs de contamination sont d'abord standardisés entre 0 et 1 avec la méthode du min-max. Ainsi pour chaque polluant, la commune qui présente le niveau de contamination le plus bas prend la valeur 0. Au contraire, la commune qui présente le niveau de contamination le plus élevé prend la valeur 1. Les indicateurs de contamination ainsi standardisés sont ensuite sommés pour évaluer la multi-contamination, selon la formule suivante :

$$\sum_{i=1}^n \frac{c_i - \min(C)}{\max(C) - \min(C)}$$

Une dernière étape est appliquée pour faciliter la lisibilité de l'indicateur final. Celui-ci est standardisé à son tour avec la méthode du min-max. Ainsi, les indices de multi-contamination dans chaque milieu s'échelonnent entre 0 et 1.

Cette méthode de calcul s'apparente à celle du score des rangs. Cependant, l'étape de standardisation des indicateurs conserve l'écart relatif entre les valeurs. La distribution statistique de chaque indicateur est conservée, elle est simplement transposée sur une échelle commune allant de 0 à 1.

c. Représentation des résultats

Tous les indices composites ont été cartographiés sur le périmètre d'étude (Nord et Pas-de-Calais). Pour chacun des milieux, quatre cartes ont donc été créées. La méthodologie retenue pour discrétiser les indices est la méthode de Jenks (Jenks, 1967). Le même code couleur a été utilisé pour tous les indices de manière à pouvoir les confronter plus facilement sur le plan de la distribution spatiale. Il est

important de noter que la couleur rouge n'a pas de signification particulière et ne vise pas à signaler un dépassement ou une anomalie.

Pour le RIM, les résultats des sous-indices ont été représentés sous forme de graphiques pour chaque milieu. Ces graphes permettent de visualiser les niveaux de dépassement des seuils pour l'ensemble des communes de la zone d'étude.

Pour l'indice ACP, les cercles de corrélation des composantes principales qui ont été utilisés dans le calcul de l'indice ont été représentés graphiquement.

2. Résultats

Tout d'abord, chaque indice composite a été interprété individuellement. Ces indices fournissent des informations différentes, et leur analyse successive a permis de mieux caractériser la multi-contamination dans l'air, l'eau et les sols sur la zone d'étude.

Ensuite, une comparaison entre les différents indices composites a été réalisée dans le but de sélectionner la méthode la plus adaptée. Si nécessaire, une vue d'ensemble présentant conjointement les douze cartographies est proposée en Figure 43.

a. Ratio d'Imprégnation Moyen (RIM)

Les trois indices obtenus ont été nommés de la façon suivante : **RIM air**, **RIM sol**, **RIM eau**. Les cartographies de ces indices sont présentées en figure 28, 30 et 32. En outre, les graphes des ratios d'imprégnation pris individuellement pour chaque polluant sont représentés en figure 29, 31 et 33.

Pour le compartiment atmosphérique, des disparités importantes sont observées sur le territoire. Un gradient est-ouest est observé avec de faibles valeurs de RIM sur la façade maritime dans le département de la Somme, et une pollution atmosphérique qui augmente à mesure que l'on se rapproche du bassin minier puis de l'agglomération lilloise. Les agglomérations urbaines sont assez facilement identifiables avec des niveaux élevés : Arras, Cambrai, Maubeuge, Boulogne-sur-Mer, Calais, Dunkerque. Mais c'est bien l'agglomération lilloise qui présente des valeurs de RIM les plus élevées. Les disparités sont donc marquées sur l'ensemble de la zone d'étude, mais elles présentent en même temps une certaine régularité dans leur répartition spatiale (zonage assez régulier). Ces observations pourraient s'expliquer notamment par la nature du milieu étudié puisque l'atmosphère est un fluide. De plus, les données de contamination utilisées sont issues de modèles de prédiction, ce qui peut également expliquer partiellement cette forme de régularité dans la répartition spatiale.

Par ailleurs, il est intéressant de noter que l'ensemble des communes présentent une valeur de RIM supérieure à 1 (le minimum est situé à 1,04). Cela signifie qu'en moyenne pour les polluants considérés, les concentrations mesurées dans toutes les communes sont supérieures aux valeurs guides de l'OMS. La valeur maximale monte à 2.46, il s'agit de la commune de La Madeleine qui est limitrophe à la ville de Lille.

En examinant les ratios d'imprégnation individuels pour chaque polluant, il s'avère que le niveau de contamination en $PM_{2.5}$ dépasse la ligne directrice de l'OMS dans toutes les communes. Le nombre de communes dépassant le seuil est également très important pour PM_{10} , mais l'ampleur du dépassement est plus faible. Enfin pour NO_2 , bien que plus de la moitié des communes présentent des valeurs supérieures au seuil, le nombre de dépassements est moins important que pour les particules.

Cependant il s'agit du polluant qui présente le ratio d'imprégnation avec la variabilité la plus importante.

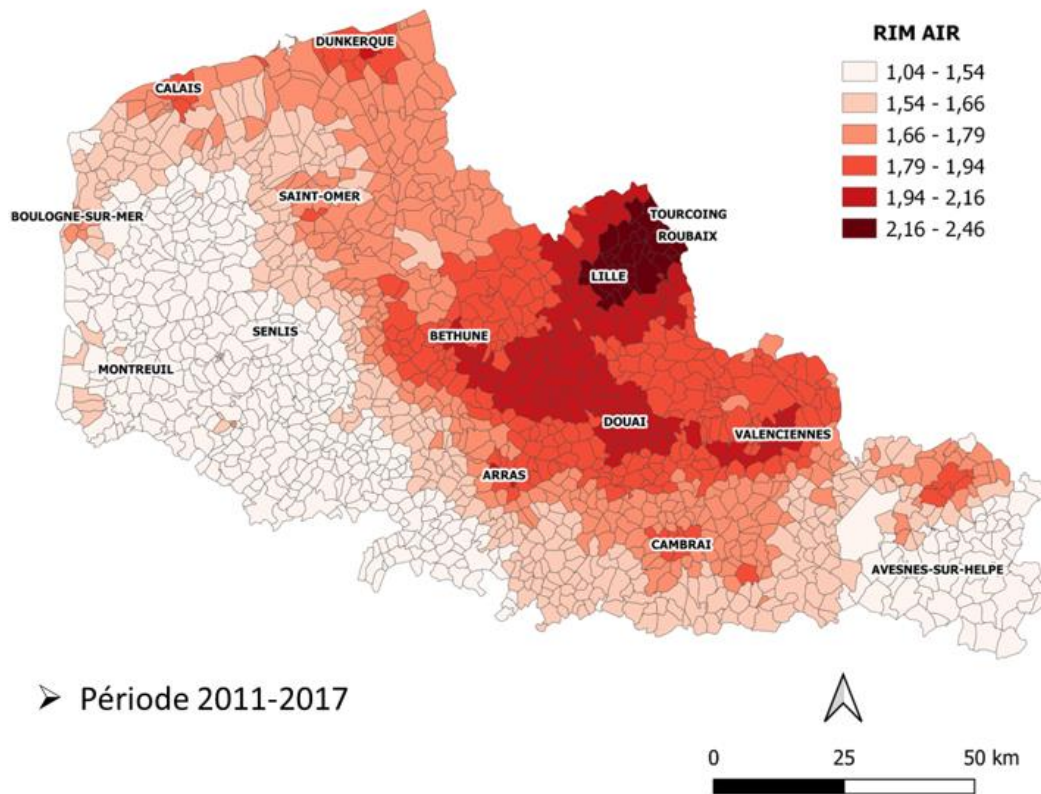


Figure 28 : Résultats obtenus pour le RIM appliqué au compartiment atmosphérique (RIM air).

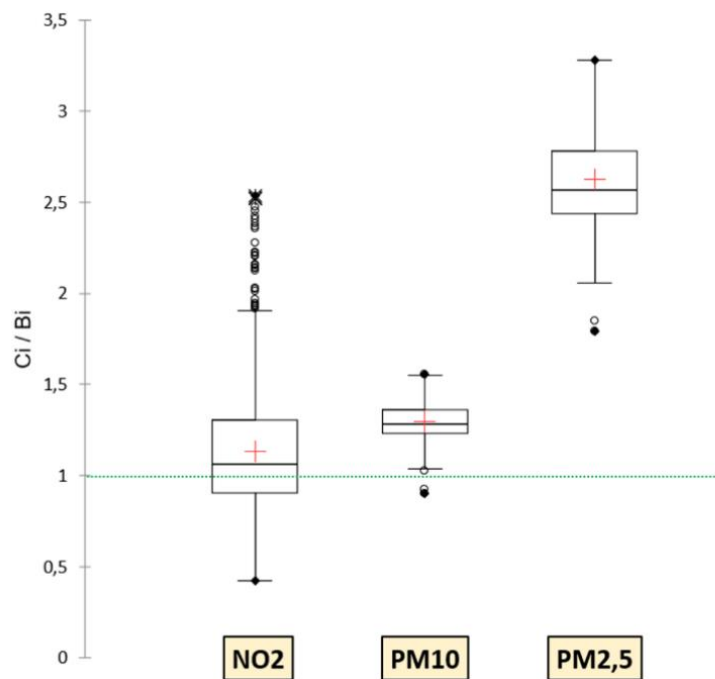


Figure 29 : Ratios d'imprégnation individuels pour chaque polluant inclus dans le calcul de RIM air.

Dans les sols, les communes qui présentent une multi-contamination métallique élevée sont assez bien agrégées sur le territoire. Il est possible d'identifier des zones où la multi-contamination métallique est particulièrement importante autour d'Arras, de Cambrai, et de Boulogne-sur-Mer notamment. D'autres îlots de communes peuvent être identifiées au sud de Lille ou de Saint-Omer par exemple. Au contraire, les communes qui présentent les valeurs de RIM les plus faibles sont très agrégées spatialement et sont situées sur la façade maritime au sud du département de la Somme. Le caractère assez régulier dans la dispersion est à relier en partie à la nature des données sources. En effet, comme évoqué dans le chapitre II, les données de contamination sont des données interpolées issues de prélèvements réalisés dans des mailles carrées régulières de 16 km de côté.

Le RIM prend des valeurs comprises entre 0.21 et 1.37, il présente donc une variabilité importante. Il est donc discriminant des situations de contamination. Ensuite, en examinant les ratios individuels pour chaque polluant, l'arsenic est le métal qui présente le plus grand nombre de dépassements parmi les communes de la zone d'étude avec plus de trois quarts des communes concernées. Au contraire, pour le cadmium et le mercure, plus des trois quarts des communes présentent des valeurs mesurées inférieures au fond géochimique du Nord-Pas-de-Calais. Les autres éléments présentent un nombre de dépassement intermédiaire.

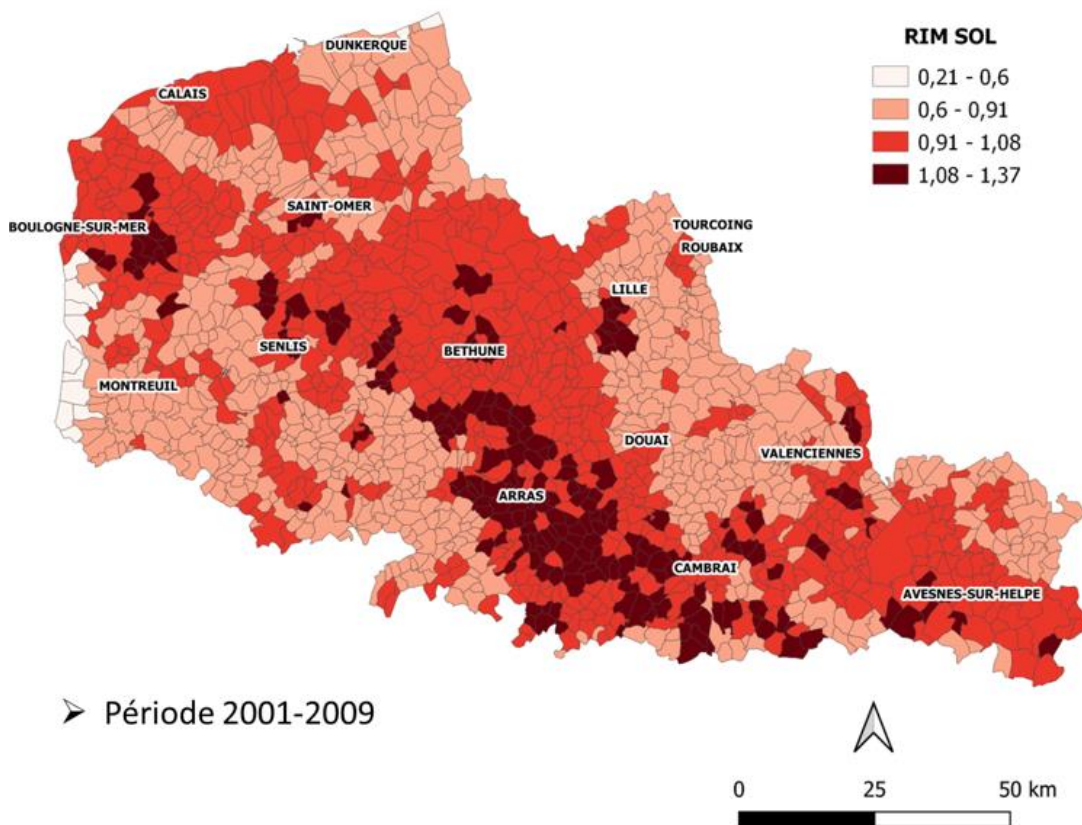


Figure 30 : Résultats obtenus pour le RIM appliqué au compartiment Sol (RIM sol).

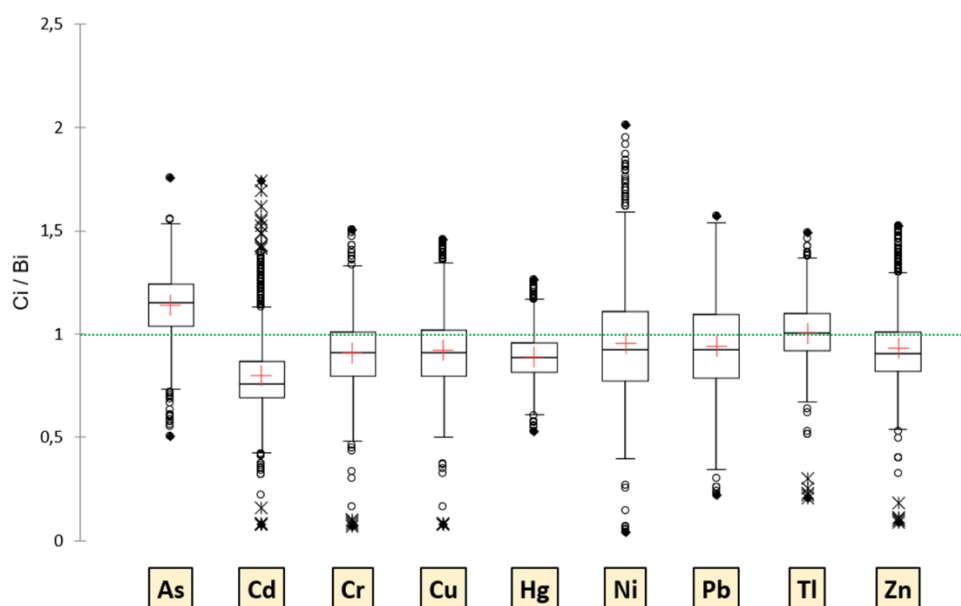


Figure 31 : Ratios d'imprégnation individuels pour chaque polluant inclus dans le calcul de RIM sol.

Pour les eaux de distribution, à la différence du compartiment atmosphérique et des sols, la répartition spatiale du RIM est plus hétérogène. Des agrégats de communes avec une multi-contamination plus élevée peuvent cependant être identifiés dans l'agglomération lilloise (bien qu'il y ait des disparités au sein même de l'agglomération), à l'est de Montreuil-sur-Mer ou au Nord de Cambrai. Concernant la répartition assez hétérogène de la multi-contamination sur le territoire, il faut souligner que même si l'eau provient initialement du milieu naturel, les prélèvements ne sont pas réalisés au même moment d'une unité de distribution à une autre, et les techniques de traitement sont potentiellement différentes également. C'est ce qui peut en partie expliquer l'allure très différente de la répartition spatiale par rapport aux sols et à l'air.

S'agissant des niveaux de contamination de l'eau potable après traitement, des valeurs de RIM assez basses étaient attendues. Les valeurs calculées de RIM sont comprises entre 0.113 et 0.446, soit beaucoup plus faibles que dans les autres milieux. Cela signifie que sur la période considérée, toutes les communes présentent des niveaux de multi-contamination en moyenne plus de deux fois inférieurs aux limites de qualité.

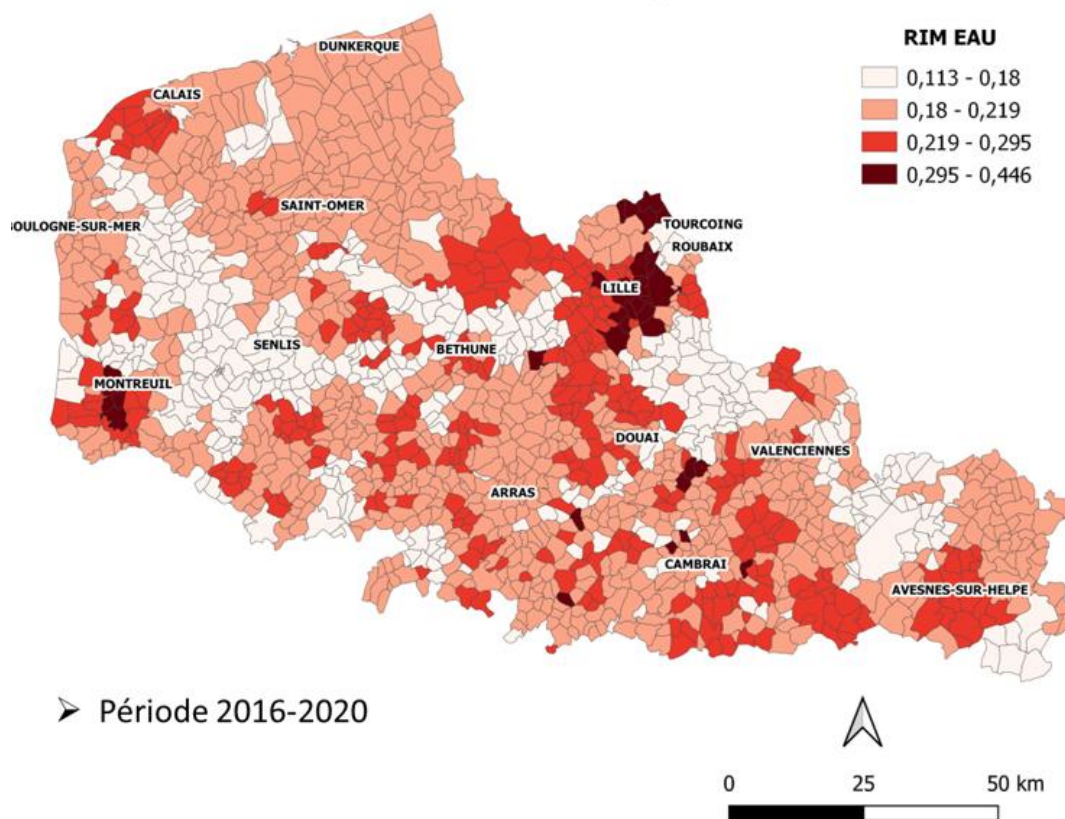


Figure 32 : Résultats obtenus pour le RIM appliqué au compartiment Eau (RIM eau).

Les zones qui apparaissent en rouge foncé sur la carte ne sont pas des zones de dépassement de seuil

Dans les eaux de distribution, pour la plupart des polluants, les niveaux de contamination au sein des communes sont très inférieurs aux limites de qualité. Les seuls paramètres qui dépassent les limites de qualité dans certaines communes sont le nickel, le plomb, les nitrates et la turbidité de l'eau. Par ailleurs, les distributions statistiques des ratios d'imprégnation sont très différentes par rapport aux autres milieux. Certains paramètres ne montrent quasiment aucune variabilité (chlorure de vinyle, chrome, nitrites, antimoine). Cela est dû à un nombre de mesures quantifiées très faible pour ces paramètres. La très grande majorité des communes prennent une valeur unique qui correspond à la limite de quantification. Cela signifie que sur les cinq années de mesures collectées, aucune n'a été quantifiée. Il est tout de même rassurant de constater que toutes les limites de quantification se situent bien en dessous des valeurs de référence (en l'occurrence, des limites de qualité des eaux destinées à la consommation humaine).

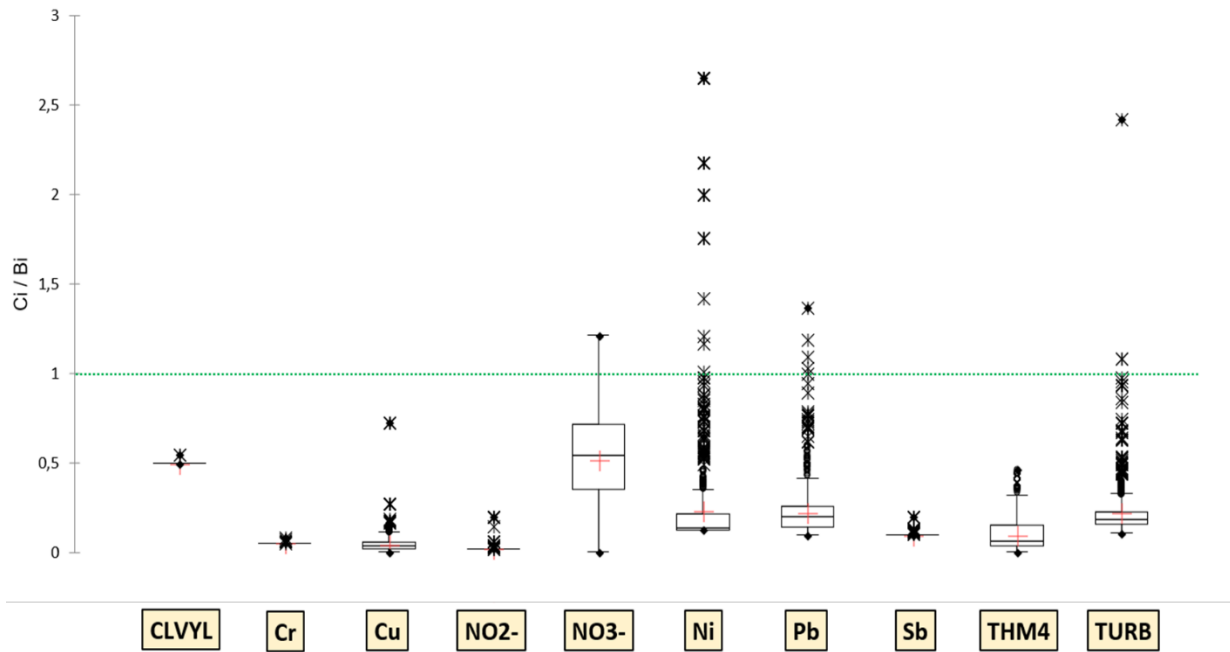


Figure 33 : Ratios d'imprégnation individuels pour chaque polluant inclus dans le calcul de RIM eau

b. Score ACP

Les trois indices obtenus ont été nommés de la façon suivante : **ACP air**, **ACP sol**, **ACP eau**. Les résultats cartographiques de ces indices sont représentés en figure 34, 35 et 36. Sur ces figures, les cercles de corrélation ont également été représentés pour les composantes principales ayant contribué au calcul de l'indice.

Le résultat cartographique de l'indice ACP air est quasiment identique à RIM air et le commentaire descriptif est similaire. Cependant, il faut noter que contrairement au RIM, les valeurs prises par l'indice ACP n'ont pas de signification particulière. Ainsi, contrairement au RIM air, il n'est pas possible de mettre en évidence, qu'en moyenne pour les trois polluants considérés, la totalité des communes dépassent les valeurs seuils recommandées par l'OMS. En revanche, l'utilisation de l'ACP permet d'apprécier les liens de corrélation entre les différents polluants. En l'occurrence, les trois polluants atmosphériques sont très bien corrélés les uns aux autres. Seule la première composante a été retenue pour le calcul de l'indice composite, elle représente à elle seule 98.66 % de la variabilité totale des données d'entrée.

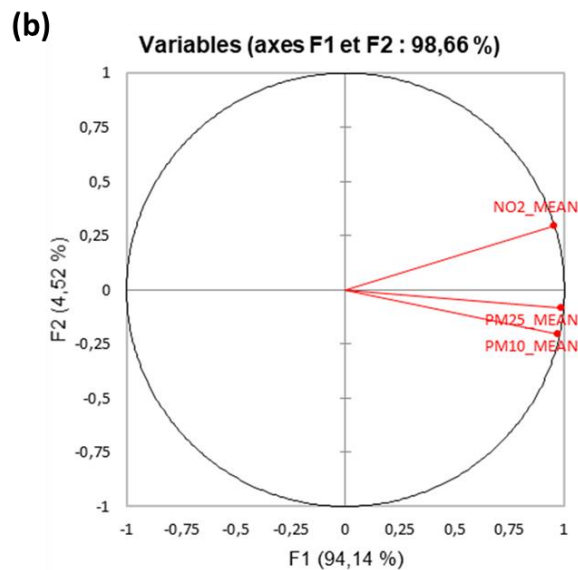
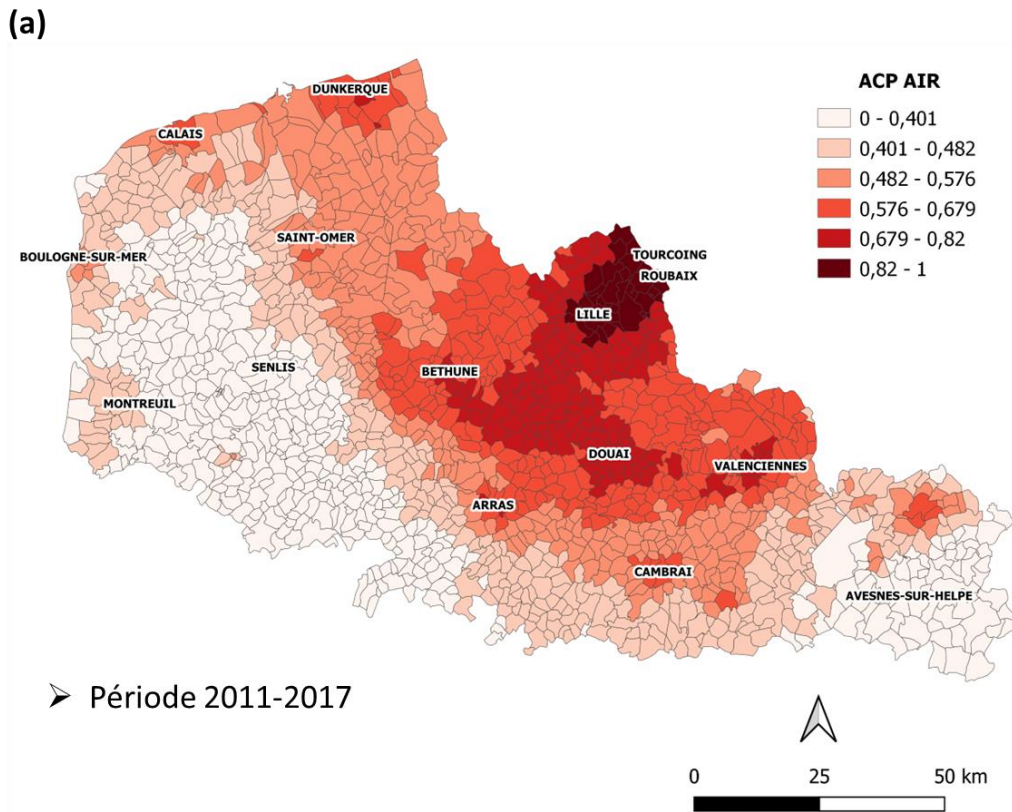
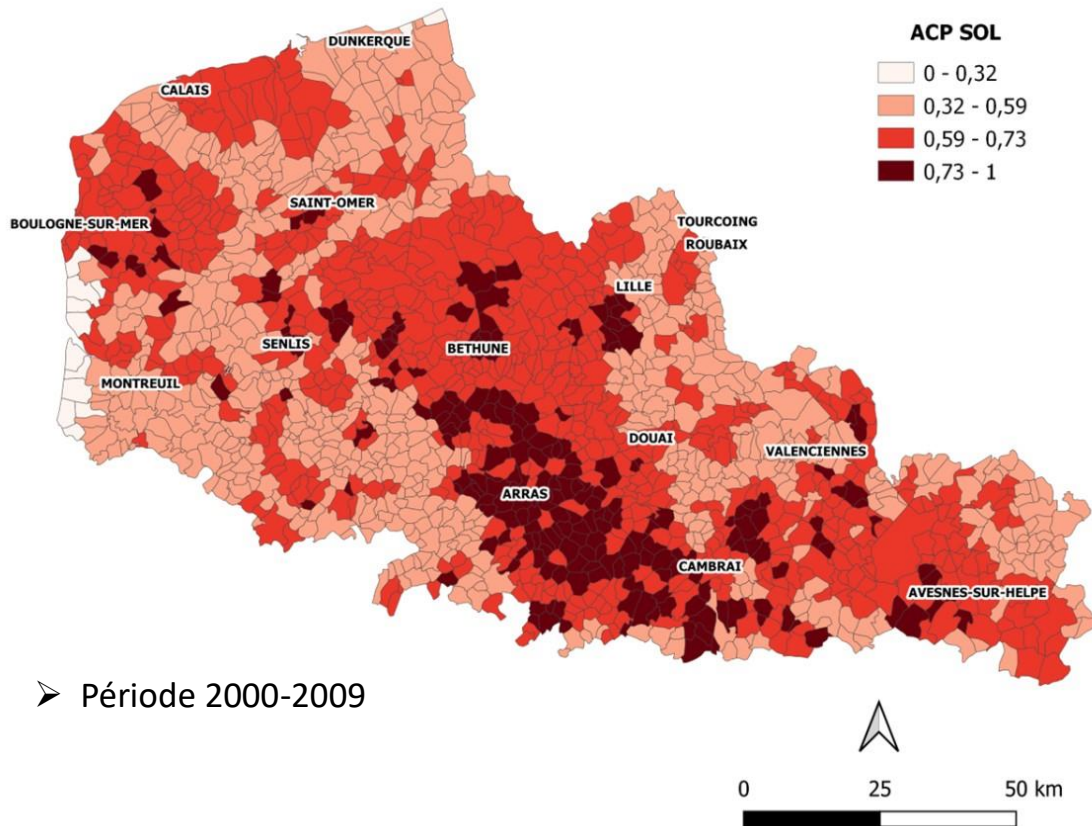


Figure 34 : Résultats obtenus pour le compartiment atmosphérique en construisant un indice composite à partir d'une ACP : (a) Cartographie ; (b) cercles de corrélation des composantes principales retenues.

Pour les sols également, la cartographie de l'indice ACP est quasi identique à celle du RIM. Les trois premières composantes ont été retenues pour le calcul de l'indice, elles représentent en cumulé 86.06 % de la variabilité totale des données d'entrée. La plupart des métaux (cuivre, zinc, arsenic, thallium, cadmium, nickel et chrome) sont relativement bien corrélés et donc bien représentés sur la première composante. En revanche le mercure, le plomb se distinguent des autres métaux et sont mieux

représentés sur la deuxième composante. De même, l'arsenic est très bien représenté sur la troisième composante principale de l'ACP.

(a)



➤ Période 2000-2009

(b)

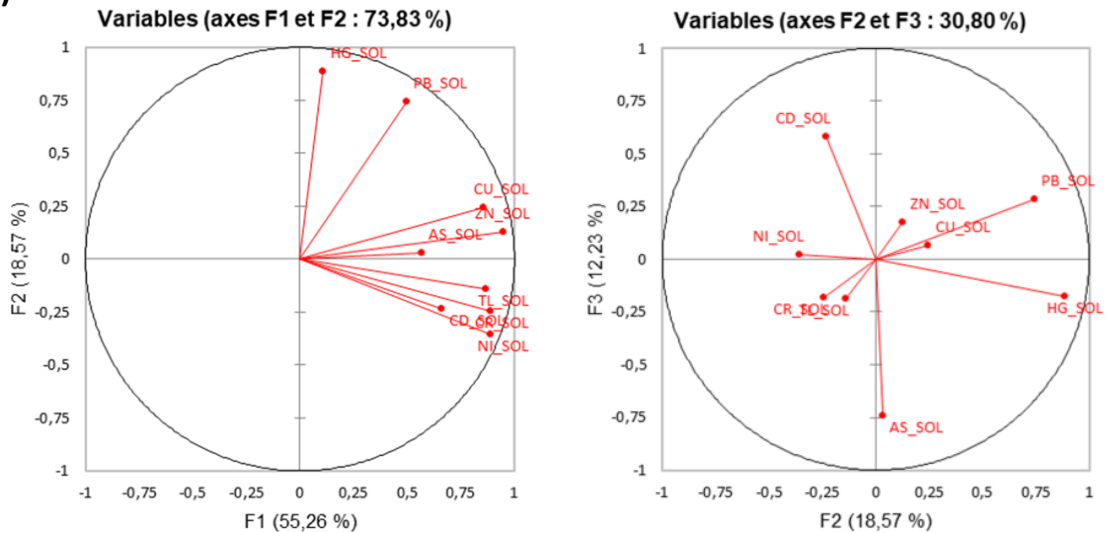
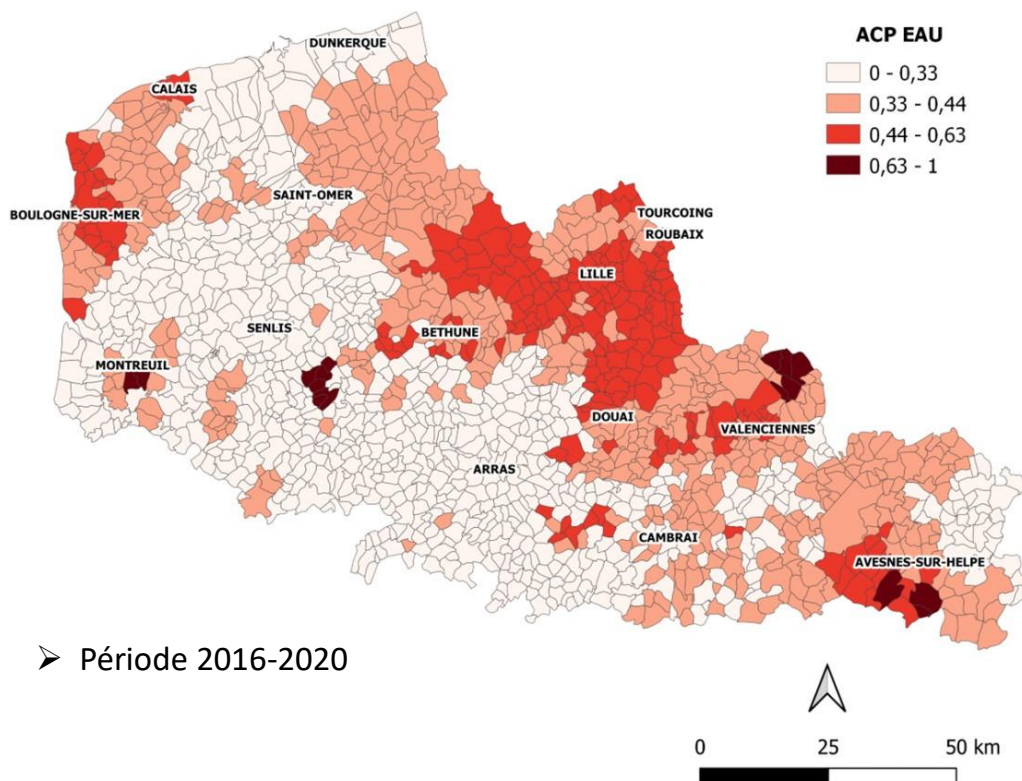


Figure 35 : Résultats obtenus pour le compartiment Sol en construisant un indice composite à partir d'une ACP : (a) Cartographie ; (b) cercles de corrélation des composantes principales retenues.

Pour les eaux de distribution, la répartition spatiale de l'indice ACP eau est extrêmement différente de celle observée avec RIM eau. Par exemple, de nombreuses communes autour de Cambrai, d'Arras ou de Senlis présentent désormais de niveaux de multi-contamination parmi les plus faibles de la zone

d'étude. Alors que la tendance inverse était observée avec la cartographie RIM eau. En examinant les cercles de corrélation, il s'avère qu'en cumulé, les quatre composantes principales retenues ne permettent d'expliquer que 55.35% de la variance totale des données d'entrée. Par ailleurs, sur la première composante, la contamination en trihalométhanes et nickel s'opposent clairement à la contamination en nitrates (marqueur de contamination plutôt rural en lien avec les activités d'élevage notamment). De plus, des contaminants comme le chlorure de vinyle sont très mal représentés sur les quatre composantes principales. Il apparaît clairement que les contaminants pour lesquels presque aucune mesure n'a été quantifiée, qui présentent donc une distribution très asymétrique, sont les moins bien représentés par l'ACP.

(a)



➤ Période 2016-2020

(b)

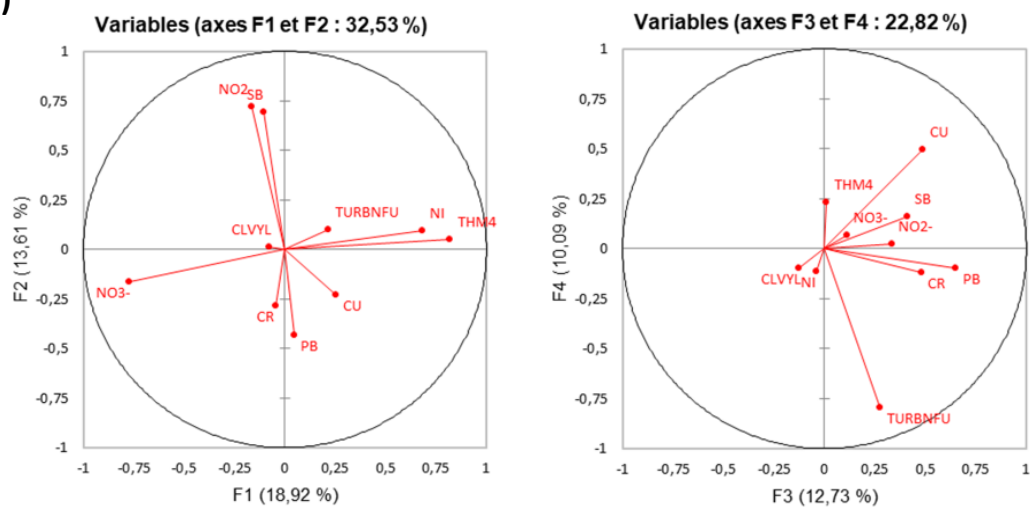


Figure 36 : Résultats obtenus pour le compartiment Eau en construisant un indice composite à partir d'une ACP : (a) Cartographie ; (b) cercles de corrélation des composantes principales retenues.

$$\sum_{i=1}^n \text{proportion}$$

c. Score Min-Max

Les trois indices obtenus ont été nommés de la façon suivante : **MinMax air**, **MinMax sol**, **MinMax eau**. Les résultats cartographiques de ces indices sont représentés en figure 37, 38 et 39.

Le score MinMax s'appuie sur la standardisation de tous les indicateurs de contamination sur une échelle commune comprise entre 0 et 1. La distribution statistique des indicateurs est conservée à la suite de cette transformation. Après agrégation, le résultat final est sensible à la présence de distributions asymétriques ou de valeurs extrêmes. Par exemple, pour un polluant en particulier, s'il existe une commune présentant une valeur extrême très élevée, celle-ci se verra affecter la valeur 1, tandis que les autres communes prendront des valeurs regroupées et proches de zéro. En conséquence, ce polluant est susceptible d'avoir moins d'influence sur le résultat final de l'indice.

Si on examine les cartographies de indices MinMax dans les trois milieux, elles sont très similaires aux deux indices précédents (RIM, ACP) pour l'air et les sols. Pour l'eau, la multi-contamination se répartit de manière hétérogène sur le territoire d'étude. La cartographie est proche de celle du RIM tandis qu'elle se démarque nettement par rapport à l'indice ACP.

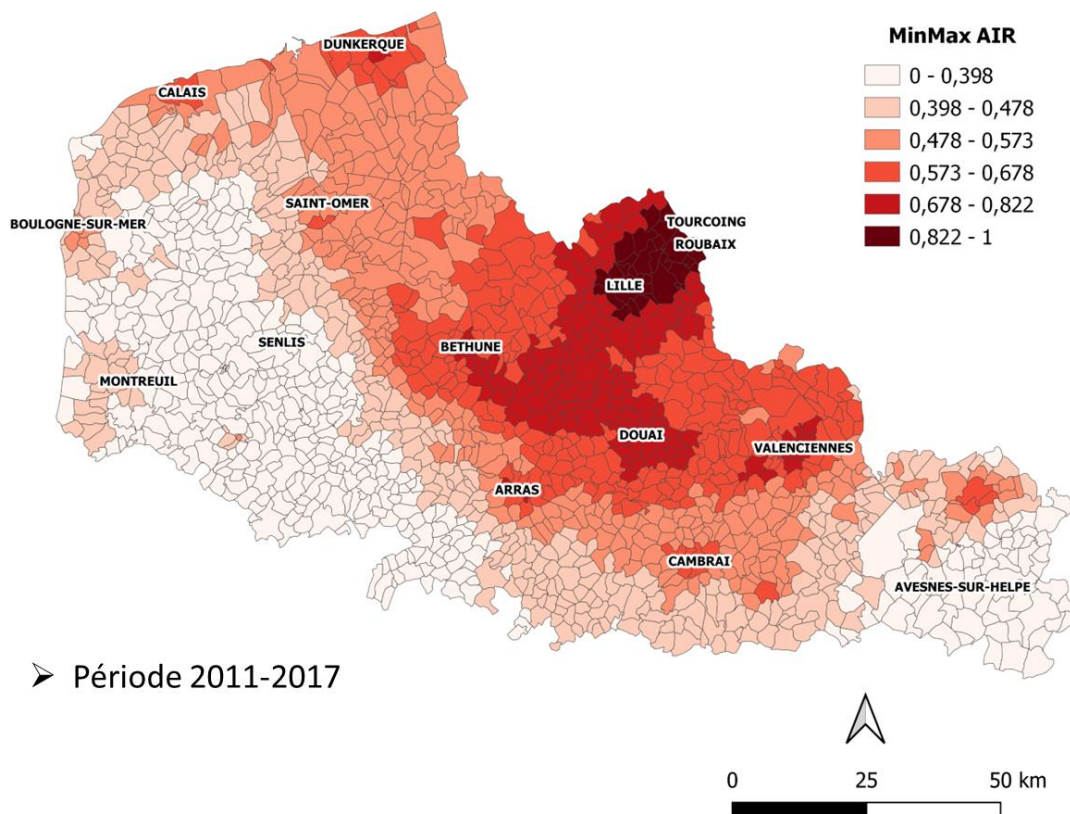


Figure 37 : Résultats obtenus pour le compartiment Air en construisant un indice composite à partir de la méthode min-max (MinMax air).

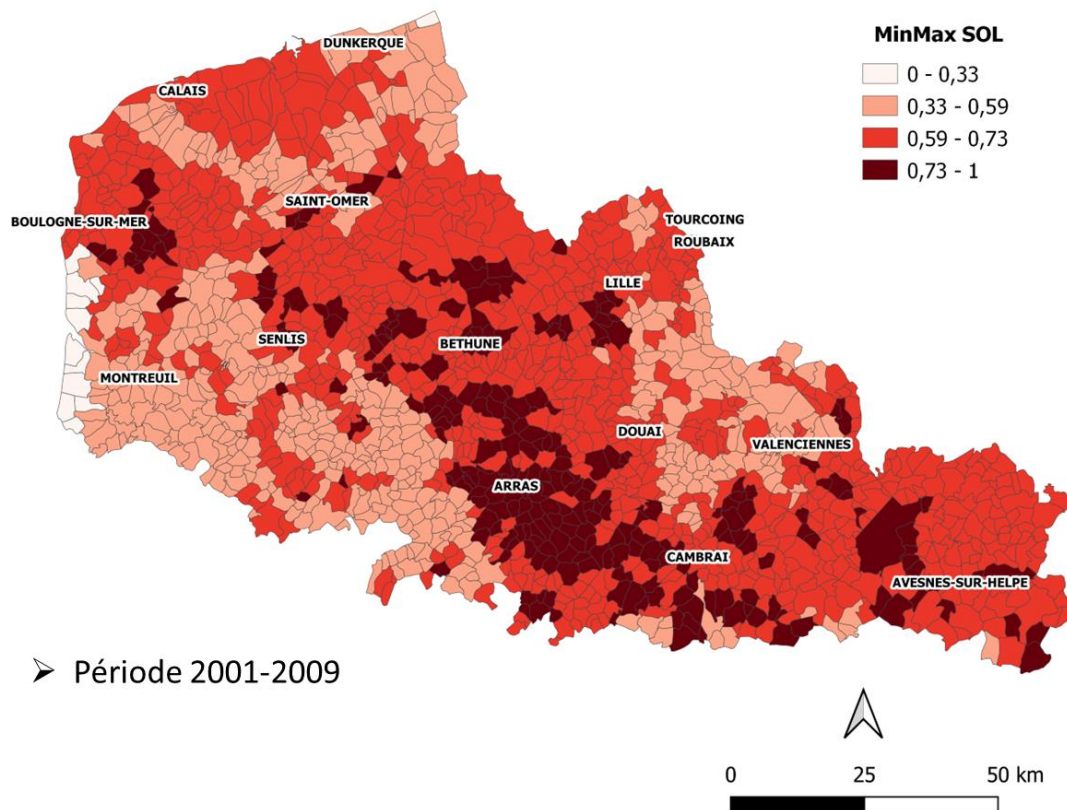


Figure 38 : Résultats obtenus pour le compartiment Sol en construisant un indice composite à partir de la méthode min-max (MinMax sol)

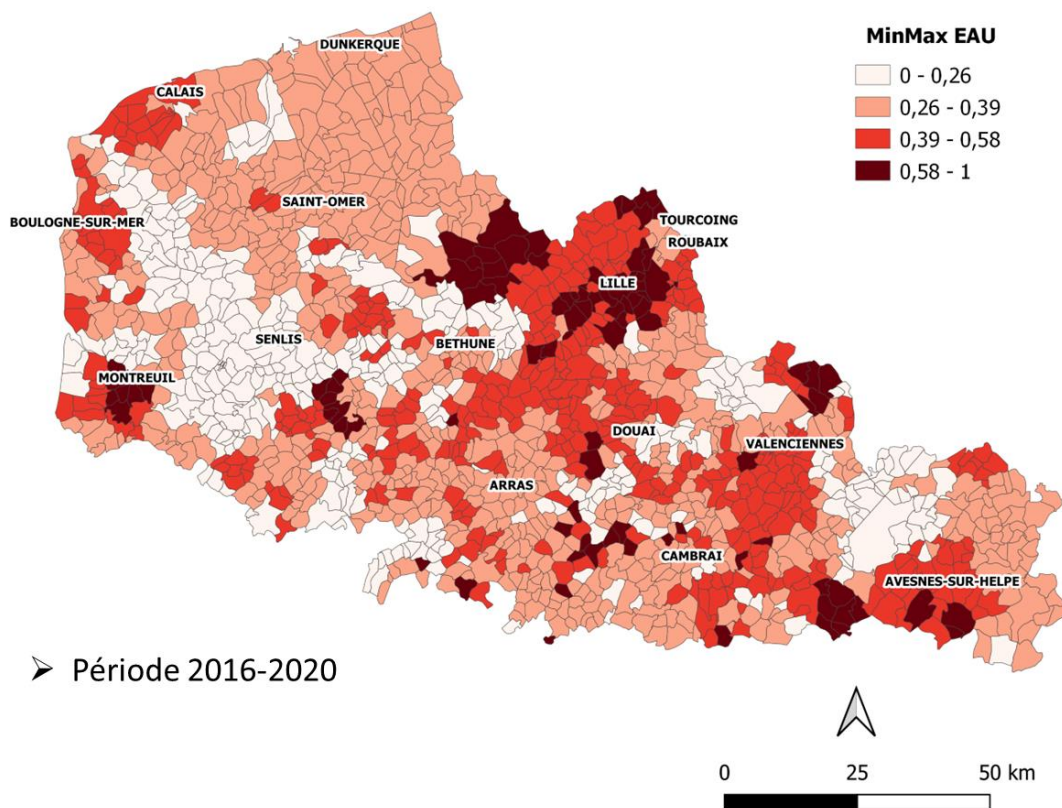


Figure 39 : Résultats obtenus pour le compartiment Eau en construisant un indice composite à partir de la méthode min-max (MinMax eau)

d. Score des Rangs

Les trois indices obtenus ont été nommés de la façon suivante : **RANG air**, **RANG sol**, **RANG eau**. Les résultats cartographiques de ces indices sont représentés en figure 40, 41 et 42.

Il est intéressant de noter que dans le compartiment atmosphérique, une commune prend la valeur 3, ce qui signifie qu'elle est caractérisée par le niveau de contamination le plus faible pour les trois contaminants considérés ($PM_{2.5}$, PM_{10} et NO_2). En l'occurrence, la commune s'appelle Moustier-en-Fagne, c'est une municipalité de 58 habitants (Insee 2020) et elle est située à l'extrême est du département du Nord, à la frontière avec la Belgique. Dans les sols et les eaux de distribution, ce cas de figure ne s'est pas présenté. La valeur minimale pour RANG eau est de 243, ce qui indique qu'aucune commune ne présente un niveau de contamination très faible pour l'ensemble des contaminants. Comme précédemment mis en évidence à travers l'ACP, cela suggère que les polluants étudiés dans l'eau ne sont que faiblement liés entre eux et présentent des distributions spatiales contrastées.

Par ailleurs, le score des rangs induit une perte d'information puisque les écarts de valeurs entre les communes ne sont pas conservés. Une des conséquences est que les valeurs extrêmes ont tendance à être écrasées dans la distribution statistique. Cependant, l'avantage principal de cette méthode est de donner un poids identique à chaque polluant dans le calcul de l'indice final. En effet, en théorie pour chaque polluant, les valeurs s'échelonnent entre 1 et 1 541 (nombre total de communes dans la zone d'étude). De cette manière, chaque indicateur présente exactement la même distribution statistique et contribue de la même façon au résultat final de l'indice. Cependant, dans la pratique, la gestion des cas d'égalité fait que pour de nombreux contaminants, la distribution statistique de l'indicateur est différente. En particulier pour l'eau, certains contaminants ont une distribution très asymétrique avec de nombreuses communes qui présentent une valeur unique qui correspond à la limite de quantification. Toutes ces communes se voient donc attribuer la valeur 1 et le contaminant en question a une très faible influence sur le résultat final de l'indice. Par exemple, pour l'antimoine dans l'eau, les rangs des communes s'échelonnent entre 1 et 12 (1 512 communes pour lesquelles aucune mesure n'a été quantifiée). La robustesse de l'indice dépend alors forcément de la capacité à quantifier les faibles concentrations. Le fait de reprendre l'ordination à 1513 a été envisagé mais cela donne alors un poids considérable à des valeurs qui peuvent rester par exemple supérieures mais très proches de la LQ.

Ainsi pour le score des rangs, des différences par rapport aux autres indices sont observées dans l'air et les sols. Et pour les raisons qui viennent d'être évoquées, ces différences sont plus marquées dans les eaux de distribution.

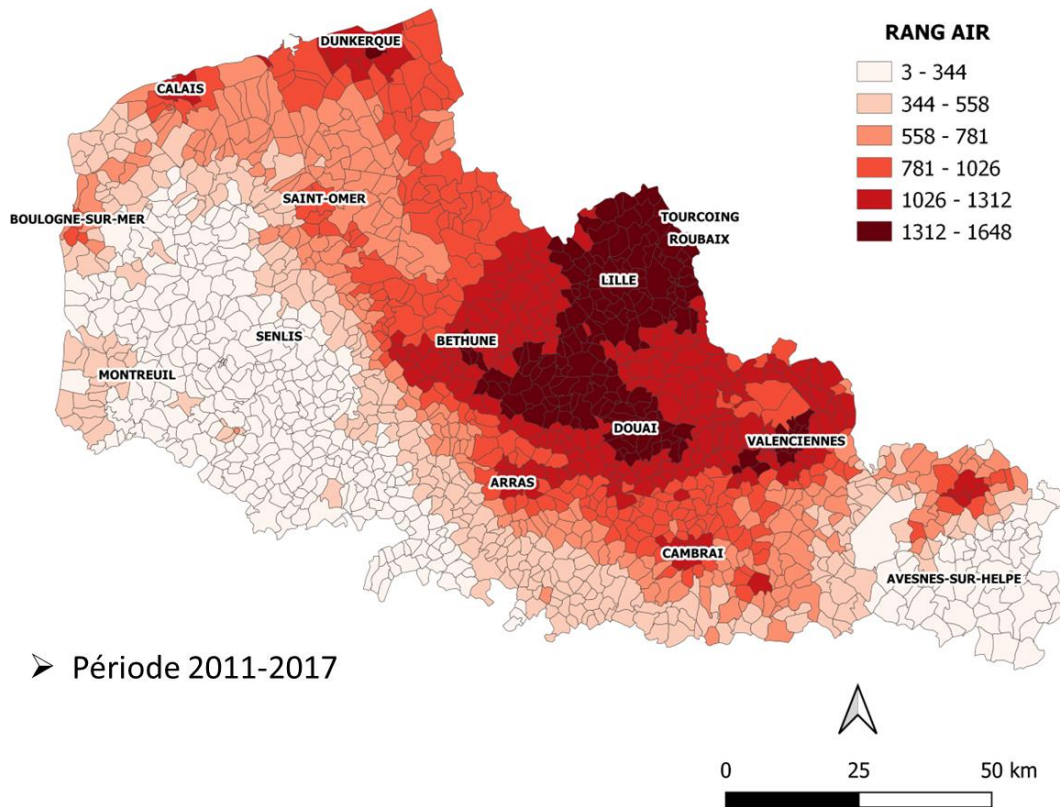


Figure 40 : Résultats obtenus pour le compartiment Air en construisant un indice composite à partir de la méthode des rangs (RANG air)

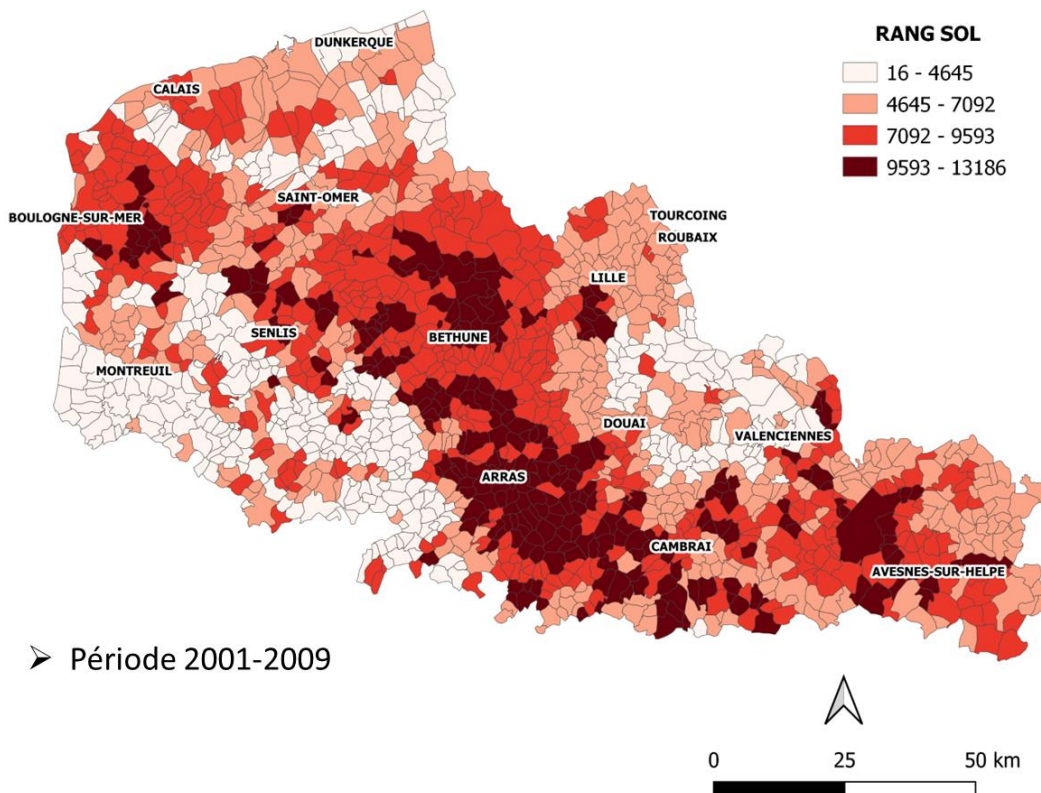


Figure 41 : Résultats obtenus pour le compartiment Sol en construisant un indice composite à partir de la méthode des rangs (RANG sol)

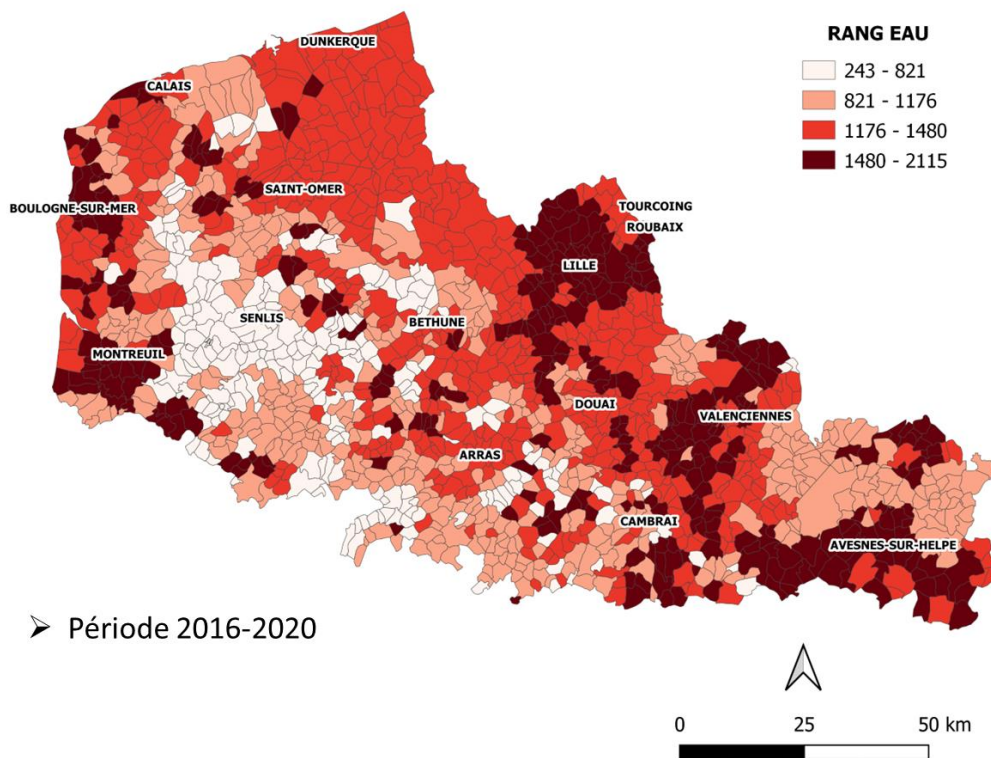


Figure 42 : Résultats obtenus pour le compartiment Eau en construisant un indice composite à partir de la méthode des rangs (RANG eau)

3. Discussion

a. Comparaison des indices composites.

Les résultats issus des quatre méthodes peuvent être confrontés. D'abord, l'ensemble des cartographies sont représentées conjointement en figure 43. Ensuite dans chaque milieu, les quatre indices ont été comparés en examinant leurs liens de corrélation. L'ensemble des nuages de points et des coefficients de corrélation de Pearson sont présentés en annexe (ANNEXE 3). Une version synthétique est présentée en figure 44. Les nuages de points et les coefficients de corrélation ont été représentés pour le RIM en lien avec les trois autres indices, à titre d'exemple.

De nouveau, la couleur rouge n'indique pas obligatoirement une contamination importante, mais un indice plus élevé que pour les autres zones. Pour le RIM, la valeur apporte une information, pour l'ACP et le min-max la couleur indique la hiérarchisation des communes. Le score des Rang indique une hiérarchisation et une valeur.

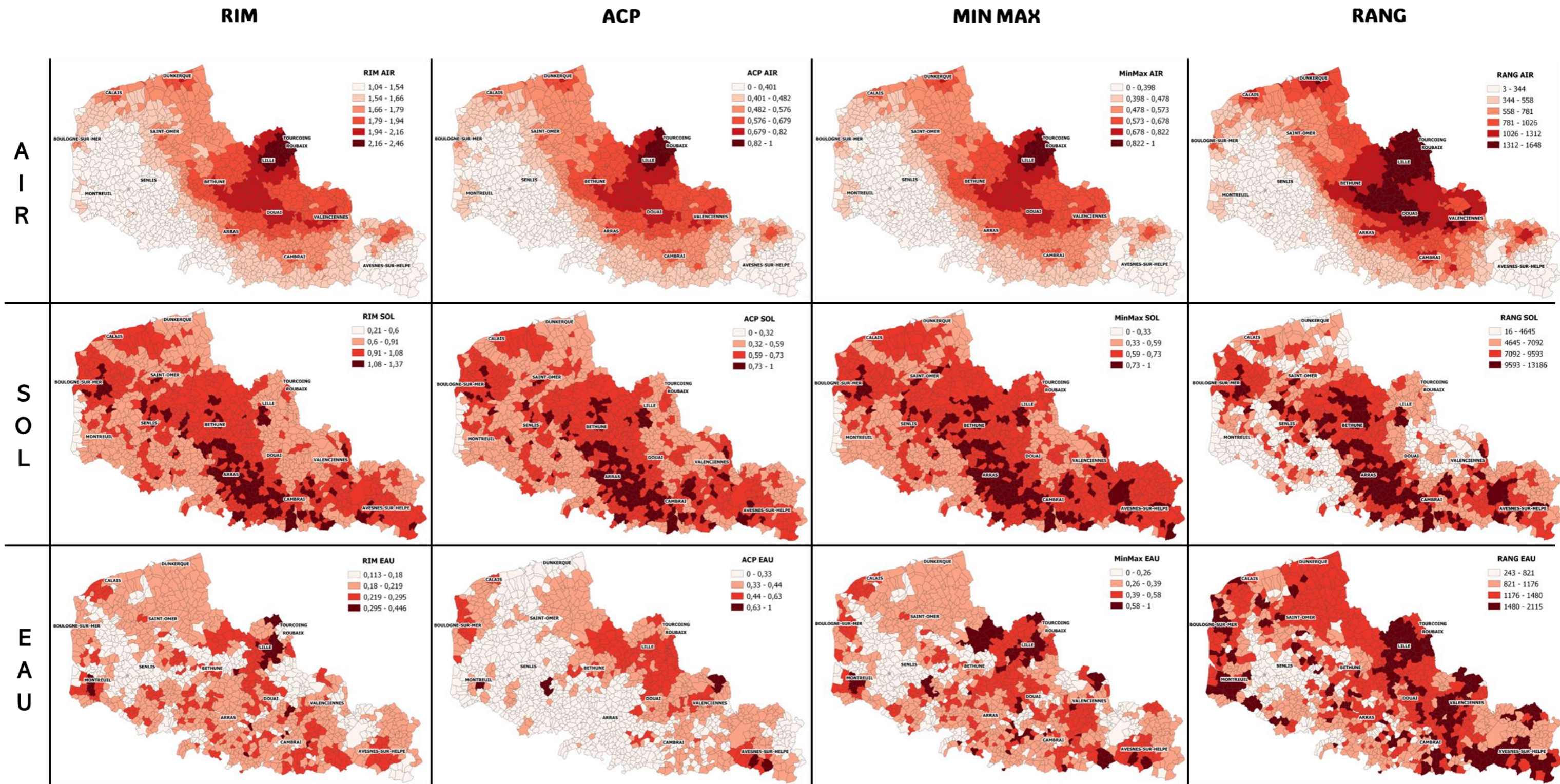


Figure 43 : Cartographies dans chacun des milieux des quatre indices composites étudiés.

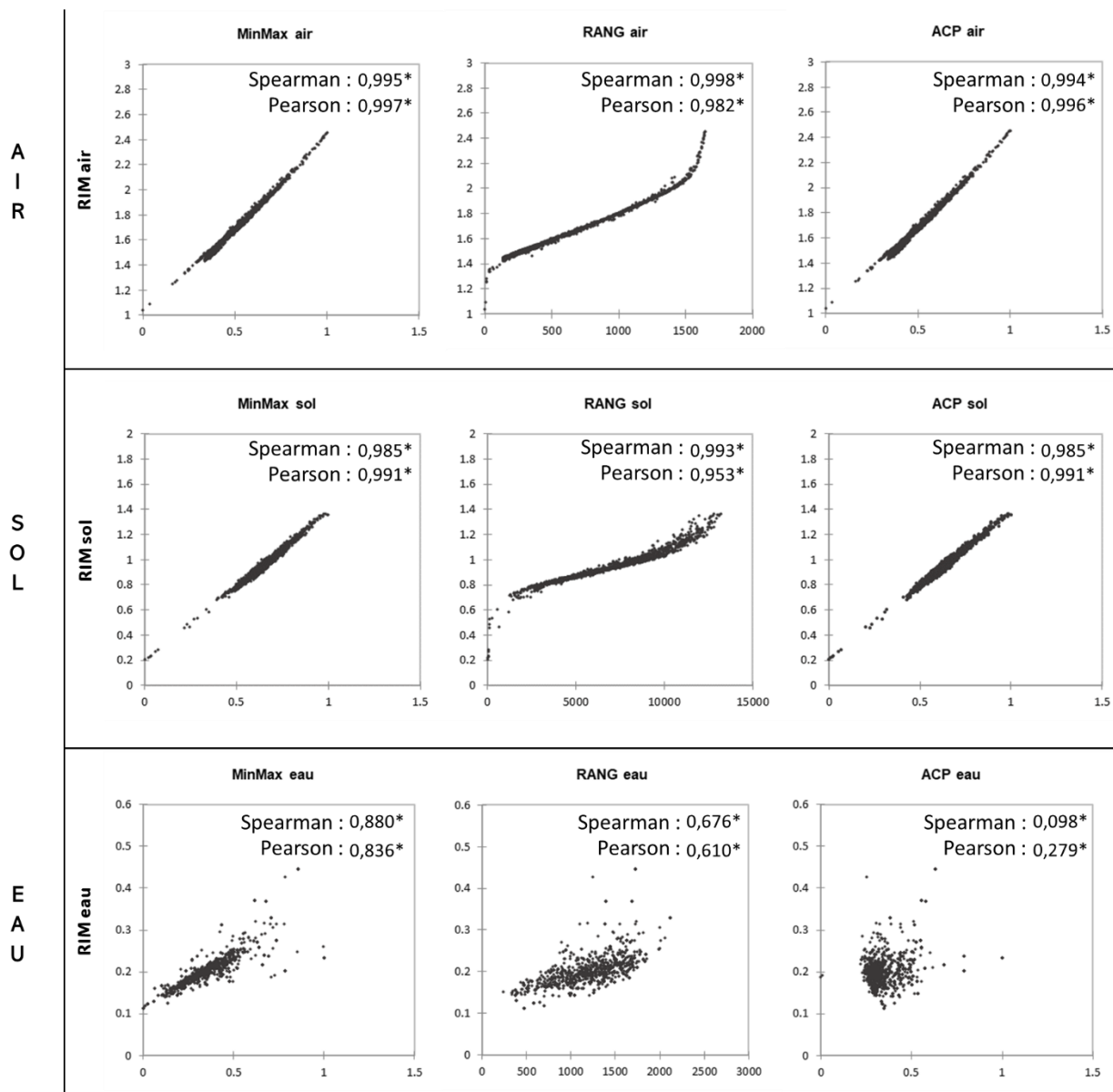


Figure 44 : Nuages de points et coefficients de corrélation (Pearson et Spearman) entre l'indice RIM et les trois autres indices composites dans chacun des milieux

Une étoile placée à côté du coefficient de corrélation indique la significativité du lien ($p < 0,05$)

Finalement, le choix de la méthode de calcul n'a que très peu d'influence sur la nature des résultats pour l'air et les sols. Cela se traduit à la fois au niveau cartographique et au niveau des liens de corrélation entre indices qui sont extrêmement élevés. En conséquence, par sa simplicité de calcul et d'interprétation, l'indicateur RIM est préféré. Toutefois, une vigilance est à maintenir quant aux valeurs de référence utilisées car il est nécessaire qu'elles soient représentatives de la contamination du milieu.

Pour les eaux de distribution, le choix de la méthode de calcul a une influence très nette sur les résultats. Les scores obtenus avec RIM eau, MinMax eau et RANG eau restent relativement similaires, même si les coefficients de corrélation sont plus faibles que dans les autres milieux. En revanche, les résultats obtenus avec l'ACP sont extrêmement différents et atypiques. L'indice de corrélation entre le RIM et l'ACP, par exemple, est de 0.098.

Plusieurs hypothèses peuvent être formulées pour expliquer les différences très marquées obtenues avec le score ACP dans le compartiment Eau. D'abord, les contaminants considérés dans l'eau sont plus nombreux et de natures très diverses (éléments traces métalliques, composés azotés, polluants organiques, etc...). De plus, si l'on examine les résultats de l'ACP, plusieurs paramètres sont corrélés négativement aux composantes principales. C'est particulièrement vrai pour la teneur en nitrates et la turbidité de l'eau (Figure 45). C'est ce qui pourrait expliquer le contraste observé par rapport aux résultats obtenus avec les autres méthodes.

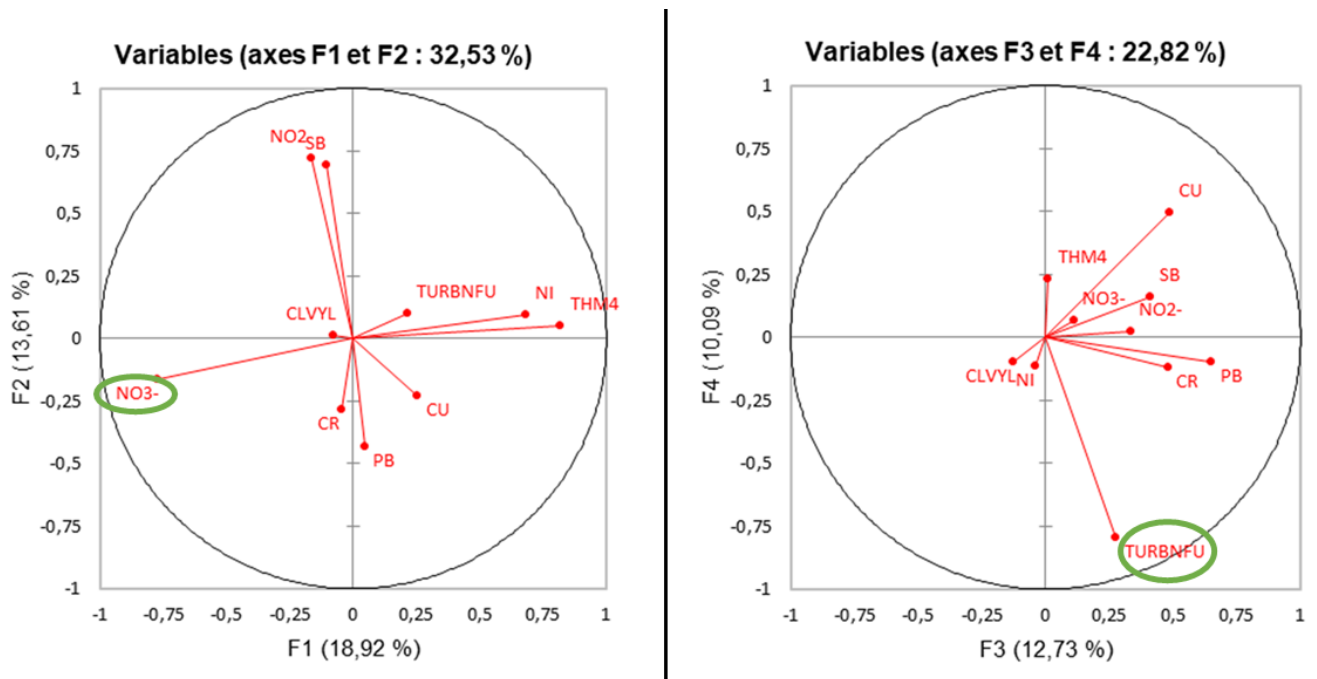


Figure 45 : Cercles des corrélations obtenus à l'issue de l'ACP appliquée au compartiment Eau.

Pour mieux comprendre ces différences, dans un premier temps un examen des cas extrêmes a été réalisé pour le compartiment eau. L'objectif était d'identifier parmi toutes les communes, celles qui présentaient les résultats les plus contrastés en fonction de la méthode de calcul retenue. Les résultats obtenus avec l'ACP ont été comparés spécifiquement aux résultats obtenus avec le RIM. Pour chacun des deux indices (RIMeau et ACPeau), les communes ont été ordonnées par ordre croissant et un rang leur a été attribué. Ensuite l'écart de rang (en valeur absolue) a été déterminé afin d'identifier les cas extrêmes. Les deux communes qui présentaient l'écart de rang le plus important étaient : Bailleul-aux-Cornailles (62070) et Frémicourt (62353). Or si l'on examine les valeurs prises par ces deux communes pour la turbidité de l'eau et la teneur en nitrates, il s'avère qu'elles sont particulièrement élevées dans les deux cas (Figure 46).

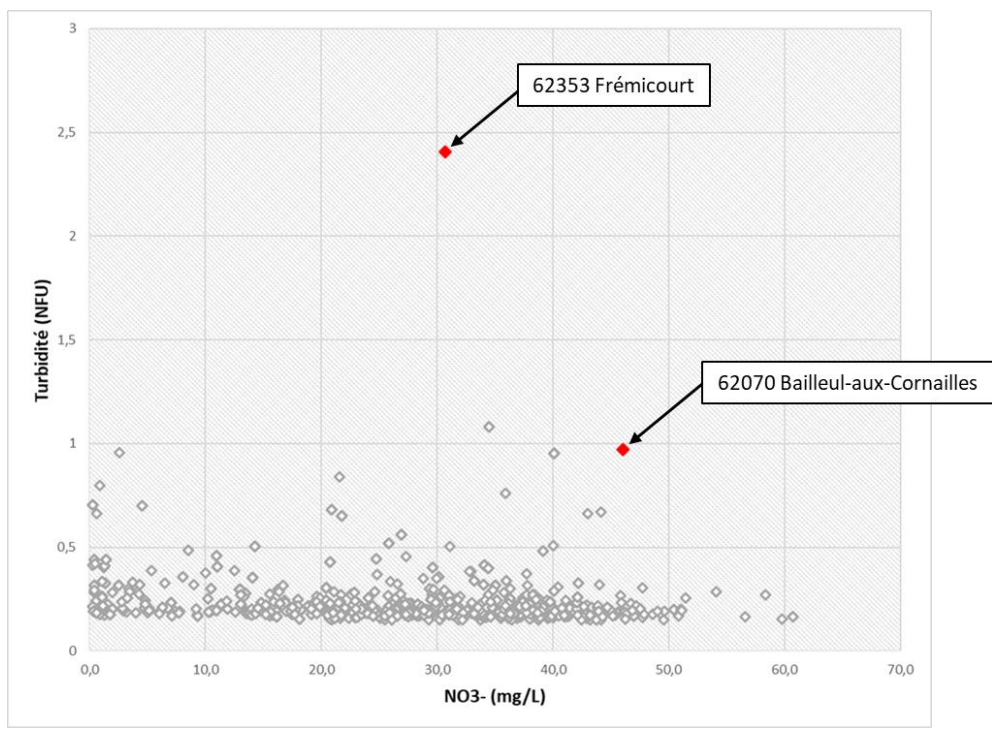


Figure 46 : Valeurs prises par l'ensemble des communes de la zone d'étude pour les paramètres Turbidité et Teneur en nitrates.

L'observation de ces cas particuliers conforte l'hypothèse formulée précédemment. Cela se confirme d'autant plus lorsqu'on examine les coefficients de corrélation entre les indices (RIMeau ou ACPeau) et les paramètres Turbidité et Teneur en nitrates (Figure 47).

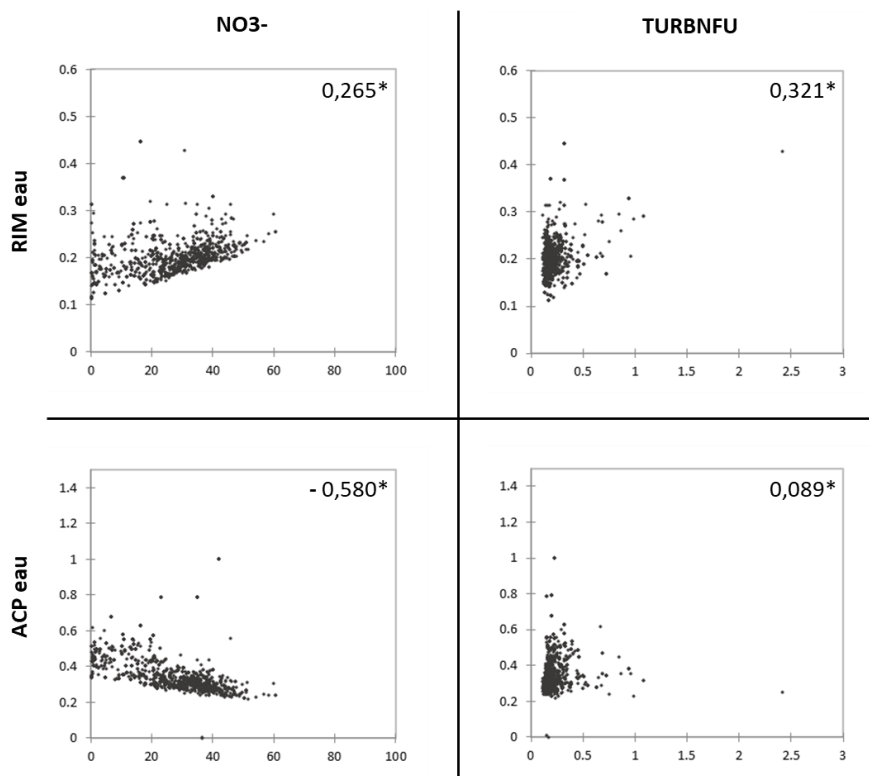


Figure 47 : Nuages de points et coefficients de corrélation (Pearson et Spearman) entre les indices et deux paramètres d'entrée : Turbidité et Teneur en nitrates.
 Une étoile placée à côté du coefficient de corrélation indique la significativité du lien (p<0,05)

En effet la teneur en nitrates est positivement corrélée à RIMEau tandis qu'elle est très nettement négativement corrélée à ACPeau. Autrement dit, plus la teneur en nitrates est importante, plus elle fait augmenter le RIM, d'où la corrélation positive. Par contre, une teneur plus élevée en nitrates est associée à un niveau d'indice plus faible pour ACPeau. Par définition l'ACP va mettre en opposition les contaminants dont la répartition dans l'espace est très différente. Pour ACPeau, cela se voit très bien sur la première composante où les nitrates sont en opposition avec le nickel. Au moment du calcul de l'indice, les nitrates seront donc sous-représentés et vont même favoriser une valeur d'indice faible, tandis que le nickel aura un poids important sur le résultat final. Les nitrates sont plutôt considérés comme des contaminants d'origine agricole. Cette sous-représentation des contaminants agricoles dans l'indice final est d'ailleurs constatée lorsqu'on étudie le lien de corrélation avec l'indicateur %SAU (pourcentage de surface communale occupée par la surface agricole utile). En effet, l'indice ACPeau est significativement anti-corrélé à %SAU (Spearman : - 0,403), tandis que ce n'est absolument pas le cas pour l'indice RIMEau (Spearman : 0,055). Dans ce contexte de multi-contamination et d'impact cumulé, la représentativité de l'indice ACPeau est donc remise en question et le mode de calcul ne semble pas particulièrement bien adapté. Cela est d'autant plus vrai dans le cadre de pollutions variées avec une répartition spatiale des contaminants très hétérogènes. En revanche, elle s'avère extrêmement précieuse pour analyser les liens de corrélation entre polluants et ainsi identifier et décrire les sources de contamination. En effet, les polluants provenant de sources similaires auront tendance à se regrouper sur une même composante principale. Cette approche est largement utilisée dans la littérature scientifique (Facchinelli et al., 2001; Huang et al., 2015; Liu et al., 2023).

Une autre particularité du milieu eau est le manque de données quantifiées pour un bon nombre de contaminants. Cela entraîne des difficultés dans le traitement des données et des déformations dans la distribution des indices. Le nombre de communes pour lesquelles les données ne sont pas quantifiées est très important pour certains contaminants : chlorure de vinyle (1539 communes non quantifiées), chrome (1532), cuivre (5), nitrites (1541), nitrates (7), nickel (739), plomb (214), antimoine (1512), trihalométhanes (16), turbidité de l'eau (0).

b. Choix de l'indice composite le plus approprié pour la suite des analyses.

L'utilisation de l'ACP pour construire un indice de multi-contamination ne semble pas être une approche adaptée. Et c'est d'autant plus vrai que les contaminants sont nombreux et divers. A partir du moment où des polluants sont faiblement corrélés ou négativement corrélés, le résultat de l'indice ACP devient difficile à interpréter et la multi-contamination n'est plus appréhendée de manière satisfaisante. Bien que cette problématique n'ait pas été rencontrée pour l'air et les sols, si des données de pollution plus diverses étaient rendues disponibles, l'usage de l'ACP serait également à proscrire. En revanche, l'ACP est très utile pour décrire les profils de pollution et identifier les sources de contamination.

Les trois autres indices appréhendent bien la notion de cumul et garantissent une meilleure prise en compte de l'ensemble des contaminants. Parmi ces indices, le RIM se distingue par sa facilité d'interprétation. Il est également adapté pour mener des comparaisons dans le temps ou entre différentes zones d'études. Le calcul du score MinMax repose sur les valeurs maximales et minimales des données d'entrée. Les résultats sont donc relatifs à la zone d'étude considérée. De même, le score des rangs s'appuie sur le classement des unités spatiales et est donc intrinsèquement dépendant de la zone d'étude.

Le RIM semble donc être l'option à privilégier. Il présente cependant l'inconvénient d'être tributaire de l'existence et de la disponibilité de valeurs seuils, qui ne sont pas encore définies pour tous les polluants. Cela peut entraîner la non prise en compte de substances potentiellement préoccupantes. Par ailleurs, des polluants comme les métaux sont des substances déjà naturellement présentes dans l'environnement. Il est donc possible de définir les concentrations attendues ou habituellement mesurées dans les sols par exemple. Cependant, pour de nombreux polluants synthétiques, il n'existe pas de concentration "naturelle". Dans de tels cas, pour surmonter cette problématique, il est possible de construire des valeurs de référence en s'appuyant sur un historique de données provenant de multiples mesures antérieures. Une autre approche consiste à utiliser des valeurs de référence établies pour déterminer à partir de quelles concentrations ces polluants peuvent présenter un risque pour la santé humaine ou l'environnement.

c. La multi-contamination des milieux environnementaux dans le Nord-Pas-de-Calais

Dans le compartiment atmosphérique, la multi-contamination présente une répartition spatiale très influencée par l'environnement urbain, du fait des polluants pris en considération et de leurs principales sources. Les niveaux de contamination dépassent largement les valeurs guides de l'OMS, en particulier pour les particules fines.

Dans les sols, la contamination métallique est particulièrement élevée autour d'Arras, Cambrai, Béthune et Boulogne-sur-Mer. Par ailleurs, une corrélation significative est constatée entre la plupart des métaux. Cependant le plomb et le mercure se distinguent des autres métaux, tout comme l'arsenic. L'arsenic est d'ailleurs l'élément qui présente le nombre de dépassements le plus important. Dans la littérature, les principales sources identifiées pour l'arsenic comprennent les activités minières, industrielles, sa présence comme impureté dans les engrais phosphatés, et son utilisation passée comme pesticide (Garelick et al., 2008). Par ailleurs, il faut rappeler que les données RMQS sont représentatives des concentrations en ETM dans les sols agricoles et forestiers. La contamination des milieux urbains et industriels est susceptible d'être sous-estimée, ce qui pourrait biaiser l'interprétation spatiale.

Dans les eaux de distribution, les corrélations entre contaminants sont beaucoup plus disparates. Cela suggère des sources de contamination multiples, contrairement aux données obtenues pour l'air qui ne focalisent que sur une source principale. Il serait d'ailleurs intéressant de poursuivre l'analyse pour analyser quels sont les principaux profils de pollution représentés, et de quelle manière ils se répartissent dans l'espace.

4. Conclusion

La construction d'un indice composite de multi-contamination nécessite donc une connaissance et une analyse critique des distributions statistiques et des structures de corrélation entre les polluants. Le choix de l'indice dépend aussi de la question posée en amont et de ce que l'on cherche à mettre en évidence (tendance globale, cumul, dépassement de seuils...). Dans notre contexte d'étude, le RIM semble être la méthode la plus adaptée pour étudier la multi-contamination, cet indice sera retenu pour la suite de l'analyse.

PARTIE 2 : Multi-contamination et variations spatiales d'incidence des pathologies étudiées

1. Matériel & Méthodes

Les indices de multi-contamination (RIMair, RIMeau, RIMsol) peuvent maintenant être confrontés aux données d'incidence de l'IRCT. Bien que l'analyse ne couvre pas l'intégralité du registre Epimad, une mise en lien avec les données d'incidence de la MC a également été menée sur la base des départements du Nord et du Pas-de-Calais.

A l'échelle écologique géographique, les associations entre l'incidence d'une pathologie et un facteur de risque nécessitent la mise en place d'un modèle de régression spatial (Besag et al., 1991b). Avant de mettre en place ce type de modèle, nous avons étudié les liens de corrélation entre les SIRs lissés de l'IRCT ou de la MC et les indices de multi-contamination dans chacun des milieux. Les coefficients de corrélation de Spearman ont été calculés pour un niveau de significativité de 5%. En parallèle, chaque indice de multi-contamination a été décomposé en quartiles (Q1, Q2, Q3, Q4). Cette opération permet de générer quatre classes de contamination croissantes. La distribution des SIRs lissés a ensuite été comparée entre ces quartiles à l'aide d'un test de Kruskal-Wallis. En cas de significativité du test ($p < 0.05$), des tests post-hoc de Dunn avec correction de Bonferroni ont été effectués.

2. Résultats

a. Association entre les indices de multi-contamination et l'incidence d'IRCT

En analysant les corrélations entre les indices de multi-contamination et les SIRs de l'IRCT, il s'avère qu'elles sont positives et significatives quel que soit le milieu considéré (Figure 48). L'association est plus prononcée pour la multi-contamination de l'air où le coefficient de corrélation de Spearman s'élève à 0.459. Concernant l'eau et les sols, le nuage de points ne montre pas véritablement de tendance à la relation linéaire.

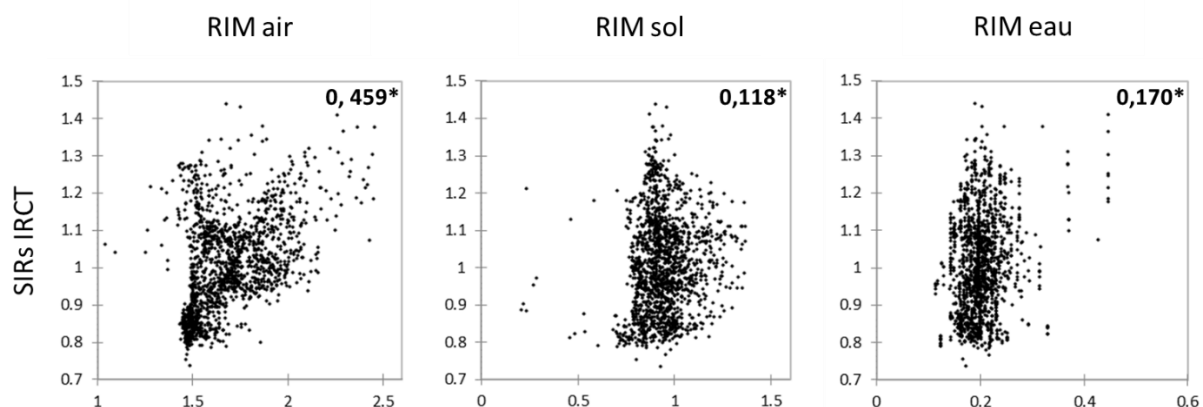


Figure 48 : Nuages de points et coefficients de corrélation de Spearman entre les indices de multi-contamination dans chaque milieu (RIM) et les SIRs de l'IRCT.

Une étoile placée à côté du coefficient de corrélation indique la significativité du lien ($p < 0,05$)

L'indice de multi-contamination de l'air a donc été découpé en quartiles et la distribution d'incidence de l'IRCT a été comparée entre ces classes de contamination croissantes (Figure 49). Pour l'eau et les sols, les niveaux de corrélation sont beaucoup plus faibles et donc difficilement interprétables. Pour ces deux milieux, les résultats sont présentés en détail dans l'annexe du mémoire (ANNEXE 4).

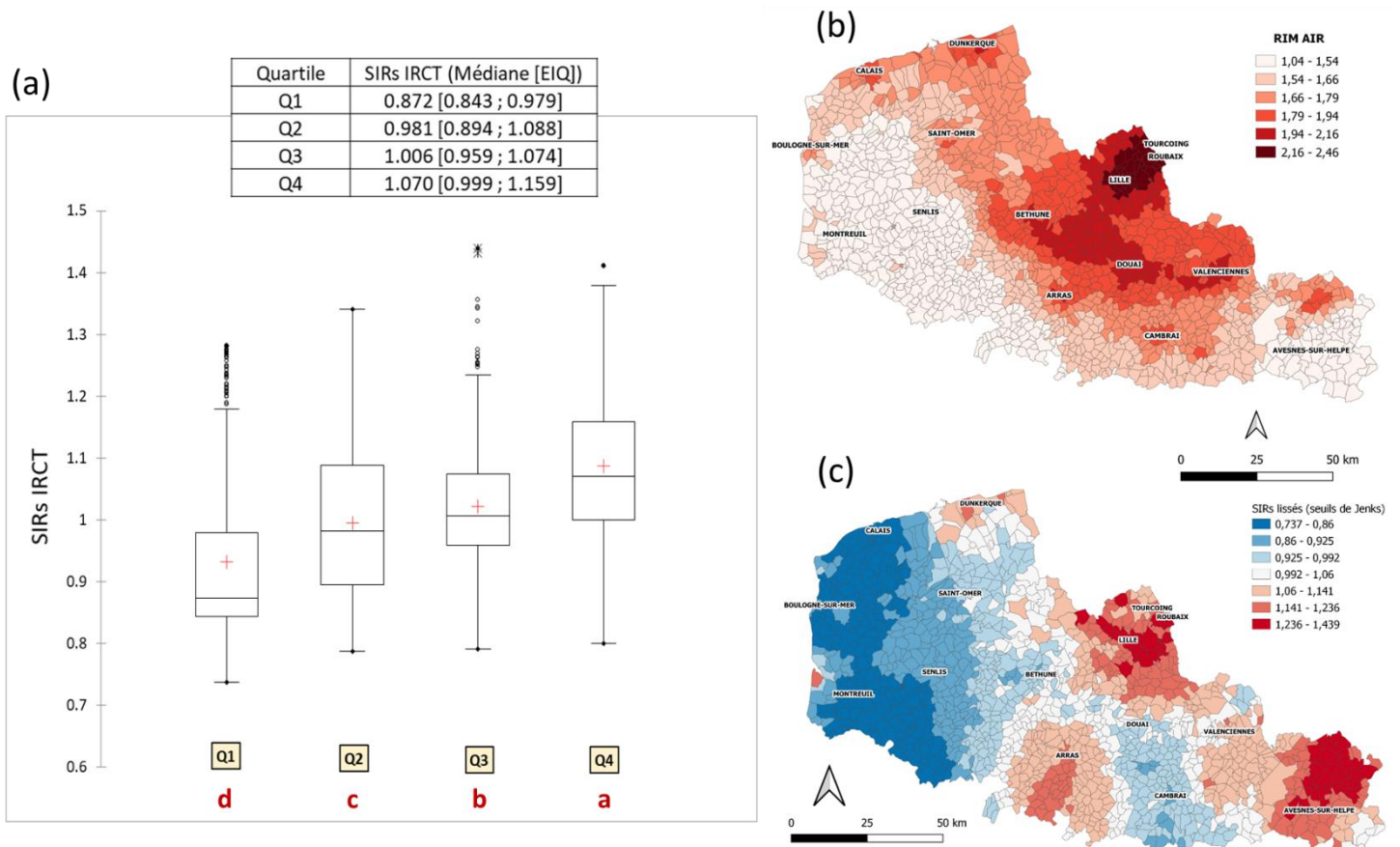


Figure 49 : (a) Graphe et Tableau de distribution des SIRs de l'IRCT par quartile croissant de multi-contamination de l'air (RIMair) ; (b) Représentation cartographique de l'indice de multi-contamination de l'air (RIMair) ; (c) Représentation cartographique de la répartition spatiale d'incidence de l'IRCT.

Si deux quartiles partagent une lettre en commun alors ils ne sont pas significativement différents, au contraire si deux quartiles ne partagent aucune lettre en commun alors ils sont significativement différents (Test de Kruskal-Wallis avec comparaisons multiples de Dunn)

Pour le compartiment atmosphérique, une différence significative est observée entre chacun des quartiles de contamination, avec un niveau d'incidence croissant du premier au quatrième quartile. Dans l'ensemble de la zone d'étude, le SIR de l'IRCT présente une médiane de 0.999 (EIQ : [0.908 ; 1.093]). Ainsi, dans le quartile de contamination de l'air le plus élevé (Q4), l'incidence présente une médiane supérieure de 7,1 % par rapport à l'ensemble de la zone d'étude. En revanche, dans le quartile de contamination de l'air le plus faible (Q1), l'incidence présente une médiane inférieure de 12,7 % par rapport à la zone d'étude dans son ensemble. Enfin, l'incidence présente une médiane supérieure de 22,7 % dans le quartile de contamination de l'air le plus élevé (Q4) par rapport au premier quartile de contamination (Q1).

b. Association entre les indices de multi-contamination et l'incidence de la MC

En analysant la corrélation de Spearman entre les indices de multi-contamination et les SIRs de la MC, les résultats varient considérablement en fonction du milieu étudié (Figure 50). La corrélation avec l'indice de multi-contamination des eaux de distribution est proche de zéro (0,094). Une corrélation négative significative est observée entre l'incidence de la MC et la multi-contamination de l'air (-0,169). En revanche, une corrélation positive significative est constatée avec l'indice de multi-contamination des sols (0,286).

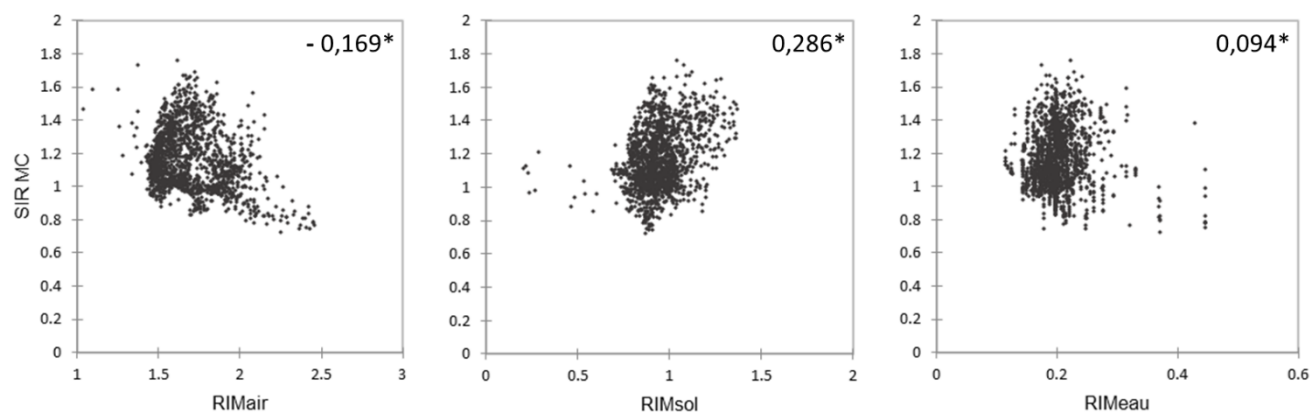


Figure 50 : Nuages de points et coefficients de corrélation de Spearman entre les indices de multi-contamination dans chaque milieu (RIM) et les SIRs de la MC.

Une étoile placée à côté du coefficient de corrélation indique la significativité du lien ($p < 0,05$)

L'indice de multi-contamination des sols a été subdivisé en quartiles et la distribution d'incidence de la MC a été comparée entre ces classes de contamination croissantes (Figure 51). Pour l'eau et l'air, les résultats sont difficiles à interpréter puisque la relation avec l'incidence est soit inverse soit quasiment nulle. Les résultats sont présentés dans le détail en annexe du mémoire (ANNEXE 4).

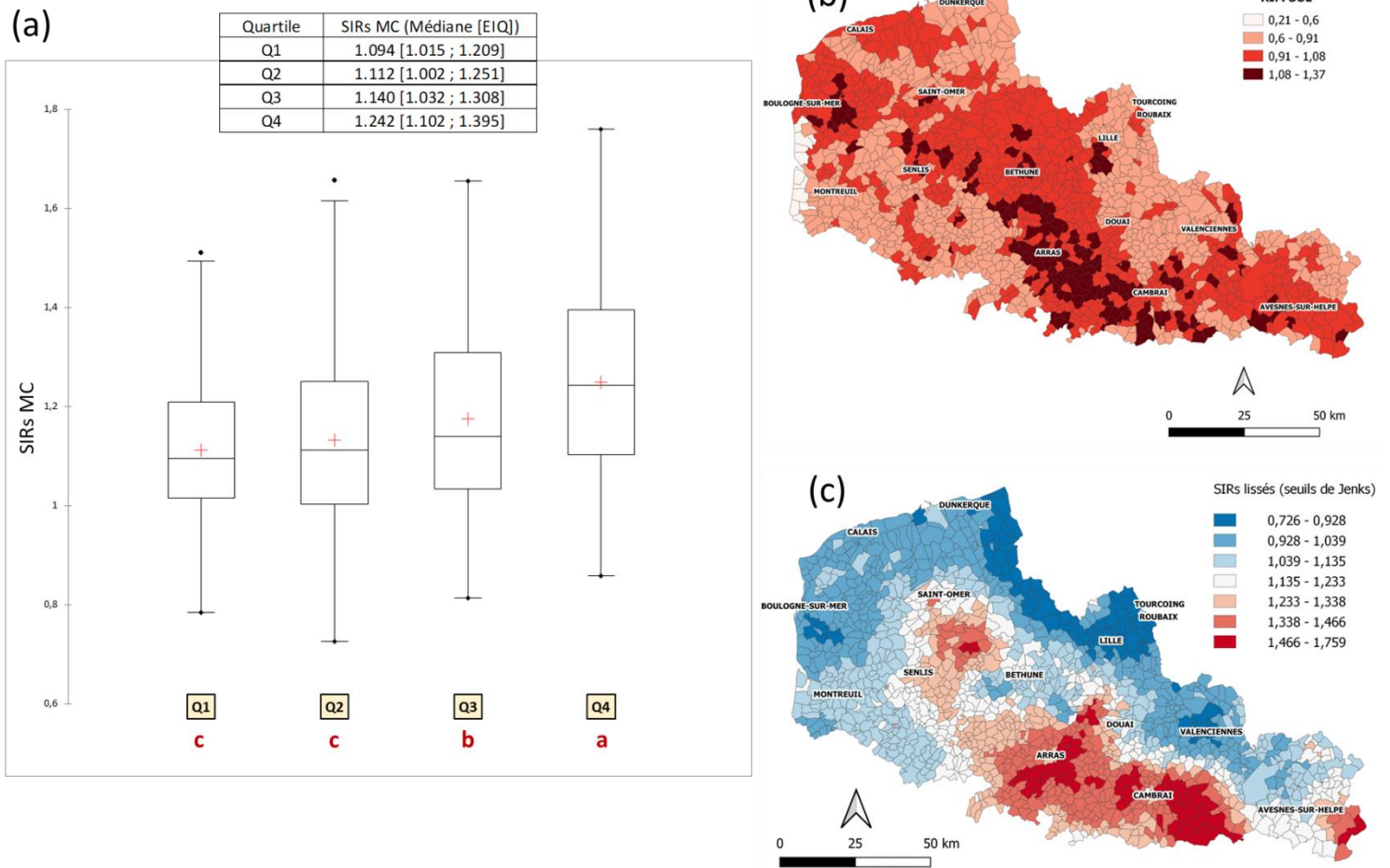


Figure 51 : (a) Graphe et Tableau de distribution des SIRs de la MC par quartile croissant de multi-contamination des sols (RIMsol) ; (b) Représentation cartographique de l'indice de multi-contamination des sols (RIMsol) ; (c) Représentation cartographique de la répartition spatiale d'incidence de la MC.

Si deux quartiles partagent une lettre en commun alors ils ne sont pas significativement différents, au contraire si deux quartiles ne partagent aucune lettre en commun alors ils sont significativement différents (Test de Kruskal-Wallis avec comparaisons multiples de Dunn)

Les deux premiers quartiles de contamination ne diffèrent pas de manière significative l'un de l'autre, mais ils présentent tous les deux le niveau d'incidence le plus faible par rapport aux deux autres quartiles. Par ailleurs, le quartile de contamination des sols le plus élevé (Q4) affiche un niveau d'incidence significativement plus élevé que le troisième quartile de contamination. Dans l'ensemble de la zone d'étude, le SIR de la MC présente une médiane de 1.139 (EIQ : [1.030 ; 1.293]). Ainsi, dans le quartile de contamination le plus élevé (Q4), l'incidence présente une médiane supérieure de 9.0 % par rapport à l'ensemble de la zone d'étude. Au contraire, dans le quartile de contamination des sols le plus faible (Q1), l'incidence présente une médiane inférieure de 4.0 % par rapport la zone d'étude dans son ensemble.

3. Discussion

a. IRCT et qualité de l'air

Les résultats obtenus suggèrent une influence de la pollution atmosphérique sur la répartition spatiale d'incidence de l'IRCT. A la lumière de la littérature existante, ce résultat semble cohérent, même si la plupart des études s'intéressent à l'insuffisance rénale chronique (IRC) et non spécifiquement au stade terminal de la maladie. En particulier, trois revues de la littérature récemment publiées ont mis en évidence une association entre une exposition à long terme à la pollution de l'air et un excès de risque pour l'IRC (M.-Y. Wu et al., 2020; Xu et al., 2018; Ye et al., 2021).

Dans la revue la plus récente (Ye et al., 2021), 13 études épidémiologiques populationnelles ont été retenues dans la méta-analyse : 8 études transversales et 5 études de cohorte. Ces études ont porté sur un nombre de participants allant de 3 345 à 2 482 737, elles ont été menées aux États-Unis, en Chine, en Corée du Sud et à Taïwan. Un excès de risque a été constaté pour une exposition aux PM_{2.5} (OR : 1.15 ; IC 95% [1.07-1.24] pour une augmentation de 10 µg/m³), PM₁₀ (OR : 1.25 ; IC 95% [1.11-1.40] pour une augmentation de 10 µg/m³), NO₂ (OR : 1.10 ; IC 95% [1.03-1.17] pour une augmentation de 10 ppb) et CO (OR : 1.04; IC 95% [1.00-1.08] pour une augmentation de 0.1 ppm). Ces résultats doivent cependant être interprétés avec prudence, le nombre d'études étant relativement limité. De plus, il est essentiel de comparer les gammes de concentrations pour lesquelles ces excès de risque ont été observés avec celles présentes dans le nord de la France. Pour PM_{2.5} par exemple, dans le périmètre d'étude du registre Néphronor, la concentration médiane s'élève à 12.84 µg/m³ avec un écart interquartile de [12.20 ; 13.90]. Parmi les 13 études sélectionnées dans la méta-analyse, cinq ont été menées sur des gammes de concentrations similaires avec des médianes ou des moyennes de concentrations comprises entre 10 et 13 µg/m³ pour PM_{2.5} (Blum et al., 2020; Bowe et al., 2018c, 2017; Bragg-Gresham et al., 2018; Paranjpe et al., 2020).

Par ailleurs, étant donné que le diabète de type II et l'hypertension constituent des facteurs de risque de l'IRC, il est également pertinent d'examiner leur association avec la pollution de l'air. Des revues de la littérature récentes ont effectivement mis en évidence une association entre l'exposition à long terme aux PM_{2.5}, au NO₂ et le développement du diabète et de l'hypertension (Yang et al., 2020, 2018). Une étude de cohorte très récente a été menée au Royaume-Uni sur 162 334 participants et sur une période médiane de suivi de 11,7 ans (Wu et al., 2023). Parmi eux, 18 496 ont développé au moins une des trois maladies étudiées : hypertension, diabète et IRC. L'étude a examiné la relation entre la pollution de l'air (PM_{2.5}, PM₁₀, NO₂, NO_x) et ces maladies, en prenant en compte plusieurs variables individuelles (âge, sexe, indice de masse corporelle, tabagisme, activité physique, revenu du ménage) et contextuelles (indice de déprivation de Townsend et accès aux espaces verts). Dans un premier modèle, regroupant les cas incidents des trois pathologies, la pollution de l'air était significativement associée à un risque accru de développer ces maladies. Dans un second modèle, où chaque pathologie était considérée individuellement, la pollution de l'air était associée à un risque accru d'hypertension et de diabète. Cependant, il n'y avait pas d'association significative entre la pollution de l'air et l'IRC.

Concernant l'association spécifique entre IRCT et pollution de l'air, une étude de cohorte menée à Taiwan sur une période de 12 ans (de 2000 à 2011) a été réalisée (Lin et al., 2020). La cohorte était composée de 161 970 citoyens taiwanais qui n'avaient aucun diagnostic préalable d'IRC ou d'IRCT avant le début de l'étude. Une association significative a été constatée entre l'exposition à la pollution atmosphérique en PM_{2.5}, NO, NO_x, SO₂ et un risque accru de développer l'IRCT. En comparant le dernier quartile de contamination au premier, les risques relatifs suivants ont été obtenus : PM_{2.5} (RR : 1.69 ; IC 95% [1.32-2.16]), NO (RR : 1.54 ; IC 95% [1.20-1.98]), NO_x (RR : 1.70 ; IC 95% [1.33-2.18]), SO₂ (RR : 1.32 ; IC 95% [1.03-1.70]). Par ailleurs, de plus en plus d'études démontrent également l'influence

négligée de la pollution de l'air sur l'évolution clinique de l'IRC. Un impact négatif sur la fonction rénale a été observé chez les personnes déjà atteintes de l'IRC (Chang et al., 2022). En France, une étude menée sur le périmètre national a permis de décrire une association significative entre la multi-contamination de l'air et le risque de décès chez les patients dialysés (Hamroun et al., 2023b).

Les mécanismes sous-jacents qui pourraient expliquer l'impact de la pollution de l'air sur la survenue et la progression de l'IRC sont complexes et multiples (Chen et al., 2021; Shubham et al., 2022). L'exposition chronique aux polluants atmosphériques pourrait entraîner un stress oxydatif et également une inflammation systémique, déclenchant ainsi des dommages vasculaires et une altération de la fonction rénale. Par ailleurs, la pollution de l'air pourrait également perturber le fonctionnement du système rénine-angiotensine-aldostérone (SRAA), qui joue un rôle clé dans la régulation de la pression artérielle et de l'équilibre hydrique dans le corps. Ainsi, une activation excessive du SRAA en réponse à la pollution de l'air pourrait contribuer à une vasoconstriction des vaisseaux rénaux, réduisant ainsi la perfusion sanguine dans les reins. Cette altération de la perfusion rénale pourrait entraîner des lésions et une altération progressive de la fonction rénale.

b. MC et contamination métallique des sols

Les données disponibles pour conduire notre étude suggèrent une association potentielle entre la pollution des sols agricoles en métaux et la répartition spatiale d'incidence de la MC. Dans la littérature existante, cette association n'a jamais été clairement mise en évidence (Tenailleau et al., 2020). En revanche, plusieurs études ont montré qu'une exposition à certains métaux pouvait impacter le microbiote intestinal et potentiellement induire des phénomènes de dysbiose (Jin et al., 2017). Les résultats allant dans ce sens sont essentiellement issus de recherches menées sur modèle murin. Des effets sur le microbiote intestinal ont été rapportés pour des métaux tels que l'aluminium (Lerner, 2007), l'arsenic (Dheer et al., 2015), le cadmium (Popov Aleksandrov et al., 2021), le plomb (Wu et al., 2016), le fer (Aamodt et al., 2008). Le déséquilibre du microbiote intestinal provoqué par une exposition aux métaux pourrait conduire de manière indirecte à une réponse inflammatoire.

Dans une étude menée au Portugal par Nair et al. (2019), l'implication éventuelle des métaux dans le développement des MICI a été examinée, en mettant l'accent sur les expositions pendant la petite enfance. Au Portugal, il est courant de conserver les dents de lait des enfants pendant de nombreuses années après leur chute. Au total, les dents de lait de 30 participants adultes ont été collectées et analysées. Parmi eux, 14 participants étaient atteints de MICI (8 MC, 6 RCH) et 16 participants non affectés. Les chercheurs ont utilisé la spectrométrie de masse avec ablation laser pour analyser les dents de lait et créer des profils temporels d'exposition aux métaux, en couvrant la période allant du deuxième trimestre de la grossesse jusqu'aux six premiers mois de la vie. Après analyse, des différences significatives ont été observées au niveau des teneurs en métaux dans les dents de lait des personnes atteintes de MICI, par rapport aux témoins. L'exposition au plomb était significativement plus élevée pendant la période intra-utérine et au cours des six premiers mois de la vie chez les patients atteints de MICI. De même, l'exposition au cuivre s'est avérée plus élevée au cours des 15 premières semaines après l'accouchement, tandis que l'exposition au chrome était plus élevée entre 10 et 15 semaines après l'accouchement. Ces résultats suggèrent qu'une exposition aux métaux et des perturbations dans leur absorption pendant la période critique du développement pourraient être associées à un risque ultérieur de développer une MICI, bien avant l'apparition des symptômes cliniques.

En épidémiologie spatiale, plusieurs études ont été menées pour étudier l'association entre incidence de la MC et contamination métallique dans les eaux de distribution. Parmi ces études, à notre

connaissance, une seule a mis en évidence une association significative. Dans le sud-est de la Norvège, Aamodt et al. (2008) ont étudié le lien entre la qualité de l'eau potable et l'incidence des MICI. Ils ont constaté qu'une augmentation de 0,1 mg/L de concentration en fer dans l'eau potable était associée à une augmentation significative de 21 % du risque relatif de développer une MICI. Les auteurs suggèrent deux mécanismes pour expliquer ces observations. Premièrement, une concentration élevée en fer agirait comme un catalyseur du stress oxydatif, ce qui pourrait produire une inflammation. Deuxièmement, la teneur en fer pourrait stimuler la croissance des bactéries et ainsi augmenter la probabilité d'une réponse immunitaire inappropriée chez les individus prédisposés génétiquement.

4. Conclusion

Dans un premier temps, quatre méthodes ont été comparées pour construire des indices de multi-contamination des milieux à l'échelle des communes. Finalement, la méthode du RIM, basée sur la confrontation à des seuils, a été retenue. Les indices de multi-contamination calculés pour chaque milieu ont ensuite été confrontés aux variations d'incidence de l'IRCT et de la MC. Deux hypothèses principales se sont dégagées de l'analyse. Tout d'abord, une corrélation entre pollution de l'air et IRCT a été constatée. Ce résultat semble particulièrement cohérent au regard de la littérature existante. De plus en plus d'études démontrent qu'une exposition à long terme à la pollution de l'air est associée à un excès de risque de développer l'IRC. Par ailleurs, les résultats suggèrent également une corrélation entre la pollution des sols en ETM et la MC. Sur ce point, la littérature existante est encore mince et les résultats contradictoires. Nos travaux constituent donc une avancée majeure sur l'identification de piste écologique, préalable à la construction de toute hypothèse étiologique. Ils renforcent également la démonstration de l'influence potentielle des sols, milieu peu considéré, dans l'exposition. Toutefois, même si ces premières approches sont prometteuses, l'ensemble mériterait, à court termes, d'être renforcé par la mise en œuvre de modèles de régression spatiale.

CHAPITRE IV – Décrire le profil environnemental des territoires : une stratégie multi-dimensionnelle pour caractériser l’environnement physique.

Introduction

Dans le premier chapitre, la notion d'environnement a été appréhendée au sens large, et les indices de vulnérabilité et de résilience ont été élaborés pour quantifier l'accumulation de facteurs environnementaux favorables ou défavorables à l'échelle des territoires sur la base de 6 dimensions. Puis, l'analyse a été recentrée sur l'environnement physique dont la puissance explicative dans le cadre des inégalités de santé est rarement explorée sans dimension sociale. En conséquence, un jeu de données environnemental complet a été construit dans le chapitre II. Ces données ont été exploitées dans le chapitre III pour générer des indices de multi-contamination permettant d'évaluer la contamination globale des milieux à l'échelle des territoires.

En focalisant l'analyse sur les données de contamination environnementale dans le chapitre III, nous avons pu émettre des hypothèses concernant le lien entre les maladies étudiées et l'environnement. Les résultats ont montré une possible influence de la pollution de l'air sur l'IRCT. Pour la MC, un lien a été suggéré avec la pollution métallique des sols. Nous avons également mis en évidence la limite des données de contamination, qui concernent seulement certains polluants. La question des produits phytosanitaires n'a par exemple pas pu être explorée directement à travers les données de contamination actuellement disponibles. Nous ne manquons cependant pas de données permettant d'approcher cette contamination de manière indirecte, *via* l'occupation des sols agricoles ou les indicateurs d'usage qui ont été développés et inclus dans le chapitre II.

Pour étudier les facteurs de risque potentiels liés au territoire, il semble aussi nécessaire d'adopter une approche plus holistique qui tienne compte à la fois de la diversité des pressions anthropiques qui s'exercent sur l'environnement, et des aménités environnementales ou de la naturalité des territoires. Dans ce chapitre, nous allons nous attacher à mobiliser l'ensemble des données construites dans le chapitre II, afin de proposer une description plus complète et intégrée de l'environnement physique.

Une réflexion doit donc être menée sur la meilleure stratégie à retenir pour mobiliser l'ensemble de ces indicateurs spatialisés. L'agrégation des données environnementales pour créer un « super indice composite », fournissant une mesure globale de l'état de l'environnement, se révèle être une première possibilité. Cette stratégie pourrait effectivement être pertinente pour établir une hiérarchie entre les territoires et pour appréhender la notion d'accumulation de facteurs favorables ou défavorables. Néanmoins, deux communes pourraient présenter des valeurs d'indice similaires mais être extrêmement différentes sur le plan environnemental. Cette limite a d'ailleurs été mise en évidence dans les précédents travaux du laboratoire (Brousriche et al., 2023). Cette piste n'a donc pas été explorée.

Dans ce chapitre, nous proposons une autre démarche permettant d'identifier et de comprendre quels sont les facteurs sous-jacents qui différencient ou qui rapprochent les communes les unes des autres.

Il n'est ainsi plus question d'agréger les données pour générer un indice composite, mais plutôt d'analyser conjointement tous les indicateurs pour parvenir à une description multivariée des communes et définir leur profil environnemental. Ce travail ne peut cependant pas être réalisé individuellement pour chaque unité spatiale, il faut parvenir à identifier les principaux profils environnementaux représentés dans la zone d'étude.

Pour exploiter ce jeu de données spatial et décrire le profil environnemental des communes, nous proposons de mettre en œuvre des méthodes de classification multivariées. Ces méthodes permettent de construire des classes (ou typologies) en regroupant les individus qui partagent des traits communs (Madhulatha, 2012). Elles peuvent être appliquées sur des ensembles de données spatiales larges et complexes. Leur atout principal est de simplifier l'information contenue dans le jeu de données, mettant ainsi en évidence des patterns géographiques et des phénomènes spatiaux qu'il aurait été difficile d'observer autrement. Ces méthodes ont été appliquées sur des données spatiales dans de nombreux domaines et dans divers contextes, tels que l'agriculture (Alvarez et al., 2018), l'économie (Segessemann and Crevoisier, 2016), la santé (Wendt, 2009) ou l'énergie (Weinand et al., 2019).

Dans le champ environnemental, ces approches ont également été couramment utilisées. Cependant, la définition du terme "environnement" varie d'un auteur à un autre. Certains considèrent l'environnement au sens large en incluant non seulement l'environnement physique, mais également des facteurs sociaux, économiques et démographiques. Par exemple, Fayet et al. (2020) ont introduit la « Geographical Classification for Health Studies » (GeoClasH), qui a été développée sur la base de dix indicateurs spatiaux décrivant l'environnement physique et social, et l'accessibilité spatiale aux soins. Parmi ces dix indicateurs, deux concernent spécifiquement l'environnement physique, plus précisément la qualité de l'air et l'occupation des sols par les cultures agricoles. Ils ont mis en évidence cinq profils de communes à l'échelle de la France métropolitaine. Dans l'étude de Brousmiche et al. (2023) l'environnement physique a été évalué par huit descripteurs spatiaux qui décrivent des facteurs associés avec les maladies cardio-vasculaires dans la littérature : la pollution de l'air, l'exposition au bruit, l'accès aux espaces verts et la monétisation des services écosystémiques. Sur le périmètre de la métropole de Lille, trois profils territoriaux ont été identifiés à l'échelle infra-communale (IRIS) à partir de ces données, ainsi qu'à partir d'indicateurs socio-économiques. Pour d'autres auteurs, l'environnement est effectivement appréhendé comme l'environnement physique et extérieur. Dans ce cas, les efforts de recherche tendent à se concentrer sur la description d'une dimension en particulier : services écosystémiques (Mullin et al., 2018; Turner et al., 2014), paysage (Pecher et al., 2013), ou climat (Praene et al., 2019; Villoslada et al., 2017). Une approche plus holistique a été introduite par Shortt et al. (2012, 2010). En considérant cinq dimensions environnementales (accès aux espaces verts, qualité de l'air, proximité aux industries polluantes, température et rayonnement ultraviolet), ils ont identifié sept profils territoriaux au Royaume-Uni et six en Nouvelle-Zélande. La dimension climatique a eu un impact significatif sur la classification obtenue. Par exemple, en Nouvelle-Zélande, trois profils à prédominance rurale ont été identifiés, ces derniers se distinguaient essentiellement sur le plan climatique. Dans la littérature existante, relativement peu d'études traitent donc, de manière exhaustive, l'environnement physique en incluant un grand nombre de descripteurs. En effet, il n'existe pas à notre connaissance de typologie territoriale construite à partir de données ouvertes et traitant simultanément de la contamination de l'air, de l'eau et du sol.

Dans ce contexte, ce travail de recherche vise à développer un cadre méthodologique permettant d'identifier et de décrire la diversité profils environnementaux représentés dans une zone d'étude. Pour ce faire, une méthodologie multivariée a été élaborée en utilisant des données ouvertes et des méthodes de classification non supervisée. L'originalité de cette approche réside dans sa focalisation sur l'environnement physique, en prenant en compte la répartition spatiale des pressions et aménités environnementales, ainsi que la multi-contamination des milieux. Ce cadre méthodologique a été appliqué à l'échelle communale sur le périmètre du registre Epimad (3 041 communes).

Les travaux de ce chapitre ont été valorisés sous la forme d'un article intitulé : « Description of the multi-dimensional environment at the territorial scale: A holistic framework using cluster analysis and open data in France » (Paumelle et al., 2023).

Matériels & Méthodes

Les principales étapes méthodologiques peuvent être résumées comme suit : collecte des données, sélection des données, préparation des données, classification non supervisée et interprétation des classes (Figure 52).

Collecte des données. Pour décrire l'environnement physique et extérieur à l'échelle des communes, un jeu de données complet comprenant 113 indicateurs a été développé. Cette étape est présentée en détail dans le second chapitre de la thèse.

Sélection des données. Avant de procéder à tout traitement statistique, les indicateurs spatiaux ont été triés afin de limiter la redondance d'information et pour accorder une importance similaire aux différents phénomènes décrits. De même, lorsque plusieurs options étaient disponibles pour décrire un même phénomène, une décision a été prise pour sélectionner l'indicateur le plus approprié et le plus représentatif. Cette étape préliminaire est cruciale, elle contribue à la qualité et à l'interprétabilité des résultats finaux.

Préparation des données. Étant donné que les indicateurs peuvent être exprimés dans des unités différentes, générés par des producteurs différents et à des fins différentes, une étape de normalisation a été engagée pour garantir la comparabilité des indicateurs (Böhringer and Jochem, 2007; Saib et al., 2015). De plus, en raison du grand nombre d'indicateurs, une étape de réduction dimensionnelle (ACP) a été appliquée avant d'initier la classification (Everitt, 1979).

Classification non supervisée. En utilisant des algorithmes de regroupement, l'objectif était d'obtenir une classification à la fois suffisamment complexe et informative, et des classes cohérentes et faciles à interpréter.

Interprétation des classes. Une fois les classes identifiées, une analyse descriptive a été menée pour décrire et caractériser le profil environnemental associé à chaque classe.

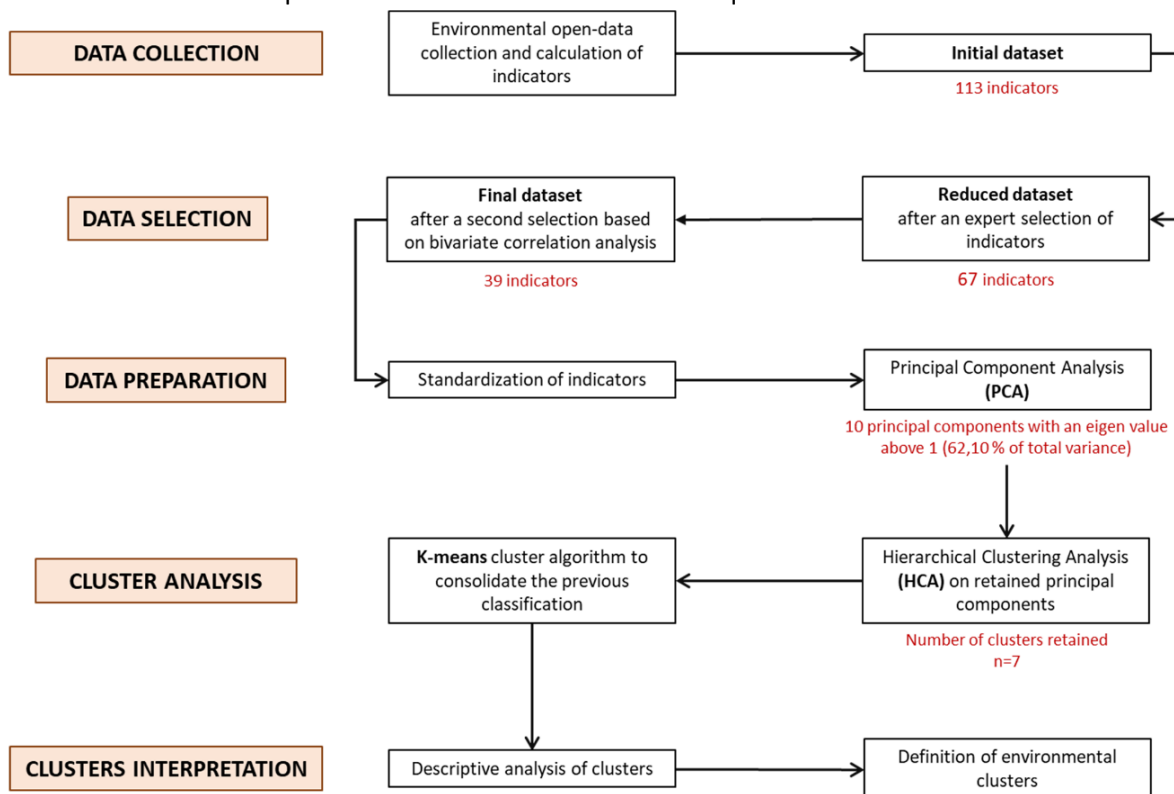


Figure 52 : Principales étapes méthodologiques pour développer une typologie territoriale. Les résultats qui concernent le cas d'étude sont représentés en rouge (Paumelle et al., 2023).

1. Sélection des données d'entrée

Sélection experte. Au préalable, une sélection experte des données a été mise en œuvre. L'objectif était d'évaluer l'équilibre et la cohérence des phénomènes décrits pour garantir une caractérisation adéquate de l'environnement physique. Lorsque plusieurs options étaient disponibles, des choix ont été faits pour sélectionner les indicateurs les plus adaptés et les plus représentatifs de l'état des milieux. L'ensemble des choix réalisés seront présentés en détail dans le premier paragraphe des résultats.

Sélection mathématique : analyse de corrélation bivariée. Ensuite, une seconde sélection basée sur l'analyse de corrélation bivariée a été réalisée pour identifier les variables redondantes et ainsi résumer l'information majeure contenue dans le jeu de données (Jolliffe, 1972; King and Jackson, 1999). Un seuil maximal pour les coefficients de corrélation de Pearson a été fixé à 0,70 (Dormann et al., 2013b; United States Environmental Protection Agency, 2020). Les indicateurs ont été triés de manière à ce qu'aucun coefficient de corrélation de Pearson ne dépasse 0,70 (en valeur absolue). Tous ces choix ont été réalisés avec la volonté de conserver les indicateurs les plus représentatifs, afin d'obtenir un jeu de données aussi complet et diversifié que possible. De nouveau, l'ensemble des choix réalisés seront présentés en détail dans le premier paragraphe des résultats.

2. Préparation des données

Normalisation des indicateurs. D'abord, chacun des indicateurs a été normalisé en utilisant la méthode recommandée du z-score (Marzi et al., 2018). De cette manière, tous les indicateurs sont placés sur une échelle commune, avec une moyenne égale à zéro et un écart-type égal à 1. Cette méthode est moins sensible aux valeurs extrêmes que d'autres méthodologies et elle préserve les structures de corrélation entre indicateurs (Kolenikov and Angeles, 2004; OECD et al., 2008a).

Analyse en composantes principales. Après normalisation, une ACP a été appliquée à l'ensemble de données final afin de définir le sous-espace qui représente au mieux la diversité des individus (en l'occurrence les communes). Ensuite, seules les composantes principales ayant une valeur propre supérieure à 1 ont été conservées pour la suite de l'analyse.

3. Méthode de classification

Les scores factoriels des communes sur les composantes principales retenues ont été utilisés comme données d'entrée pour initier la classification. La distance euclidienne a été utilisée pour mesurer la dissimilarité les communes.

Dans un premier temps, une classification ascendante hiérarchique (CAH) a été appliquée en utilisant le critère de Ward. Ce critère consiste à minimiser l'inertie intra-classe à chaque nouveau regroupement effectué lors de la CAH (Ward, 1963). Ensuite, pour déterminer le nombre optimal de classes à retenir, la méthode du coude a été employée (Tibshirani et al., 2001). Cette méthode implique d'analyser comment les gains d'inertie inter-classe évoluent à mesure que le nombre de classes augmente. Lorsque le gain d'inertie inter-classe diminue fortement, cela suggère que l'ajout d'une classe supplémentaire n'apporterait qu'une amélioration marginale à la qualité de la classification. En parallèle, des critères qualitatifs comme l'interprétabilité des classes ont également été pris en compte

pour guider le choix du nombre de classes à retenir. L'objectif était d'obtenir une classification à la fois suffisamment complexe et informative, tout en conservant la cohérence et la facilité d'interprétation des classes.

Dans un deuxième temps, un algorithme des k-means a été appliqué (Hartigan and Wong, 1979). Les centroïdes des classes obtenues lors de la CAH ont été utilisés pour initialiser l'algorithme des k-means. L'information hiérarchique issue de la CAH est perdue mais cette étape permet de consolider la classification (Praene et al., 2019). Toutes les analyses statistiques ont été réalisées à l'aide du logiciel R (R Core Team, 2021).

4. Interprétation des profils des communes

Pour décrire le profil environnemental associé à chaque classe, nous avons d'abord calculé la médiane et l'écart interquartile de chaque indicateur d'entrée au sein des différentes classes. Pour analyser les différences entre les classes, les indicateurs d'entrée ont été soumis au test de Kruskal-Wallis ($p < 0,05$) (McKight and Najab, 2010). En cas de significativité du test, des tests post hoc de Dunn avec ajustement de Bonferroni ont été menés pour analyser dans le détail les différences entre classes (Dinno, 2015).

En parallèle, pour faciliter l'interprétation des résultats, tous les indicateurs d'entrée ont été normalisés et placés à la même échelle grâce à la méthode du z-score. Cette transformation permet ensuite d'examiner la distribution des z-scores au sein de chaque classe. Un graphique unique a été généré pour chaque classe, permettant ainsi de visualiser d'un coup d'œil la distribution des z-scores de tous les indicateurs d'entrée. Cette représentation globale permet d'identifier les indicateurs qui affichent des valeurs atypiques par rapport à l'ensemble de la zone d'étude, c'est-à-dire les indicateurs qui caractérisent le mieux chaque classe. Plus concrètement, si une commune présente un z-score de zéro pour un indicateur donné, cela correspond exactement à la moyenne de cet indicateur sur l'ensemble de la zone d'étude. Si une commune présente un z-score de +1 (ou -1), cela signifie qu'elle se situe à un écart-type au-dessus (ou au-dessous) de la moyenne.

Enfin, la classification obtenue a été cartographiée pour visualiser la répartition spatiale des profils environnementaux au sein de la zone d'étude.

Résultats

1. Sélection des données d'entrée

Sélection d'experts. Le jeu de données initial contenant 113 indicateurs a été trié sur la base d'une sélection experte. Les principales décisions qui ont été prises sont détaillées ci-dessous.

Pour analyser la contamination métallique des sols, les données RMQS (Arrouays et al., 2003) ont été utilisées. Elles fournissent à la fois les concentrations totales et les concentrations de la phase extractible pour la plupart des métaux étudiés. Nous avons calculé à l'échelle communale les concentrations moyennes pour les deux protocoles d'analyses. Cependant, pour plusieurs métaux, les données sur la phase extractible étaient manquantes. Pour assurer la comparabilité des indicateurs, nous avons donc uniquement retenu les indicateurs décrivant la concentration totale.

Pour étudier la proximité aux stations d'épuration, nous avons développé deux indicateurs : l'un prenant en compte toutes les stations, qu'elles soient conformes ou non, et l'autre ne prenant en compte que les stations non conformes. Cependant, l'information sur la non-conformité n'était

disponible que pour l'année 2017. Ainsi, dans un souci de représentativité, nous avons choisi de retenir l'indicateur général qui prend en compte toutes les stations.

Pour évaluer la pression phytosanitaire, les données de la BNVD (Système d'Information sur l'Eau and Office Français de la Biodiversité, 2023). ont été utilisées. À partir de ces données, nous avons développé 15 indicateurs de pression phytosanitaire à l'échelle communale. Un indicateur global prend en compte l'ensemble des substances actives référencées, trois indicateurs spécifiques ont été calculés par type d'usage (herbicides, fongicides, insecticides), et 11 indicateurs ont été calculés pour une substance active en particulier. Ces 11 indicateurs spécifiques ont été éliminés au profit des quatre indicateurs plus généraux.

Concernant les installations de traitement et de gestion des déchets, nous avons initialement dix indicateurs décrivant la proximité à des installations spécifiques (incinérateurs, installations de stockage, centres de tri, plateformes de compostage, etc...). Cependant, en raison de leur distribution spatiale très dispersée, nous avons choisi de retenir un seul indicateur de proximité général qui prend en compte sans distinction tous les types d'installations de traitement et de gestion des déchets.

Une autre décision majeure a été prise, à savoir l'exclusion des données climatiques du jeu de données. Cette démarche visait à prévenir toute influence des zones climatiques de la région sur la typologie territoriale obtenue, afin d'éviter de mettre en évidence, par exemple, des profils fortement influencés par la proximité avec la façade maritime. Cette problématique avait été soulevée dans des études antérieures (Shortt et al., 2012;2010). D'autre part, il s'agit plutôt ici de caractériser la qualité des milieux environnementaux.

Enfin, dans certains cas, nous disposons de plusieurs sources de données pour étudier le même phénomène. Dans ces situations, nous avons sélectionné l'indicateur généré à partir de la base de données la plus complète.

Finalement, parmi les 113 indicateurs initialement inclus dans l'ensemble de données, 67 ont été retenus.

Sélection mathématique : analyse de corrélation bivariée.

Cette étape sert à éliminer les informations redondantes. Nous avons d'abord traité les corrélations entre les indicateurs de contamination. Les trois polluants atmosphériques considérés ($PM_{2.5}$, PM_{10} et NO_2) étant fortement corrélés, nous avons choisi de retenir l'indicateur de contamination $PM_{2.5}$. A la lumière des connaissances scientifiques actuelles, il semblait plus informatif en raison de la diversité des sources d'émissions et de leurs implications très documentées sur la santé publique (Nan et al., 2023). De même, pour la contamination des sols, des corrélations significatives ont été trouvées entre les différents métaux. Nous avons choisi le zinc comme marqueur synthétique de quatre autres éléments très fortement corrélés : le cuivre, le cadmium, le chrome et le nickel. Ainsi, nous avons finalement retenu quatre indicateurs de contamination des sols : arsenic, mercure, zinc et thallium.

Pour les niveaux d'émissions atmosphériques, nous avons exclu en priorité les polluants déjà décrits par les données de contamination. Nous avons donc retenu uniquement les données sur les polluants suivants : dioxyde de soufre, composés organiques volatils non méthaniques, ammoniac, arsenic et gaz à effet de serre.

Concernant l'occupation des sols, nous avons conservé la proportion de territoires artificialisés et non la proportion de terres agricoles. En effet, l'utilisation des sols agricoles est décrite de manière

beaucoup plus précise via les indicateurs développés à partir du registre parcellaire graphique qui est intégré dans l'analyse.

Pour évaluer la pression phytosanitaire, nous avons conservé l'indicateur global prenant en compte l'ensemble des substances actives et l'indicateur relatif aux substances insecticides. Les indicateurs relatifs aux herbicides et aux fongicides ont été écartés en raison de leur forte corrélation avec l'indicateur global.

Le jeu de données final comprend 39 indicateurs.

2. Jeu de données final et réduction dimensionnelle (ACP)

Les 39 indicateurs spatiaux retenus ont été répartis entre les sous-dimensions comme suit : niveaux de contamination (n=10) ; niveaux d'émission (n=5) ; localisation des sources d'émission (n=5) ; utilisation des sols (n=11) ; pratiques agricoles (n=5) ; naturalité (n=3). La liste détaillée de ces indicateurs est fournie en Tableau 11.

Par ailleurs, à l'issue de l'ACP, les dix premières composantes principales ont été retenues pour la suite de l'analyse. Ces dix premières composantes principales représentent 62,10 % de la variance totale contenue dans le jeu de données.

Tableau 11 : Jeu de données final retenu pour la détermination des profils environnementaux. (modifié à partir de Paumelle et al., 2023).

SUB-DIMENSION	INDICATOR	DESCRIPTION
CONTAMINATION LEVELS	PM2.5_AIR	Mean concentration of PM2.5 in ambient air ($\mu\text{g}/\text{m}^3$)
	AS_SOIL	Arsenic concentration in soils (mg/kg)
	HG_SOIL	Mercury concentration in soils (mg/kg)
	ZN_SOIL	Zinc concentration in soils (mg/kg)
	TL_SOIL	Thallium concentration in soils (mg/kg)
	ENTERO_WATER	Enumeration of enterococci (CFU/100mL)
	PH_WATER	Mean pH in drinking water
	CONDUCT_WATER	Mean conductivity in drinking water ($\mu\text{S}/\text{cm}$)
	TURBIDITY_WATER	Mean turbidity in drinking water (NFU)
CL2TOT_WATER	Mean concentration of total chlorine in drinking water (mg/L)	
EMISSION LEVELS	GHGs_EM	Greenhouse gas atmospheric emissions (thousand metric tons of CO2 equivalent)
	SO2_EM	Sulfur dioxide atmospheric emissions (kg/ha/an)
	NMVOC_EM	Non-methane volatile organic compound atmospheric emissions (kg/ha/an)
	NH3_EM	Ammoniac atmospheric emissions (kg/ha/an)
	AS_EM	Arsenic atmospheric emissions (kg/ha/an)
LOCATION OF EMISSION SOURCES	POLLUT_SITES	Density of contaminated sites and soils (number per hectare)
	CLASSIF_FACILITIES	Density of facilities classified for environmental protection (number per hectare)
	INDUST_ACCIDENTS	Density of industrial accidents recorded since 1960 (number per hectare)
	WASTE_PROX	Distance to the closest waste management installation (km)
	SEWAGE_PROX	Distance to the closest wastewater treatment plant (km)
LAND USE	ARTIF_LAND	Percentage of municipality's surface area occupied by artificial lands (%)
	HUMID_LAND	Percentage of municipality's surface area occupied by wetlands (%)
	WATER_LAND	Percentage of municipality's surface area occupied by water surfaces (%)
	GRASSLAND	Percentage of municipality's surface area occupied by grasslands (%)
	PERENNIAL_CROPS	Percentage of municipality's surface area occupied by perennial crops (%)
	CEREALS	Percentage of municipality's surface area occupied by cereals (%)
	OILSEEDS	Percentage of municipality's surface area occupied by oil crops (%)
	PROTEIN_CROPS	Percentage of municipality's surface area occupied by protein crops (%)
	VEGETABLES	Percentage of municipality's surface area occupied by vegetables and flowers (%)
INDUSTRIAL_CROPS	Percentage of municipality's surface area occupied by industrial crops (%)	
ARTIF_2009/2019	Percentage of municipality's surface area artificialized between 2009 and 2019 (%)	
AGRICULTURAL PRACTICES	TOTPEST	Total phytosanitary pressure index (kg/ha/an)
	INSECTICIDE	Insecticide pressure index (kg/ha/an)
	LIVESTOCK	Number of livestock units per hectare of municipality's surface area (livestock units per hectare)
	HNV	High Natural Value indicator
	AGRI_PRACTICES	Indicator of extensification of farming practices
NATURALNESS	NATURALNESS	Potential naturalness indicator
	BIODIV_RESERVOIRS	Percentage of the population having access to a green space within 400 meters (%)
	GREEN_SPACES	Percentage of municipality's surface area occupied by biodiversity reservoirs (%)

3. Classification obtenue

Le nombre optimal de classes a été déterminé en se basant sur les gains d'inertie inter-classe obtenus avec la CAH, comme illustré en figure 53(c). En complément, des critères qualitatifs ont également été pris en compte. Le fait de retenir 2, 3 ou 4 classes aurait été approprié sur le plan statistique. Cependant, l'un des objectifs de l'analyse est de dépasser la dichotomie urbain-rural et d'obtenir des profils environnementaux plus informatifs et complexes. Pour atteindre cet objectif, l'analyse a donc été poussée plus loin. Le gain d'inertie inter-classe restait relativement stable lors du passage de 4 à 5 classes, de 5 à 6 classes ou de 6 à 7 classes (1,169, 0,997 et 0,963, respectivement). Ensuite, le gain d'inertie chute à 0,689 lors du passage de 7 à 8 classes. En conséquence, nous avons retenu 7 classes comme nombre optimal pour la classification. Le dendrogramme de la CAH est illustré à la figure 53(b). Les sept classes ainsi obtenues à l'issue de la CAH ont ensuite été utilisées pour initier l'algorithme des K-means.

L'algorithme des K-means a permis d'améliorer la classification initiale obtenue avec la CAH. L'inertie inter-classe a significativement augmenté de 10,79 à 11,90 (soit une augmentation de 10,3%) pour un nombre constant de classes ($n=7$). Une comparaison entre les résultats obtenus avant et après la consolidation des K-Means est fournie en annexe (ANNEXE 5). Finalement, la classification finale obtenue est représentée en figure 53(a).

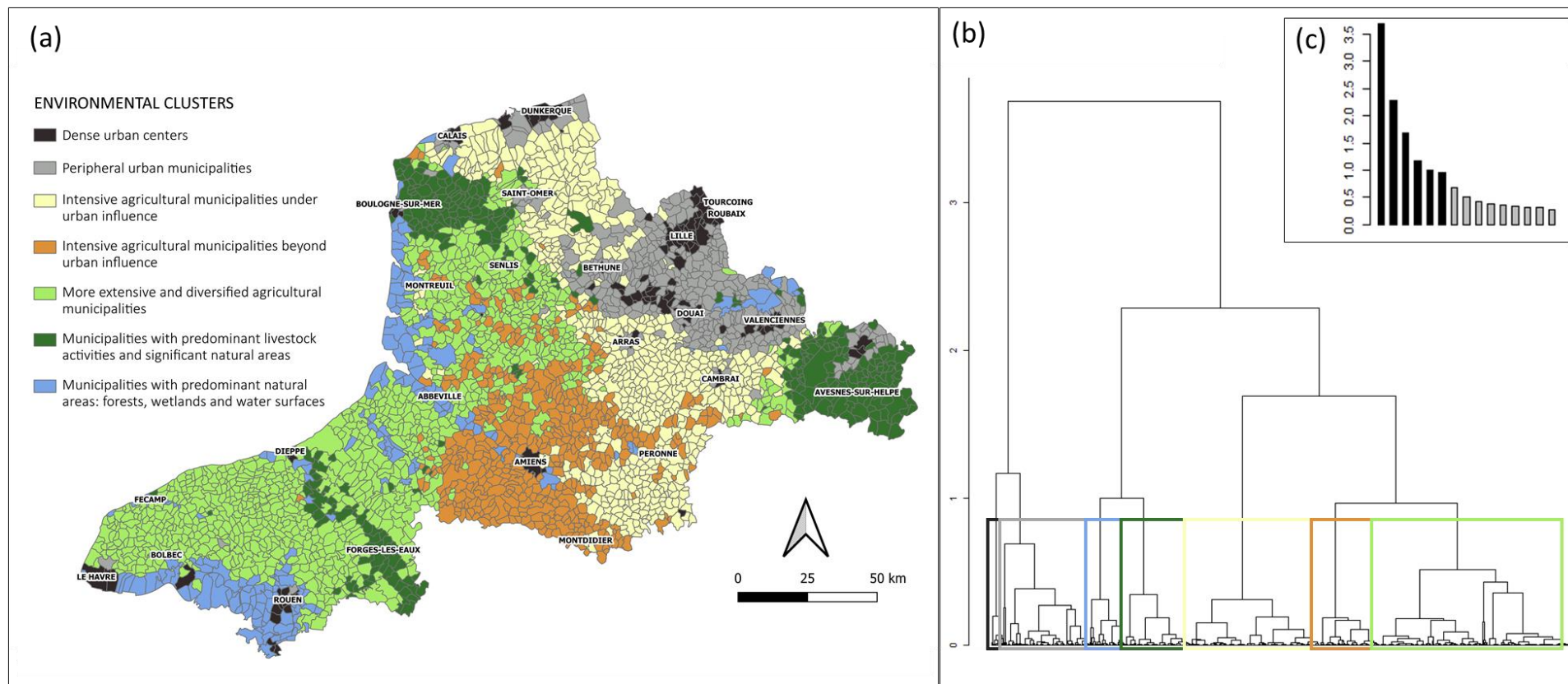
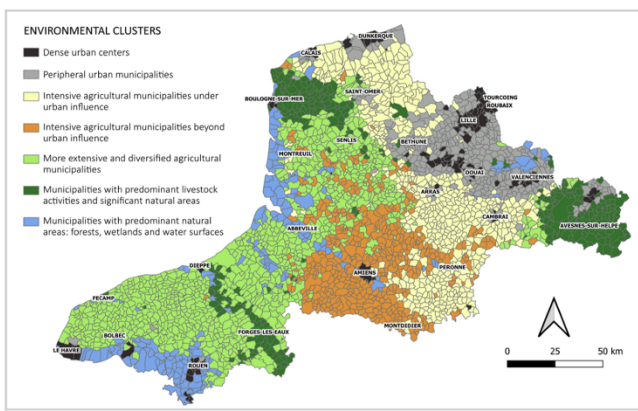


Figure 53 : (a) Représentation cartographique de la classification obtenue après CAH et consolidation des k-means ; (b) Dendrogramme issu de la CAH ; (c) Gains d'inertie inter-classe obtenus avec la CAH selon le nombre de classes retenues (c) (Paumelle et al., 2023).

4. Description et interprétation des profils de communes

Pour chacun des 39 indicateurs d'entrée, nous avons calculé la valeur médiane et l'intervalle interquartile dans chaque classe, comme présenté dans le Tableau 12. Dans le but d'identifier les indicateurs les plus atypiques associés à chaque classe, nous avons également représenté les résultats des tests post hoc de Dunn. Lorsqu'un indicateur présentait des valeurs significativement plus élevées dans une classe par rapport à toutes les autres, la cellule correspondante du tableau a été colorée en rouge vif. Si plusieurs classes n'étaient pas significativement différentes les unes des autres, mais présentaient des valeurs significativement plus élevées par rapport à toutes les autres classes, les cellules correspondantes ont été représentées en rouge pâle. À l'inverse, nous avons utilisé le bleu clair et le bleu pâle pour mettre en évidence les valeurs notablement inférieures sur le plan statistique.

De plus, pour chacune des sept classes, nous avons créé un graphique représentant les distributions des z-scores de tous les indicateurs d'entrée (Figure 54). Les résultats des tests post hoc de Dunn ont également été représentés sur ces graphiques à l'aide d'étoiles en utilisant le même code couleur que dans le tableau 12.



- ★ This indicator is significantly higher in that cluster than in any other.
- ★ This indicator is significantly lower in that cluster than in any other.
- ★ This indicator is significantly higher in that cluster than in all others, except those sharing the same color.
- ★ This indicator is significantly lower in that cluster than in all others, except those sharing the same color.

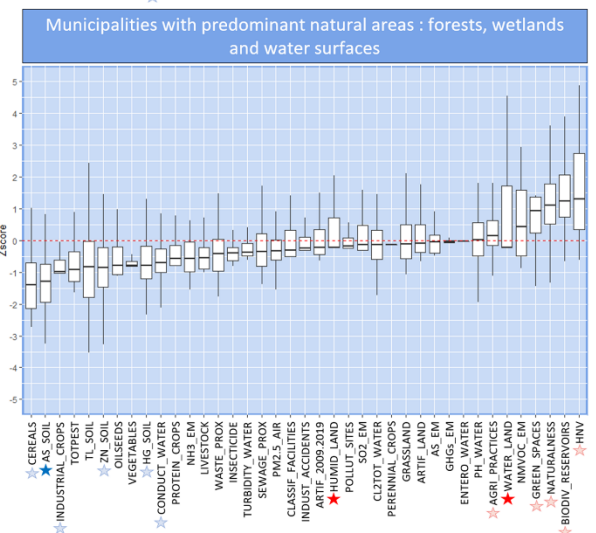
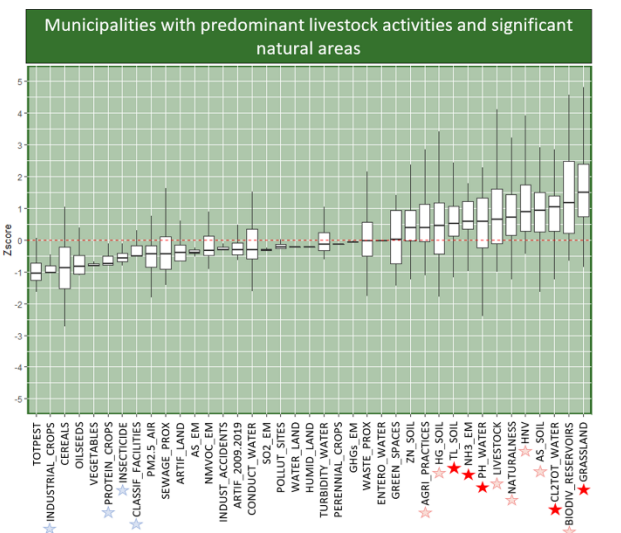
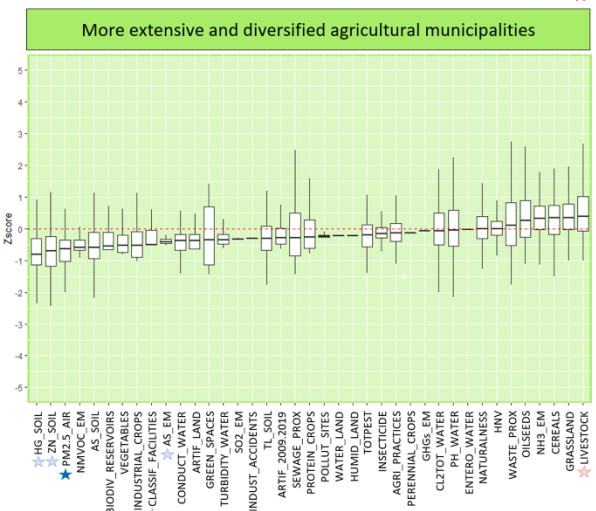
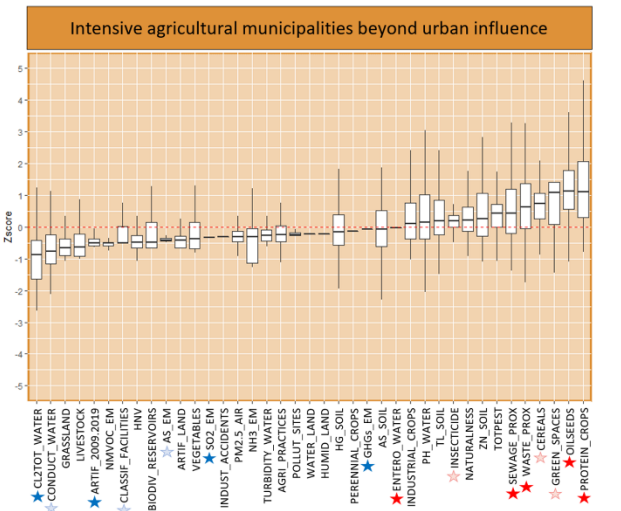
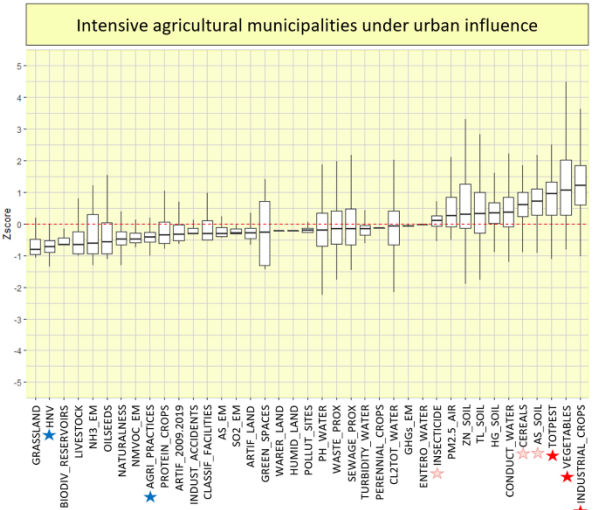
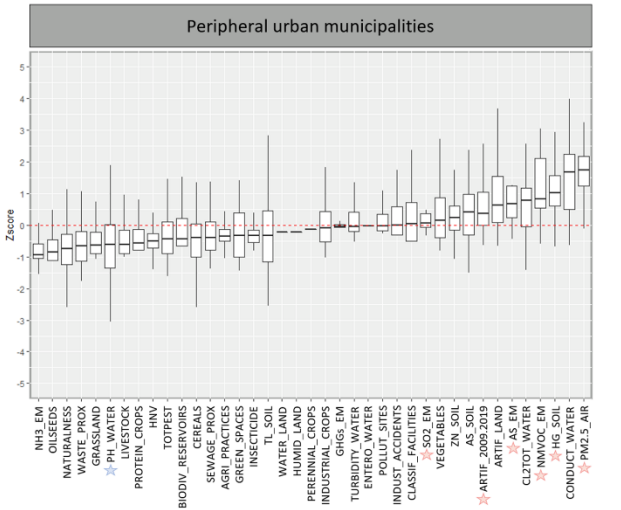
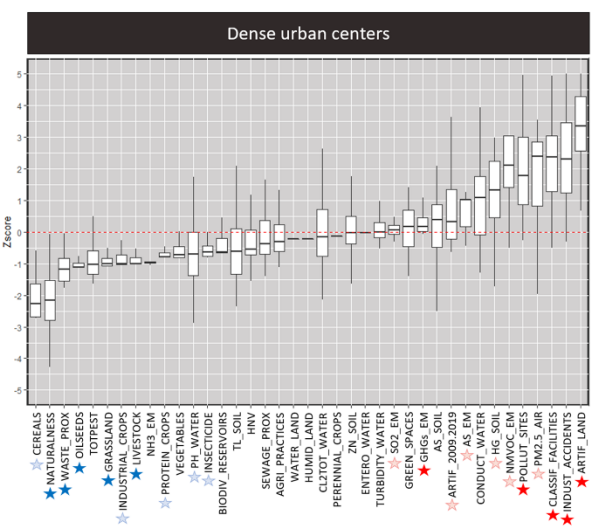


Figure 54 : Boxplots représentant la distribution des z-scores pour les 39 indicateurs d'entrée au sein de chaque profil (Paumelle et al., 2023).

Pour simplifier le graphe, les valeurs extrêmes n'ont pas été représentées.

Centres urbains denses. Cette classe compte le plus petit nombre de communes (n=103) et le plus grand nombre d'indicateurs dont le z-score médian est supérieur à 1 ou inférieur à -1 (n=15). Ainsi, comparé à l'ensemble de la zone d'étude, ce profil présente les caractéristiques les plus atypiques. En particulier, ces communes sont fortement soumises à la pression urbaine sous toutes ses formes. Tout d'abord, 71% [53;87] (médiane [EIQ]) de la surface de la commune est occupée par des territoires artificialisés. De plus, les indicateurs suivants sont également significativement plus élevés dans cette classe : les émissions de gaz à effet de serre, la densité de sites et sols pollués, la densité d'installations classées pour la protection de l'environnement, la densité d'accidents industriels, la concentration de PM_{2,5} dans l'air ambiant, la concentration de mercure dans les sols, les émissions atmosphériques de dioxyde de soufre, les émissions atmosphériques de composés organiques volatils non méthaniques, les émissions atmosphériques d'arsenic, la proportion de la surface communale artificialisée entre 2009 et 2019. Outre ces caractéristiques distinctives, les communes regroupées dans cette classe correspondent très clairement aux principaux centres urbains de la zone d'étude : Lille, Dunkerque, Calais, Amiens, Rouen, Le Havre, Maubeuge et Valenciennes. Enfin, ces communes comptent 2 390 028 habitants (INSEE, 2018), ce qui représente 40,5 % de la population totale de la zone d'étude.

Communes urbaines périphériques. Ces communes sont également fortement soumises à la pression urbaine. Comme pour les « Centres urbains denses », les indicateurs suivants sont significativement plus élevés dans ce profil : la concentration de PM_{2,5} dans l'air ambiant, les concentrations de mercure dans les sols, les émissions atmosphériques de dioxyde de soufre, les émissions atmosphériques de composés organiques volatils non méthaniques, les émissions atmosphériques d'arsenic, la proportion de la surface communale artificialisée entre 2009 et 2019. Par ailleurs, les indicateurs suivants ont la deuxième médiane la plus élevée après « Centres urbains denses » : la proportion de surface communale occupée par des territoires artificialisés (20% [11;34]), les émissions de gaz à effet de serre, la densité de sites et sols pollués, la densité d'installations classées pour la protection de l'environnement, la densité d'accidents industriels. Par ailleurs, près d'un quart (23,7 %) de la population de la zone d'étude se concentre dans ces communes, ce qui représente 1 400 648 habitants (INSEE, 2018). La grande majorité de ces communes sont situées dans les départements du Nord (245 communes) et du Pas-de-Calais (102 communes) qui sont les départements les plus densément peuplés de la zone d'étude. De plus, ces communes sont réparties en périphérie des grands centres urbains : Lille, Dunkerque, Calais, Maubeuge, Valenciennes.

Il est également intéressant de noter qu'un meilleur accès aux espaces verts est constaté dans les « Centres urbains denses » par rapport aux « Communes urbaines périphériques ». Seulement 39% [15;64] de la population a accès à un espace vert dans un rayon de 400 mètres dans les « Communes urbaines périphériques », contre 56% [34;75] dans les « Centres urbains denses ».

Communes agricoles intensives sous influence urbaine. Cette classe est constituée de communes rurales à l'agriculture intensive et spécialisée. Elles sont principalement situées autour d'Arras, Cambrai, Saint-Omer, mais aussi entre Lille et Dunkerque. Les territoires artificialisés ne représentent que 6% [3;8] de la superficie communale. L'occupation des sols est fortement marquée par les grandes cultures. Plus précisément, les cultures suivantes occupent une part nettement plus importante de la surface communale par rapport à tous les autres profils : légumes et fleurs (13% [8;20]), et cultures industrielles (9,4% [6,8;12,0]). Les cultures céréalières couvrent également 47% [42;53] de la superficie de la commune. En ce qui concerne les pratiques agricoles, l'indice de pression phytosanitaire global est significativement le plus élevé parmi toutes les classes, tandis que l'indicateur de haute valeur naturelle (HVN) est le plus faible. Par rapport à la zone d'étude, les activités d'élevage sont peu

développées. De plus, seulement 3% [1;7] de la surface communale est couverte par des prairies et 0% [0;4] par des réservoirs de biodiversité.

En parallèle, ces communes sont également impactées par la pression urbaine. Tout d'abord, la concentration en PM_{2,5} dans l'air ambiant est la troisième plus élevée après les « Centres urbains denses » et les « Communes urbaines périphériques ». De plus, par rapport à l'ensemble de la zone d'étude, les concentrations en métaux dans les sols sont plus élevées, notamment pour l'arsenic.

Communes agricoles intensives hors influence urbaine. Dans ce groupe de communes, les territoires artificialisés ne couvrent que 4 % [0;6] de la superficie de la commune. La majorité de la surface est consacrée à l'agriculture. En particulier, les cultures suivantes couvrent une proportion significativement plus élevée de la surface communale par rapport à tous les autres profils : les cultures oléagineuses (7,5% [5,6;9,7]), et les cultures protéagineuses (2,72% [1,55;4,11]). Les céréales couvrent également 49% [42;54] de la surface de la commune. Comme dans les « Communes agricoles intensives sous influence urbaine », l'indice de pression relatif aux insecticides est significativement plus élevé que dans les autres classes. Ceci pourrait être attribué à des cultures spécifiques comme le colza (oléagineux) ou le pois protéagineux (protéagineux), dont les indices de fréquence de traitement aux insecticides sont parmi les plus élevés en France (Agreste, 2020b). L'indice de pression phytosanitaire global est également très élevé par rapport à la zone d'étude. Enfin, seulement 5 % [2;8] de la surface communale est couverte par des prairies et 3 % [0;15] par des réservoirs de biodiversité.

Contrairement aux « Communes agricoles intensives sous influence urbaine », ces communes sont beaucoup moins touchées par la pression urbaine. Tout d'abord, 376 des 453 communes sont situées dans le département de la Somme, qui a la plus faible densité de population de la zone d'étude et qui est le plus en retrait des grands centres urbains. De plus, ces communes sont significativement les plus éloignées des stations d'épuration et des installations de traitement des déchets. La dynamique d'artificialisation des sols est significativement la plus faible de tous les profils avec seulement 0,15% [0,06;0,32] de la surface communale artificialisée entre 2009 et 2019. Enfin, par rapport aux « Communes agricoles intensive sous influence urbaine », ce profil présente un meilleur accès aux espaces verts et une plus forte naturalité. Cela témoigne d'une plus grande alternance entre parcelles agricoles et petites zones forestières.

Communes où l'agriculture est plus extensive et diversifiée. Ce profil regroupe le plus grand nombre de communes (1 008 communes sur les 3 041 de la zone d'étude). Logiquement, les indicateurs présentent des valeurs moins atypiques par rapport à l'ensemble de la zone d'étude. C'est le seul profil où aucun indicateur ne présente un z-score médian supérieur à 1 ou inférieur à -1. Cependant, des indicateurs spécifiques caractérisent fortement ce profil. Tout d'abord, la concentration en PM_{2,5} dans l'air ambiant est significativement la plus faible parmi tous les profils. Deuxièmement, trois marqueurs des activités d'élevage présentent des valeurs élevées : le nombre d'unités de gros bétail par hectare de surface communale, la proportion de la superficie communale couverte par des prairies, et les émissions atmosphériques d'ammoniac. Les territoires artificialisés ne représentent que 4 % [0;7] de la surface communale et l'utilisation des sols est principalement agricole avec 43 % [36;49] de la surface occupée par des céréales et 17 % [12;22] par des prairies. Contrairement aux deux classes précédentes marquées par des pratiques agricoles très intensives, l'indice de pression phytosanitaire global est plus modéré. Ce profil correspond à des communes rurales, avec une agriculture plus extensive et diversifiée, en particulier avec des activités d'élevage importantes.

Communes où les activités d'élevage sont prédominantes et les espaces naturels importants. Par rapport à la zone d'étude, ces communes abritent une proportion significativement plus élevée de réservoirs de biodiversité (35% [17;60]) et présentent un niveau de naturalité significativement plus élevé. Par ailleurs, la prédominance des activités d'élevage est clairement mise en évidence avec les indicateurs suivants qui prennent des valeurs significativement plus élevées : proportion de la surface communale occupée par des prairies (31% [22;42]), nombre d'unités de gros bétail par hectare de surface communale, et émissions atmosphériques d'ammoniac. En revanche, ce profil présente des valeurs significativement inférieures pour l'indice de pression relatif aux insecticides et pour la densité d'installations classées pour la protection de l'environnement. Néanmoins, par rapport aux autres profils, les sols sont plus fortement concentrés en métaux avec des valeurs significativement plus élevées observées pour le thallium, l'arsenic et le mercure.

Communes où les espaces naturels sont prédominants : forêts, zones humides et surfaces en eau. Comme dans la classe précédente, ces communes présentent un indice de naturalité significativement plus élevé et abritent une plus grande proportion de réservoirs de biodiversité (36 % [26;52]). Cependant, elles diffèrent du profil précédent à plusieurs égards. Tout d'abord, la teneur en métaux dans les sols est nettement plus faible que dans les autres profils, en particulier pour l'arsenic, le mercure et le zinc. Deuxièmement, les activités d'élevage sont beaucoup moins importantes. Une autre spécificité est la présence significative de surfaces en eau. Dans cette classe, 148 communes sur 190 abritent au moins une surface en eau et/ou une zone humide. Par ailleurs, un grand nombre de ces communes sont situées soit le long de la côte, soit en bordure de la Seine ou de la Somme. Enfin, par rapport aux autres profils, les surfaces agricoles ne représentent qu'une faible proportion de la surface communale. Enfin, l'indice de pression phytosanitaire global est faible, et des valeurs très élevées sont observées pour l'indicateur haute valeur naturelle (HVN).

Discussion

Ce travail de recherche visait à développer un cadre robuste et transposable pour distinguer et décrire les profils environnementaux des territoires et ainsi générer une typologie territoriale. Une méthodologie multivariée utilisant des données ouvertes et des méthodes de classification non supervisée a été développée et appliquée à l'échelle des communes sur le périmètre du registre Epimad. Néanmoins, l'analyse pourrait être étendue au territoire métropolitain, les bases de données utilisées ont été choisies pour cela. Sur la base de 39 indicateurs spatiaux, sept profils environnementaux ont été identifiés dans la zone d'étude. Les indicateurs d'entrée retenus pour l'analyse décrivent les niveaux de contamination des milieux (air, eau et sol), le niveau des émissions polluantes, la localisation des sources d'émission, l'utilisation des sols, l'intensité des pratiques agricoles et le degré de naturalité des territoires.

1. Une description multi-dimensionnelle des territoires

Les méthodes statistiques mobilisées dans ce chapitre sont reconnues et très largement utilisées dans l'analyse de données spatiales. Parmi les méthodes de classification existantes, la CAH et la partition des K-means sont les deux alternatives principales. Dans ce travail, les deux méthodes ont d'ailleurs été combinées. En amont, une ACP a également été appliquée pour réduire le nombre de dimensions et optimiser ainsi le résultat de la classification. Les composantes principales retenues pour la CAH expliquaient 62% de la variance totale contenue dans le jeu de données. En comparaison, Pecher et al. (2013) avaient pu expliquer 71% de la variance totale pour définir des profils paysagers dans les Alpes. Mais leur étude visait à définir des profils de paysage et les indicateurs d'entrée étaient donc moins nombreux et moins diversifiés.

A la suite des défis associés à la collecte et à l'utilisation des données ouvertes, le jeu de données utilisé fournit une description assez complète de l'environnement physique. Il comporte des informations sur la contamination des milieux, les émissions polluantes, la proximité aux sources d'émission, l'occupation des sols, les pratiques agricoles et le degré de naturalité. Grâce à cette exhaustivité, une classification qui va bien au-delà de la dichotomie urbain/rural a pu être décrite. Le plus remarquable est probablement d'avoir réussi à distinguer cinq profils de communes rurales. Nous avons pu mettre en évidence des particularités environnementales qui ne sont pas directement liées à la densité de population, mais à la présence de sources d'émissions particulières, de pratiques agricoles, ou d'aménités environnementales. Cette approche semble donc particulièrement pertinente dans le contexte de l'évaluation de l'effet combiné de différents facteurs environnementaux.

Finalement, les méthodes de classification utilisées permettent d'ordonner un jeu de données complexe et redondant pour obtenir une représentation issue des données efficaces, au sens de l'information qu'elles apportent.

Même si chaque commune possède ses propres caractéristiques, cette méthode permet d'extraire les tendances principales et les facteurs clés qui distinguent les unités spatiales les unes des autres, en faisant ainsi un outil efficace pour décrire les inégalités spatiales en matière d'environnement. Les méthodes de classification mises en œuvre dans ce cadre fournissent une vue globale et multidimensionnelle de l'environnement physique et extérieur, permettant l'analyse simultanée des différentes caractéristiques territoriales. Il devient alors possible d'identifier la combinaison de facteurs environnementaux qui façonnent un territoire.

2. Perspectives de recherche en santé-environnement

Pour parvenir à mieux comprendre les inégalités territoriales de santé, l'importance d'adopter des approches multidimensionnelles pour décrire l'environnement est régulièrement soulignée (Brousliche et al., 2023; Genin et al., 2022b; Greco et al., 2020; Lanier et al., 2019b; Li, 2022). La création de typologies territoriales peut s'avérer utile pour traiter certaines questions en environnement-santé. Par exemple, ce type d'approche pourrait contribuer à renseigner l'étiologie de maladies multifactorielles dont les facteurs de risque environnementaux restent encore indéterminés, comme pour la maladie de Crohn par exemple. Analyser si la sur-incidence d'une maladie est associée à un profil territorial spécifique serait une première étape dans ce processus. De plus, cette méthode présente l'avantage d'être relativement peu coûteuse, étant déployée à l'échelle des territoires et

reposant sur des données ouvertes. Par conséquent, elle pourrait permettre d'orienter les recherches et de mieux cibler les facteurs de risques environnementaux qui méritent d'être examinées plus attentivement, à l'échelle individuelle notamment.

Par ailleurs, la typologie environnementale développée dans ce chapitre s'appuie sur une vision intégrée de l'environnement physique, où les indicateurs d'entrée n'ont pas été sélectionnés en fonction d'un événement de santé en particulier. Si cette approche est utile pour étudier les maladies dont les facteurs de risque environnementaux sont inconnus, une sélection plus ciblée des données peut être opérée pour étudier des maladies dont l'étiologie est mieux renseignée. Par exemple, Brousmiche et al. (2023) ont construit une typologie environnementale au niveau des quartiers (IRIS) sur le périmètre de la métropole lilloise. Les données d'entrée ont été sélectionnées sur la base des facteurs de risque potentiels associés aux maladies coronariennes. Les auteurs ont identifié trois profils environnementaux qui différaient principalement sur le plan de la défaveur sociale, de la pollution de l'air et de l'accessibilité aux espaces verts. Après analyse, deux des trois profils environnementaux présentaient un taux d'incidence standardisé anormalement élevé pour les maladies coronariennes.

En dehors du champ de la santé, ces typologies peuvent également être utilisées pour étudier d'autres problématiques. Par exemple, Mullin et al. (2018) ont mené une étude au Royaume-Uni à l'échelle des quartiers, l'objectif était d'étudier l'inégale répartition des services écosystémiques et du capital naturel. À l'aide de 14 indicateurs spatiaux, ils ont élaboré une typologie appelée « Natural Capital Classification », qui distingue six profils de quartiers. Après avoir relié ces profils au niveau de désavantage social, des différences significatives ont été constatées. Ces résultats mettent ainsi en évidence un accès inégal au capital naturel en fonction du niveau de désavantage social des populations. Aux États-Unis, Tran et al. (2010) ont développé une typologie à l'échelle du bassin versant pour évaluer le profil de vulnérabilité des cours d'eau. Le jeu de données utilisé est particulièrement riche, avec 50 indicateurs spatiaux décrivant l'occupation des sols, l'usage de produits phytosanitaires, la contamination des eaux de surface, ou la distance à certaines sources d'émission. Leur objectif était d'identifier les leviers potentiels permettant d'améliorer la qualité de l'eau au niveau des bassins versants. Dans de futures études, ces méthodes pourraient par exemple être déployées en ciblant uniquement les données de contamination, avec pour objectif de distinguer des profils d'exposition spécifiques.

Conclusion

Sur le périmètre du registre Epimad, une typologie environnementale a été développée à l'échelle des communes, en se focalisant sur l'environnement physique. A partir de 39 indicateurs spatiaux soigneusement sélectionnés, sept profils environnementaux ont été identifiés et décrits. La classification obtenue permet d'aller plus loin que la simple description du continuum urbain-rural. Elle fournit une perspective plus nuancée sur la manière dont les pressions et aménités environnementales se répartissent dans l'espace et se superposent les unes aux autres.

Désormais, l'enjeu consiste à évaluer l'intérêt de cette approche pour étudier des maladies dont l'étiologie et les facteurs de risque environnementaux sont en grande partie inconnus. Pour ce faire, la typologie environnementale va être mise en lien avec la répartition d'incidence de la MC et de l'IRCT.

CHAPITRE V – L’apport des profils environnementaux des territoires pour expliquer les variabilités d’incidence des maladies chroniques.

Dans le chapitre IV, une typologie territoriale a été développée et sept profils environnementaux ont été décrits et analysés à l’échelle des communes. Les communes du groupe "Centres urbains denses" représentent les principales agglomérations de la zone d’étude et font face à des pressions environnementales d’origine urbaine et/ou industrielle. Les "Communes urbaines périphériques" présentent un profil similaire, mais avec des niveaux de contamination moindres. Les cinq autres profils correspondent à des communes rurales, se distinguant essentiellement par l’intensité des pratiques agricoles, le degré de naturalité et les pressions exercées par les zones urbaines avoisinantes (pollution atmosphérique, pollution des sols, proximité aux sources d’émission, etc...). Les deux premiers profils ruraux, à savoir les "Communes agricoles intensives sous influence urbaine" et "Communes agricoles intensives hors influence urbaine", sont fortement influencés par une agriculture intensive et spécialisée, nécessitant une utilisation importante d’intrants chimiques. Ces communes ont un caractère très peu naturel, avec des indicateurs de naturalité qui prennent des valeurs comparables aux « Centres urbains denses ». Cependant, ces deux profils diffèrent par leur degré d’influence urbaine et le type de cultures agricoles. Le premier profil combine à la fois pressions agricoles et urbaines, tandis que le second est principalement influencé par les activités agricoles. De plus, les "Communes agricoles intensives sous influence urbaine" sont fortement marquées par les cultures industrielles (essentiellement betterave) et les cultures de légumes (essentiellement pommes de terre), tandis que les "Communes agricoles intensives hors influence urbaine" sont davantage influencées par les céréales et les cultures destinées à l’alimentation animale. Les "Communes où l’agriculture est plus extensive et diversifiée" sont les plus éloignées des sources de pollution urbaine. Leurs pratiques agricoles sont plus diversifiées et extensives, ce qui se traduit par une utilisation plus modérée des produits phytosanitaires. Ce profil est également fortement influencé par les activités d’élevage, qui constituent des sources d’émissions d’ammoniac. Les deux dernières classes sont moins dispersées dans l’espace. Les « Communes où les activités d’élevage sont prédominantes et les espaces naturels importants » comportent des espaces naturels qui correspondent à deux grands parcs naturels régionaux de la zone d’étude (Caps et Marais d’Opale, Avesnois). Bien que la pollution atmosphérique y soit faible en comparaison aux autres profils, les émissions d’ammoniac liées à l’élevage sont élevées, tout comme la présence de métaux dans les sols. Enfin, les "Communes où les espaces naturels sont prédominants : forêts, zones humides et surfaces en eau" sont caractérisées par des territoires à haute valeur écologique, moins exposés aux pressions urbaines et agricoles.

La question porte maintenant sur l’apport de ces résultats pour renseigner l’étiologie des maladies chroniques. Sa pertinence a déjà été illustrée par plusieurs exemples dans la littérature. Fayet et al. (2020) ont développé la « Geographical Classification for Health Studies » à l’échelle communale sur toute la France métropolitaine. Les données d’entrée décrivent l’environnement physique et social, ainsi que l’accès aux soins. Ils ont identifié cinq profils territoriaux distincts et ont observé des différences significatives au niveau des taux de mortalité standardisés sur l’âge. Une sous-mortalité significative a été observée dans les zones métropolitaines riches et dans les périphéries résidentielles, tandis qu’une sur-mortalité significative a été constatée dans les quartiers à population précaire, les plaines agricoles et industrielles et les marges rurales. Sur la base d’indicateurs décrivant la qualité de l’environnement physique, Shortt et al. (2012) ont élaboré une classification territoriale appelée « Multiple Environmental Deprivation Classification ». La méthode a été appliquée en Nouvelle-

Zélande à l'échelle des unités spatiales de recensement (Census Area Unit). Elle a permis de décrire six profils environnementaux distincts. Par la suite, indépendamment du niveau de défaveur sociale, des associations ont été constatées entre certains profils environnementaux et le taux de mortalité pour certaines causes spécifiques (en particulier pour les maladies respiratoires et le cancer du poumon).

La typologie environnementale des communes permet-elle d'identifier de nouvelles pistes étiologiques pour la MC et l'IRCT ? Dans ce chapitre, les profils environnementaux des communes ont donc été mis en lien avec les données d'incidence de la MC et de l'IRCT. L'objectif est de déterminer si certains profils environnementaux présentent des niveaux d'incidence plus élevés ou plus faibles pour l'une ou l'autre de ces maladies.

Association entre les profils environnementaux et l'incidence de la MC

1. Matériel et méthodes

Pour la MC, deux types de régressions écologiques ont été mises en œuvre. D'abord, une régression a été menée pour pouvoir évaluer les associations entre l'incidence de la MC et les sept profils environnementaux. Pour ce faire, une des modalités (en l'occurrence un des profils) a été sélectionnée comme référence. Puis les risques relatifs (RR) associés aux six autres profils ont été estimés comparativement à cette référence. Le profil sélectionné comme référence correspond aux communes où les espaces naturels sont prédominants (en bleu sur les cartographies). Il s'agit du groupe de communes soumis au niveau d'anthropisation le plus faible. L'estimation du RR associé à chaque profil a été réalisée avec un intervalle de confiance de 95%. Si un profil environnemental présente un RR significativement supérieur à 1, cela signifie que ce profil est associé à un risque accru de développer la MC, en comparaison au profil de référence.

Dans un second temps, des régressions écologiques univariées ont été réalisées pour examiner individuellement l'association entre l'incidence de la MC et les 39 indicateurs environnementaux qui ont servi à définir les profils. Ainsi, pour chaque indicateur, une estimation du RR a été réalisée avec un intervalle de confiance de 95%.

Les régressions écologiques mentionnées ci-dessus ont été effectuées à l'aide d'un modèle de Poisson hiérarchique bayésien (Besag et al., 1991c).

2. Résultats

Dans un premier temps, une analyse comparative a été entreprise en juxtaposant la cartographie des profils environnementaux avec celle des SIRs de la MC (Figure 55). En comparant ces deux cartes, la sur-incidence de la MC semble se concentrer principalement au sein des profils de communes où les activités agricoles sont prédominantes. Au contraire, les profils de communes caractérisés par un environnement plus urbain ou naturel semblent davantage se superposer à une sous-incidence de la MC.

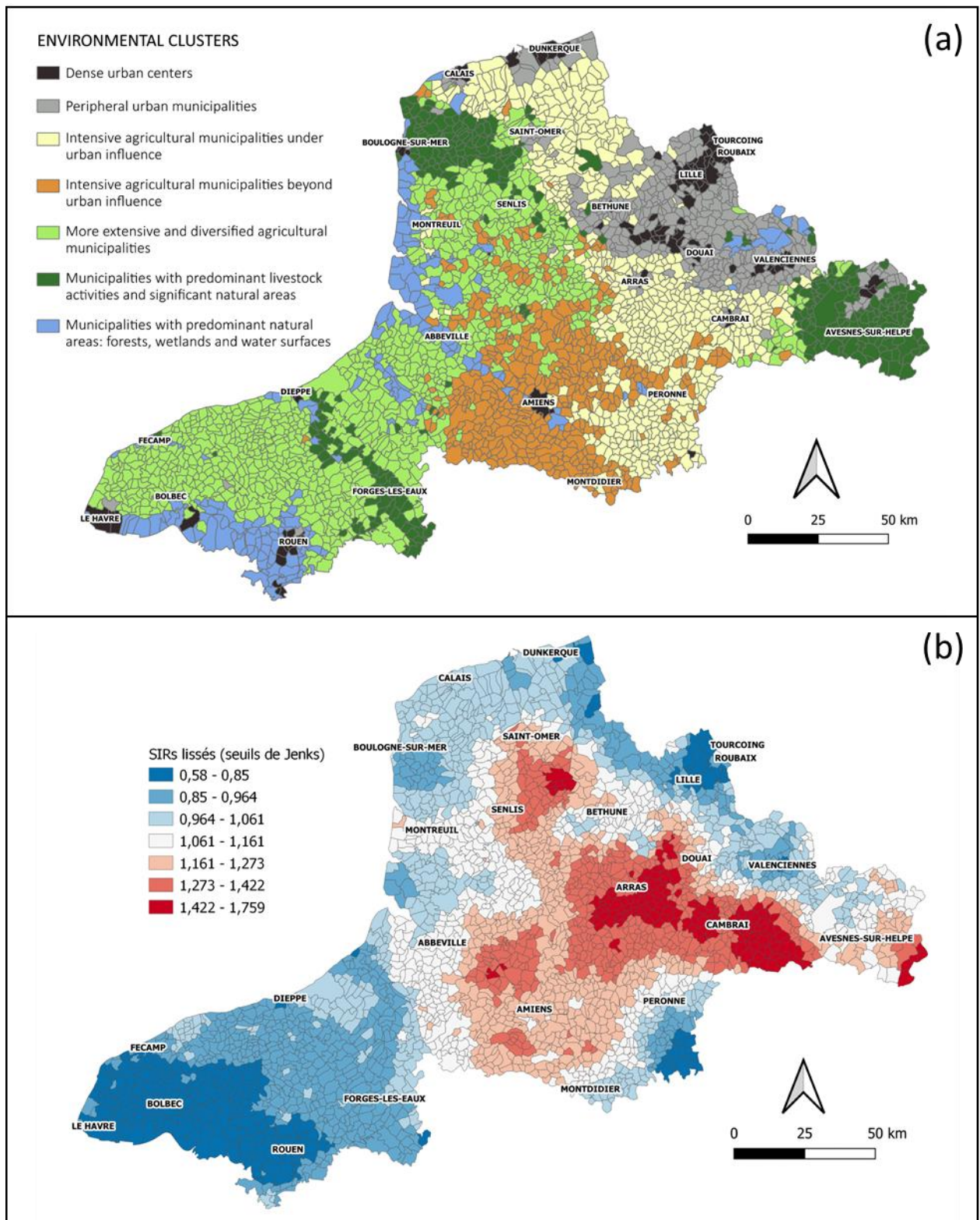
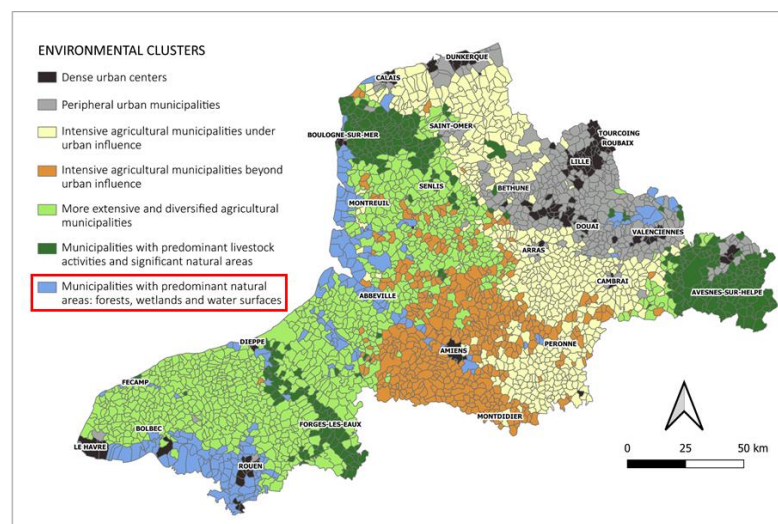


Figure 55 : (a) Représentation cartographique de la typologie environnementale des communes ; (b) Représentation cartographique de la répartition spatiale d'incidence de la MC.

Par la suite, l'incidence de la MC a été comparée au sein des sept profils environnementaux, les résultats issus de la régression écologique sont présentés en figure 56. Pour rappel, le profil environnemental correspondant aux communes où les espaces naturels sont prédominants (en bleu sur la cartographie des profils) a été choisi comme référence. En comparaison, les « communes urbaines périphériques », les « centres urbains denses » et les « communes où les activités d'élevage sont prédominantes et les espaces naturels importants » ne présentent pas de risque significativement différent puisque le RR est très proche de 1 (respectivement 0,999 [0,889 ; 1,122], 0,98 [0,876 ; 1,097] et 0,961 [0,823 ; 1,122]). A l'inverse, les RR sont supérieurs à 1 pour les trois profils de communes rurales où les activités agricoles sont prédominantes. Ces RR sont à la limite de la significativité pour les profils « communes agricoles hors influence urbaine » et « communes agricoles sous influence urbaine » (respectivement 1,152 [0,982 ; 1,353] et 1,12 [0,983 ; 1,275]). Le profil « communes où l'agriculture est plus extensive et diversifiée » présente lui un RR significatif (1,138 [1,014 ; 1,277]). Ainsi, les résultats de la régression écologique confirment les premières observations qui avaient été formulées lors de la comparaison entre les deux cartes. Les trois profils de communes où l'agriculture est prédominante présentent des niveaux d'incidence plus élevés.



Variable

RR

95% CI

Variable	RR	95% CI
More extensive and diversified agricultural municipalities	1.138	[1.014; 1.277]
Intensive agricultural municipalities beyond urban influence	1.152	[0.982; 1.353]
Intensive agricultural municipalities under urban influence	1.12	[0.983; 1.275]
Predominant livestock activities and significant natural areas	0.961	[0.823; 1.122]
Dense urban centers	0.98	[0.876; 1.097]
Peripheral urban municipalities	0.999	[0.889; 1.122]

Figure 56: Résultats issus de la régression écologique comparant les RR associés à chaque profil environnemental.

Pour la régression, un des profils a été pris comme référence. Il s'agit des « communes où les espaces naturels sont prédominants » (profil encadré en rouge dans la légende de la carte).

Enfin, des régressions écologiques univariées ont été menées spécifiquement sur les 39 indicateurs ayant servi à la détermination des profils environnementaux. L'ensemble de ces résultats sont

présentés en annexe (ANNEXE 6). Toutefois, les indicateurs pour lesquels le RR était significativement différent de 1 (supérieur ou inférieur) ont été synthétisés dans le Tableau 13. En supplément, quelques indicateurs ont également été ajoutés lorsque le RR était très proche d'être significativement différent de 1.

Tableau 13 : Principaux résultats issus des régressions écologiques univariées menées sur les 39 indicateurs d'entrée.

Les indicateurs sont représentés en gras si le RR associé est significativement différent de 1.

RR > 1		RR < 1	
INDUSTRIAL_CROPS ¹	1.079 [1.040 ; 1.119]	PM2.5_AIR ¹¹	0.919 [0.871 ; 0.970]
OILSEEDS ²	1.075 [1.034 ; 1.117]	NMVOC_EM ¹²	0.956 [0.924 ; 0.989]
CEREALS ³	1.050 [1.025 ; 1.076]	HUMID_LAND ¹³	0.965 [0.937 ; 0.995]
AS_SOIL ⁴	1.049 [1.005 ; 1.095]	BIODIV_RESERVOIR ¹⁴	0.970 [0.942 ; 0.998]
ZN_SOIL ⁵	1.042 [1.008 ; 1.078]	P_ARTIF ¹⁵	0.978 [0.962 ; 0.993]
TOTPEST ⁶	1.029 [1.001 ; 1.057]	AS_EM ¹⁶	0.978 [0.955 ; 1.002]
PERENNIAL_CROPS ⁷	1.031 [0.996 ; 1.067]	HVN ¹⁷	0.982 [0.956 ; 1.010]
VEGETABLES ⁸	1.022 [0.987 ; 1.058]		
INSECTICIDE ⁹	1.019 [0.994 ; 1.045]		
PH_WATER ¹⁰	1.015 [0.990 ; 1.041]		

¹ INDUSTRIAL_CROPS - Percentage of municipality's surface area occupied by industrial crops (%); ² OILSEEDS - Percentage of municipality's surface area occupied by oil crops (%); ³ CEREALS - Percentage of municipality's surface area occupied by cereals (%); ⁴ AS_SOIL - Arsenic concentration in soils (mg/kg); ⁵ ZN_SOIL - Zinc concentration in soils (mg/kg); ⁶ TOTPEST - Total phytosanitary pressure index (kg/ha/an); ⁷ PERENNIAL - Percentage of municipality's surface area occupied by perennial crops (%); ⁸ VEGETABLES - Percentage of municipality's surface area occupied by vegetables and flowers (%); ⁹ INSECTICIDE - Insecticide pressure index (kg/ha/an); ¹⁰ PH_WATER - Mean pH in drinking water; ¹¹ PM2.5_AIR - Mean concentration of PM2.5 in ambient air ($\mu\text{g}/\text{m}^3$); ¹² NMVOC_EM - Non-methane volatile organic compound atmospheric emissions (kg/ha/an); ¹³ HUMID_LAND - Percentage of municipality's surface area occupied by wetlands (%); ¹⁴ BIODIV_RESERVOIRS - Percentage of municipality's surface area occupied by biodiversity reservoirs (%); ¹⁵ ARTIF_LAND - Percentage of municipality's surface area occupied by artificial lands (%); ¹⁶ AS_EM - Arsenic atmospheric emissions (kg/ha/an); ¹⁷ HNV - High Natural Value indicator.

Plusieurs marqueurs d'une agriculture intensive sont associés à un excès de risque significatif pour la MC. Trois de ces indicateurs décrivent l'occupation des sols agricoles et correspondent au pourcentage de surface communale occupée par un groupe de culture en particulier : cultures industrielles (INDUSTRIAL_CROPS), oléagineux (OILSEEDS), céréales (CEREALS). En parallèle, l'indice de pression phytosanitaire global (TOTPEST) est également associé à l'incidence de MC. Cet indicateur prend en compte l'ensemble des substances actives et s'exprime en kg de substance active achetée par hectare de surface communale et par an. En outre, d'autres indicateurs présentent un excès de risque qui n'est cependant pas significatif. Il s'agit à nouveau d'indicateurs d'occupation des sols agricoles (PERENNIAL_CROPS : cultures pérennes, et VEGETABLES : légumes & fleurs) et d'un indice de pression phytosanitaire calculé spécifiquement pour les insecticides (INSECTICID). Enfin, le niveau de

contamination des sols en arsenic (AS_SOIL) et zinc (ZN_SOIL) est également associé à un excès de RR pour la MC.

A l'inverse, un risque plus faible et significatif est constaté pour des marqueurs de contamination plutôt urbains ou industriels : PM2.5, NMVOC_EM. Les émissions atmosphériques d'arsenic, marqueur industriel, montrent également un risque plus faible, mais non-significatif (AS_EM), tout comme la part de surface artificialisée (P_ARTIF), marqueur du milieu urbain.

Enfin, un risque significativement plus faible est observé pour des indicateurs qui témoignent d'une naturalité importante des territoires : part de la surface communale couverte par des zones humides (HUMID_LAND), part de la surface communale couverte par des réservoirs de biodiversité (BIODIV_RESERVOIR). Cette même tendance mais non significative est observée pour le score HVN qui traduit des pratiques agricoles à faible pression environnementale.

3. Discussion

L'analyse a mis en évidence trois profils environnementaux qui présentent un excès de risque pour la MC, en considérant le profil le moins anthropisé du territoire. Ces trois profils sont différents les uns des autres mais ils correspondent globalement à des communes rurales où les activités agricoles sont prédominantes. Cet excès de risque est significatif pour le profil correspondant à des activités agricoles extensives et diversifiées. De plus, un excès de risque significatif a également été constaté pour plusieurs marqueurs agricoles pris individuellement. Ces marqueurs décrivent essentiellement la présence de grandes cultures et la pression phytosanitaire qui s'exerce sur ces territoires. A notre connaissance, ce résultat est assez inédit dans la littérature existante sur la MC.

Si l'on interprète plus généralement la balance de l'incidence de MC entre urbain et rural, une tendance urbaine ressort de la plupart des études. Une méta-analyse avait été conduite sur le sujet en 2012 (Soon et al., 2012). Sur la base de 9 études cas-témoins et 21 études de cohorte, une association positive avait été constatée entre milieu urbain et MC (RR : 1.42 ; IC 95% [1.26-1.60]). Une revue de la littérature plus récente a étudié la question spécifiquement dans les pays en voie de développement (Rajbhandari et al., 2020). Cette tendance urbaine a été confirmée mais des résultats contradictoires ont été observés dans certains pays comme le Sri Lanka. Dans l'état du Victoria en Australie, une étude rétrospective sur 31 ans a examiné l'incidence de la maladie de Crohn chez les enfants de moins de 16 ans (Phavichitr et al., 2003). Au-delà de mettre en évidence une augmentation importante de l'incidence, cette étude rapporte une sur-représentation significative d'enfants vivant en milieu urbain par rapport au milieu rural (RR : 1.66 ; IC 95% [1.28-2.16]). Une étude de cohorte menée dans quatre provinces du Canada visait à étudier l'association entre les MICI et le statut rural/urbain au cours des cinq premières années de vie (Benchimol et al., 2017). Les résultats ont montré un effet protecteur de l'environnement urbain sur la survenue des MICI (RR : 0.76 ; IC 95% [0.51-1.00]). Concentrons-nous désormais sur les études menées en France. Nerich et al. (2010) ont utilisé les données de l'assurance maladie pour évaluer l'incidence de la MC à l'échelle des zones d'emploi (découpage Insee), sur tout le périmètre national entre 2000 et 2002. Les résultats ont montré une association inverse entre la proportion de travailleurs agricoles et la MC. La proportion de travailleurs agricoles a été découpée en terciles, le premier tercile (faible proportion) a été comparé au troisième tercile (forte proportion). Ils ont finalement obtenu un RR de 1.18 (IC 95% [1.26-1.72]). En conséquence, les auteurs ont suggéré que le degré d'urbanisation pouvait constituer un facteur de risque potentiel de la MC. A l'échelle de

la France métropolitaine, Genin et al. (2020) ont cartographié la prévalence de la MC en utilisant les données d'hospitalisation sur la période 2007-2014. Une association positive a été observée entre la prévalence de la MC et un fort degré d'urbanisation, modélisé par la proportion de surfaces artificialisées (1.09 ; IC 95% [1.04-1.14]).

Cependant, malgré cette tendance urbaine constatée dans la littérature, un regard critique doit être porté sur les études au regard des données de santé utilisées. Par exemple, dans les deux études menées en France, les données de santé utilisées ne sont pas issues d'un registre de santé comme c'est le cas pour le registre Epimad. Nerich et al. (2010) ont évalué l'incidence de la MC à partir des données de l'assurance maladie, Genin et al. (2020) ont calculé la prévalence de la MC à partir des données d'hospitalisation, données tout de même validées par celles du registre Epimad. Par ailleurs, Deepak et al. (2023) abordent la question des disparités spatiales d'accès aux soins pour les patients atteints de MICI. Ils évoquent un accès plus limité aux spécialistes et des retards de diagnostic plus fréquents en milieu rural. Par extension, cette inégale répartition de l'accès aux soins pourrait potentiellement induire une sous-évaluation des cas de MC en milieu rural.

Pour finir, des résultats contradictoires émergent également dans la littérature. En Inde, une étude récente n'a pas constaté de différence urbain/rural, alors même que des travaux antérieurs indiquaient le contraire quelques années auparavant (Banerjee et al., 2023). En France, avec les données du registre Epimad, l'incidence de la MC avait été précédemment étudiée à l'échelle des cantons sur la période 1990-2003 (Declercq et al., 2010b). A cette occasion, les auteurs avaient constaté une sur-incidence de la MC concentrée essentiellement en milieu rural et péri-urbain. Van Kruiningen et al. (2005) ont mené une étude populationnelle en Belgique pour évaluer les déterminants environnementaux et familiaux liés à la MC. Vingt-et-une familles comptant au moins trois parents au premier degré atteints de la MC ont été inclus dans l'étude, ainsi que dix familles témoins. Au total, l'étude comptabilise 74 patients atteints de la MC, 84 membres de la famille non atteints et 59 témoins. Bien que ce passage de l'article relève davantage du témoignage que d'un élément tangible, il m'a semblé intéressant de le traduire et de le citer ici : « *Étant donné que la littérature rapporte une fréquence plus élevée de la MC dans les zones urbaines, il était inattendu qu'à une ou deux exceptions près, les maisons que nous avons visitées soient entourées de terres agricoles intensivement labourées. Celles-ci s'étendaient presque jusqu'aux murs extérieurs, du moins jusqu'aux limites très proches des petites parcelles de terrain sur lesquelles étaient situées les maisons* ».

Enfin, une étude publiée dans Nature a conçu une méthodologie très novatrice pour tenter d'identifier les facteurs environnementaux susceptibles d'induire une inflammation intestinale (Sanmarco et al., 2022). Une approche intégrée a été développée, elle combine des bases de données publiques, des essais de criblage chimique sur poissons-zèbres et du machine learning. Plus concrètement, des méthodes d'apprentissage automatique ont été utilisées pour modéliser l'activité biologique des produits chimiques et identifier des candidats susceptibles de moduler l'inflammation intestinale. Le modèle a été construit en utilisant un ensemble d'apprentissage comprenant des produits chimiques ayant des effets connus sur l'inflammation intestinale ainsi que des produits chimiques témoins. Les résultats ont ensuite été validés en pratiquant des bioessais sur poissons-zèbres. Ces tests ont permis de vérifier que les substances chimiques identifiées par le modèle stimulaient effectivement l'inflammation intestinale chez le poisson-zèbre. Sur le top 20 des candidats identifiés, 11 sont des substances chimiques utilisées en agriculture.

A l'inverse, parmi les résultats obtenus dans ce travail de thèse, un RR plus faible a été observé pour la pollution atmosphérique en $PM_{2.5}$ et pour les émissions d'arsenic et de carbone organique volatil non méthanique. La même tendance a également été constatée pour l'indicateur P_ARTIF qui décrit la proportion de surfaces artificialisées. Pour $PM_{2.5}$, un résultat similaire est rapporté dans une étude de cohorte menée dans plusieurs pays européens : Danemark, France, Pays-Bas et Royaume-Uni (Opstelten et al., 2016). La cohorte compte 227 620 participants pour 142 cas incidents de MICI. Les résultats ont mis en évidence une association inverse entre pollution atmosphérique en $PM_{2.5}$ et MICI (OR : 0.24 ; IC 95% [0.07-0.81] pour une augmentation de $5 \mu\text{g}/\text{m}^3$). Un résultat similaire a également été rapporté aux Etats-Unis (Lochhead et al., 2018). Malgré des résultats comparables obtenus dans ces études, il est difficilement envisageable d'attribuer un rôle bénéfique à l'exposition aux $PM_{2.5}$. En revanche, l'indicateur de contamination aux $PM_{2.5}$ peut être considéré comme un marqueur urbain sous faible influence agricole.

Enfin, dans ce chapitre, un dernier résultat concerne le RR significatif plus faible constaté pour la part de surface communale couverte par des zones humides, et la part de la surface communale couverte par des réservoirs de biodiversité. D'abord, de manière générale, sans lien spécifique avec la MC, plusieurs publications suggèrent un effet bénéfique de la biodiversité sur la régulation du système immunitaire (Aerts et al., 2018; Hanski et al., 2012; Rook, 2013b). Une étude menée au Brésil a permis de cartographier la prévalence des MICI (da Luz Moreira et al., 2022). Sur cette base, des clusters régionaux ont été identifiés et des régressions écologiques ont été menées. Dans certaines zones, la perte de biodiversité était associée à une forte prévalence des MICI. Elten et al. (2021) ont mené une étude de cohorte au Canada pour évaluer l'association entre les espaces verts et la survenue pédiatrique des MICI. La population d'étude comprenait 2 715 318 couples mère-enfant avec des naissances s'échelonnant de 1991 à 2014. Par ailleurs, 3 444 diagnostics de MICI ont été posés pendant la période de suivi. L'exposition aux espaces verts pendant la période de l'enfance a été mesurée à l'aide de l'indice de végétation par différence normalisée (NDVI). Elle était associée à un risque réduit de développer une MICI pédiatrique (OR ajusté par augmentation de la plage interquartile du NDVI : 0,77 ; IC 95% [0,74–0,81]). Une analyse longitudinale menée au Royaume-Uni a également obtenu des résultats allant dans ce sens (Zhang et al., 2022). Au total, 216 868 participants sans MICI au départ ont été inclus dans l'étude, avec un suivi moyen de 11.7 ans. Au cours du suivi, 1 271 cas incidents de MICI ont été identifiés. Après analyse, l'exposition aux espaces verts, aux espaces bleus (surfaces en eau) et aux espaces naturels était associée à un risque réduit de survenue des MICI. L'exposition aux espaces verts a été évaluée en calculant le pourcentage d'espaces verts dans une zone tampon de 300 mètres autour des lieux de résidence des participants. Le même procédé a été utilisé pour les espaces bleus et les espaces naturels. Ensuite, le risque d'incidence des MICI a été comparé entre le troisième et le premier tercile d'exposition, les résultats obtenus pour espaces verts étaient un Hazard Ratio (HR) de 0.846 (IC 95% [0.736-0.972]), pour les espaces bleus un HR de 0.785 (IC 95% [0.685-0.899]), et pour les espaces naturels un HR de 0.803 (IC 95% [0.697-0.926]). A noter que ces résultats ont été ajustés sur un certain nombre de facteurs confondants identifiés dans la littérature : âge, sexe, origine ethnique, niveau d'éducation, situation professionnelle, niveau de revenu, indice de masse corporelle, tabagisme, consommation d'alcool, activité physique, consommation de fruits et légumes, prise de médicaments anti-inflammatoires.

Association entre les profils environnementaux et l'incidence de l'IRCT

1. Matériel et méthodes

A ce stade d'avancement de la thèse, nous avons mis en relation les mêmes profils environnementaux, définis sur le périmètre du registre Epimad, avec l'incidence de l'IRCT à l'échelle du Nord et Pas-de-Calais. La distribution statistique des SIRs lissés au sein des sept profils de communes a été comparée. Ces comparaisons ont été effectuées en utilisant le test de Kruskal-Wallis, suivi de tests post-hoc de Dunn avec ajustement de Bonferroni. Cette étape est une étape préalable avant la mise en œuvre de méthodes de régressions spatiales.

2. Résultats

Dans un premier temps, la cartographie des profils environnementaux et la cartographie des SIRs de l'IRCT ont été mises en parallèle (Figure 57). Les zones de sur-incidence semblent se concentrer principalement au niveau de certaines agglomérations urbaines de la zone d'étude : Lille, Maubeuge, Dunkerque et dans une moindre mesure Arras. Au contraire, la sous-incidence se répartit majoritairement à l'est du Pas-de-Calais. Dans cette zone, les profils environnementaux sont essentiellement ruraux et agricoles. Enfin, quelques agglomérations urbaines comme Boulogne sur Mer et Calais se distinguent également par un faible niveau d'incidence.

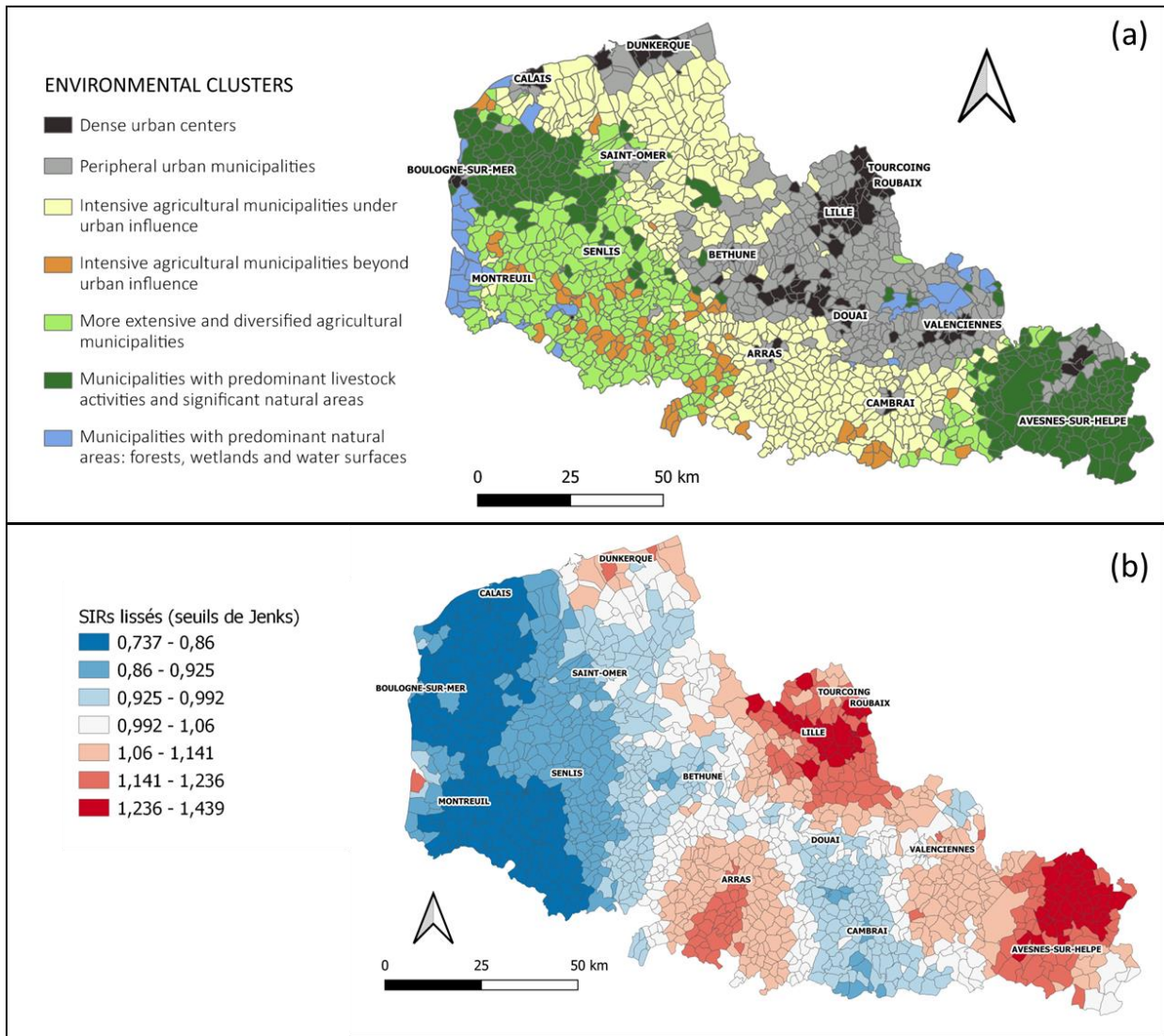
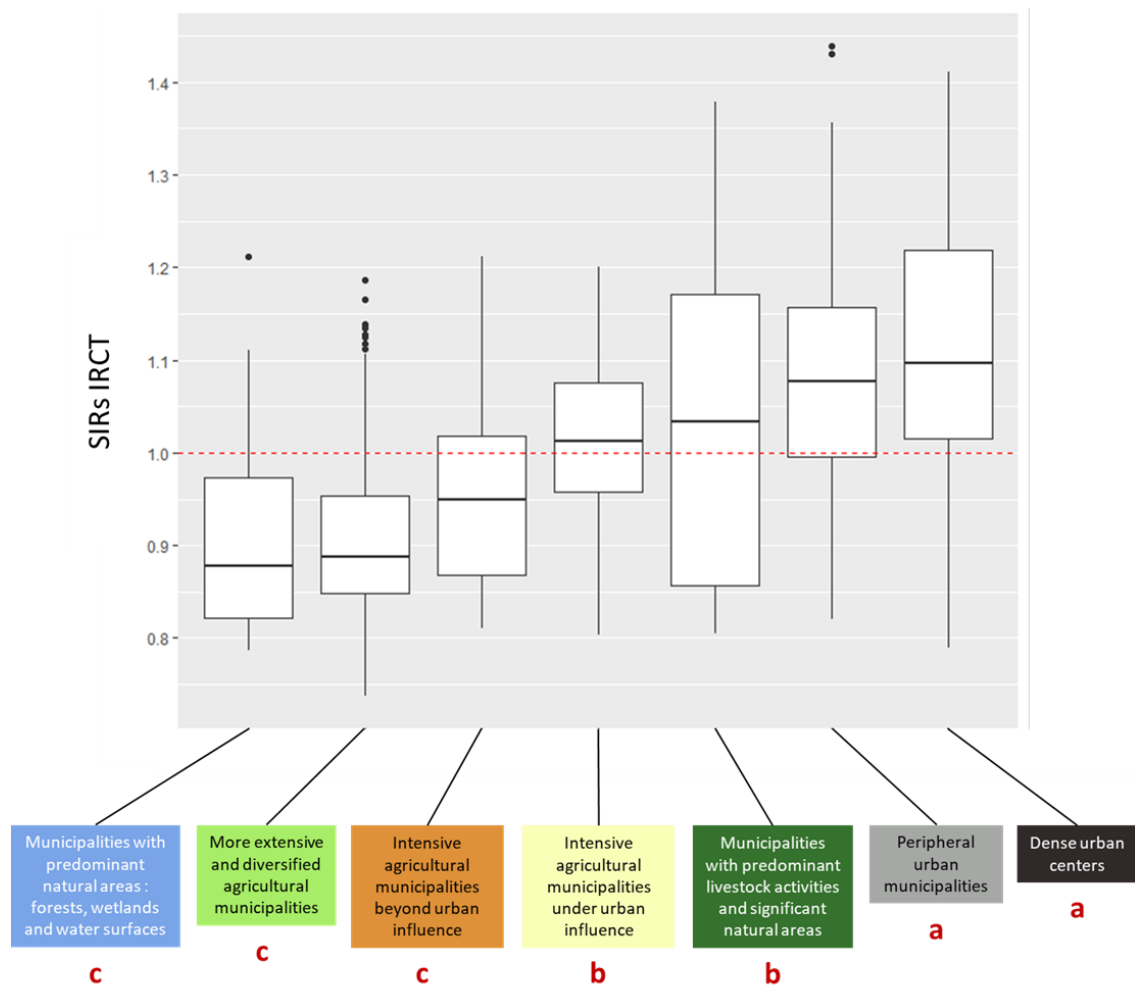


Figure 57 : (a) Représentation cartographique de la typologie environnementale des communes ; (b) Représentation cartographique de la répartition spatiale d'incidence de l'IRCT.

La distribution des SIRs de l'IRCT au sein des sept profils environnementaux a été examinée et représentée en figure 58.



Profils environnementaux	Effectifs	SIRs IRCT (Médiane [EIQ])
Municipalities with predominant natural areas : forests, wetlands and water surfaces	41	0.878 [0.822 ; 0.973]
More extensive and diversified agricultural municipalities	313	0.887 [0.848 ; 0.953]
Intensive agricultural municipalities beyond urban influence	76	0.949 [0.868 ; 1.018]
Intensive agricultural municipalities under urban influence	439	1.013 [0.958 ; 1.075]
Municipalities with predominant livestock activities and significant natural areas	242	1.034 [0.857 ; 1.171]
Peripheral urban municipalities	347	1.077 [0.996 ; 1.157]
Dense urban centers	83	1.096 [1.016 ; 1.219]

Figure 58 : Graphe et Tableau de distribution des SIRs de l'IRCT au sein des sept profils environnementaux identifiés.

Si deux profils partagent une lettre en commun alors ils ne sont pas significativement différents, au contraire si deux profils ne partagent aucune lettre en commun alors ils sont significativement différents (Test de Kruskal-Wallis avec comparaisons multiples de Dunn)

D'abord, il est intéressant de constater qu'une forte hétérogénéité d'incidence est observée entre les différents profils environnementaux. Le point le plus frappant concerne les « Centres urbains denses » et les « Communes urbaines périphériques » qui présentent des ratios d'incidence plus élevés et qui diffèrent significativement de tous les autres profils. Si l'on revient à la description de la typologie environnementale, ces deux profils sont également ceux qui présentent la pollution atmosphérique en $PM_{2.5}$ la plus élevée en comparaison à tous les autres profils (voir tableau 11). Ce même constat peut être fait pour les indicateurs d'émission atmosphérique suivants : arsenic, dioxyde de soufre et carbone organique volatil non méthanique.

A l'inverse, trois profils affichent une incidence médiane plus faible et se distinguent significativement des autres. Ce sont les « Communes où les espaces naturels sont prédominants », les « Communes où l'agriculture est plus extensive et diversifiée » et les « Communes agricoles intensives hors influence urbaine ». Le profil de « Communes où les espaces naturels sont prédominants » est le groupe soumis au niveau d'anthropisation le plus faible, il a d'ailleurs été pris comme référence lors des régressions écologiques menées pour la MC. Les deux autres profils sont également assez préservés des pressions urbaines notamment. Ils présentent la plus faible densité de sites et sols pollués et d'accidents industriels. Les « Communes où l'agriculture est plus extensive et diversifiée » affichent la pollution atmosphérique en $PM_{2.5}$ la plus faible parmi tous les profils. La pollution métallique des sols y est également l'une des plus basses pour le mercure et le zinc. Enfin, les « Communes agricoles hors influence urbaine » présentent les niveaux d'émission les plus faibles pour le dioxyde de soufre et l'arsenic.

3. Discussion

L'incidence de l'IRCT est nettement supérieure dans les « Centres urbains denses » et « Communes urbaines périphériques » qui cumulent une multitude de pressions dont la pollution atmosphérique, mais pas uniquement. Ces communes font également face à une densité particulièrement élevée de sites et sols pollués, d'installations classées pour la protection de l'environnement, d'accidents industriels, et présentent une proximité significativement plus marquée avec les installations de stockage et de traitement des déchets.

Les résultats du chapitre III portant sur la multi-contamination des milieux suggèrent un lien entre IRCT et pollution de l'air. Les résultats obtenus ici sont cohérents et renforcent cette hypothèse. Si nécessaire, la mise en perspective de ces résultats vis-à-vis de la littérature existante a été effectuée dans la discussion du chapitre III. Dans cette discussion, nous avons mis en avant que cette association avec la pollution de l'air semble de plus en plus étayée dans la littérature. L'impact de la pollution atmosphérique (en particulier des particules fines) sur la survenue et sur l'évolution clinique de l'insuffisance rénale semble être une piste de plus en plus crédible.

Il convient de noter que les résultats obtenus pour l'IRCT sont issus d'une simple comparaison de la distribution des SIRs au sein des différents profils environnementaux. Contrairement à la MC pour laquelle des régressions écologiques ont été menées, les analyses effectuées pour l'IRCT ne prennent pas en compte la dépendance spatiale entre unités spatiales. Le travail engagé doit donc être poursuivi pour confirmer ces observations.

Conclusion

La description des profils environnementaux des territoires offre une vision synthétique, mais également complexe, des particularités territoriales. En appréhendant les combinaisons de facteurs qui distinguent ces territoires les uns des autres, cette méthode permet de décrire les principaux profils environnementaux présents dans la zone d'étude. Cette approche revêt un intérêt certain pour caractériser le lien environnement-santé. Elle a été appliquée à deux maladies chroniques en particulier. D'abord, dans les deux cas, une hétérogénéité d'incidence est constatée entre les différents profils environnementaux. Il est intéressant de remarquer que les trois profils qui affichent l'incidence médiane la plus faible pour l'IRCT sont également ceux qui présentent le RR le plus élevé pour la MC. La répartition spatiale de ces deux maladies chroniques est très contrastée. Il est important de souligner que l'objectif n'était pas de démontrer de lien de causalité. Mais l'approche s'avère particulièrement utile pour orienter vers des pistes étiologiques. En l'occurrence, un lien est suggéré entre la MC et les pratiques agricoles, ainsi que la naturalité des territoires. Pour l'IRCT, un lien est suggéré avec les pressions urbaines, en particulier la pollution de l'air.

Discussion générale

Portée des données environnementales ouvertes pour la recherche en santé-environnement.

Pour identifier des pistes environnementales impliqués dans l'étiologie des maladies chroniques, le défi consistait à utiliser des données ouvertes pour caractériser l'environnement au sens du milieu de vie extérieur. Une avancée significative a été réalisée sur ce plan. 24 sources de données ont été identifiées, exploitées et 113 indicateurs spatialisés ont été générés. Le jeu de données élaboré s'étend sur une vaste zone géographique, englobant quatre départements, avec une résolution spatiale fine au niveau des communes. Chaque indicateur est disponible pour l'ensemble des 3 041 communes de la zone d'étude. Le jeu de données ainsi constitué permet de décrire le niveau de contamination des milieux (air, eau, sols) ; le niveau des émissions polluantes ; la localisation des sources d'émissions ; l'occupation des sols ; les pratiques agricoles ; la naturalité des territoires ; le climat.

Les données ouvertes disponibles pour caractériser directement les niveaux de contamination dans les milieux sont relativement peu exhaustives. Elles concernent essentiellement les polluants réglementés dans l'air, les ETM dans les sols et les polluants surveillés en routine dans les eaux de distribution d'eau potable. Se contenter de ces données n'aurait pas permis d'engager une véritable description intégrée de l'environnement physique afin de tester de nombreuses hypothèses étiologiques. Ainsi, pour pouvoir caractériser une plus grande diversité de phénomènes, il a fallu passer par des approches indirectes qui retracent l'occupation des sols, la localisation des sources, l'intensité des activités humaines (agriculture, transport, industries) et des pressions environnementales associées (pression phytosanitaire, artificialisation des sols, émissions polluantes, etc...). De plus, l'environnement est à considérer également comme générateur d'effets potentiellement bénéfiques sur la santé des populations. Une description des aménités environnementales a été engagée en intégrant des indicateurs qui caractérisent la naturalité des territoires (espaces verts, continuité spatiale des milieux naturels, réservoirs de biodiversité, etc...). Enfin, des indicateurs climatiques ont également été générés. Il s'agit d'une avancée certaine par rapport au jeu de données de départ qui ne comportait que 5 indicateurs pour caractériser l'environnement physique. L'ensemble de ces données permet finalement d'appréhender un spectre large de phénomènes environnementaux susceptibles de générer des inégalités de santé. Elles s'avèrent pertinentes pour la recherche en santé-environnement lorsqu'il s'agit de tester différentes hypothèses étiologiques à l'échelle des territoires, en particulier dans le cadre de l'étude de maladies dont l'influence de l'environnement est suspectée mais pour lesquels les facteurs de risque restent indéterminés.

Une discussion approfondie a été menée dans le chapitre 2 pour évaluer la qualité des données collectées et la robustesse des indicateurs produits. De nombreuses pistes d'amélioration ont été évoquées pour optimiser les indicateurs développés. De plus, nous avons discuté des perspectives liées à la mise à disposition de nouvelles données ouvertes dans les années à venir. À présent, il convient d'évaluer la portée de ces données environnementales et les principaux manques qui peuvent être identifiés dans la caractérisation de l'environnement physique.

De mon point de vue, la priorité doit être mise sur les données de contamination. Dans l'air, les données existantes concernent les polluants réglementés, qui sont étroitement corrélés les uns aux autres et émis à partir de sources globalement similaires. Dans les sols, les données de contamination disponibles concernent les ETM. Bien qu'il y ait une plus grande diversité dans les sources d'émission

par rapport aux polluants atmosphériques, le même constat peut tout de même être formulé. Ils sont particulièrement rémanents dans les sols et ont impact sanitaire avéré, donc ce choix des métaux est parfaitement logique. Mais d'autres polluants pourraient être investigués, notamment les pesticides dans les sols. Plusieurs travaux récents démontrent une persistance inattendue des molécules de pesticides dans l'environnement, bien au-delà de leur temps de dégradation théorique et à des concentrations supérieures à celles attendues (Froger et al., 2023b). Certains composés (même théoriquement non persistants) ont été mesurés dans les sols des années ou des décennies après leur application (Riedo et al., 2023). Dans les eaux de distribution, une plus grande diversité de polluants sont surveillés. Mais pour le coup, des difficultés ont été rencontrées dans l'exploitation et la spatialisation des données brutes mises à disposition. Un travail en amont doit être consenti par les producteurs de ces données, pour optimiser leur collecte et leur structuration. Par ailleurs, il serait intéressant d'avoir accès à des données sur d'autres types de polluants, notamment les molécules pharmaceutiques.

En outre, il existe un déséquilibre dans la description des pressions environnementales, avec une tendance à la sur-représentation des pressions urbaines et industrielles dans les données ouvertes disponibles. Pour étudier les pressions agricoles par exemple, nous avons dû effectuer un travail souvent plus poussé et recourir à davantage d'approches indirectes. La prise en compte de la naturalité des territoires est également essentielle. Le même constat peut être formulé puisque la description de cette thématique a été principalement abordée sous le prisme de l'occupation des sols.

En parallèle de l'utilisation de bases de données ouvertes, d'autres pistes pourraient être explorées, notamment la télédétection. Bien que cela requière des compétences spécifiques, cette approche offre des possibilités très intéressantes pour la caractérisation de l'environnement. Par exemple, l'utilisation des méthodes de télédétection dans le suivi de la biodiversité est une tendance (Luque et al., 2018). Elle est par exemple utilisée pour évaluer la continuité spatiale des milieux naturels, la diversité des habitats ou des espèces végétales (Wang and Gamon, 2019).

Au-delà des limites intrinsèques des données environnementales ouvertes, le principal facteur limitant est avant tout leur disponibilité. En réalité, de nombreuses données existent, mais la mise en place d'une véritable dynamique de partage des données prend du temps. Par exemple, des données sur la contamination des sols sont produites en routine dans le cadre de projets de développement urbain. Ces données existent bel et bien mais elles ne passent pas dans le domaine public. Un autre exemple assez frappant concerne la BNVD (registre des achats et ventes de produits phytosanitaires), qui a été utilisée dans cette thèse pour construire des indicateurs de pression phytosanitaire. Un article de Mediapart (Defer et al., 2019) relate que ces données ont été une première fois publiées par le ministère de la transition écologique en juin 2018, avant d'être retirées quelques jours plus tard, puis remises en ligne en juillet 2019. Entre temps, les noms des distributeurs agréés de produits phytosanitaires avaient été retirés du jeu de données. Il n'était donc plus possible de calculer par exemple les quantités annuelles de pesticides vendues par chaque distributeur. Après dans la majorité des cas, pas forcément volontaire. Cependant, cet exemple demeure un cas particulier. En réalité, dans la majorité des cas, les données disponibles sont mises en ligne par les autorités publiques. Mais les bases de données ne sont simplement pas suffisamment pensées pour la réutilisation dans d'autres contextes. Et très souvent, elles ne permettent pas de générer un indicateur spatialisé quantitatif sur l'ensemble du territoire national. Pour relever ces défis, il est nécessaire de faire preuve d'inventivité, comme l'a illustré, par exemple, l'Observatoire national de l'artificialisation des sols en détournant la

base de données des fichiers fonciers (d'origine fiscale), pour évaluer les tendances d'artificialisation des sols.

La construction d'indicateurs pertinents dans une optique de santé-environnement à partir des bases de données environnementales ouvertes constitue un véritable défi. Ces difficultés sont liées, comme nous l'avons constaté, à la disponibilité limitée de ces données, mais également à leur grande hétérogénéité. D'abord bien souvent, ces données sont de nature différentes, proviennent de sources diverses et ont été générées par des communautés indépendantes. Elles n'ont généralement pas été conçues pour être réutilisées, et il y a eu peu d'anticipation à cet égard, bien que la situation évolue. Il est impératif de déployer des efforts pour favoriser une plus grande interopérabilité entre ces bases de données. Pour cela, il est nécessaire de mettre en place des référentiels communs et des métadonnées suffisamment précises pour faciliter leur réutilisation. Les principes FAIR (Findable, Accessible, Interoperable and Reusable) peuvent être utilisés comme point de repère méthodologique. Tout cela requiert la définition et la mise en œuvre de normes communes visant à simplifier l'acquisition, le traitement et la gestion des données. Il est également nécessaire de promouvoir de nouvelles pratiques de partage des données par tous les acteurs, pour le bénéfice de la communauté scientifique et de la société en général.

Dans ce contexte, le jeu de données que nous avons développé offre une perspective assez inédite. Il faut également souligner que ce jeu de données sera enrichi et amélioré au fil du temps. En effet, la disponibilité des données ouvertes est en constante évolution et ce travail d'identification, de collecte et de traitement devra être pérennisé dans le laboratoire à l'issue de la thèse. Des initiatives comme le Green Data for health (GD4H ; <https://gd4h.ecologie.gouv.fr/>) devraient permettre de faciliter le suivi et la centralisation de ces données en créant une plateforme dédiée. Cette initiative s'inscrit dans le cadre du PNSE4 dont l'un des objectifs prioritaires est de faciliter l'accès aux données environnementales à l'échelle des territoires. Ainsi, étant donné les dynamiques qui sont à l'œuvre pour le partage des données environnementales, les possibilités offertes par ce type d'approche devraient s'élargir de plus en plus dans les années à venir.

A l'issue de la thèse, le jeu de données construit sera partagé soit via un data article, soit en le rendant accessible sur une plateforme de dépôt type Zenodo. Cette démarche semble être la suite logique, contribuant ainsi à la transparence de nos recherches et favorisant la réappropriation de ces données par d'autres acteurs.

Approches multi-dimensionnelles et description intégrée de l'environnement physique

Le second volet de la thèse consistait à mobiliser ces indicateurs environnementaux pour mettre en œuvre des approches multivariées et pour permettre de décrire l'environnement sous un prisme multi-dimensionnel. En effet, la caractérisation de l'environnement et son lien à la santé est souvent appréhendée de manière cloisonnée dans les études en population. Pour mieux approcher la complexité du lien environnement-santé, il est crucial d'adopter des approches multi-dimensionnelles. Il paraît en effet plus pertinent de procéder par étapes en commençant par des approches globales avant éventuellement de cibler un ou des facteurs de risque environnementaux spécifiques. Trois approches multi-dimensionnelles distinctes ont pu être appliquées au cours de la thèse.

En premier lieu, des indices de vulnérabilité et de résilience ont été calculés, en s'appuyant sur une méthodologie qui a été développée au cours d'un précédent travail de thèse (Brousmiche, 2022). Cette méthodologie a été reprise ici et déployée sur une plus large zone d'étude pour couvrir celle des registres de santé. Ces deux indices mesurent le cumul de déterminants environnementaux favorables ou défavorables à la santé à l'échelle des territoires. Les données utilisées pour le calcul englobent bien plus que la simple caractérisation de l'environnement physique, elles intègrent six dimensions distinctes : environnement physique, économie, social, politique, santé et services. Initialement, ces indices ont notamment été conçus à destination des gestionnaires de territoires comme un outil de diagnostic territorial. Dans le cadre de cette thèse, ils ont été utilisés pour tester l'influence globale de l'environnement sur la répartition spatiale d'incidence des maladies chroniques.

Les données de contamination de l'air, de l'eau et des sols ont été utilisées pour explorer la piste de la multi-contamination des milieux. Quatre méthodes de calcul identifiées dans la littérature ont été testées et comparées. La méthode du RIM a finalement été retenue pour caractériser spatialement la multi-contamination des milieux. Contrairement aux autres indices, le RIM offre l'avantage non négligeable de permettre des comparaisons dans le temps ou entre différentes zones d'études. La multi-contamination a pour l'instant été traitée individuellement dans chacun des milieux. Certaines études tentent d'aller plus loin en créant des indices de multi-contamination multi-milieux (Golge et al., 2013). Cependant, la question de la hiérarchisation entre milieux est très épineuse (en particulier pour étudier des maladies dont l'étiologie est encore largement inconnue). La multi-contamination des milieux a pu être évaluée à partir des données ouvertes disponibles. Mais de nombreux phénomènes environnementaux incontournables n'ont pu être décrits qu'à partir de données indirectes. Il était également indispensable d'inclure dans l'analyse des descripteurs des aménités environnementales. Ainsi, pour approfondir la compréhension et la description de l'environnement physique, il était primordial de compléter l'analyse avec une approche plus holistique.

Une autre méthodologie a donc été conçue et mise en œuvre. Elle permet une description intégrée de l'environnement physique et consiste à construire des typologies territoriales. En l'occurrence, les profils environnementaux des communes ont été décrits. Ces résultats permettent d'avoir une vision beaucoup plus complexe des territoires et d'appréhender les combinaisons de facteurs environnementaux qui les distinguent les uns des autres. Sur le plan environnemental, les résultats obtenus vont beaucoup plus loin que la simple description du continuum urbain-rural. Les profils territoriaux identifiés permettent d'appréhender la manière dont les pressions environnementales se répartissent dans l'espace et se superposent les unes aux autres. De plus, cette approche apporte une vision complémentaire par rapport aux indices composites en intégrant conjointement facteurs de vulnérabilité et facteurs de résilience des territoires. Pour étudier des pathologies dont l'étiologie est en grande partie inconnue, il semble important de prendre en compte l'environnement de manière exhaustive, sans préjugés particuliers, pour capturer l'ensemble des influences potentielles. En cela, les résultats participent à mieux appréhender la complexité des relations entre environnement et santé.

Construire des profils de territoires et combiner cette approche avec des indices composites pourrait s'avérer pertinent. En effet, plusieurs publications ont soulevé la question de la comparabilité des indices composites lorsque les unités spatiales étudiées sont trop radicalement différentes les unes

des autres (Abel et al., 2016; Bertin et al., 2014; Owusu et al., 2022; Zoeteman et al., 2016). Les profils environnementaux pourraient permettre de comparer entre eux des territoires présentant un profil similaire. Ainsi, en calculant des indices composites spécifiquement pour chaque profil de territoire, il serait possible d'accroître leur pertinence et leur pouvoir explicatif. Owusu et al. (2022) ont testé cette approche en calculant un indice de pression environnemental au niveau des unités de recensement sur tout le périmètre des États-Unis. Ils ont ensuite construit une typologie partir de données d'occupation des sols. L'indice de pression environnemental a été recalculé, mais cette fois-ci de manière stratifiée en ne prenant en compte que les unités spatiales appartenant au même profil. La stratification du calcul de ce score s'est avérée très bénéfique, elle a permis de mieux identifier les points noirs environnementaux, non seulement dans les zones urbaines, mais également dans les zones rurales. De même, en matière de santé, il serait très intéressant de considérer un profil territorial en particulier et d'examiner spécifiquement l'association entre un facteur de risque environnemental et une pathologie. L'hypothèse sous-jacente est que les facteurs de risque et leur influence peuvent varier en fonction du profil de territoire considéré. Ce type d'approche, appelé analyse stratifiée, a fait l'objet de plusieurs publications (Cohen et al., 2018; Gray et al., 2018; Jagai et al., 2020, 2017; Rappazzo et al., 2015; Wood et al., 2016; Zhang et al., 2014). Cependant, ces analyses stratifiées n'ont jusqu'à présent été initiées que sur la base de typologies simples distinguant essentiellement les territoires urbains et ruraux.

Il serait également possible de faire le raisonnement inverse en utilisant par exemple les six dimensions utilisées pour les indices de vulnérabilité et de résilience : environnement physique, économie, social, politique, santé, services. Après avoir construit un indice composite spécifique pour chacune de ces dimensions, il serait ensuite envisageable de construire des profils territoriaux sur cette base. Procéder de cette manière permettrait d'une part de mieux maîtriser le poids attribué à chaque phénomène ou dimension. D'autre part, les profils de territoires pourraient alors plus facilement intégrer à la fois la dimension environnementale, sociale, économique, etc... La faisabilité d'une telle approche doit bien sûr être évaluée, mais elle pourrait s'avérer très intéressante pour mieux comprendre les liens complexes entre environnement et santé.

Identification de pistes environnementales impliquées dans l'étiologie des maladies chroniques.

Le dernier volet de cette thèse consistait à exploiter les résultats de ces approches décrivant l'environnement sous un angle multidimensionnel pour éclairer l'étiologie des maladies chroniques. Les résultats ont été donc associés aux variations spatiales d'incidence des maladies chroniques par le biais d'approches écologiques menées à l'échelle territoriale. Dans le cadre de cette thèse, ce travail a été réalisé pour la MC et l'IRCT.

Pour l'IRCT, les résultats obtenus suggèrent un lien avec la pression urbaine et en particulier avec la pollution atmosphérique en particules fines. Ces conclusions ont été obtenues par l'intermédiaire de deux approches complémentaires : indices de multi-contamination et profils environnementaux des communes. Comme évoqué dans les discussions précédentes, ces résultats sont en cohérence avec la littérature existante qui établit avec de plus en plus de certitude une association entre pollution atmosphérique et insuffisance rénale. Plusieurs méta-analyses ont précédemment confirmé cette tendance (Wu et al., 2020; Xu et al., 2018; Ye et al., 2021). Ainsi, les résultats obtenus donnent du crédit aux approches développées et mobilisées dans cette thèse. C'est d'ailleurs un des intérêts d'étudier

deux maladies chroniques, une pour laquelle les déterminants environnementaux sont inconnus (MC), et une autre qui présente une étiologie un peu plus établie (IRCT).

Pour la MC, des résultats très intéressants ont également été obtenus. D'abord, un lien est suggéré entre la répartition spatiale d'incidence de la MC et les pratiques agricoles. Ces résultats sont assez inédits dans la littérature existante, même si quelques publications vont également dans ce sens. Une influence de la naturalité des territoires est aussi une des pistes qui ressort des analyses. Cette hypothèse trouve écho dans la littérature puisque l'accès aux espaces verts (ou naturels) et l'exposition à la biodiversité semblent avoir un effet positif sur la régulation du système immunitaire. Enfin, une association a également été suggérée avec la pollution métallique des sols.

Avant tout, il s'agit de rappeler que les données d'incidence sont agrégées à l'échelle des communes. Ainsi, les résultats obtenus sont issus d'approches écologiques menées à l'échelle du territoire qui ne tiennent pas compte des paramètres individuels. Ces approches présentent un certain nombre de biais écologiques qui ne permettent absolument pas de transposer directement les résultats à l'échelle individuelle. Cependant, un des objectifs initiaux de ce travail de thèse a été atteint. Ces résultats contribuent à mieux comprendre les déterminants environnementaux susceptibles d'influencer la survenue des maladies chroniques à l'échelle des territoires. L'ambition était d'identifier des pistes qui pourront orienter les futures recherches dans ce domaine, ce a été réalisé pour les deux pathologies étudiées.

La possibilité de disposer de données issues de deux registres de santé constitue une chance unique. Cependant, comme pour les données environnementales, il est important d'adopter un regard critique sur ces données également.

Pour l'IRCT, dans le cadre de la thèse de Delphine Brousmiche (Brousmiche, 2022), un travail important a été mené en collaboration avec des néphrologues et des statisticiens pour ajuster les données d'incidence sur un certain nombre de facteurs confondants : défaveur sociale, accessibilité aux centres de dialyse, pratiques de soins. Malgré ce travail considérable, certaines limites dans les données persistent. En particulier, d'autres paramètres individuels (e.g. diabète et hypertension) n'ont pas pu être pris en compte.

Les disparités spatiales d'incidence de la MC ont été détectés par le biais d'une standardisation indirecte, exprimant le RR de développer la maladie par rapport au risque global de la population étudiée, en l'occurrence celle du registre EPIMAD, couvrant quatre départements sur la période 1990-2014. En 2020, une carte nationale a été publiée illustrant une prévalence élevée dans le nord de la France par rapport à l'ensemble du territoire (figure 59) (Genin et al., 2020). Bien que cette carte ait été élaborée à partir de données hospitalières (prévalence au lieu d'incidence) issues du PMSI, sa comparaison avec celle du registre sur les quatre départements indique sa robustesse et sa capacité à refléter la tendance spatiale nationale.

Toute la zone du registre correspond en réalité à une zone de prévalence élevée à l'échelle nationale. Ainsi, les zones de sur-incidence identifiées dans notre étude constituent de véritables zones à risque. Les RR pourraient même être sous-estimés par rapport à l'ensemble du territoire national, étant donné qu'ils ont été calculés de manière relative au risque de la population du Nord de la France, déjà en sur-

incidence. En d'autres termes, nous observons des zones où la sur-incidence est prononcée au sein de régions déjà à risque élevé. Au contraire, les zones de sous-incidence identifiées grâce au registre EPIMAD le sont à l'échelle régionale mais sont probablement moins représentatives à l'échelle nationale. Ces zones pourraient plutôt correspondre à une incidence moyenne, voire à une légère sur-incidence par rapport à d'autres régions françaises. Ainsi, pour consolider nos résultats, il serait particulièrement intéressant de décliner le même type d'approche à l'échelle nationale ou en focalisant l'analyse sur un périmètre où la variabilité d'incidence est plus importante.

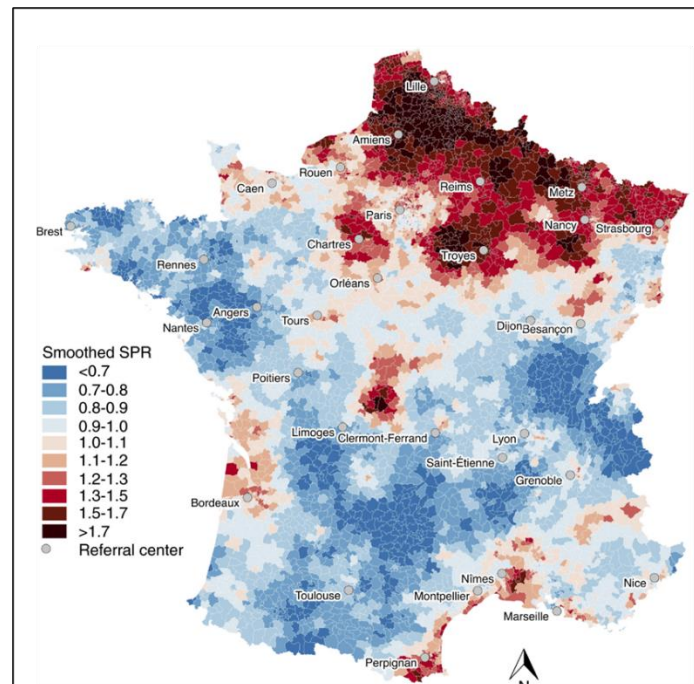


Figure 59 : Répartition spatiale des taux de prévalence de la MC standardisés sur l'âge et le sexe (2007-2014).

Perspectives

Les pistes étiologiques identifiées doivent être validées en menant des analyses plus approfondies. Par ailleurs, d'autres approches doivent être envisagées pour éprouver ces hypothèses.

D'une part, en lien avec l'IRCT, certains résultats présentés dans ce travail doivent encore être consolidés. Par exemple, des régressions écologiques doivent être menées pour confirmer le lien suggéré entre IRCT et pollution atmosphérique. Pour le moment, la distribution d'incidence a été comparée entre les quartiles de contamination croissante, il s'agit donc désormais de confirmer ces observations en prenant en compte la dépendance spatiale entre communes.

Par ailleurs, sur la base des résultats obtenus au cours de la thèse, le projet Maladie Rénale Chronique et Profils Environnementaux (MRC-PE) a récemment été lauréat de l'appel à projet commun Health Data Hub (HDH) et Green Data for Health (GD4H). Les objectifs principaux, sont précisément de

transposer la méthodologie de construction des profils environnementaux à l'échelle des communes sur le territoire national, et de les mettre en relation avec l'incidence de l'IRCT à partir du registre REIN national. Une fois les profils décrits, ce sera également l'occasion de les mettre en relation avec les données nationales de prévalence de la MC.

Pour parvenir à dresser le profil environnemental des communes à l'échelle nationale, de nombreuses réflexions doivent être menées en amont. Ce travail sera l'occasion de tester la validité et la pertinence de la méthode sur un périmètre aussi large. Le premier défi sera d'étendre le jeu de données sur tout le territoire. Des méthodes d'imputation des données manquantes plus performantes pourront éventuellement être envisagées si nécessaire. Par ailleurs, à l'échelle nationale, une des principales difficultés sera de décrire des profils de territoires qui présentent un réel intérêt du point de vue du lien environnement-santé. Il sera essentiel de veiller à ne pas simplement reproduire des profils "paysagers" fortement influencés par des caractéristiques climatiques préétablies (méditerranéen, océanique, continental, montagnard). Afin de ne pas tomber dans cet écueil et dans cette simplification excessive, une attention particulière devra être portée sur la sélection des données d'entrée. Une approche potentiellement fructueuse serait de cibler une thématique environnementale spécifique ou de choisir les données en fonction de la pathologie examinée. Par exemple, dans le cas de l'IRCT, l'élaboration de profils de pollution atmosphérique pourrait se révéler pertinente. Enfin, il pourrait être judicieux de combiner l'approche des profils environnementaux avec celle des indices composites. Cette stratégie permettrait de mener des comparaisons sur des communes aux profils similaires, contribuant ainsi à optimiser le pouvoir explicatif de certains indices composites. Cela permettrait par exemple d'examiner le lien entre la pollution de l'air et l'incidence de l'IRCT dans les différents profils environnementaux identifiés pour caractériser les communes rurales.

Pour finir, ce travail s'inscrit également dans le cadre du projet de recherche « CROhn disease and Pollution of Soils » (Project CROPS / ANR-20-CE34-0015) financé par l'Agence Nationale de la Recherche. Il s'agit d'un projet interdisciplinaire dont l'objectif principal est de renseigner l'étiologie de la MC. Le projet s'appuie sur les données du registre Epimad et sur l'étude des disparités spatiales d'incidence de la maladie. Il s'articule en plusieurs volets. Un premier volet consiste à utiliser des données environnementales ouvertes pour constituer un jeu de données spatial et tenter d'expliquer les variations spatiales d'incidence de la MC. Cette première manière de raisonner correspond au travail de thèse présenté dans ce mémoire. Mais en parallèle, plusieurs approches très complémentaires ont également été mises en œuvre.

D'abord, des campagnes de prélèvements ont été menées afin de caractériser les niveaux de contamination dans les trois compartiments environnementaux (air, eau, sols), au sein des clusters spatiaux d'incidence de la MC. Ces campagnes offrent l'opportunité d'étudier des polluants qui ne sont pas ou peu décrits dans les bases de données ouvertes. Les trois milieux ont pu être étudiés : niveaux de contamination dans les sols, niveaux de contamination dans les lichens épiphytes (biosurveillance de la qualité de l'air), niveaux de contamination dans les eaux de distribution.

Pour rappel, quatre clusters de sur-incidence et quatre clusters de sous-incidence ont été détectés sur le périmètre du registre Epimad. Un plan d'échantillonnage identique a été utilisé pour les trois milieux environnementaux. Les sites de prélèvement ont été répartis au sein des huit clusters de la MC, en utilisant une maille hexagonale et en ciblant l'épicentre des clusters (c'est-à-dire la zone où le risque relatif est le plus atypique). La taille des mailles est de 3.64 km de largeur (diamètre du cercle circonscrit

de l'hexagone). Dans chaque cluster, environ 30 prélèvements ont été réalisés, soit un total d'environ 240 prélèvements pour chacun des milieux. Le plan d'échantillonnage en question est présenté en figure 60.

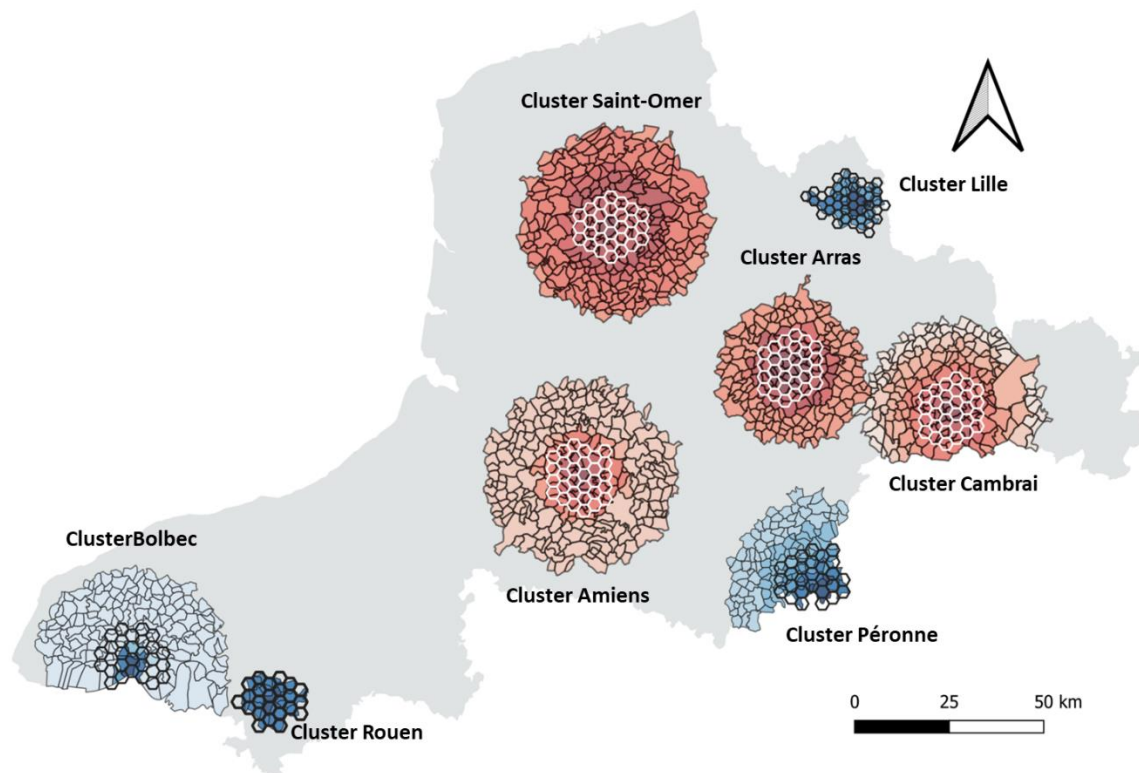


Figure 60 : Plan d'échantillonnage utilisé pour les campagnes de prélèvements réalisés au sein des huit clusters spatiaux d'incidence de la MC.

La contamination des sols a été étudiée par Lara-Maria Wakim qui est doctorante au laboratoire LGCgE. Dans les sols, ses recherches se sont axées principalement sur l'étude des hormones (œstrogènes) et antibiotiques. Elle a mis au point un protocole expérimental pour doser ces contaminants dans les sols. Leur analyse présente un intérêt majeur car ces molécules sont évoquées dans certaines études comme possiblement impliquées dans l'étiologie de la MC (Axelrad et al., 2020; Jacenik et al., 2019; Martin et al., 2015; Meyer et al., 2020; Nie et al., 2018; Tenailleau et al., 2020). En parallèle, des polluants plus couramment surveillés dans les sols ont également été analysés (HAPs, métaux, pesticides, PCBs, etc...). Dans le cadre de ma thèse, j'ai participé à l'organisation et à la réalisation des campagnes de biosurveillance lichénique et de prélèvements d'eau potable dans les réseaux de distribution. Pour les lichens, les polluants ciblés sont les métaux. Dans les eaux de distribution, un grand nombre de polluants ont été sélectionnés : métaux, pesticides, HAPs, PCBs, COVs. A l'heure actuelle, tous les prélèvements ont été effectués et les résultats des trois campagnes sont en cours d'analyse.

Ces résultats ouvrent de nombreuses perspectives puisqu'ils vont permettre de comparer la qualité des milieux entre les clusters d'incidence. L'objectif est toujours le même, à savoir identifier des pistes étiologiques pour la MC. Une attention particulière sera donnée aux contaminations d'origine agricole, étant donné les résultats obtenus dans ce mémoire. Par ailleurs, ces résultats constituent aussi une opportunité pour confronter les données de terrain avec les résultats issus de l'exploitation des

données ouvertes. Ce travail pourra être réalisé pour les polluants étudiés en « doublon » (métaux dans les sols notamment).

Un autre doctorant (Léo Heuguebart), qui fait partie intégrante du projet CROPS, se consacre à la caractérisation historique des clusters spatiaux d'incidence. Ses recherches se sont particulièrement concentrées sur l'analyse des pratiques agricoles au sein de ces clusters au cours de la seconde moitié du XXe siècle. Cette approche historique offre une perspective précieuse pour mieux comprendre l'évolution des facteurs environnementaux qui pourraient expliquer la répartition spatiale d'incidence actuelle de la MC.

La dernière phase du projet sera menée en étroite collaboration avec des médecins et toxicologues. L'objectif principal de cette phase consistera à recouper et synthétiser les résultats obtenus à partir des différentes approches mobilisées. Sur la base des connaissances actuelles sur la MC, les différentes pistes environnementales identifiées seront hiérarchisées et des réflexions seront menées sur les orientations à donner pour la suite des recherches.

Références bibliographiques

- Aamodt, G., Bukholm, G., Jahnsen, J., Moum, B., Vatn, M.H., the IBSEN Study Group, 2008. The Association Between Water Supply and Inflammatory Bowel Disease Based on a 1990–1993 Cohort Study in Southeastern Norway. *American Journal of Epidemiology* 168, 1065–1072. <https://doi.org/10.1093/aje/kwn218>
- Abel, G.A., Barclay, M.E., Payne, R.A., 2016. Adjusted indices of multiple deprivation to enable comparisons within and between constituent countries of the UK including an illustration using mortality rates. *BMJ Open* 6, e012750. <https://doi.org/10.1136/bmjopen-2016-012750>
- Aerts, R., Honnay, O., Van Nieuwenhuysse, A., 2018. Biodiversity and human health: mechanisms and evidence of the positive health effects of diversity in nature and green spaces. *British Medical Bulletin* 127, 5–22. <https://doi.org/10.1093/bmb/ldy021>
- Aghighi, M., Mahdavi Mazdeh, M., Zamyadi, M., Heydari Rouchi, A. r, Rajolani, H., Nourouzi, S., 2009. Changing epidemiology of end-stage renal disease in last ten years in Iran 3, 192–196.
- Agrawal, M., Burisch, J., Colombel, J.-F., C. Shah, S., 2020. Viewpoint: Inflammatory Bowel Diseases Among Immigrants From Low- to High-Incidence Countries: Opportunities and Considerations. *Journal of Crohn’s and Colitis* 14, 267–273. <https://doi.org/10.1093/ecco-jcc/jjz139>
- Agreste, 2020a. Pratiques culturelles en grandes cultures 2017.
- Agreste, 2020b. Pratiques culturelles en grandes cultures 2017 : IFT et nombre de traitements. (Bilan périodique), Agreste Chiffres et Données. Ministère de l’agriculture et de l’alimentation.
- Almaguer, M., Herrera, R., Orantes, C.M., 2014. Chronic Kidney Disease of Unknown Etiology in Agricultural Communities. *MEDICC Review* 16, 9. <https://doi.org/10.37757/MR2014.V16.N2.3>
- Alvarez, S., Timler, C.J., Michalscheck, M., Paas, W., Descheemaeker, K., Tittonell, P., Andersson, J.A., Groot, J.C.J., 2018. Capturing farm diversity with hypothesis-based typologies: An innovative methodological framework for farming system typology development. *PLOS ONE* 13, e0194757. <https://doi.org/10.1371/journal.pone.0194757>
- Andre, J.-M., Bouchard, D., Druart, A., Durand, A., Gavel, A., Gueguen, C., Jeannot, C., Mathias, E., Nicco, L., Bort, R., Robert, C., Serveau, L., Taieb, N., Ringuet, J., Vieira Da Rocha, T., 2018. Rapport OMINEA – 15ème édition. Organisation et méthodes des inventaires nationaux des émissions atmosphériques en France. CITEPA.
- Angulo, E., 1996. The Tomlinson Pollution Load Index applied to heavy metal, ‘Mussel-Watch’ data: a useful index to assess coastal pollution. *Science of The Total Environment* 187, 19–56. [https://doi.org/10.1016/0048-9697\(96\)05128-5](https://doi.org/10.1016/0048-9697(96)05128-5)
- Armitage, E.L., Aldhous, M.C., Anderson, N., Drummond, H.E., Riemersma, R.A., Ghosh, S., Satsangi, J., 2004. Incidence of juvenile-onset Crohn’s disease in Scotland: Association with northern latitude and affluence. *Gastroenterology* 127, 1051–1057. <https://doi.org/10.1053/j.gastro.2004.06.024>
- Arrouays, D., Jolivet, C., Boulonne, L., Bodineau, G., Saby, N., Grolleau, É., 2003. Le Réseau de Mesures de la Qualité des Sols (RMQS) de France. *Etude et Gestion des Sols* 10, 241–250.
- Axelrad, J.E., Bazarbashi, A., Zhou, J., Castañeda, D., Gujral, A., Sperling, D., Glass, J., Agrawal, M., Hong, S., Lawlor, G., Hudesman, D., Chang, S., Shah, S., Yajnik, V., Ananthakrishnan, A., Khalili, H., Colombel, J.-F., Itzkowitz, S., 2020. Hormone Therapy for Cancer Is a Risk Factor for Relapse of Inflammatory Bowel Diseases. *Clinical Gastroenterology and Hepatology* 18, 872–880.e1. <https://doi.org/10.1016/j.cgh.2019.06.042>
- Banerjee, R., Pal, P., Patel, R., Godbole, S., Komawar, A., Mudigonda, S., Akki, Y., Gaddam, A., Pasula, N.P., Joseph, S., Valluari, S., Mekala, D., Khalil, M., Kanaganti, S., Sekaran, A., Reddy, D.N.,

- Travis, S.P.L., 2023. Inflammatory bowel disease (IBD) in rural and urban India: results from community colonoscopic evaluation of more than 30,000 symptomatic patients. *The Lancet Regional Health - Southeast Asia* 100259. <https://doi.org/10.1016/j.lanseas.2023.100259>
- Barbieri, M., 2016. The Importance of Enrichment Factor (EF) and Geoaccumulation Index (Igeo) to Evaluate the Soil Contamination. *J Geol Geophys* 5. <https://doi.org/10.4172/2381-8719.1000237>
- Benchimol, E.I., Kaplan, G.G., Otley, A.R., Nguyen, G.C., Underwood, F.E., Guttman, A., Jones, J.L., Potter, B.K., Catley, C.A., Nugent, Z.J., Cui, Y., Tanyingoh, D., Mojaverian, N., Bitton, A., Carroll, M.W., deBruyn, J., Dummer, T.J.B., El-Matary, W., Griffiths, A.M., Jacobson, K., Kuenzig, M.E., Leddin, D., Lix, L.M., Mack, D.R., Murthy, S.K., Sánchez, J.N.P., Singh, H., Targownik, L.E., Vutcovici, M., Bernstein, C.N., 2017. Rural and Urban Residence During Early Life is Associated with Risk of Inflammatory Bowel Disease: A Population-Based Inception and Birth Cohort Study. *Am J Gastroenterol* 112, 1412–1422. <https://doi.org/10.1038/ajg.2017.208>
- Benmarhnia, T., Kihal-Talantikite, W., Ragetti, M.S., Deguen, S., 2017. Small-area spatiotemporal analysis of heatwave impacts on elderly mortality in Paris: A cluster analysis approach. *Science of The Total Environment* 592, 288–294. <https://doi.org/10.1016/j.scitotenv.2017.03.102>
- Bertin, M., Chevrier, C., Pelé, F., Serrano-Chavez, T., Cordier, S., Viel, J.-F., 2014. Can a deprivation index be used legitimately over both urban and rural areas? *International Journal of Health Geographics* 13, 22. <https://doi.org/10.1186/1476-072X-13-22>
- Besag, J., York, J., Mollié, A., 1991a. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math* 43, 1–20. <https://doi.org/10.1007/BF00116466>
- Besag, J., York, J., Mollié, A., 1991b. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math* 43, 1–20. <https://doi.org/10.1007/BF00116466>
- Besag, J., York, J., Mollié, A., 1991c. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 43, 1–20.
- Billon, L., Gregoire, A., Duchene, C., Mustiere, S., Lombard, A., Sordello, R., 2017. Réalisation d'une carte de synthèse nationale des continuités écologiques régionales (Rapport méthodologique). UMS Patrinat.
- Bishoi, B., Prakash, A., Jain, V.K., 2009. A Comparative Study of Air Quality Index Based on Factor Analysis and US-EPA Methods for an Urban Environment. *Aerosol Air Qual. Res.* 9, 1–17. <https://doi.org/10.4209/aaqr.2008.02.0007>
- Blum, M.F., Surapaneni, A., Stewart, J.D., Liao, D., Yanosky, J.D., Whitsel, E.A., Power, M.C., Grams, M.E., 2020. Particulate Matter and Albuminuria, Glomerular Filtration Rate, and Incident CKD. *Clin J Am Soc Nephrol* 15, 311–319. <https://doi.org/10.2215/CJN.08350719>
- Bocquet, M., Pierreuse, E., Dupré, O., Lory, P., 2022. Mesure de la consommation d'espaces à l'aide des Fichiers Fonciers : Définitions, précisions méthodologiques, limites et précautions d'interprétation. Cerema.
- Böhringer, C., Jochem, P.E.P., 2007. Measuring the immeasurable — A survey of sustainability indices. *Ecological Economics* 63, 1–8. <https://doi.org/10.1016/j.ecolecon.2007.03.008>
- Bowe, B., Xie, Y., Li, T., Yan, Y., Xian, H., Al-Aly, Z., 2018a. Particulate Matter Air Pollution and the Risk of Incident CKD and Progression to ESRD. *JASN* 29, 218–230. <https://doi.org/10.1681/ASN.2017030253>
- Bowe, B., Xie, Y., Li, T., Yan, Y., Xian, H., Al-Aly, Z., 2018b. Particulate Matter Air Pollution and the Risk of Incident CKD and Progression to ESRD. *Journal of the American Society of Nephrology* 29, 218. <https://doi.org/10.1681/ASN.2017030253>

- Bowe, B., Xie, Y., Li, T., Yan, Y., Xian, H., Al-Aly, Z., 2018c. Particulate Matter Air Pollution and the Risk of Incident CKD and Progression to ESRD. *J Am Soc Nephrol* 29, 218–230. <https://doi.org/10.1681/ASN.2017030253>
- Bowe, B., Xie, Y., Li, T., Yan, Y., Xian, H., Al-Aly, Z., 2017. Associations of ambient coarse particulate matter, nitrogen dioxide, and carbon monoxide with the risk of kidney disease: a cohort study. *The Lancet Planetary Health* 1, e267–e276. [https://doi.org/10.1016/S2542-5196\(17\)30117-1](https://doi.org/10.1016/S2542-5196(17)30117-1)
- Boyle, P.J., Kudlac, H., Williams, A.J., 1996. Geographical variation in the referral of patients with chronic end stage renal failure for renal replacement therapy. *QJM: An International Journal of Medicine* 89, 151–158. <https://doi.org/10.1093/qjmed/89.2.151>
- Bragg-Gresham, J., Morgenstern, H., McClellan, W., Saydah, S., Pavkov, M., Williams, D., Powe, N., Tuot, D., Hsu, R., Saran, R., System, for the C. for D.C. and P.C.S., 2018. County-level air quality and the prevalence of diagnosed chronic kidney disease in the US Medicare population. *PLOS ONE* 13, e0200612. <https://doi.org/10.1371/journal.pone.0200612>
- Braveman, P.A., Cubbin, C., Egerter, S., Williams, D.R., Pamuk, E., 2010. Socioeconomic Disparities in Health in the United States: What the Patterns Tell Us. *Am J Public Health* 100, S186–S196. <https://doi.org/10.2105/AJPH.2009.166082>
- Bréhon, D., Clair, P., Gueguen, C., Luneau, G., Mazurais, S., Malacarne, R., Perrussel, O., Piga, D., Racher, A., Real, E., 2018. Guide méthodologique pour l'élaboration des inventaires territoriaux des émissions atmosphériques. Ministère chargé de l'environnement.
- Brender, J.D., Maantay, J.A., Chakraborty, J., 2011. Residential Proximity to Environmental Hazards and Adverse Health Outcomes. *Am J Public Health* 101, S37–S52. <https://doi.org/10.2105/AJPH.2011.300183>
- Brousmiche, D., 2022. Caractérisation de la multi-exposition environnementale : exploration des potentiels de résilience et de vulnérabilité face aux risques sanitaires (These de doctorat). Université de Lille.
- Brousmiche, D., Genin, M., Occelli, F., Frank, L., Deram, A., Cuny, D., Lanier, C., 2021. Data for the assessment of vulnerability and resilience in the field of environmental health in the north of France. Data in Brief 37, 107220. <https://doi.org/10.1016/j.dib.2021.107220>
- Brousmiche, D., Genin, M., Occelli, F., Frank, L., Deram, A., Cuny, D., Lanier, C., 2020a. How can we analyze environmental health resilience and vulnerability? A joint analysis with composite indices applied to the north of France. *Science of The Total Environment* 142983. <https://doi.org/10.1016/j.scitotenv.2020.142983>
- Brousmiche, D., Lanier, C., Cuny, D., Frevent, C., Genin, M., Blanc-Garin, C., Amouyel, P., Deram, A., Occelli, F., Meirhaeghe, A., 2023. How do territorial characteristics affect spatial inequalities in the risk of coronary heart disease? *Science of The Total Environment* 867, 161563. <https://doi.org/10.1016/j.scitotenv.2023.161563>
- Brousmiche, D., Occelli, F., Genin, M., Cuny, D., Deram, A., Lanier, C., 2020b. Spatialized composite indices to evaluate environmental health inequalities: Meeting the challenge of selecting relevant variables. *Ecological Indicators* 111, 106023. <https://doi.org/10.1016/j.ecolind.2019.106023>
- Bukalasa, J.S., Brunekreef, B., Brouwer, M., Vermeulen, R., de Jongste, J.C., van Rossem, L., Vonk, J.M., Wijga, A., Huss, A., Gehring, U., 2017. Proximity to agricultural fields as proxy for environmental exposure to pesticides among children: The PIAMA birth cohort. *Science of The Total Environment* 595, 515–520. <https://doi.org/10.1016/j.scitotenv.2017.03.269>
- Burgass, M.J., Halpern, B.S., Nicholson, E., Milner-Gulland, E.J., 2017. Navigating uncertainty in environmental composite indicators. *Ecological Indicators* 75, 268–278. <https://doi.org/10.1016/j.ecolind.2016.12.034>

- Candela, S., Bonvicini, L., Ranzi, A., Baldacchini, F., Broccoli, S., Cordioli, M., Carretta, E., Luberto, F., Angelini, P., Evangelista, A., Marzaroli, P., Giorgi Rossi, P., Forastiere, F., 2015. Exposure to emissions from municipal solid waste incinerators and miscarriages: A multisite study of the MONITER Project. *Environment International* 78, 51–60. <https://doi.org/10.1016/j.envint.2014.12.008>
- Chang, P.-Y., Li, Y.-L., Chuang, T.-W., Chen, S.-Y., Lin, L.-Y., Lin, Y.-F., Chiou, H.-Y., 2022. Exposure to ambient air pollutants with kidney function decline in chronic kidney disease patients. *Environmental Research* 215, 114289. <https://doi.org/10.1016/j.envres.2022.114289>
- Chayre, A., Pointereau, P., 2023. Évaluation de l'usage territorialisé des pesticides en France métropolitaine.
- Chayre, A., Pointereau, P., Mezilet, Y., 2023. Méthodologie de calcul de l'indicateur de fréquence de traitement phytosanitaire en agriculture par commune.
- Chen, Y., Cao, F., Xiao, J.-P., Fang, X.-Y., Wang, X.-R., Ding, L.-H., Wang, D.-G., Pan, H.-F., 2021. Emerging role of air pollution in chronic kidney disease. *Environ Sci Pollut Res* 28, 52610–52624. <https://doi.org/10.1007/s11356-021-16031-6>
- Cheng, Y.-Y., Chang, Y.-T., Cheng, H.-L., Shen, K.-H., Sung, J.-M., Guo, H.-R., 2018. Associations between arsenic in drinking water and occurrence of end-stage renal disease with modifications by comorbidities: A nationwide population-based study in Taiwan. *Science of The Total Environment* 626, 581–591. <https://doi.org/10.1016/j.scitotenv.2018.01.043>
- Cheng, Y.-Y., Huang, N.-C., Chang, Y.-T., Sung, J.-M., Shen, K.-H., Tsai, C.-C., Guo, H.-R., 2017. Associations between arsenic in drinking water and the progression of chronic kidney disease: A nationwide study in Taiwan. *Journal of Hazardous Materials* 321, 432–439. <https://doi.org/10.1016/j.jhazmat.2016.09.032>
- Chien, L.-C., Guo, Y., Zhang, K., 2016. Spatiotemporal analysis of heat and heat wave effects on elderly mortality in Texas, 2006–2011. *Science of The Total Environment* 562, 845–851. <https://doi.org/10.1016/j.scitotenv.2016.04.042>
- Cohen, S.A., Cook, S.K., Sando, T.A., Sabik, N.J., 2018. What Aspects of Rural Life Contribute to Rural-Urban Health Disparities in Older Adults? Evidence From a National Survey. *The Journal of Rural Health* 34, 293–303. <https://doi.org/10.1111/jrh.12287>
- Collins, A.J., Foley, R.N., Gilbertson, D.T., Chen, S.-C., 2015. United States Renal Data System public health surveillance of chronic kidney disease and end-stage renal disease. *Kidney International Supplements, Dialysis Registries Worldwide: Proceedings from the 59th Annual Meeting of the Japanese Society for Dialysis Therapy* 5, 2–7. <https://doi.org/10.1038/kisup.2015.2>
- Collister, D., Ferguson, T., Komenda, P., Tangri, N., 2016. The Patterns, Risk Factors, and Prediction of Progression in Chronic Kidney Disease: A Narrative Review. *Seminars in Nephrology* 36, 273–282. <https://doi.org/10.1016/j.semnephrol.2016.05.004>
- Cornelis, C., Schoeters, G., Kellen, E., Buntinx, F., Zeegers, M., 2009. Development of a GIS-based indicator for environmental pesticide exposure and its application to a Belgian case-control study on bladder cancer. *International Journal of Hygiene and Environmental Health* 212, 172–185. <https://doi.org/10.1016/j.ijheh.2008.06.001>
- Coste, A., Goujon, S., Faure, L., Hémon, D., Clavel, J., 2020. Agricultural crop density in the municipalities of France and incidence of childhood leukemia: An ecological study. *Environmental Research* 187, 109517. <https://doi.org/10.1016/j.envres.2020.109517>
- Couchoud, C., Guihenneuc, C., Bayer, F., Stengel, B., on behalf of the REIN registry, 2010. The timing of dialysis initiation affects the incidence of renal replacement therapy. *Nephrology Dialysis Transplantation* 25, 1576–1578. <https://doi.org/10.1093/ndt/gfp675>
- Couchoud, C., Stengel, B., Landais, P., Aldigier, J.-C., de Cornelissen, F., Dabot, C., Maheut, H., Joyeux, V., Kessler, M., Labeeuw, M., Isnard, H., Jacquelinet, C., 2006. The renal epidemiology and

- information network (REIN): a new registry for end-stage renal disease in France. *Nephrol Dial Transplant* 21, 411–418. <https://doi.org/10.1093/ndt/gfi198>
- Counil, É., Cherni, N., Kharrat, M., Achour, A., Trimech, H., 2008. Trends of Incident Dialysis Patients in Tunisia Between 1992 and 2001. *American Journal of Kidney Diseases* 51, 463–470. <https://doi.org/10.1053/j.ajkd.2007.10.032>
- da Luz Moreira, A., de Campos Lobato, L.F., de Lima Moreira, J.P., Luiz, R.R., Elia, C., Fiocchi, C., de Souza, H.S.P., 2022. Geosocial Features and Loss of Biodiversity Underlie Variable Rates of Inflammatory Bowel Disease in a Large Developing Country: A Population-Based Study. *Inflammatory Bowel Diseases* 28, 1696–1708. <https://doi.org/10.1093/ibd/izab346>
- Dahlgren, G., Whitehead, M., 1991. Policies and strategies to promote social equity in health. Background document to WHO – strategy paper for Europe. Institute for futures studies.
- Davoli, E., Fattore, E., Paiano, V., Colombo, A., Palmiotto, M., Rossi, A.N., Il Grande, M., Fanelli, R., 2010. Waste management health risk assessment: A case study of a solid waste landfill in South Italy. *Waste Management* 30, 1608–1613. <https://doi.org/10.1016/j.wasman.2009.10.013>
- De la Fuente, F., Saldías, M.A., Cubillos, C., Mery, G., Carvajal, D., Bowen, M., Bertoglia, M.P., 2021. Green Space Exposure Association with Type 2 Diabetes Mellitus, Physical Activity, and Obesity: A Systematic Review. *International Journal of Environmental Research and Public Health* 18, 97. <https://doi.org/10.3390/ijerph18010097>
- Declercq, C., Gower-Rousseau, C., Vernier-Massouille, G., Salleron, J., Baldé, M., Poirier, G., Lerebours, É., Dupas, J.L., Merle, V., Marti, R., Duhamel, A., Cortot, A., Salomez, J.-L., Colombel, J.-F., 2010a. Mapping of inflammatory bowel disease in northern France: Spatial variations and relation to affluence. *Inflammatory Bowel Diseases* 16, 807–812. <https://doi.org/10.1002/ibd.21111>
- Declercq, C., Gower-Rousseau, C., Vernier-Massouille, G., Salleron, J., Baldé, M., Poirier, G., Lerebours, É., Dupas, J.L., Merle, V., Marti, R., Duhamel, A., Cortot, A., Salomez, J.-L., Colombel, J.-F., 2010b. Mapping of inflammatory bowel disease in northern France: Spatial variations and relation to affluence. *Inflammatory Bowel Diseases* 16, 807–812. <https://doi.org/10.1002/ibd.21111>
- Deepak, P., Barnes, E.L., Shaukat, A., 2023. Health Disparities in Inflammatory Bowel Disease Care Driven by Rural Versus Urban Residence: Challenges and Potential Solutions. *Gastroenterology* 165, 11–15. <https://doi.org/10.1053/j.gastro.2023.05.017>
- Defer, A., Gautier, M., Georges, A., Girard, D., Hadrane, S., Irrien, A., Julliard, B., Pigeon, H., Pineau, N., 2019. Commune par commune, la carte de France des pesticides [WWW Document]. Mediapart. URL <https://www.mediapart.fr/journal/france/040719/commune-par-commune-la-carte-de-france-des-pesticides> (accessed 9.17.23).
- Dheer, R., Patterson, J., Dudash, M., Stachler, E.N., Bibby, K.J., Stolz, D.B., Shiva, S., Wang, Z., Hazen, S.L., Barchowsky, A., Stolz, J.F., 2015. Arsenic induces structural and compositional colonic microbiome change and promotes host nitrogen and amino acid metabolism. *Toxicology and Applied Pharmacology* 289, 397–408. <https://doi.org/10.1016/j.taap.2015.10.020>
- Dinno, A., 2015. Nonparametric pairwise multiple comparisons in independent groups using Dunn’s test. *The Stata Journal* 15, 292–300.
- Domingo, J.L., Marquès, M., Mari, M., Schuhmacher, M., 2020. Adverse health effects for populations living near waste incinerators with special attention to hazardous waste incinerators. A review of the scientific literature. *Environmental Research* 187, 109631. <https://doi.org/10.1016/j.envres.2020.109631>
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013a. Collinearity: a review of

- methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013b. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Eggers, P.W., 2011. Has the incidence of end-stage renal disease in the USA and other countries stabilized? *Current Opinion in Nephrology and Hypertension* 20, 241–245. <https://doi.org/10.1097/MNH.0b013e3283454319>
- Elliott, P., Briggs, D., Morris, S., Hoogh, C. de, Hurt, C., Jensen, T.K., Maitland, I., Richardson, S., Wakefield, J., Jarup, L., 2001. Risk of adverse birth outcomes in populations living near landfill sites. *BMJ* 323, 363–368. <https://doi.org/10.1136/bmj.323.7309.363>
- Elliott, P., Shaddick, G., Kleinschmidt, I., Jolley, D., Walls, P., Beresford, J., Grundy, C., 1996. Cancer incidence near municipal solid waste incinerators in Great Britain. *Br J Cancer* 73, 702–710. <https://doi.org/10.1038/bjc.1996.122>
- Elten, M., Benchimol, E.I., Fell, D.B., Kuenzig, M.E., Smith, G., Kaplan, G.G., Chen, H., Crouse, D., Lavigne, E., 2021. Residential Greenspace in Childhood Reduces Risk of Pediatric Inflammatory Bowel Disease: A Population-Based Cohort Study. *Official journal of the American College of Gastroenterology | ACG* 116, 347. <https://doi.org/10.14309/ajg.0000000000000990>
- Everitt, B.S., 1979. Unresolved Problems in Cluster Analysis. *Biometrics* 35, 169–181. <https://doi.org/10.2307/2529943>
- Facchinelli, A., Sacchi, E., Mallen, L., 2001. Multivariate statistical and GIS-based approach to identify heavy metal sources in soils. *Environmental Pollution* 114, 313–324. [https://doi.org/10.1016/S0269-7491\(00\)00243-8](https://doi.org/10.1016/S0269-7491(00)00243-8)
- Fayet, Y., Praud, D., Fervers, B., Ray-Coquard, I., Blay, J.-Y., Ducimetiere, F., Fagherazzi, G., Faure, E., 2020. Beyond the map: evidencing the spatial dimension of health inequalities. *International Journal of Health Geographics* 19, 46. <https://doi.org/10.1186/s12942-020-00242-0>
- Froger, C., Jolivet, C., Budzinski, H., Pierdet, M., Caria, G., Saby, N.P.A., Arrouays, D., Bispo, A., 2023a. Pesticide Residues in French Soils: Occurrence, Risks, and Persistence. *Environ. Sci. Technol.* 57, 7818–7827. <https://doi.org/10.1021/acs.est.2c09591>
- Froger, C., Jolivet, C., Budzinski, H., Pierdet, M., Caria, G., Saby, N.P.A., Arrouays, D., Bispo, A., 2023b. Pesticide Residues in French Soils: Occurrence, Risks, and Persistence. *Environ. Sci. Technol.* 57, 7818–7827. <https://doi.org/10.1021/acs.est.2c09591>
- Fuller, R., Landrigan, P.J., Balakrishnan, K., Bathan, G., Bose-O'Reilly, S., Brauer, M., Caravanos, J., Chiles, T., Cohen, A., Corra, L., Cropper, M., Ferraro, G., Hanna, J., Hanrahan, D., Hu, H., Hunter, D., Janata, G., Kupka, R., Lanphear, B., Lichtveld, M., Martin, K., Mustapha, A., Sanchez-Triana, E., Sandilya, K., Schaeffli, L., Shaw, J., Seddon, J., Suk, W., Téllez-Rojo, M.M., Yan, C., 2022. Pollution and health: a progress update. *The Lancet Planetary Health* 0. [https://doi.org/10.1016/S2542-5196\(22\)00090-0](https://doi.org/10.1016/S2542-5196(22)00090-0)
- Fumery, M., Savoye, G., Pariente, B., Turck, D., Gower-Rousseau, C., 2018. Epidemiology and natural history of inflammatory bowel disease: 30 years of epimad registry. *Hépatogastro & Oncologie Digestive* 25, 145–152. <https://doi.org/10.1684/hpg.2017.1572>
- G, K.E., Osama, E.M., 2010. ENVIRONMENTAL FACTORS INCRIMINATED IN THE DEVELOPMENT OF END STAGE RENAL DISEASE IN EL-MINIA GOVERNORATE, UPPER EGYPT 2, 431–437.
- Garcia, J., Colosio, J., 2001. Les indices de qualité de l'air: élaboration, usages et comparaisons internationales. Presses des MINES.

- Garelick, H., Jones, H., Dybowska, A., Valsami-Jones, E., 2008. Arsenic Pollution Sources, in: *Reviews of Environmental Contamination Volume 197: International Perspectives on Arsenic Pollution and Remediation*, *Reviews of Environmental Contamination and Toxicology*. Springer, New York, NY, pp. 17–60. https://doi.org/10.1007/978-0-387-79284-2_2
- Gatti, M.G., Bechtold, P., Campo, L., Barbieri, G., Quattrini, G., Ranzi, A., Sucato, S., Olgiati, L., Polledri, E., Romolo, M., Iacuzio, L., Carrozzi, G., Lauriola, P., Goldoni, C.A., Fustinoni, S., 2017. Human biomonitoring of polycyclic aromatic hydrocarbons and metals in the general population residing near the municipal solid waste incinerator of Modena, Italy. *Chemosphere* 186, 546–557. <https://doi.org/10.1016/j.chemosphere.2017.07.122>
- Genin, M., Duhamel, A., Preda, C., Fumery, M., Savoye, G., Peyrin-Biroulet, L., Salleron, J., Lerebours, E., Vasseur, F., Cortot, A., Colombel, J.-F., Gower-Rousseau, C., 2013. Space-time clusters of Crohn's disease in northern France. *J Public Health* 21, 497–504. <https://doi.org/10.1007/s10389-013-0580-9>
- Genin, M., Fumery, M., Occelli, F., Savoye, G., Pariente, B., Dauchet, L., Giovannelli, J., Vignal, C., Body-Malapel, M., Sarter, H., Gower-Rousseau, C., Ficheur, G., 2020. Fine-scale geographical distribution and ecological risk factors for Crohn's disease in France (2007-2014). *Aliment Pharmacol Ther* 51, 139–148. <https://doi.org/10.1111/apt.15512>
- Genin, M., Lecoivre, A., Cuny, D., Subtil, D., Chevalier, G., Ficheur, G., Occelli, F., Garabedian, C., 2022a. The association between the incidence of preterm birth and overall air pollution: A nationwide, fine-scale, spatial study in France from 2012 to 2018. *Environmental Pollution* 311, 120013. <https://doi.org/10.1016/j.envpol.2022.120013>
- Genin, M., Lecoivre, A., Cuny, D., Subtil, D., Chevalier, G., Ficheur, G., Occelli, F., Garabedian, C., 2022b. The association between the incidence of preterm birth and overall air pollution: A nationwide, fine-scale, spatial study in France from 2012 to 2018. *Environ Pollut* 311, 120013. <https://doi.org/10.1016/j.envpol.2022.120013>
- Ghione, S., Sarter, H., Fumery, M., Mathurin, A., Armengol-Debeir, L., Savoye, G., Guillaume, L., Ley, D., Spyckerelle, C., Pariente, B., Peyrin-Biroulet, L., Turck, D., Gower-Rousseau, C., Andre, J.M., Antonietti, M., Aouakli, A., Armand, A., Aroichane, I., Assi, F., Aubert, J.P., Auxenfans, E., Ayafi-Ramelot, F., Bankovski, D., Barbry, B., Bardoux, N., Baron, P., Baudet, A., Bazin, B., Bebahani, A., Becqwert, J.P., Benet, V., Benali, H., Benguigui, C., Soussan, B.E., Bental, A., Berkelmans, I., Bernet, J., Bernou, K., Bernou-Dron, C., Bertot, P., Bertiaux-Vandaële, N., Bertrand, V., Billoud, E., Biron, N., Bismuth, B., Bleuet, M., Blondel, F., Blondin, V., Bohon, P., Boniface, E., Bonnière, P., Bonvarlet, E., Bonvarlet, P., Boruchowicz, A., Bostvironnois, R., Boualit, M., Bouche, B., Boudaillez, C., Bourgeaux, C., Bourgeois, M., Bourguet, A., Bourienne, A., Branche, J., Bray, G., Brazier, F., Breban, P., Brihier, H., Brung-Lefebvre, V., Bulois, P., Burgiere, P., Butel, J., Canva, J.Y., Canva-Delcambre, V., Capron, J.P., Cardot, F., Carpentier, P., Cartier, E., Cassar, J.F., Cassagnou, M., Castex, J.F., Catala, P., Cattan, S., Catteau, S., Caujolle, B., Cayron, G., Chandelier, C., Chantre, M., Charles, J., Charneau, T., Chavance-Thelu, M., Chirita, D., Choteau, A., Claerbout, J.F., Clergue, P.Y., Coevoet, H., Cohen, G., Collet, R., Colombel, J.F., Coopman, S., Corvisart, J., Cortot, A., Couttenier, F., Crinquette, J.F., Crombe, V., Dadamessi, I., Dapvril, V., Davion, T., Dautreme, S., Debas, J., Degrave, N., Dehont, F., Delatre, C., Delcenserie, R., Delette, O., Delgrange, T., Delhoustal, L., Delmotte, J.S., Demmane, S., Deregnaucourt, G., Descombes, P., Desechalliers, J.P., Desmet, P., Desreumaux, P., Desseaux, G., Desurmont, P., Devienne, A., Devouge, E., Devred, M., Devroux, A., Dewailly, A., Dharancy, S., Di Fiore, A., Djeddi, D., Djedir, R., Dreher-Duwat, M.L., Dubois, R., Dubuque, C., Ducatillon, P., Duclay, J., Ducrocq, B., Ducrot, F., Ducrotte, P., Dufilho, A., Duhamel, C., Dujardin, D., Dumant-Forest, C., Dupas, J.L., Dupont, F., Durantou, Y., Duriez, A., El Achkar, K., El Farisi, M., Elie, C., Elie-Legrand, M.C., Elkhaki, A., Eoche, M., Evrard, D., Evrard, J.P., Fatome, A., Filoche, B., Finet, L., Flahaut, M., Flamme, C.,

Foissey, D., Fournier, P., Foutrein-Comes, M.C., Foutrein, P., Fremond, D., Frere, T., Fumery, M., Gallet, P., Gamblin, C., Ganga-Zandzou, P.S., Gérard, R., Geslin, G., Gheysens, Y., Ghossini, N., Ghrib, S., Gilbert, T., Gillet, B., Godard, D., Godard, P., Godchaux, J.M., Godchaux, R., Goegebeur, G., Gorla, O., Gottrand, F., Gower, P., Grandmaison, B., Groux, M., Guedon, C., Guillard, J.F., Guillem, L., Guillemot, F., Guimber, D., Haddouche, B., Hakim, S., Hanon, D., Hautefeuille, V., Heckestweiller, P., Hecquet, G., Hedde, J.P., Hellal, H., Henneresse, P.E., Heyman, B., Heraud, M., Herve, S., Hochain, P., Houssin-Bailly, L., Houcke, P., Huguenin, B., Iobagiu, S., Ivanovic, A., Iwanicki-Caron, I., Janicki, E., Jarry, M., Jeu, J., Joly, J.P., Jonas, C., Katherin, F., Kerleveo, A., Khachfe, A., Kiriakos, A., Kiriakos, J., Klein, O., Kohut, M., Kornhauser, R., Koutsomanis, D., Laberene, J.E., Laffineur, G., Lagarde, M., Lannoy, P., Lapchin, J., Lapprand, M., Laude, D., Leblanc, R., Lecieux, P., Leclerc, N., Le Couteux, C., Ledent, J., Lefebvre, J., Lefiliatre, P., Legrand, C., Le Grix, A., Lelong, P., Leluyer, B., Lenaerts, C., Lepilleur, L., Leplat, A., Lepoutre-Dujardin, E., Leroi, H., Leroy, M.Y., Lesage, J.P., Lesage, X., Lesage, J., Lescanne-Darchis, I., Lescut, J., Lescut, D., Leurent, B., Levy, P., Lhermie, M., Lion, A., Lisambert, B., Loire, F., Louf, S., Louvet, A., Luciani, M., Lucidarme, D., Lugand, J., Macaigne, O., Maetz, D., Maillard, D., Mancheron, H., Manolache, O., Marks-Brunel, A.B., Marti, R., Martin, F., Martin, G., Marzloff, E., Mathurin, P., Mauillon, J., Maunoury, V., Maupas, J.L., Mesnard, B., Metayer, P., Methari, L., Meurisse, B., Meurisse, F., Michaud, L., Mirmaran, X., Modaine, P., Monthe, A., Morel, L., Mortier, P.E., Moulin, E., Mouterde, O., Mudry, J., Nachury, M., Khac, N.E., Notteghem, B., Ollevier, V., Ostyn, A., Ouraghi, A., Ouvry, D., Paillot, B., Panien-Claudot, N., Paoletti, C., Papazian, A., Parent, B., Pariente, B., Paris, J.C., Patrier, P., Paupart, L., Pauwels, B., Pauwels, M., Petit, R., Piat, M., Piotte, S., Plane, C., Plouvier, B., Pollet, E., Pommelet, P., Pop, D., Pordes, C., Pouchain, G., Prades, P., Prevost, A., Prevost, J.C., Quesnel, B., Queuniet, A.M., Quinton, J.F., Rabache, A., Rabelle, P., Raclot, G., Ratajczyk, S., Rault, D., Razemon, V., Reix, N., Revillon, M., Richez, C., Robinson, P., Rodriguez, J., Roger, J., Roux, J.M., Rudelli, A., Saber, A., Savoye, G., Schlosseberg, P., Segrestin, M., Seguy, D., Serin, M., Seryer, A., Sevenet, F., Shekh, N., Silvie, J., Simon, V., Spyckerelle, C., Talbodec, N., Techy, A., Thelu, J.L., Thevenin, A., Thiebault, H., Thomas, J., Thorel, J.M., Tielman, G., Tode, M., Toisin, J., Tonnel, J., Touchais, J.Y., Touze, Y., Tranvouez, J.L., Triplet, C., Turck, D., Uhlen, S., Vaillant, E., Valmage, C., Vanco, D., Vandamme, H., Vanderbecq, E., Eecken, V.E., Vandermolen, P., Vandevenne, P., Vandeville, L., Vandewalle, A., Vandewalle, C., Vaneslander, P., Vanhoove, J.P., Vanrenterghem, A., Varlet, P., Vasies, I., Verbiese, G., Vernier-Massouille, G., Vermelle, P., Verne, C., Vezilier-Cocq, P., Vigneron, B., Vincendet, M., Viot, J., Voiment, Y.M., Wacrenier, A., Waeghemaecker, L., Wallez, J.Y., Wantiez, M., Wartel, F., Weber, J., Willocquet, J.L., Wizla, N., Wolschies, E., Zalar, A., Zaouri, B., Zellweger, A., Ziade, C., Group, T.E., 2018. Dramatic Increase in Incidence of Ulcerative Colitis and Crohn's Disease (1988–2011): A Population-Based Study of French Adolescents. *Official journal of the American College of Gastroenterology* | ACG 113, 265–272.

<https://doi.org/10.1038/ajg.2017.228>

Golge, M., Yenilmez, F., Aksoy, A., 2013. Development of pollution indices for the middle section of the Lower Seyhan Basin (Turkey). *Ecological Indicators* 29, 6–17.

<https://doi.org/10.1016/j.ecolind.2012.11.021>

Gower-Rousseau, C., Salomez, J.L., Dupas, J.L., Marti, R., Nuttens, M.C., Votte, A., Lemahieu, M., Lemaire, B., Colombel, J.F., Cortot, A., 1994. Incidence of inflammatory bowel disease in northern France (1988-1990). *Gut* 35, 1433–1438. <https://doi.org/10.1136/gut.35.10.1433>

Gower-Rousseau, C., Vasseur, F., Fumery, M., Savoye, G., Salleron, J., Dauchet, L., Turck, D., Cortot, A., Peyrin-Biroulet, L., Colombel, J.F., 2013. Epidemiology of inflammatory bowel diseases: New insights from a French population-based registry (EPIMAD). *Digestive and Liver Disease* 45, 89–94. <https://doi.org/10.1016/j.dld.2012.09.005>

- Gray, C.L., Lobdell, D.T., Rappazzo, K.M., Jian, Y., Jagai, J.S., Messer, L.C., Patel, A.P., DeFlorio-Barker, S.A., Lyttle, C., Solway, J., Rzhetsky, A., 2018. Associations between environmental quality and adult asthma prevalence in medical claims data. *Environmental Research* 166, 529–536. <https://doi.org/10.1016/j.envres.2018.06.020>
- Greco, S., Ishizaka, A., Resce, G., Torrìsi, G., 2020. Measuring well-being by a multidimensional spatial model in OECD Better Life Index framework. *Socio-Economic Planning Sciences* 70, 100684. <https://doi.org/10.1016/j.seps.2019.01.006>
- Green, C., Elliott, L., Beaudoin, C., Bernstein, C.N., 2006. A Population-based Ecologic Study of Inflammatory Bowel Disease: Searching for Etiologic Clues. *American Journal of Epidemiology* 164, 615–623. <https://doi.org/10.1093/aje/kwj260>
- Guetté, A., Carruthers-Jones, J., Carver, S.J., 2021. *Projet CARTNAT Cartographie de la Naturalité (Notice technique)*.
- Guetté, A., Carruthers-Jones, J., Godet, L., Robin, M., 2018. « Naturalité » : concepts et méthodes appliqués à la conservation de la nature. *Cybergeo: European Journal of Geography*. <https://doi.org/10.4000/cybergeo.29140>
- Habran, S., Philippart, C., Jacquemin, P., Remy, S., 2022. Mapping agricultural use of pesticides to enable research and environmental health actions in Belgium. *Environmental Pollution* 301, 119018. <https://doi.org/10.1016/j.envpol.2022.119018>
- Hamroun, A., Génin, M., Glowacki, F., De Courrèges, A., Dauchet, L., Gauthier, V., Amouyel, P., Couchoud, C., Lassalle, M., Occelli, F., 2023a. #4724 MULTIPLE AIR POLLUTANT EXPOSURE AND RISK OF ALL-CAUSE MORTALITY IN DIALYSIS PATIENTS: A FRENCH REGISTRY-BASED NATIONWIDE STUDY. *Nephrology Dialysis Transplantation* 38, gfad063b_4724. https://doi.org/10.1093/ndt/gfad063b_4724
- Hamroun, A., Génin, M., Glowacki, F., De Courrèges, A., Dauchet, L., Gauthier, V., Amouyel, P., Couchoud, C., Lassalle, M., Occelli, F., 2023b. #4724 MULTIPLE AIR POLLUTANT EXPOSURE AND RISK OF ALL-CAUSE MORTALITY IN DIALYSIS PATIENTS: A FRENCH REGISTRY-BASED NATIONWIDE STUDY. *Nephrology Dialysis Transplantation* 38, gfad063b_4724. https://doi.org/10.1093/ndt/gfad063b_4724
- Han, X., Liu, X., Gao, D., Ma, B., Gao, X., Cheng, M., 2022. Costs and benefits of the development methods of drinking water quality index: A systematic review. *Ecological Indicators* 144, 109501. <https://doi.org/10.1016/j.ecolind.2022.109501>
- Hanski, I., von Hertzen, L., Fyhrquist, N., Koskinen, K., Torppa, K., Laatikainen, T., Karisola, P., Auvinen, P., Paulin, L., Mäkelä, M.J., Vartiainen, E., Kosunen, T.U., Alenius, H., Haahtela, T., 2012. Environmental biodiversity, human microbiota, and allergy are interrelated. *Proceedings of the National Academy of Sciences* 109, 8334–8339. <https://doi.org/10.1073/pnas.1205624109>
- Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28, 100–108. <https://doi.org/10.2307/2346830>
- Hellström, L., Elinder, C.-G., Dahlberg, B., Lundberg, M., Järup, L., Persson, B., Axelson, O., 2001. Cadmium exposure and end-stage renal disease. *American Journal of Kidney Diseases* 38, 1001–1008. <https://doi.org/10.1053/ajkd.2001.28589>
- Higgs, G., Langford, M., 2009. GIScience, environmental justice, & estimating populations at risk: The case of landfills in Wales. *Applied Geography* 29, 63–76. <https://doi.org/10.1016/j.apgeog.2008.07.006>
- Hommel, K., Rasmussen, S., Kamper, A.-L., Madsen, M., 2010. Regional and social inequalities in chronic renal replacement therapy in Denmark. *Nephrology Dialysis Transplantation* 25, 2624–2632. <https://doi.org/10.1093/ndt/gfq110>

- Horálek, J., Schreiberová, M., Vlasáková, L., Ďoubalová, J., Tognet, F., Schneider, P., Kurfürst, P., Schovánková, J., 2021. European air quality maps for 2018. <https://doi.org/10.5281/zenodo.4638651>
- Hu, D., Ren, J., Wang, G., Gu, G., Liu, S., Wu, X., Chen, J., Ren, H., Hong, Z., Li, J., 2014. Geographic Mapping of Crohn's Disease and Its Relation to Affluence in Jiangsu Province, an Eastern Coastal Province of China. *Gastroenterology Research and Practice* 2014, e590467. <https://doi.org/10.1155/2014/590467>
- Huang, Y., Li, T., Wu, C., He, Z., Japenga, J., Deng, M., Yang, X., 2015. An integrated approach to assess heavy metal source apportionment in peri-urban agricultural soils. *Journal of Hazardous Materials* 299, 540–549. <https://doi.org/10.1016/j.jhazmat.2015.07.041>
- Hugot, J.P., Chamaillard, M., Zouali, H., Lesage, S., Cézard, J.P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C.A., Gassull, M., Binder, V., Finkel, Y., Cortot, A., Modigliani, R., Laurent-Puig, P., Gower-Rousseau, C., Macry, J., Colombel, J.F., Sahbatou, M., Thomas, G., 2001. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 411, 599–603. <https://doi.org/10.1038/35079107>
- INERIS, 2018. Caractérisation des inégalités environnementales : inventaire des bases de données nationales environnementales et spatialisées. Mise à jour dans le cadre de la préparation du PNSE 4. INERIS.
- INERIS, 2016. Caractérisation des inégalités environnementales : inventaire des bases de données nationales environnementales et spatialisées. INERIS.
- Jacenic, D., Cygankiewicz, A.I., Mokrowiecka, A., Małecká-Panas, E., Fichna, J., Krajewska, W.M., 2019. Sex- and Age-Related Estrogen Signaling Alteration in Inflammatory Bowel Diseases: Modulatory Role of Estrogen Receptors. *International Journal of Molecular Sciences* 20, 3175. <https://doi.org/10.3390/ijms20133175>
- Jagai, J.S., Krajewski, A.K., Shaikh, S., Lobdell, D.T., Sargis, R.M., 2020. Association between environmental quality and diabetes in the USA. *Journal of Diabetes Investigation* 11, 315–324. <https://doi.org/10.1111/jdi.13152>
- Jagai, J.S., Messer, L.C., Rappazzo, K.M., Gray, C.L., Grabich, S.C., Lobdell, D.T., 2017. County-level cumulative environmental quality associated with cancer incidence. *Cancer* 123, 2901–2908. <https://doi.org/10.1002/cncr.30709>
- Jenks, G., 1967. The Data Model Concept in Statistical Mapping. *International Yearbook of Cartography* 7, 186–190.
- Ji, G., Chen, Q., Ding, Z., Gu, J., Guo, M., Shi, L., Yu, H., Sun, H., 2022. High mortality and high PCDD/Fs exposure among residents downwind of municipal solid waste incinerators: A case study in China. *Environmental Pollution* 294, 118635. <https://doi.org/10.1016/j.envpol.2021.118635>
- Jia, P., Cao, X., Yang, H., Dai, S., He, P., Huang, G., Wu, T., Wang, Y., 2021. Green space access in the neighbourhood and childhood obesity. *Obesity Reviews* 22. <https://doi.org/10.1111/obr.13100>
- Jin, Y., Wu, S., Zeng, Z., Fu, Z., 2017. Effects of environmental pollutants on gut microbiota. *Environmental Pollution* 222, 1–9. <https://doi.org/10.1016/j.envpol.2016.11.045>
- Jolliffe, I.T., 1972. Discarding Variables in a Principal Component Analysis. I: Artificial Data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 21, 160–173. <https://doi.org/10.2307/2346488>
- Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., Essers, J., Mitrovic, M., Ning, K., Cleynen, I., Theatre, E., Spain, S.L., Raychaudhuri, S., Goyette, P., Wei, Z., Abraham, C., Achkar, J.-P., Ahmad, T., Amininejad, L., Ananthakrishnan, A.N., Andersen, V., Andrews, J.M., Baidoo, L., Balschun, T., Bampton, P.A., Bitton, A., Boucher, G., Brand, S., Büning, C., Cohain, A., Cichon, S., D'Amato, M., De Jong, D., Devaney, K.L., Dubinsky, M., Edwards, C., Ellinghaus, D., Ferguson, L.R.,

- Franchimont, D., Fransen, K., Geary, R., Georges, M., Gieger, C., Glas, J., Haritunians, T., Hart, A., Hawkey, C., Hedl, M., Hu, X., Karlsen, T.H., Kupcinskas, L., Kugathasan, S., Latiano, A., Laukens, D., Lawrance, I.C., Lees, C.W., Louis, E., Mahy, G., Mansfield, J., Morgan, A.R., Mowat, C., Newman, W., Palmieri, O., Ponsioen, C.Y., Potocnik, U., Prescott, N.J., Regueiro, M., Rotter, J.I., Russell, R.K., Sanderson, J.D., Sans, M., Satsangi, J., Schreiber, S., Simms, L.A., Sventoraityte, J., Targan, S.R., Taylor, K.D., Tremelling, M., Verspaget, H.W., De Vos, M., Wijmenga, C., Wilson, D.C., Winkelmann, J., Xavier, R.J., Zeissig, S., Zhang, B., Zhang, C.K., Zhao, H., International IBD Genetics Consortium (IBDGC), Silverberg, M.S., Annese, V., Hakonarson, H., Brant, S.R., Radford-Smith, G., Mathew, C.G., Rioux, J.D., Schadt, E.E., Daly, M.J., Franke, A., Parkes, M., Vermeire, S., Barrett, J.C., Cho, J.H., 2012. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 491, 119–124. <https://doi.org/10.1038/nature11582>
- Kaiser, H.F., 1958. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 23, 187–200. <https://doi.org/10.1007/BF02289233>
- Kanchan, K., Gorai, A., Goyal, P., 2015. A Review on Air Quality Indexing System. *Asian Journal of Atmospheric Environment* 9, 101–113. <https://doi.org/10.5572/ajae.2015.9.2.101>
- King, J.R., Jackson, D.A., 1999. Variable selection in large environmental data sets using principal components analysis. *Environmetrics* 10, 67–77. [https://doi.org/10.1002/\(SICI\)1099-095X\(199901/02\)10:1<67::AID-ENV336>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1099-095X(199901/02)10:1<67::AID-ENV336>3.0.CO;2-0)
- Knox, E., 2000. Childhood cancers, birthplaces, incinerators and landfill sites. *International Journal of Epidemiology* 29, 391–397. <https://doi.org/10.1093/ije/29.3.391>
- Kolenikov, S., Angeles, G., 2004. The Use of Discrete Data in PCA: Theory, Simulations, and Applications to Socioeconomic Indices. University of North Carolina 60.
- Kowalska, J.B., Mazurek, R., Gąsiorek, M., Zaleski, T., 2018. Pollution indices as useful tools for the comprehensive evaluation of the degree of soil contamination—A review. *Environ Geochem Health* 40, 2395–2420. <https://doi.org/10.1007/s10653-018-0106-z>
- Kudsk, P., Jørgensen, L.N., Ørum, J.E., 2018. Pesticide Load—A new Danish pesticide risk indicator with multiple applications. *Land Use Policy* 70, 384–393. <https://doi.org/10.1016/j.landusepol.2017.11.010>
- Kulldorff, M., 1997. A spatial scan statistic. *Communications in Statistics - Theory and Methods* 26, 1481–1496. <https://doi.org/10.1080/03610929708831995>
- Lanier, C., Deram, A., Cuny, M.-A., Cuny, D., Occelli, F., 2019a. Spatial analysis of environmental inequalities caused by multiple air pollutants: A cumulative impact screening method, applied to the north of France. *Ecological Indicators* 99, 91–100. <https://doi.org/10.1016/j.ecolind.2018.12.011>
- Lanier, C., Deram, A., Cuny, M.-A., Cuny, D., Occelli, F., 2019b. Spatial analysis of environmental inequalities caused by multiple air pollutants: A cumulative impact screening method, applied to the north of France. *Ecological Indicators* 99, 91–100. <https://doi.org/10.1016/j.ecolind.2018.12.011>
- Lebov, J.F., Engel, L.S., Richardson, D., Hogan, S.L., Hoppin, J.A., Sandler, D.P., 2016. Pesticide use and risk of end-stage renal disease among licensed pesticide applicators in the Agricultural Health Study. *Occup Environ Med* 73, 3–12. <https://doi.org/10.1136/oemed-2014-102615>
- Lerner, A., 2007. Aluminum Is a Potential Environmental Factor for Crohn's Disease Induction. *Annals of the New York Academy of Sciences* 1107, 329–345. <https://doi.org/10.1196/annals.1381.035>
- Lerner, A., Jeremias, P., Matthias, T., 2016. The World Incidence and Prevalence of Autoimmune Diseases is Increasing. *IJCD* 3, 151–155. <https://doi.org/10.12691/ijcd-3-4-8>

- Li, C., 2022. China's multi-dimensional ecological well-being performance evaluation: A new method based on coupling coordination model. *Ecological Indicators* 143, 109321. <https://doi.org/10.1016/j.ecolind.2022.109321>
- Li, X., Sundquist, J., Hamano, T., Sundquist, K., 2019. Neighborhood Deprivation and Risks of Autoimmune Disorders: A National Cohort Study in Sweden. *Int J Environ Res Public Health* 16, 3798. <https://doi.org/10.3390/ijerph16203798>
- Lilliefors, H.W., 1967. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of the American Statistical Association* 62, 399–402. <https://doi.org/10.1080/01621459.1967.10482916>
- Lim, K.K., Kwan, Y.H., Tan, C.S., Low, L.L., Chua, A.P., Lee, W.Y., Pang, L., Tay, H.Y., Chan, S.Y., Ostbye, T., 2017. The association between distance to public amenities and cardiovascular risk factors among lower income Singaporeans. *Preventive Medicine Reports* 8, 116–121. <https://doi.org/10.1016/j.pmedr.2017.09.004>
- Lin, S.-Y., Ju, S.-W., Lin, C.L., Hsu, W.-H., Lin, C.-C., Ting, I.-W., Kao, C.-H., 2020. Air pollutants and subsequent risk of chronic kidney disease and end-stage renal disease: A population-based cohort study. *Environmental Pollution* 261, 114154. <https://doi.org/10.1016/j.envpol.2020.114154>
- Liu, J., Kang, H., Tao, W., Li, H., He, D., Ma, L., Tang, H., Wu, S., Yang, K., Li, X., 2023. A spatial distribution – Principal component analysis (SD-PCA) model to assess pollution of heavy metals in soil. *Science of The Total Environment* 859, 160112. <https://doi.org/10.1016/j.scitotenv.2022.160112>
- Lochhead, P., Hart, J.E., Ananthakrishnan, A., Khalili, H., Laden, F., Chan, A.T., 2018. 58 - Air Pollution and Risk of Inflammatory Bowel Diseases. *Gastroenterology* 154, S-20. [https://doi.org/10.1016/S0016-5085\(18\)30547-X](https://doi.org/10.1016/S0016-5085(18)30547-X)
- Luque, S., Pettorelli, N., Vihervaara, P., Wegmann, M., 2018. Improving biodiversity monitoring using satellite remote sensing to provide solutions towards the 2020 conservation targets. *Methods in Ecology and Evolution* 9, 1784–1786. <https://doi.org/10.1111/2041-210X.13057>
- Madhulatha, T.S., 2012. An overview on clustering methods. *IOSRJEN* 02, 719–725. <https://doi.org/10.9790/3021-0204719725>
- Marot, G., 2019. Le bassin minier : un territoire densément peuplé confronté à diverses fragilités sociales., *Insee Flash Hauts-de-France*. Insee.
- Martin, T.D., Chan, S.S.M., Hart, A.R., 2015. Environmental Factors in the Relapse and Recurrence of Inflammatory Bowel Disease: A Review of the Literature. *Dig Dis Sci* 60, 1396–1405. <https://doi.org/10.1007/s10620-014-3437-3>
- Marzi, S., Mysiak, J., Santato, S., 2018. Comparing adaptive capacity index across scales: The case of Italy. *Journal of Environmental Management* 223, 1023–1036. <https://doi.org/10.1016/j.jenvman.2018.06.060>
- McCormick, R., 2017. Does Access to Green Space Impact the Mental Well-being of Children: A Systematic Review. *Journal of Pediatric Nursing* 37, 3–7. <https://doi.org/10.1016/j.pedn.2017.08.027>
- McKight, P.E., Najab, J., 2010. Kruskal-Wallis Test, in: *The Corsini Encyclopedia of Psychology*. pp. 1–1. <https://doi.org/10.1002/9780470479216.corpsy0491>
- Meyer, A., Drouin, J., Weill, A., Carbonnel, F., Dray-Spira, R., 2020. Pregnancy in women with inflammatory bowel disease: a French nationwide study 2010-2018. *Alimentary Pharmacology & Therapeutics* 52, 1480–1490. <https://doi.org/10.1111/apt.16074>
- Michel, P., St-Onge, L., Lowe, A.-M., Bigras-Poulin, M., Brassard, P., 2010. Geographical variation of Crohn's disease residual incidence in the Province of Quebec, Canada. *International Journal of Health Geographics* 9, 22. <https://doi.org/10.1186/1476-072X-9-22>

- Ministère des Solidarités et de la Santé, 2023. Résultats du contrôle sanitaire de l'eau du robinet - data.gouv.fr [WWW Document]. URL <https://www.data.gouv.fr/fr/datasets/resultats-du-controle-sanitaire-de-leau-du-robinet/> (accessed 3.16.23).
- Mitchell, R., Popham, F., 2008. Effect of exposure to natural environment on health inequalities: an observational population study. *The Lancet* 372, 1655–1660. [https://doi.org/10.1016/S0140-6736\(08\)61689-X](https://doi.org/10.1016/S0140-6736(08)61689-X)
- Morton, R.L., Schlackow, I., Mihaylova, B., Staplin, N.D., Gray, A., Cass, A., 2016. The impact of social disadvantage in moderate-to-severe chronic kidney disease: an equity-focused systematic review*. *Nephrology Dialysis Transplantation* 31, 46–56. <https://doi.org/10.1093/ndt/gfu394>
- Mullin, K., Mitchell, G., Nawaz, N.R., Waters, R.D., 2018. Natural capital and the poor in England: Towards an environmental justice analysis of ecosystem services in a high income country. *Landscape and Urban Planning* 176, 10–21. <https://doi.org/10.1016/j.landurbplan.2018.03.022>
- Nair, N., Austin, C., Rocha, M., Gouveia, C., Curtin, P., Eisele, C., Colombel, J., Torres, J., Peter, I., Arora, M., 2019. P756 Assessment of metal exposures in deciduous teeth of patients with inflammatory bowel disease. *Journal of Crohn's and Colitis* 13, S499–S500. <https://doi.org/10.1093/ecco-jcc/jjy222.880>
- Nan, N., Yan, Z., Zhang, Y., Chen, R., Qin, G., Sang, N., 2023. Overview of PM2.5 and health outcomes: Focusing on components, sources, and pollutant mixture co-exposure. *Chemosphere* 323, 138181. <https://doi.org/10.1016/j.chemosphere.2023.138181>
- Nerich, V., Monnet, E., Weill, A., Vallier, N., Vanbockstael, V., Auleley, G.-R., Balaire, C., Dubost, P., Rican, S., Allemand, H., Carbonnel, F., 2010. Fine-scale geographic variations of inflammatory bowel disease in France: Correlation with socioeconomic and house equipment variables. *Inflammatory Bowel Diseases* 16, 813–821. <https://doi.org/10.1002/ibd.21122>
- Ng, S.C., Bernstein, C.N., Vatn, M.H., Lakatos, P.L., Loftus, E.V., Tysk, C., O'Morain, C., Moum, B., Colombel, J.-F., Disease (IOIBD), on behalf of the E. and N.H.T.F. of the I.O. of I.B., 2013. Geographical variability and environmental risk factors in inflammatory bowel disease. *Gut* 62, 630–649. <https://doi.org/10.1136/gutjnl-2012-303661>
- Nie, X., Xie, R., Tuo, B., 2018. Effects of Estrogen on the Gastrointestinal Tract. *Dig Dis Sci* 63, 583–596. <https://doi.org/10.1007/s10620-018-4939-1>
- Occelli, F., Bavdek, R., Deram, A., Hellequin, A.-P., Cuny, M.-A., Zwarterook, I., Cuny, D., 2016. Using lichen biomonitors to assess environmental justice at a neighbourhood level in an industrial area of Northern France. *Ecological Indicators* 60, 781–788. <https://doi.org/10.1016/j.ecolind.2015.08.026>
- Occelli, F., Deram, A., Génin, M., Noël, C., Cuny, D., Glowacki, F., Network, on behalf of the N., 2014. Mapping End-Stage Renal Disease (ESRD): Spatial Variations on Small Area Level in Northern France, and Association with Deprivation. *PLOS ONE* 9, e110132. <https://doi.org/10.1371/journal.pone.0110132>
- Occelli, F., Lanier, C., Cuny, D., Deram, A., Dumont, J., Amouyel, P., Montaye, M., Dauchet, L., Dallongeville, J., Genin, M., 2020a. Exposure to multiple air pollutants and the incidence of coronary heart disease: A fine-scale geographic analysis. *Science of The Total Environment* 714, 136608. <https://doi.org/10.1016/j.scitotenv.2020.136608>
- Occelli, F., Lanier, C., Cuny, D., Deram, A., Dumont, J., Amouyel, P., Montaye, M., Dauchet, L., Dallongeville, J., Genin, M., 2020b. Exposure to multiple air pollutants and the incidence of coronary heart disease: A fine-scale geographic analysis. *Science of The Total Environment* 714, 136608. <https://doi.org/10.1016/j.scitotenv.2020.136608>
- OECD, European Union, Joint Research Centre - European Commission, 2008a. Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD. <https://doi.org/10.1787/9789264043466-en>

- OECD, Union, E., Commission, J.R.C.-E., 2008b. Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD Publishing.
- Ogura, Y., Bonen, D.K., Inohara, N., Nicolae, D.L., Chen, F.F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R.H., Achkar, J.P., Brant, S.R., Bayless, T.M., Kirschner, B.S., Hanauer, S.B., Nuñez, G., Cho, J.H., 2001. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 411, 603–606. <https://doi.org/10.1038/35079114>
- OMS, 2022. Noncommunicable diseases: progress monitor 2022. OMS.
- OMS, 2021. Air quality guidelines.
- Opstelten, J.L., Beelen, R.M.J., Leenders, M., Hoek, G., Brunekreef, B., van Schaik, F.D.M., Siersema, P.D., Eriksen, K.T., Raaschou-Nielsen, O., Tjønneland, A., Overvad, K., Boutron-Ruault, M.-C., Carbonnel, F., de Hoogh, K., Key, T.J., Luben, R., Chan, S.S.M., Hart, A.R., Bueno-de-Mesquita, H.B., Oldenburg, B., 2016. Exposure to Ambient Air Pollution and the Risk of Inflammatory Bowel Disease: A European Nested Case–Control Study. *Dig Dis Sci* 61, 2963–2971. <https://doi.org/10.1007/s10620-016-4249-4>
- Orholm, M., Munkholm, P., Langholz, E., Nielsen, O.H., Sørensen, T.I., Binder, V., 1991. Familial occurrence of inflammatory bowel disease. *N Engl J Med* 324, 84–88. <https://doi.org/10.1056/NEJM199101103240203>
- Owusu, C., Flanagan, B., Lavery, A.M., Mertzlufft, C.E., McKenzie, B.A., Kolling, J., Lewis, B., Dunn, I., Hallisey, E., Lehnert, E.A., Fletcher, K., Davis, R.T., Conn, M., Owen, L.R., Smith, M.M., Dent, A., 2022. Developing a granular scale environmental burden index (EBI) for diverse land cover types across the contiguous United States. *Science of The Total Environment* 838, 155908. <https://doi.org/10.1016/j.scitotenv.2022.155908>
- Panés, J., Gomollón, F., Taxonera, C., Hinojosa, J., Clofent, J., Nos, V., 2007. Crohn's Disease. *Drugs* 67, 2511–2537. <https://doi.org/10.2165/00003495-200767170-00005>
- Paranjpe, I., Chaudhary, K., Paranjpe, M., O'Hagan, R., Manna, S., Jaladanki, S., Kapoor, A., Horowitz, C., DeFelice, N., Cooper, R., Glicksberg, B., Bottinger, E.P., Just, A.C., Nadkarni, G.N., 2020. Association of APOL1 Risk Genotype and Air Pollution for Kidney Disease. *Clin J Am Soc Nephrol* 15, 401–403. <https://doi.org/10.2215/CJN.11921019>
- Parisse, S., Antoni, V., Aouir, C., Arnaud, M., Bonnet, A., Besancon, M., Bonnard, P., Boughaba, J., Colas, S., Denoyer, G., Dubois, A., Duvernoy, J., Fèvre, C., Garcia, B., Gendre, C., Ghewy, X., Giroux, G., Grevet, A., Grassart, L., Gomez, S., Hébrail, V., Hirschler, O., Irz, P., Jédor, B., Moullec, A.L., Leenhardt, S., Lemarquis, D., Levêque, A., Lorge, A., Magnier, C., Molinié, L., Morard, V., Pasquier, J.-L., Pautard, E., Pons, C., Siniscalco, S., Soleilhavoup, M., Triquenot, A., Thual, J., Vey, F., Veyrac, C., Aillery, F., 2018. Environnement & agriculture - Les chiffres clés. Ministère de l'agriculture et de l'alimentation.
- Paumelle, M., Occelli, F., Wakim, L.M., Brousmiche, D., Bouhadj, L., Ternynck, C., Lanier, C., Cuny, D., Deram, A., 2023. Description of the multi-dimensional environment at the territorial scale: A holistic framework using cluster analysis and open data in France. *Ecological Indicators* 154, 110562. <https://doi.org/10.1016/j.ecolind.2023.110562>
- Pecher, C., Tasser, E., Walde, J., Tappeiner, U., 2013. Typology of Alpine region using spatial-pattern indicators. *Ecological Indicators* 24, 37–47. <https://doi.org/10.1016/j.ecolind.2012.05.025>
- Phavichitr, N., Cameron, D.J.S., Catto-Smith, A., 2003. Increasing incidence of Crohn's disease in Victorian children. *Journal of Gastroenterology and Hepatology* 18, 329–332. <https://doi.org/10.1046/j.1440-1746.2003.02975.x>
- PNSE4, 2021. Un environnement, une santé : 4e Plan National Santé Environnement. Ministère de la transition écologique.
- Pointereau, P., Coulon, F., Jiguet, F., Doxa, A., Paracchini, M.-L., Terres, J.-M., 2010. Les systèmes agricoles à haute valeur naturelle en France métropolitaine 16.

- Popov Aleksandrov, A., Mirkov, I., Tucovic, D., Kulas, J., Zeljkovic, M., Popovic, D., Ninkov, M., Jankovic, S., Kataranovski, M., 2021. Immunomodulation by heavy metals as a contributing factor to inflammatory diseases and autoimmune reactions: Cadmium as an example. *Immunology Letters* 240, 106–122. <https://doi.org/10.1016/j.imlet.2021.10.003>
- Poux, X., Pointereau, P., 2014. L'agriculture à "haute valeur naturelle" en France métropolitaine. Un indicateur pour le suivi de la biodiversité et l'évaluation de la politique de développement rural. Rapport d'étude au Ministère de l'agriculture de l'agroalimentaire et de la forêt.
- Praene, J.P., Malet-Damour, B., Radanielina, M.H., Fontaine, L., Rivière, G., 2019. GIS-based approach to identify climatic zoning: A hierarchical clustering on principal component analysis. *Building and Environment* 164, 106330. <https://doi.org/10.1016/j.buildenv.2019.106330>
- Prüss-Üstün, A., Wolf, J., Corvalán, C., Organization, W.H., Bos, R., Neira, D.M., 2016. Preventing Disease Through Healthy Environments: A Global Assessment of the Burden of Disease from Environmental Risks. World Health Organization.
- QGIS Development Team, 2021. QGIS Geographic Information System.
- R Core Team, 2021. R: A language and environment for statistical computing.
- Rajbhandari, R., Blakemore, S., Gupta, N., Adler, A.J., Noble, C.A., Mannan, S., Nikolli, K., Yih, A., Joshi, S., Bukhman, G., 2020. Crohn's disease in low and lower-middle income countries: A scoping review. *World J Gastroenterol* 26, 6891–6908. <https://doi.org/10.3748/wjg.v26.i43.6891>
- Rappazzo, K.M., Messer, L.C., Jagai, J.S., Gray, C.L., Grabich, S.C., Lobdell, D.T., 2015. The associations between environmental quality and preterm birth in the United States, 2000–2005: a cross-sectional analysis. *Environmental Health* 14, 50. <https://doi.org/10.1186/s12940-015-0038-3>
- Reiss, F., 2013. Socioeconomic inequalities and mental health problems in children and adolescents: A systematic review. *Social Science & Medicine* 90, 24–31. <https://doi.org/10.1016/j.socscimed.2013.04.026>
- Rey, G., Jouglu, E., Fouillet, A., Hémon, D., 2009. Ecological association between a deprivation index and mortality in France over the period 1997 – 2001: variations with spatial scale, degree of urbanicity, age, gender and cause of death. *BMC Public Health* 9, 33. <https://doi.org/10.1186/1471-2458-9-33>
- Richardson, E.A., Mitchell, R., 2010. Gender differences in relationships between urban green space and health in the United Kingdom. *Social Science & Medicine* 71, 568–575. <https://doi.org/10.1016/j.socscimed.2010.04.015>
- Richardson, M.J., Madrigal, D.S., Wilkie, A., Wong, M., Roberts, E., 2017. Environmental Health Tracking Improves Pesticide Use Data to Enable Research and Inform Public Health Actions in California. *Journal of Public Health Management and Practice* 23, S97–S104.
- Riedo, J., Wächter, D., Gubler, A., Wettstein, F.E., Meuli, R.G., Bucheli, T.D., 2023. Pesticide residues in agricultural soils in light of their on-farm application history. *Environmental Pollution* 331, 121892. <https://doi.org/10.1016/j.envpol.2023.121892>
- Roderick, P., Clements, S., Stone, N., Martin, D., Diamond, I., 1999. What Determines Geographical Variation in Rates of Acceptance onto Renal Replacement Therapy in England? *J Health Serv Res Policy* 4, 139–146. <https://doi.org/10.1177/135581969900400304>
- Rook, G.A., 2013a. Regulation of the immune system by biodiversity from the natural environment: An ecosystem service essential to health. *Proceedings of the National Academy of Sciences* 110, 18360–18367. <https://doi.org/10.1073/pnas.1313731110>
- Rook, G.A., 2013b. Regulation of the immune system by biodiversity from the natural environment: an ecosystem service essential to health. *Proc Natl Acad Sci U S A* 110, 18360–18367. <https://doi.org/10.1073/pnas.1313731110>
- Rosansky, S.J., Huntsberger, T.L., Jackson, K., Eggers, P., 1990. Comparative Incidence Rates of End-Stage Renal Disease Treatment by State. *AJN* 10, 198–204. <https://doi.org/10.1159/000168081>

- Sabath, E., Robles-Osorio, M.L., 2012. Renal health and the environment: heavy metal nephrotoxicity. *Nefrología (English Edition)* 32, 279–286. <https://doi.org/10.3265/Nefrologia.pre2012.Jan.10928>
- Saib, M.-S., Caudeville, J., Beauchamp, M., Carré, F., Ganry, O., Trugeon, A., Cicolella, A., 2015. Building spatial composite indicators to analyze environmental health inequalities on a regional scale. *Environmental Health* 14, 68. <https://doi.org/10.1186/s12940-015-0054-3>
- Sanmarco, L.M., Chao, C.-C., Wang, Y.-C., Kenison, J.E., Li, Z., Rone, J.M., Rejano-Gordillo, C.M., Polonio, C.M., Gutierrez-Vazquez, C., Piester, G., Plasencia, A., Li, L., Giovannoni, F., Lee, H.-G., Faust Akl, C., Wheeler, M.A., Mascanfroni, I., Jaronen, M., Alsuwailm, M., Hewson, P., Yeste, A., Andersen, B.M., Franks, D.G., Huang, C.-J., Ekwudo, M., Tjon, E.C., Rothhammer, V., Takenaka, M., de Lima, K.A., Linnerbauer, M., Guo, L., Covacu, R., Queva, H., Fonseca-Castro, P.H., Bladi, M.A., Cox, L.M., Hodgetts, K.J., Hahn, M.E., Mildner, A., Korzenik, J., Hauser, R., Snapper, S.B., Quintana, F.J., 2022. Identification of environmental factors that promote intestinal inflammation. *Nature* 611, 801–809. <https://doi.org/10.1038/s41586-022-05308-6>
- Santé Publique France, 2021. Impact de la pollution de l'air ambiant sur la mortalité en France métropolitaine : réduction en lien avec le confinement du printemps 2020 et nouvelles données sur le poids total pour la période 2016-2019.
- Segessemann, A., Crevoisier, O., 2016. Beyond Economic Base Theory: The Role of the Residential Economy in Attracting Income to Swiss Regions. *Regional Studies* 50, 1388–1403. <https://doi.org/10.1080/00343404.2015.1018882>
- Shaw, C., Pitcher, D., Pruthi, R., Fogarty, D., 2013. UK Renal Registry 16th annual report: chapter 2 UK RRT prevalence in 2012: national and centre-specific analyses. *Nephron Clin Pract* 125, 29–53. <https://doi.org/10.1159/000360021>
- Sheridan, S.C., Lin, S., 2014. Assessing Variability in the Impacts of Heat on Health Outcomes in New York City Over Time, Season, and Heat-Wave Duration. *EcoHealth* 11, 512–525. <https://doi.org/10.1007/s10393-014-0970-7>
- Shortt, N.K., Richardson, E.A., Mitchell, R., Pearce, J., 2010. Re-engaging with the physical environment: a health-related environmental classification of the UK: Re-engaging with the physical environment. *Area* no-no. <https://doi.org/10.1111/j.1475-4762.2010.00960.x>
- Shortt, N.K., Richardson, E.A., Pearce, J., Mitchell, R.J., 2012. Mortality inequalities by environment type in New Zealand. *Health & Place* 18, 1132–1136. <https://doi.org/10.1016/j.healthplace.2012.04.008>
- Shubham, S., Kumar, M., Sarma, D.K., Kumawat, M., Verma, V., Samartha, R.M., Tiwari, R.R., 2022. Role of air pollution in chronic kidney disease: an update on evidence, mechanisms and mitigation strategies. *Int Arch Occup Environ Health* 95, 897–908. <https://doi.org/10.1007/s00420-021-01808-6>
- Soon, I.S., Molodecky, N.A., Rabi, D.M., Ghali, W.A., Barkema, H.W., Kaplan, G.G., 2012. The relationship between urban environment and the inflammatory bowel diseases: a systematic review and meta-analysis. *BMC Gastroenterology* 12, 51. <https://doi.org/10.1186/1471-230X-12-51>
- Soulhac, L., Garbero, V., Salizzoni, P., Mejean, P., Perkins, R.J., 2009. Flow and dispersion in street intersections. *Atmospheric Environment* 43, 2981–2996. <https://doi.org/10.1016/j.atmosenv.2009.02.061>
- Sterckeman, T., Douay, F., Baize, D., Fourrier, H., N., P., Schwartz, C., 2002. Référentiel pédo-géochimique du Nord-pas-de-Calais. Méthodes et principaux resultants. *Etude Gest Sols* 14, 153–168.
- Stiglitz, J.E., Sen, A., Fitoussi, J.-P., 2009. Rapport de la commission sur la mesure des performances économiques et du progrès social.

- Sun, Z., Yang, L., Bai, X., Du, W., Shen, G., Fei, J., Wang, Y., Chen, A., Chen, Y., Zhao, M., 2019. Maternal ambient air pollution exposure with spatial-temporal variations and preterm birth risk assessment during 2013–2017 in Zhejiang Province, China. *Environment International* 133, 105242. <https://doi.org/10.1016/j.envint.2019.105242>
- Système d'Information sur l'Eau, Office Français de la Biodiversité, 2023. Achats de pesticides par code postal - data.gouv.fr [WWW Document]. URL <https://www.data.gouv.fr/fr/datasets/achats-de-pesticides-par-code-postal/> (accessed 3.16.23).
- Tanner, R.M., Gutiérrez, O.M., Judd, S., McClellan, W., Bowling, C.B., Bradbury, B.D., Safford, M.M., Cushman, M., Warnock, D., Muntner, P., 2013. Geographic Variation in CKD Prevalence and ESRD Incidence in the United States: Results From the Reasons for Geographic and Racial Differences in Stroke (REGARDS) Study. *American Journal of Kidney Diseases* 61, 395–403. <https://doi.org/10.1053/j.ajkd.2012.10.018>
- Tenailleau, Q.M., Lanier, C., Gower-Rousseau, C., Cuny, D., Deram, A., Occelli, F., 2020. Crohn's disease and environmental contamination: Current challenges and perspectives in exposure evaluation. *Environmental Pollution* 263, 114599. <https://doi.org/10.1016/j.envpol.2020.114599>
- Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63, 411–423. <https://doi.org/10.1111/1467-9868.00293>
- Townsend, P., 1987. Deprivation. *Journal of Social Policy* 16, 125–146. <https://doi.org/10.1017/S0047279400020341>
- Tran, L.T., O'Neill, R.V., Smith, E.R., 2010. Spatial pattern of environmental vulnerability in the Mid-Atlantic region, USA. *Applied Geography* 30, 191–202. <https://doi.org/10.1016/j.apgeog.2009.05.003>
- Tsai, C.-C., Wu, C.-L., Kor, C.-T., Lian, I.-B., Chang, C.-H., Chang, T.-H., Chang, C.-C., Chiu, P.-F., 2018. Prospective associations between environmental heavy metal exposure and renal outcomes in adults with chronic kidney disease. *Nephrology* 23, 830–836. <https://doi.org/10.1111/nep.13089>
- Turner, K.G., Odgaard, M.V., Bøcher, P.K., Dalgaard, T., Svenning, J.-C., 2014. Bundling ecosystem services in Denmark: Trade-offs and synergies in a cultural landscape. *Landscape and Urban Planning* 125, 89–104. <https://doi.org/10.1016/j.landurbplan.2014.02.007>
- Tysk, C., Lindberg, E., Järnerot, G., Flodérus-Myrhed, B., 1988. Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut* 29, 990–996. <https://doi.org/10.1136/gut.29.7.990>
- United States Environmental Protection Agency, 2020. Environmental Quality Index 2006-2010 Technical Report. United States Environmental Protection Agency (USEPA).
- U.S. Environmental Protection Agency, 2020. Environmental Quality Index - Technical Report (2006-2010) (Final, 2020). [WWW Document]. URL <https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=350545> (accessed 4.30.23).
- Usami, T., Koyama, K., Takeuchi, O., Morozumi, K., Kimura, G., 2000. Regional Variations in the Incidence of End-Stage Renal Failure in Japan. *JAMA* 284, 2622–2624. <https://doi.org/10.1001/jama.284.20.2622>
- USEPA, 1999. Guideline for reporting of daily air quality—air quality index (AQI) (No. EPA-454/R-99-010.). U.S. Environmental Protection Agency.
- van de Luijngaarden, M.W.M., Noordzij, M., Tomson, C., Couchoud, C., Cancarini, G., Ansell, D., Bos, W.-J.W., Dekker, F.W., Gorriz, J.L., Iatrou, C., Garneata, L., Wanner, C., Cala, S., Stojceva-Taneva, O., Finne, P., Stel, V.S., van Biesen, W., Jager, K.J., 2012. Factors Influencing the Decision to Start Renal Replacement Therapy: Results of a Survey Among European

- Nephrologists. *American Journal of Kidney Diseases* 60, 940–948.
<https://doi.org/10.1053/j.ajkd.2012.07.015>
- Van Kruiningen, H.J., Joossens, M., Vermeire, S., Joossens, S., Debeugny, S., Gower-Rousseau, C., Cortot, A., Colombel, J.-F., Rutgeerts, P., Vlietinck, R., 2005. Environmental Factors in Familial Crohn's Disease in Belgium. *Inflammatory Bowel Diseases* 11, 360–365.
<https://doi.org/10.1097/01.MIB.0000158536.31557.90>
- Villoslada, M., Bunce, R.G.H., Sepp, K., Jongman, R.H.G., Metzger, M.J., Kull, T., Raet, J., Kuusemets, V., Kull, A., Leito, A., 2017. A framework for habitat monitoring and climate change modelling: construction and validation of the Environmental Stratification of Estonia. *Reg Environ Change* 17, 335–349. <https://doi.org/10.1007/s10113-016-1002-7>
- Wakai, K., Nakai, S., Kikuchi, K., Iseki, K., Miwa, N., Masakane, I., Wada, A., Shinzato, T., Nagura, Y., Akiba, T., 2004. Trends in incidence of end-stage renal disease in Japan, 1983–2000: age-adjusted and age-specific rates by gender and cause. *Nephrology Dialysis Transplantation* 19, 2044–2052. <https://doi.org/10.1093/ndt/gfh317>
- Walker, C., Allamneni, C., Orr, J., Yun, H., Fitzmorris, P., Xie, F., Malik, T.A., 2018. Socioeconomic Status and Race are both Independently associated with Increased Hospitalization Rate among Crohn's Disease Patients. *Sci Rep* 8, 4028. <https://doi.org/10.1038/s41598-018-22429-z>
- Wang, C.-C., Zhang, Q.-C., Yan, C.-A., Tang, G.-Y., Zhang, M.-Y., Ma, L.Q., Gu, R.-H., Xiang, P., 2023. Heavy metal(loid)s in agriculture soils, rice, and wheat across China: Status assessment and spatiotemporal analysis. *Science of The Total Environment* 882, 163361.
<https://doi.org/10.1016/j.scitotenv.2023.163361>
- Wang, R., Gamon, J.A., 2019. Remote sensing of terrestrial plant biodiversity. *Remote Sensing of Environment* 231, 111218. <https://doi.org/10.1016/j.rse.2019.111218>
- Ward, J.H., 1963. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 58, 236–244. <https://doi.org/10.1080/01621459.1963.10500845>
- Ward, M.H., Nuckols, J.R., Weigel, S.J., Maxwell, S.K., Cantor, K.P., Miller, R.S., 2000. Identifying populations potentially exposed to agricultural pesticides using remote sensing and a Geographic Information System. *Environ Health Perspect* 108, 5–12.
<https://doi.org/10.1289/ehp.001085>
- Wardle, R.A., Wardle, A.J., Charadva, C., Ghosh, S., Moran, G.W., 2017. Literature review: impacts of socioeconomic status on the risk of inflammatory bowel disease and its outcomes. *European Journal of Gastroenterology & Hepatology* 29, 879–884.
<https://doi.org/10.1097/MEG.0000000000000899>
- Wasana, H.M.S., Aluthpatabendi, D., Kularatne, W.M.T.D., Wijekoon, P., Weerasooriya, R., Bandara, J., 2016. Drinking water quality and chronic kidney disease of unknown etiology (CKDu): synergic effects of fluoride, cadmium and hardness of water. *Environ Geochem Health* 38, 157–168. <https://doi.org/10.1007/s10653-015-9699-7>
- Weinand, J.M., McKenna, R., Fichtner, W., 2019. Developing a municipality typology for modelling decentralised energy systems. *Utilities Policy* 57, 75–96.
<https://doi.org/10.1016/j.jup.2019.02.003>
- Wendt, C., 2009. Mapping European healthcare systems: a comparative analysis of financing, service provision and access to healthcare. *Journal of European Social Policy* 19, 432–445.
<https://doi.org/10.1177/0958928709344247>
- Wimmer, F., Oberaigner, W., Kramar, R., Mayer, G., 2003. Regional variability in the incidence of end-stage renal disease: an epidemiological approach. *Nephrology Dialysis Transplantation* 18, 1562–1567. <https://doi.org/10.1093/ndt/gfg184>

- Wood, S.L., Demougin, P.R., Higgins, S., Husk, K., Wheeler, B.W., White, M., 2016. Exploring the relationship between childhood obesity and proximity to the coast: A rural/urban perspective. *Health & Place* 40, 129–136. <https://doi.org/10.1016/j.healthplace.2016.05.010>
- Wu, C.-D., Chern, Y.-R., Pan, W.-C., Lung, S.-C.C., Yao, T.-C., Tsai, H.-J., Spengler, J.D., 2020a. Effects of surrounding environment on incidence of end stage renal disease. *Science of The Total Environment* 723, 137915. <https://doi.org/10.1016/j.scitotenv.2020.137915>
- Wu, C.-D., Chern, Y.-R., Pan, W.-C., Lung, S.-C.C., Yao, T.-C., Tsai, H.-J., Spengler, J.D., 2020b. Effects of surrounding environment on incidence of end stage renal disease. *Science of The Total Environment* 723, 137915. <https://doi.org/10.1016/j.scitotenv.2020.137915>
- Wu, G., Cai, M., Wang, C., Zou, H., Wang, X., Hua, J., Lin, H., 2023. Ambient air pollution and incidence, progression to multimorbidity and death of hypertension, diabetes, and chronic kidney disease: A national prospective cohort. *Science of The Total Environment* 881, 163406. <https://doi.org/10.1016/j.scitotenv.2023.163406>
- Wu, J., Wen, X.W., Faulk, C., Boehnke, K., Zhang, H., Dolinoy, D.C., Xi, C., 2016. Perinatal Lead Exposure Alters Gut Microbiota Composition and Results in Sex-specific Bodyweight Increases in Adult Mice. *Toxicological Sciences* 151, 324–333. <https://doi.org/10.1093/toxsci/kfw046>
- Wu, M.-Y., Lo, W.-C., Chao, C.-T., Wu, M.-S., Chiang, C.-K., 2020. Association between air pollutants and development of chronic kidney disease: A systematic review and meta-analysis. *Science of The Total Environment* 706, 135522. <https://doi.org/10.1016/j.scitotenv.2019.135522>
- Xu, P., Chen, Z., Wu, L., Chen, Y., Xu, D., Shen, H., Han, J., Wang, X., Lou, X., 2019. Health risk of childhood exposure to PCDD/Fs emitted from a municipal waste incinerator in Zhejiang, China. *Science of The Total Environment* 689, 937–944. <https://doi.org/10.1016/j.scitotenv.2019.06.425>
- Xu, X., Nie, S., Ding, H., Hou, F.F., 2018. Environmental pollution and kidney diseases. *Nat Rev Nephrol* 14, 313–324. <https://doi.org/10.1038/nrneph.2018.11>
- Yang, B.-Y., Fan, S., Thiering, E., Seissler, J., Nowak, D., Dong, G.-H., Heinrich, J., 2020. Ambient air pollution and diabetes: A systematic review and meta-analysis. *Environmental Research* 180, 108817. <https://doi.org/10.1016/j.envres.2019.108817>
- Yang, B.-Y., Qian, Z., Howard, S.W., Vaughn, M.G., Fan, S.-J., Liu, K.-K., Dong, G.-H., 2018. Global association between ambient air pollution and blood pressure: A systematic review and meta-analysis. *Environmental Pollution* 235, 576–588. <https://doi.org/10.1016/j.envpol.2018.01.001>
- Yang, Yang, Ruan, Z., Wang, X., Yang, Yin, Mason, T.G., Lin, H., Tian, L., 2019. Short-term and long-term exposures to fine particulate matter constituents and health: A systematic review and meta-analysis. *Environmental Pollution* 247, 874–882. <https://doi.org/10.1016/j.envpol.2018.12.060>
- Ye, J.-J., Wang, S.-S., Fang, Y., Zhang, X.-J., Hu, C.-Y., 2021. Ambient air pollution exposure and risk of chronic kidney disease: A systematic review of the literature and meta-analysis. *Environmental Research* 195, 110867. <https://doi.org/10.1016/j.envres.2021.110867>
- Zhang, X., Holt, J.B., Lu, H., Onufrak, S., Yang, J., French, S.P., Sui, D.Z., 2014. Neighborhood commuting environment and obesity in the United States: An urban–rural stratified multilevel analysis. *Preventive Medicine* 59, 31–36. <https://doi.org/10.1016/j.ypmed.2013.11.004>
- Zhang, Z., Chen, L., Qian, Z. (Min), Li, H., Cai, M., Wang, X., McMillin, S.E., Vaughn, M.G., Liu, K., Shao, Z., Lin, H., 2022. Residential green and blue space associated with lower risk of adult-onset inflammatory bowel disease: Findings from a large prospective cohort study. *Environment International* 160, 107084. <https://doi.org/10.1016/j.envint.2022.107084>

- Zoccali, C., Kramer, A., Jager, K.J., 2010. Chronic kidney disease and end-stage renal disease—a review produced to contribute to the report ‘the status of health in the European union: towards a healthier Europe.’ *NDT Plus* 3, 213–224. <https://doi.org/10.1093/ndtplus/sfp127>
- Zoeteman, K., Mommaas, H., Dagevos, J., 2016. Are larger cities more sustainable? Lessons from integrated sustainability monitoring in 403 Dutch municipalities. *Environmental Development* 17, 57–72. <https://doi.org/10.1016/j.envdev.2015.08.003>

ANNEXE 1 : Tableau de présentation des indicateurs retenus pour la construction des indices de vulnérabilité
et de résilience

ECONOMIE (10 indicateurs)

	Acronyme	Nom de l'indicateur	Précisions éventuelles	Unité	Source(s)	Période temporelle	Données manquantes (%)	Donnée calculée OU donnée brute	Médiane [Q1;Q3] OU modalités (%)
ECONOMIE	TX_CR_ENT	Taux de création d'entreprises	<i>Calculé par l'observatoire des territoires à partir du répertoire des entreprises et des établissements (REE-Sirene).</i>	Pourcentage	INSEE Observatoire des territoires Répertoire des entreprises et établissements	2018	1,85	Donnée brute	12,5 [5,7;20]
	V_EMP_0813	Evolution de la population active de 15 à 64 ans		Pourcentage	INSEE	Entre 2008 et 2013	0,00	Donnée calculée	3,36 [-2,26;10,29]
	TX_EMP	Taux d'emploi des 15-64 ans	<i>Taux d'emploi au sens du recensement de la population, proportion de personnes disposant d'un emploi (population active occupée) parmi celles en âge de travailler (population des 15 à 64 ans).</i>	Pourcentage	INSEE Observatoire des territoires	2017	0,86	Donnée brute	67,1 [62,8;70,7]
	REV_MED_UC_13	Revenu médian par unité de consommation		€/unité de consommation	INSEE Observatoire des territoires	2013	0,35	Donnée brute	19752 [18130;21501]
	EV_POP_ACT_0914	Evolution de la part des actifs occupés de 15 ans et plus parmi la population totale		Pourcentage	INSEE	Entre 2009 et 2014	0,00	Donnée calculée	-1,3 [-5,04;3,02]
	EV_EMPL_PR	Evolution de l'emploi pour les activités présentesielles	<i>Les activités présentesielles sont les activités mises en œuvre localement pour la production de biens et de services visant la satisfaction des besoins de personnes résidentes ou touristes de la zone</i>	Pourcentage	INSEE	Entre 2009 et 2014	0,82	Donnée calculée	0 [-0,26;0,33]
	SHANNON_CSP	Indice de Shannon-Weaver de représentation des catégories socio-professionnelles	<i>Utilisation de la méthodologie de l'indice de Shannon Weaver pour mesurer la diversité des catégories socio-professionnelles représentées.</i>	Indice	INSEE	2014	0,86	Donnée calculée	0,76 [0,51;0,94]
	T_CHOM	Taux de chômage parmi les 15-64 ans	<i>Taux de chômage au sens du recensement de la population, c'est-à-dire le rapport entre le nombre de chômeurs et la population active.</i>	Pourcentage	INSEE	2017	0,86	Donnée brute	11,1 [8,5;14,5]
	T_PREC	Taux de précarité parmi les 15-64 ans	<i>Part des salariés de 15 et plus en CDD, intérim, emplois aidés et en apprentissage parmi le nombre de personnes salariées de 15 ans et plus.</i>	Pourcentage	INSEE	2014	0,00	Donnée calculée	5,49 [0;10,83]
	P_AC_OCC_MMCOM	Part des actifs occupés de 15 ans et plus vivant et travaillant dans la même commune		Pourcentage	INSEE	2014	0,00	Donnée calculée	15,57 [11,82;20,51]

ENVIRONNEMENT (15 indicateurs)

	Acronyme	Nom de l'indicateur	Précisions éventuelles	Unité	Source(s)	Période temporelle	Données manquantes (%)	Donnée calculée OU donnée brute	Médiane [Q1;Q3] OU modalités (%)
ENVIRONNEMENT	AC_EV	Part de la population communale ayant accès à un espace vert dans un rayon de 400 mètres	Utilisation de la couche « Land Use » construite par Open Street Map (version 2020). Dans cette couche « Land Use », les catégories suivantes ont été retenues pour représenter les espaces verts : « forest », « park » et « recreation ground ».	Pourcentage	INSEE Open Street Map	2020	0,00	Donnée calculée	62,55 [22,46;93,23]
	NAT_HAB	m² d'espaces verts ou espaces naturels par habitant	Utilisation de trois couches Open Street Map : couches « Land Use », « Points of Interest » et « Bodies of Water ». Différentes catégories retenues au sein de ces couches : « forest », « park », « recreation_ground », « cemetery », « grass », « playground », « golf course », « stadium », « camp site », « zoo », « water », « river », « wetland ».	m² par habitant	INSEE Open Street Map	2020	0,00	Donnée calculée	1109,58 [210,81;3715,44]
	H_ESV	Monétarisation par habitant des services écosystémiques	Utilisation de la couche Corine Land Cover et création de quatre grandes catégories d'occupation du sol : urbain, terres cultivées et/ou arables, prairies et forêts. Attribution d'une valeur monétaire par hectare aux services écosystémiques fournis par ces quatre catégories d'occupation du sol.	USD/ha/habitant	INSEE, Corine Land Cover	2012	0,00	Donnée calculée	5,6 [2,61;11,1]
	Emis_GES	Emission des gaz à effet de serre hors puits (PRG)	Le pouvoir de réchauffement global (PRG) exprime en milliers de tonnes équivalent CO2, l'effet cumulé des substances émises dans l'air qui contribuent à l'accroissement de l'effet de serre. Le PRG est exprimé ici sans tenir compte des puits de carbone.	Milliers de teq CO2	Observatoire des territoires	2012	0,75	Donnée brute	4 [3;8]
	BASIAS	Densité de sites BASIAS par commune		nombre/ha	Base de données BASIAS	2020	0,00	Donnée calculée	0,0027 [0;0,0092]
	BASOL	Densité de sites BASOL par commune		nombre/ha	Base de données BASOL	2020	0,00	Donnée calculée	0 [0;0]
	ICPE	Densité de sites ICPE par commune		nombre/ha	Base de données S3IC	2020	0,00	Donnée calculée	0 [0;0,0017]
	INTERSEC	Densité d'intersections routières par commune		nombre/ha	BD TOPO (IGN)	2020	0,00	Donnée calculée	0,059 [0,038;0,101]
	P_ARTI	Proportion de territoires artificialisés		Proportion	Corine Land Cover	2018	0,00	Donnée calculée	0,06 [0,03;0,1]
	P_FOR_NAT	Proportion de forêts et milieux semi-naturels		Proportion	Corine Land Cover	2018	0,00	Donnée calculée	0,06 [0;0,17]
	P_HUMID	Proportion de zones humides		Proportion	Corine Land Cover	2018	0,00	Donnée calculée	0 [0;0]
	ARTIF_0919	Pourcentage de la surface communale convertie en surface artificialisée entre 2009 et 2019.	Calcul réalisé par l'observatoire de l'artificialisation à partir des Fichiers fonciers. Il s'agit du rapport entre le flux d'artificialisation et la surface communale. Le flux d'artificialisation correspond au nombre d'hectares passant du statut "Naturel, Agricole ou Forestier" à "Artificialisé" sur une période donnée.	Pourcentage	Fichiers Fonciers, Observatoire national de l'artificialisation	Entre 2009 et 2019	0,86	Donnée brute	0,31 [0,11;0,72]
	NO2_MEAN	Moyenne interannuelle de la pollution atmosphérique en NO2		µg/m³	Agence Européenne pour l'environnement	2011 à 2017	0,00	Donnée calculée	9,1 [8,2;10,51]
	PM10_MEAN	Moyenne interannuelle de la pollution atmosphérique en PM10		µg/m³	Agence Européenne pour l'environnement	2011 à 2017	0,00	Donnée calculée	18,31 [17,4;18,84]
	PM25_MEAN	Moyenne interannuelle de la pollution atmosphérique en PM2,5		µg/m³	Agence Européenne pour l'environnement	2011 à 2017	0,00	Donnée calculée	12,19 [11,72;12,69]

SOCIAL (11 indicateurs)

Acronyme	Nom de l'indicateur	Précisions éventuelles	Unité	Source(s)	Période temporelle	Données manquantes (%)	Donnée calculée OU donnée brute	Médiane [Q1;Q3] OU modalités (%)
TXSCO_1824	Taux de poursuite d'études des 18-24 ans		Pourcentage	INSEE	2013	0,55	Donnée calculée	39,16 [29,68;49,53]
EMP_ED_FORM_14	Emploi dans l'éducation et la formation pour 1000 habitants		Nombre pour 1000 habitant	INSEE	2014	0,00	Donnée calculée	0 [0;11,56]
NB_EC_PUB_1000H	Nombre d'écoles publiques pour 1000 habitants	<i>Ecoles maternelles publiques, écoles élémentaires publiques et écoles élémentaires publiques avec classe pré-élémentaire.</i>	Nombre pour 1000 habitants	INSEE BPE GEOIDD	2014	0,00	Donnée calculée	1,51 [0;2,87]
PEU_DIPL_1564	Part des peu diplômés parmi les 15-64 ans	<i>Une personne peu diplômée est ici définie comme une personne non scolarisée de 15 ans ou plus titulaire d'aucun diplôme ou au plus d'un BEPC, brevet des collèges ou DNB.</i>	Pourcentage	INSEE	2014	0,24	Donnée calculée	33 [28;39]
P_COUV_AL	Part de la population couverte par une prestation de la branche famille	<i>Par exemple les allocations familiales, la PAJE (prestation d'accueil du jeune enfant), le complément familial, allocation d'éducation de l'enfant handicapé, allocation de soutien familial, allocation de rentrée scolaire, allocation de logement à caractère familial, allocation de logement à caractère social.</i>	Pourcentage	CNAF	2014	1,85	Donnée calculée	44,74 [39,76;49,71]
P_MONOP	Part des familles monoparentales parmi les ménages	<i>Une famille monoparentale comprend un parent isolé et un ou plusieurs enfants célibataires (n'ayant pas d'enfant).</i>	Pourcentage	Observatoire des territoires	2012	1,10	Donnée brute	15,4 [8,35;22,4]
Densite_pop	Densité de population		Nombre d'habitants par kilomètre carré	INSEE	2017	0,88	Donnée brute	61,8 [34,3;131,9]
GRIL_COM_DENS	Grille communale de densité	<i>La grille communale de densité s'appuie sur la distribution de la population à l'intérieur de la commune en découpant le territoire en carreaux d'un kilomètre de côté. Elle repère ainsi des zones agglomérées. C'est l'importance de ces zones agglomérées au sein des communes qui va permettre de les caractériser (et non la densité communale habituelle)</i>	Catégories : 1 : espaces très peu denses 2 : espaces peu denses 3 : espaces de densité intermédiaires 4 : espaces densément peuplés	INSEE Observatoire des territoires	2019	0,88	Donnée brute	1 : 25,91% 2 : 58,61% 3 : 12,64% 4 : 2,84%
P_PERS_plus65	Part de la population de plus de 65 ans dans la population totale		Pourcentage	INSEE	2015	0,35	Donnée calculée	16,33 [13,59;19,3]
FDEP_17	Indice de défaveur sociale Fdep	<i>Indicateur spatialisé du désavantage social spécifiquement adapté aux études de santé sur la population Française. Indice composite construit à partir du taux de chômage, du taux d'ouvriers, du taux de bacheliers et du revenu médian par ménage.</i>	Nombre	INSEE	2017	0,00	Donnée calculée	0,02 [-1,01;1,04]
P_IMM_15	Part des immigrés dans la population totale	<i>Selon la définition adoptée par le Haut Conseil à l'Intégration, un immigré est une personne née étrangère à l'étranger et résidant en France.</i>	Pourcentage	INSEE	2015	0,35	Donnée calculée	1,44 [0,75;2,59]

SANTE (6 indicateurs)

	Acronyme	Nom de l'indicateur	Précisions éventuelles	Unité	Source(s)	Période temporelle	Données manquantes (%)	Donnée calculée OU donnée brute	Médiane [Q1;Q3] OU modalités (%)
SANTÉ	APL_MG_2013	Accessibilité potentielle localisée aux médecins généralistes	<i>L'indicateur d'accessibilité potentielle localisée (APL) vise à mesurer l'accessibilité des populations à l'offre de soins. Cet indicateur repose sur trois variables : la demande, l'offre, la distance. Il est calculé à l'échelle communale en prenant en compte l'offre et la demande des communes voisines.</i>	Equivalents temps plein pour 100 000 habitants	Direction de la recherche, des études, de l'évaluation et des statistiques (DREES)	2013	0,02	Donnée brute	48,07 [32,76;67,29]
	APL_PHAR_2013	Accessibilité potentielle localisée aux pharmaciens	<i>L'indicateur d'accessibilité potentielle localisée (APL) vise à mesurer l'accessibilité des populations à l'offre de soins. Cet indicateur repose sur trois variables : la demande, l'offre, la distance. Il est calculé à l'échelle communale en prenant en compte l'offre et la demande des communes voisines.</i>	Equivalents temps plein pour 100 000 habitants	Direction de la recherche, des études, de l'évaluation et des statistiques (DREES)	2013	0,02	Donnée brute	22,41 [15,17;32,11]
	P_SANTE_1000H	Nombre de professionnels de santé pour 1000 habitants	<i>Nombre de professionnels de santé toutes spécialités confondues.</i>	Nombre pour 1000 habitants	INSEE	2016	0,24	Donnée calculée	0 [0;2,99]
	P_NAISS_DOM_HAB_14	Part des naissances domiciliées au domicile de la mère par habitant	<i>Les naissances domiciliées correspondent aux naissances dénombrées à partir des bulletins d'état civil, et comptabilisées au domicile de la mère.</i>	Pourcentage	INSEE	2014	0,00	Donnée calculée	1,1 [0,74;1,47]
	P_DECES_DOM_HAB_14	Part des décès domiciliés par habitant	<i>Les statistiques annuelles concernent les décès domiciliés ainsi que les jugements déclaratifs de décès. Le lieu retenu pour le décès est le domicile du défunt et non pas le lieu du décès.</i>	Pourcentage	INSEE	2014	0,37	Donnée calculée	0,72 [0,41;1,07]
	SHAN_MED	Indice de Shannon-Weaver pour les professions médicales et paramédicales	<i>Utilisation de la méthodologie de l'indice de Shannon Weaver pour mesurer la diversité des professions médicales représentées dans la commune.</i>	Indice	INSEE	2016	0,02	Donnée calculée	0 [0;0,27]

SERVICES (7 indicateurs)

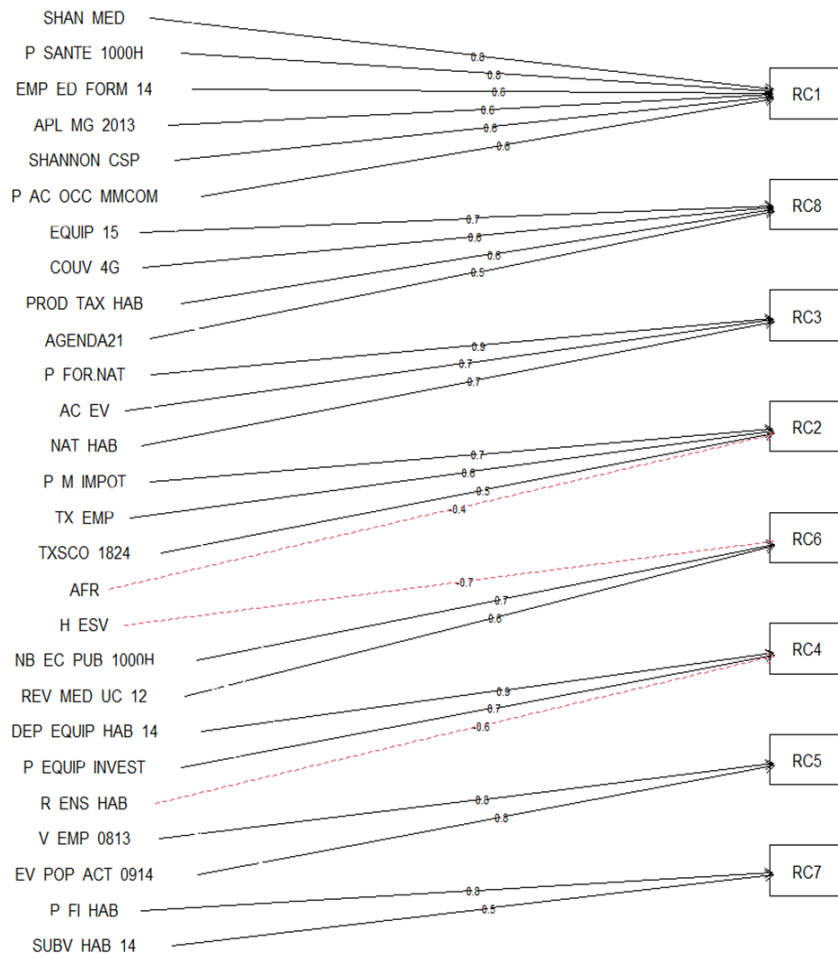
	Acronyme	Nom de l'indicateur	Précisions éventuelles	Unité	Source(s)	Période temporelle	Données manquantes (%)	Donnée calculée OU donnée brute	Médiane [Q1;Q3] OU modalités (%)
SERVICES	EQUIP_15	Part de la population ayant accès en moyenne aux 12 équipements de la gamme intermédiaire en 15 minutes ou moins	<i>Utilisation du distancier METRIC de l'INSEE, et repose sur la classification des gammes d'équipements de l'INSEE. Calcul d'une accessibilité géographique qui ne prend pas en compte par exemple les temps d'attente ou les transports en commun.</i>	Pourcentage	INSEE Observatoire des territoires	2011	0,86	Donnée brute	94,1 [85,5;99,3]
	COUV_4G	Part de surface couverte en 4G par l'opérateur mieux disant		Pourcentage	ARCEP	2016	0,00	Donnée brute	91 [56;99]
	HLM	Nombre de résidences principales HLM pour 1000 habitants		Nombre pour 1000 habitants	INSEE	2013	0,00	Donnée calculée	0 [0;20,83]
	P_PROP	Part des propriétaires dans les résidences principales		Pourcentage	INSEE	2014	0,18	Donnée calculée	81,68 [74,11;86,83]
	P_LOCA_HLM	Part des locataires en HLM dans les résidences principales		Pourcentage	INSEE	2014	0,24	Donnée calculée	0 [0;5,45]
	P_LOCA	Part des locataires dans les résidences principales (dont HLM)		Pourcentage	INSEE	2014	0,24	Donnée calculée	17,65 [11,9;28,55]
	LNG_RT_H	Longueur des routes primaires par habitant	<i>La route primaire est définie comme une portion de voie de communication destinée aux automobiles, aux piétons, aux cycles ou aux animaux, homogène pour l'ensemble des attributs et des relations qui la concerne. Cette classe est un sous-ensemble de la classe ROUTE, et comprend uniquement les tronçons de route d'importance 1 ou 2. Cela permet de ne prendre en compte que le réseau dit principal.</i>	m par habitant	IGN INSEE	2020	0,00	Donnée calculée	0 [0;4,02]

POLITIQUE (12 indicateurs)

Acronyme	Nom de l'indicateur	Précisions éventuelles	Unité	Source(s)	Période temporelle	Données manquantes (%)	Donnée calculée OU donnée brute	Médiane [Q1;Q3] OU modalités (%)
P_M_IMPOT	Part des ménages fiscaux imposés		Pourcentage	INSEE Observatoire des territoires	2014	0,00	Donnée brute	52,7 [49,3;57,2]
AGENDA21	Part de la population couverte par un agenda 21 communal ou infra-départemental	<i>Ces agendas sont élaborés à l'échelle des collectivités ou à l'échelle infradépartementale, comme mesure du développement durable. Elle traduit d'un engagement volontaire des collectivités dans le développement durable au-delà de ce qui est imposé par la loi dans les plans locaux d'urbanisme (PLU) et les schémas de cohérence territoriale (SCOT).</i>	Catégories : 0 : pas de couverture par un agenda21 1 : couverture par agenda21 communal ou infra-départemental 2 : couverture par agenda21 communal <u>et</u> infra-départemental	GEOIDD MEDDE site agenda21france.org	2014	0,00	Donnée brute	0 : 79,07% 1 : 19,47% 2 : 1,46%
P_FI_HAB	Potentiel financier par habitant	<i>Cette variable permet de mesurer la richesse d'une commune en comptabilisant ses ressources financières issues de sa fiscalité et des aides qu'elle perçoit de l'état.</i>	€/habitant	Observatoire des territoires, Direction Générale des Collectivités Locales INSEE	2017	0,37	Donnée brute	566 [475,5;722]
DEP_EQUIP_HAB_14	Dépenses pour les équipements de la commune, par habitant	<i>Ces dépenses correspondent aux dépenses effectuées par la commune pour des immobilisations corporelles, incorporelles ou des immobilisations en cours (par exemple construction d'une école, d'une ligne de tramway ou d'une installation sportive), rapportées au nombre d'habitants.</i>	€/habitant	Direction Générale des Finances Publiques, INSEE	2014	0,02	Donnée calculée	145,45 [62,05;309,07]
SUBV_HAB_14	Montant des subventions versées par la commune, par habitant		€/habitant	Direction Générale des Finances Publiques, INSEE	2014	0,02	Donnée calculée	13,74 [6,97;25,11]
PROD_TAX_HAB	Produit de la taxe d'habitation rapporté au nombre de résidences principales et de résidences secondaires		€/habitant	Observatoire des territoires	2015	0,20	Donnée brute	453,8 [370,7;551,75]
POT_FISC_H	Potentiel fiscal par habitant	<i>Le potentiel fiscal est un indicateur qui peut permettre de comparer la richesse fiscale potentielle des collectivités les unes par rapport aux autres.</i>	€/habitant	Observatoire des territoires Direction Générale des Collectivités Locales	2017	0,37	Donnée brute	468,5 [377,2;626,45]
R_ENS_HAB	Résultat d'ensemble comptable de la commune, par habitant		€/habitant	Direction Générale des Finances Publiques, INSEE	2014	0,02	Donnée calculée	108,7 [8,64;250]
ANN_DET_H	Annuité de la dette de la commune, par habitant	<i>Elle mesure le poids de l'endettement à moyen et long terme de la commune.</i>	€/habitant	Direction Générale des Finances Publiques, INSEE	2014	0,02	Donnée calculée	44,12 [17,62;76,92]
P_EQUIP_EMP_INVES T	Part des dépenses pour les équipements parmi le total des dépenses d'investissements	<i>Les emplois d'investissement correspondent aux investissements de la commune (équipements, aménagements urbains, ...) et les remboursements d'emprunts.</i>	Pourcentage	Direction Générale des Finances Publiques, INSEE	2014	0,64	Donnée calculée	81,48 [60,71;93,41]
P_SUBV_CH_FONC	Part des subventions versées parmi les charges de fonctionnement	<i>Les charges de fonctionnement comprennent principalement les charges de personnel, les participations obligatoires notamment au service départemental d'incendie et de secours, les intérêts de la dette et les subventions versées.</i>	Pourcentage	Direction Générale des Finances Publiques, INSEE	2014	0,02	Donnée calculée	2,67 [1,45;4,45]
AFR	Zones éligibles au zonage d'aide à finalité régionale	<i>Les communes identifiées dans ce zonage peuvent à ce titre bénéficier d'aides octroyées par les pouvoirs publics pour les entreprises réalisant des investissements et créant des emplois.</i>	Catégories : 0 : non éligible 1 : Partiellement éligible 2 : intégralement éligible	Observatoire des territoires	2014	0,00	Donnée brute	0 : 83,57% 1 : 0,09% 2 : 16,34%

ANNEXE 2 : Interprétation donnée aux composantes principales retenues pour le calcul des indices

INDICE DE RÉSILIENCE



RC1 : Composante liée principalement à des variables mesurant la quantité d'offre de soins : nombre de professionnels de santé pour 1000 habitants, accessibilité potentielle localisée aux médecins généralistes, indice de Shannon pour les professions médicales et paramédicales. La composante semble également assez urbaine au vu des autres liens de corrélation : indice de Shannon de représentation des catégories socio-professionnelles, part des actifs occupés vivant et travaillant dans la même commune et emploi dans l'éducation et la formation pour 1000 habitants.

RC8 : Composante principalement corrélée à COUV 4G (part de la surface communale couverte en 4G par l'opérateur mieux disant) et EQUIP 15 (part de la population ayant accès en moyenne aux 12 équipements de la gamme intermédiaire en 15 minutes ou moins). Cette composante est donc à priori très urbaine mais reste difficile à caractériser.

RC3 : Composante liée à l'accès aux espaces verts et espaces naturels. Elle est fortement corrélée avec les variables : proportion de forêts et milieux semi-naturels, m² d'espaces naturels par habitant, part de la population communale ayant accès à un espace vert dans un rayon de 400 mètres.

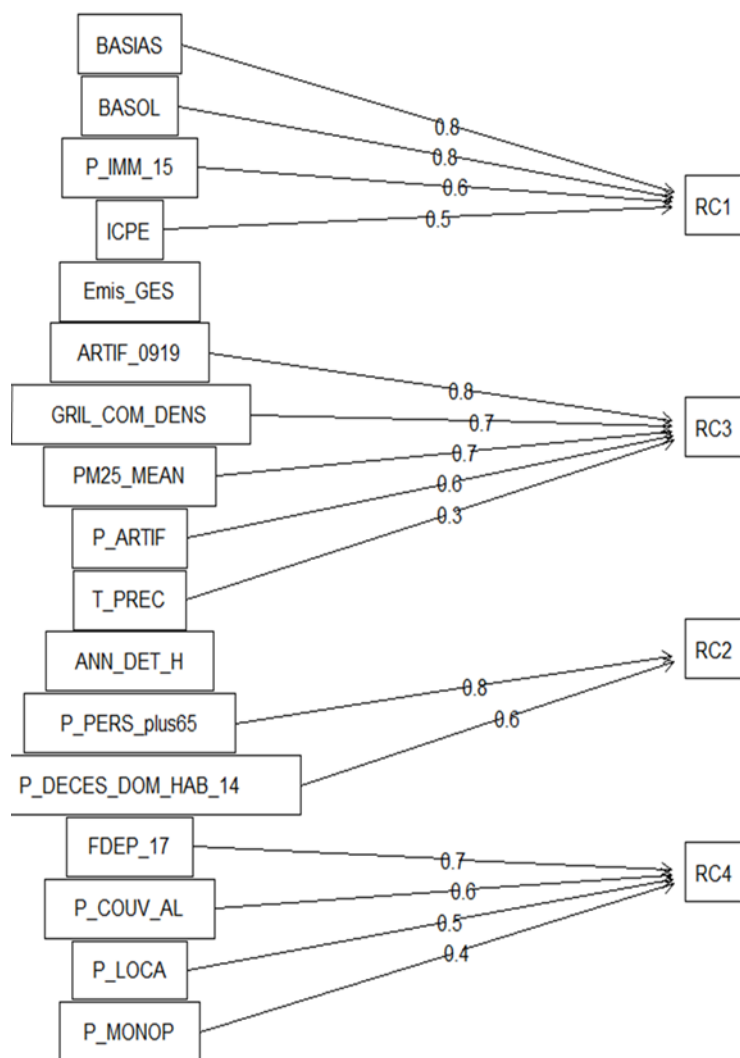
RC2 : Composante liée au dynamisme économique et au niveau d'éducation.

RC6 : Corrélée négativement à H_ESV (monétarisation par habitant des services écosystémiques) et corrélée positivement au nombre d'écoles publiques pour 1000 habitants et au revenu médian par unité de consommation. Pour l'instant difficile d'émettre des hypothèses sur sa signification.

RC4 : Composante corrélée positivement aux dépenses communales pour les équipements et de manière assez logique corrélée négativement aux résultats d'ensemble comptable des communes.

RC5 : Composante reliée à l'évolution de la population active.

RC7 : Composante liée à la richesse de la commune, corrélée au potentiel financier par habitant et au montant des subventions versées par habitant.



INDICE DE VULNÉRABILITÉ

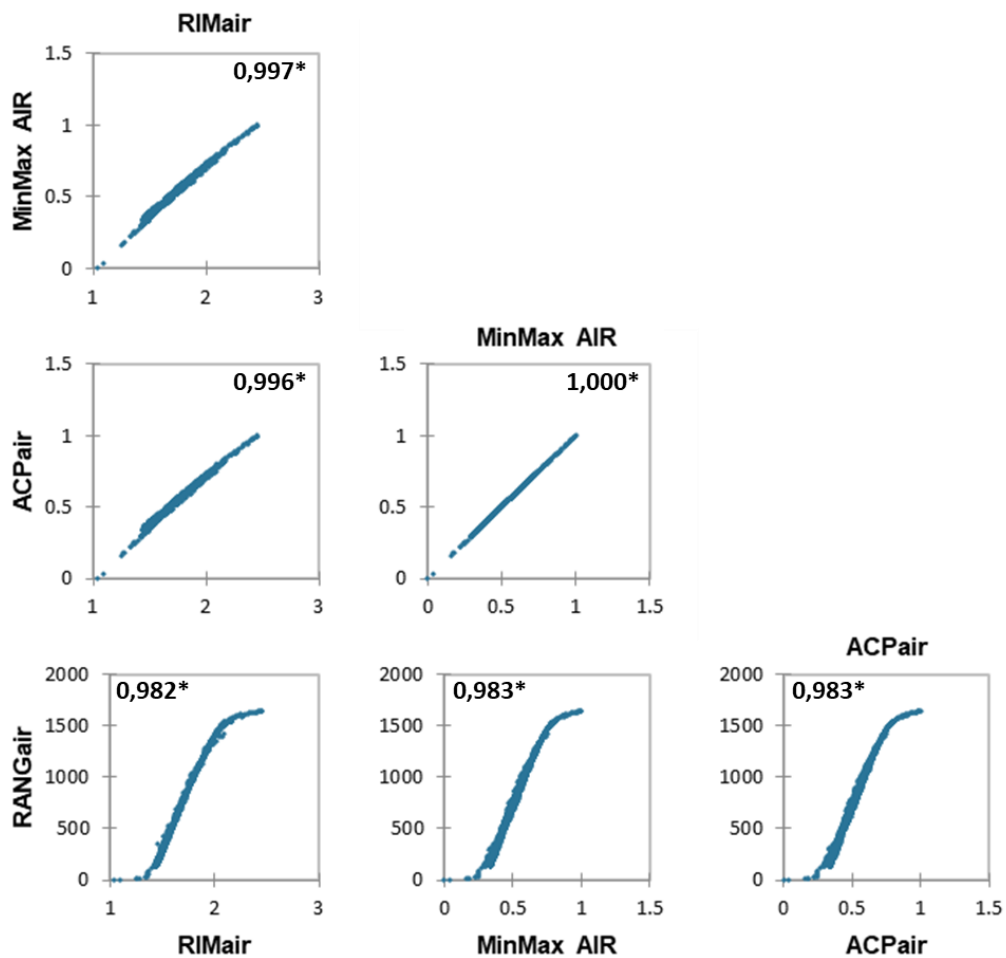
RC1 : composante liée aux activités industrielles présentes ou passées. Corrélée positivement à la densité de sites et sols pollués, à la densité de sites classés pour la protection de l'environnement. Également liée à la part d'immigrés dans la population totale.

RC3 : composante plutôt urbaine liée positivement à la densité de population, au pourcentage de territoires artificialisés, à la pollution de l'air.

RC2 : composante qui semble liée au vieillissement de la population (part de la population de plus de 65 ans et part des décès domiciliés par habitant).

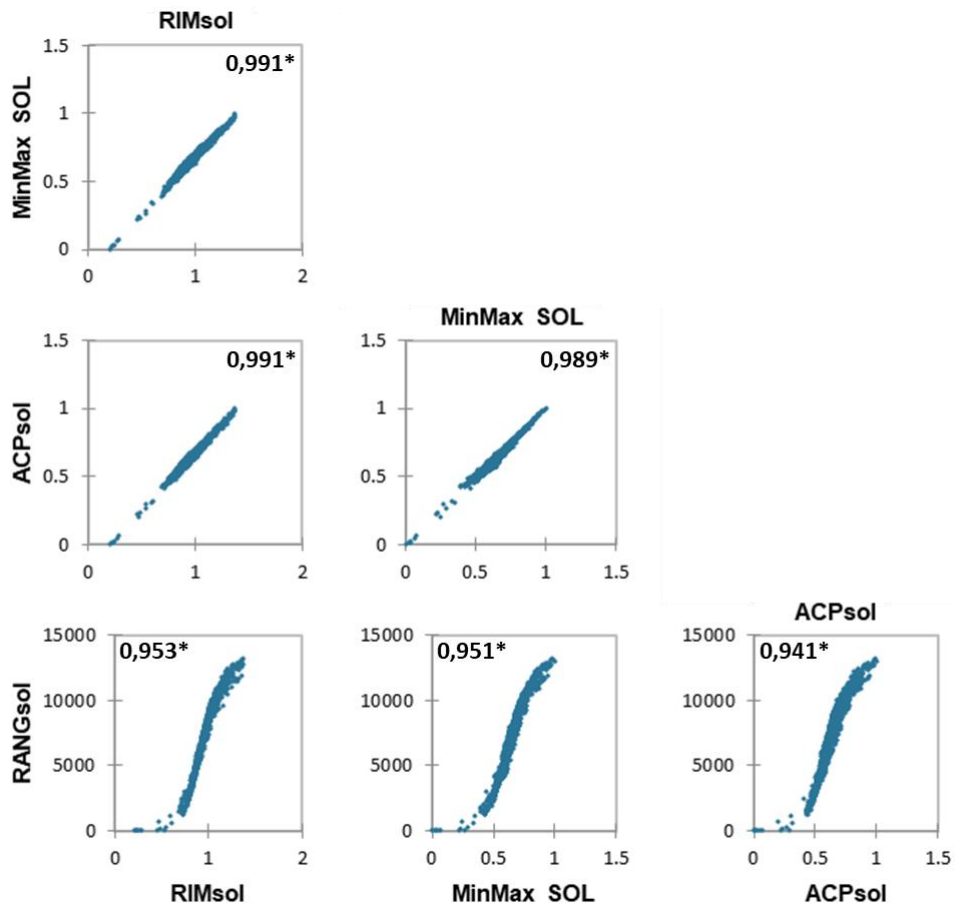
RC4 : composante de précarité, de défaveur sociale liée positivement aux variables suivantes : FDep, part de la population couverte par une prestation de la branche famille, part des locataires dans les résidences principales, part des familles monoparentales parmi les ménages.

ANNEXE 3 : Nuages de points et liens de corrélation entre les différents indices composites de multi-contamination.



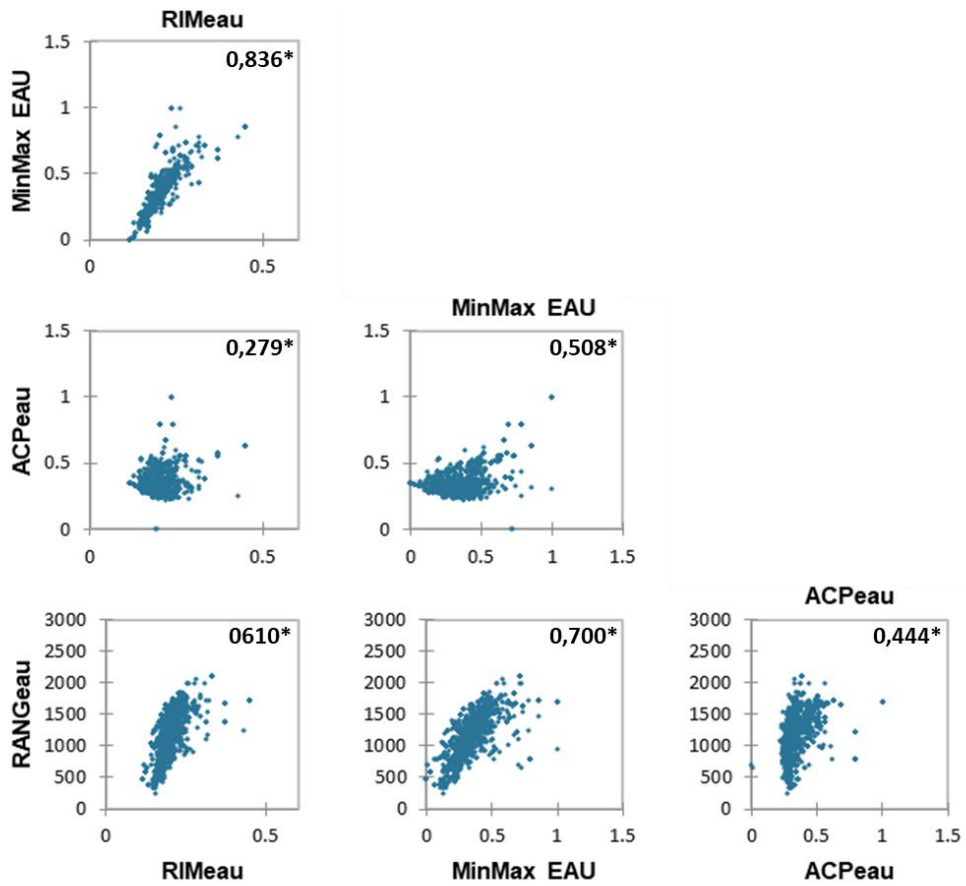
Nuages de points et coefficients de corrélation de Pearson les indices composites de multi-contamination dans l'air.

Une étoile placée à côté du coefficient de corrélation indique la significativité du lien ($p < 0,05$)



Nuages de points et coefficients de corrélation de Pearson les indices composites de multi-contamination dans les sols.

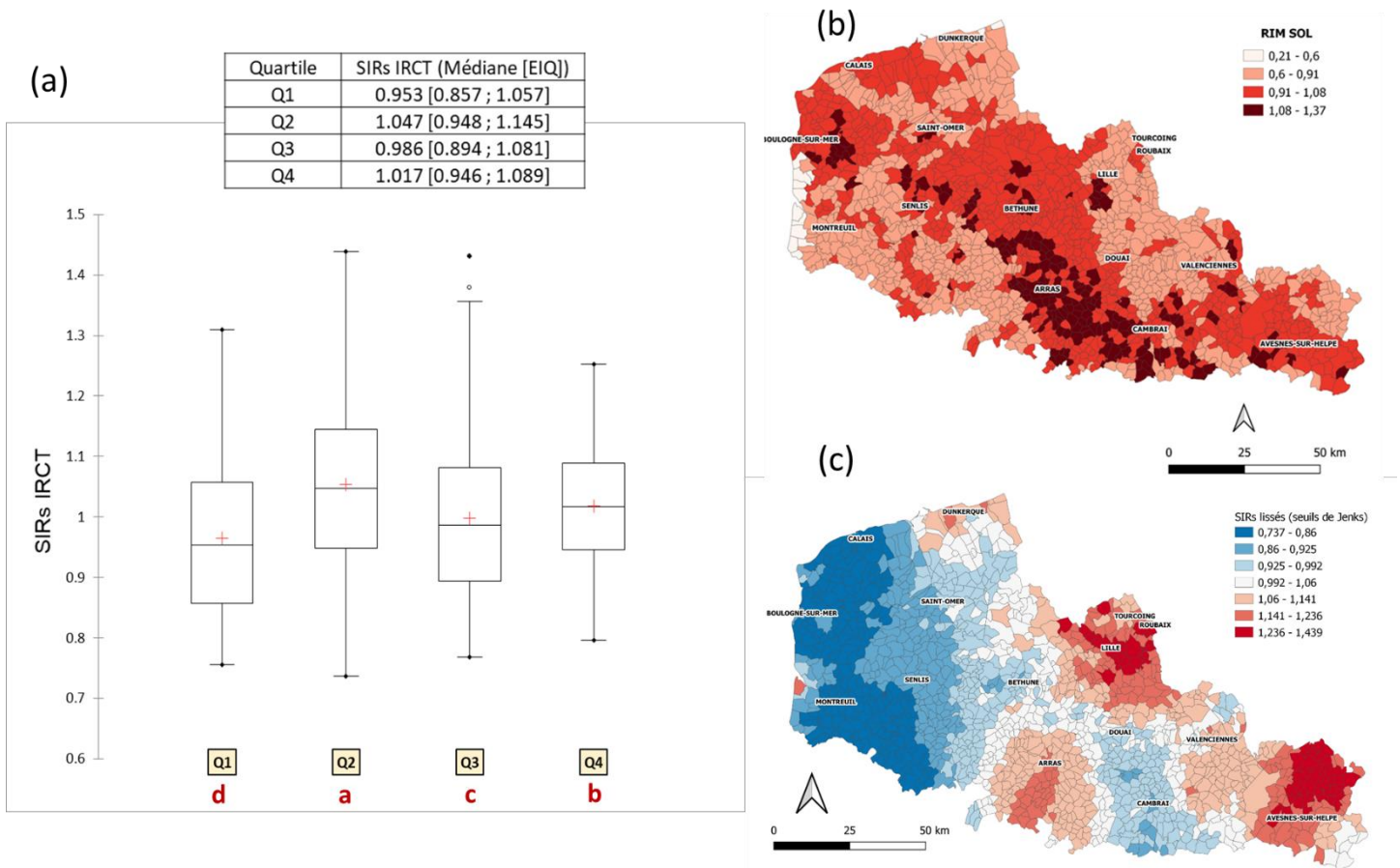
Une étoile placée à côté du coefficient de corrélation indique la significativité du lien ($p < 0,05$)



Nuages de points et coefficients de corrélation de Pearson les indices composites de multi-contamination dans les eaux de distribution.

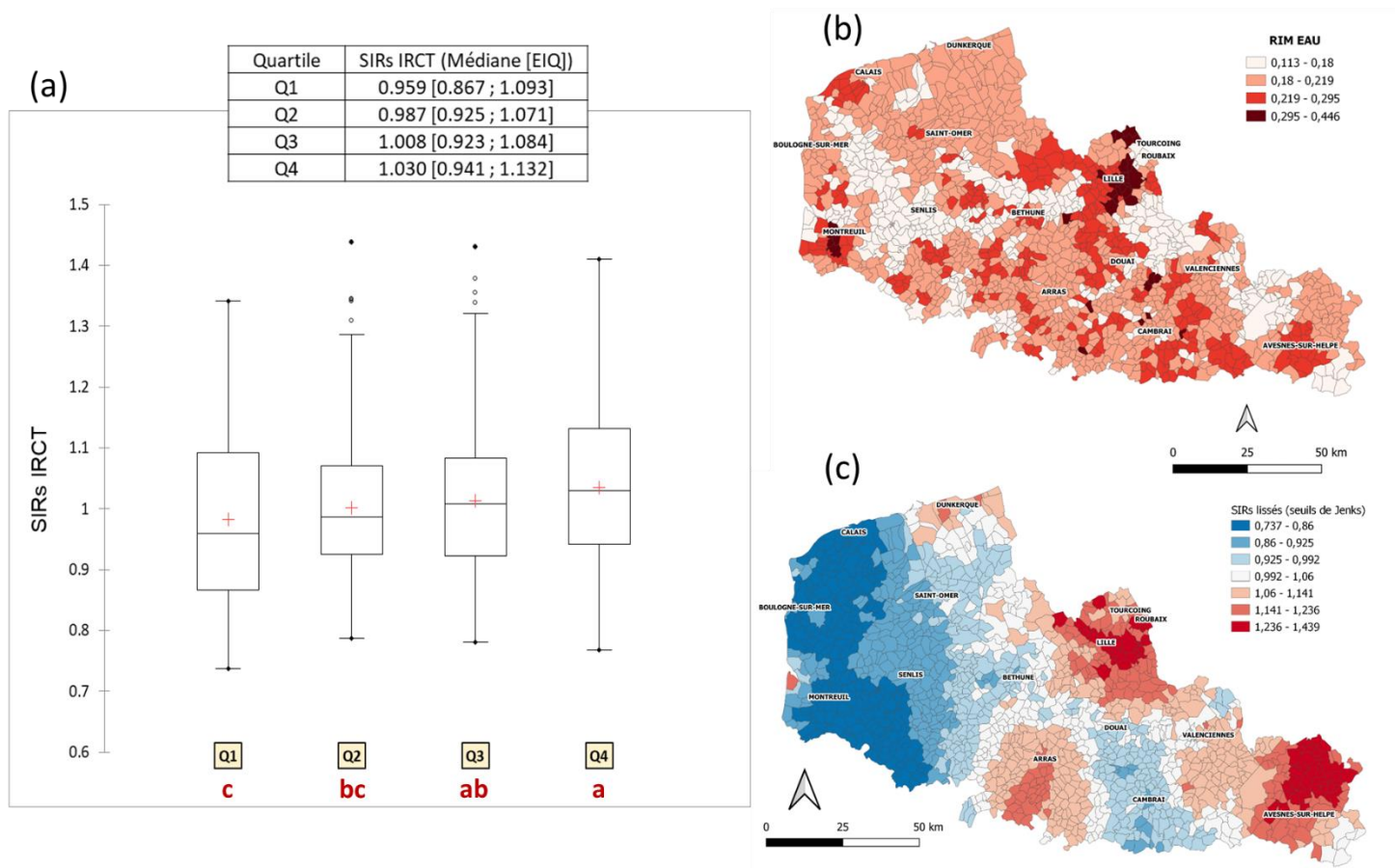
Une étoile placée à côté du coefficient de corrélation indique la significativité du lien ($p < 0,05$)

ANNEXE 4 : Multi-contamination et variations spatiales d'incidence des pathologies étudiées



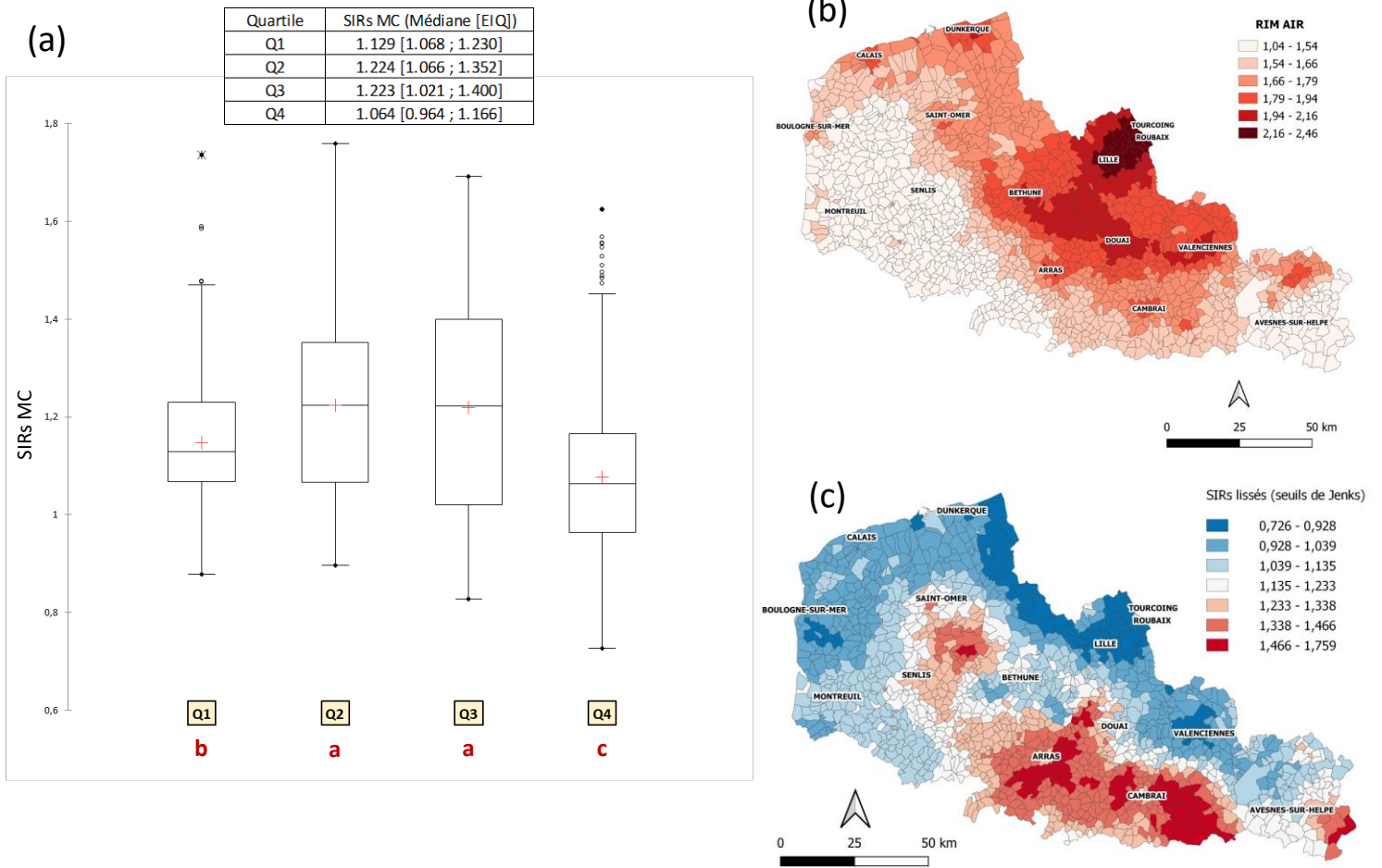
Graphes et Tableau de distribution des SIRs de l'IRCT par quartile croissant de multi-contamination des sols (RIMsol)

Si deux quartiles partagent une lettre en commun alors ils ne sont pas significativement différents, au contraire si deux quartiles ne partagent aucune lettre en commun alors ils sont significativement différents (Test de Kruskal-Wallis avec comparaisons multiples de Dunn)



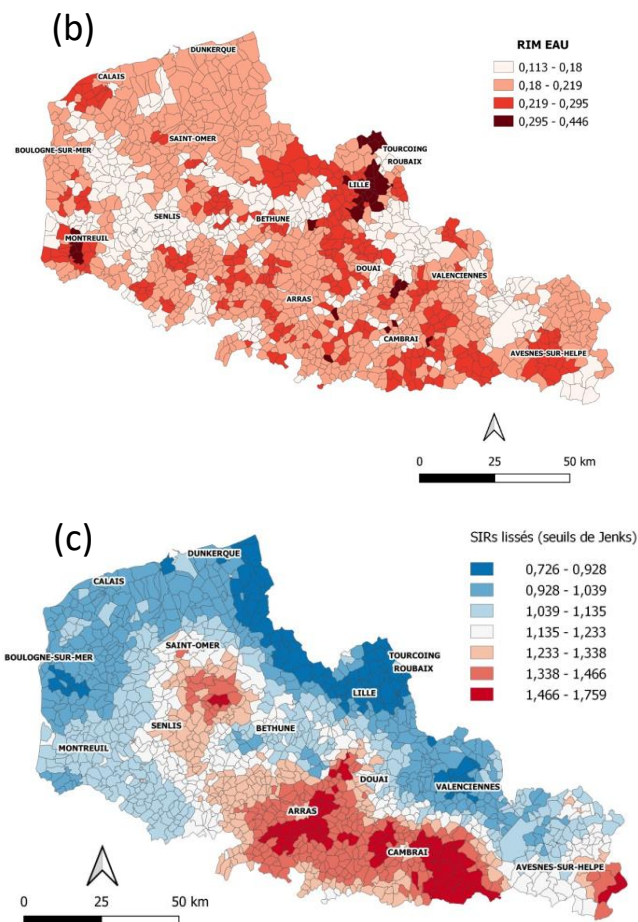
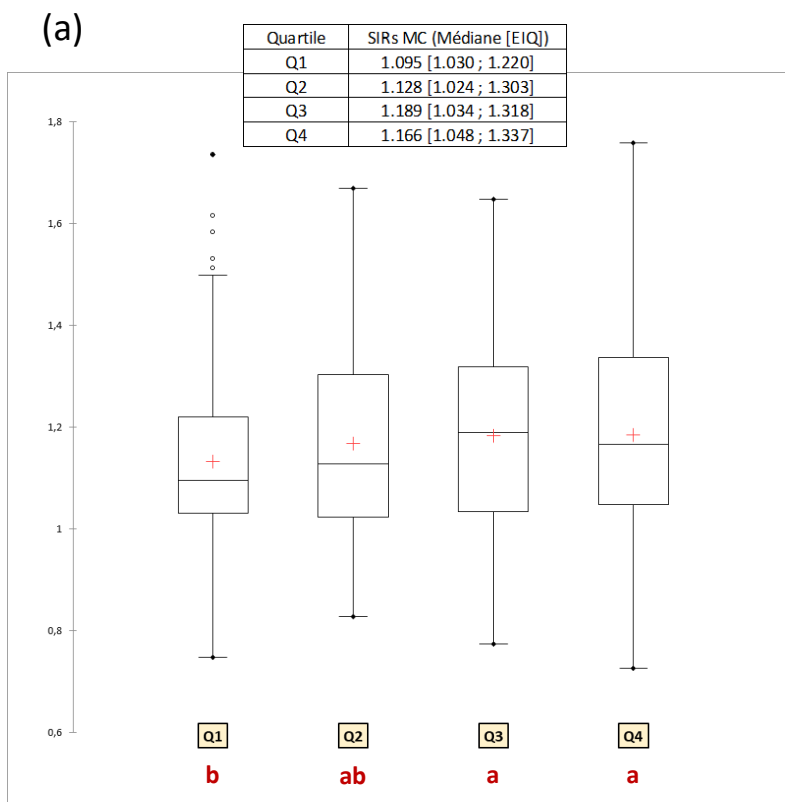
Graphe et Tableau de distribution des SIRs de l'IRCT par quartile croissant de multi-contamination de l'eau (RIMEau)

Si deux quartiles partagent une lettre en commun alors ils ne sont pas significativement différents, au contraire si deux quartiles ne partagent aucune lettre en commun alors ils sont significativement différents (Test de Kruskal-Wallis avec comparaisons multiples de Dunn)



Graphique et Tableau de distribution des SIRs de la MC par quartile croissant de multi-contamination de l'air (RIMair)

Si deux quartiles partagent une lettre en commun alors ils ne sont pas significativement différents, au contraire si deux quartiles ne partagent aucune lettre en commun alors ils sont significativement différents (Test de Kruskal-Wallis avec comparaisons multiples de Dunn)

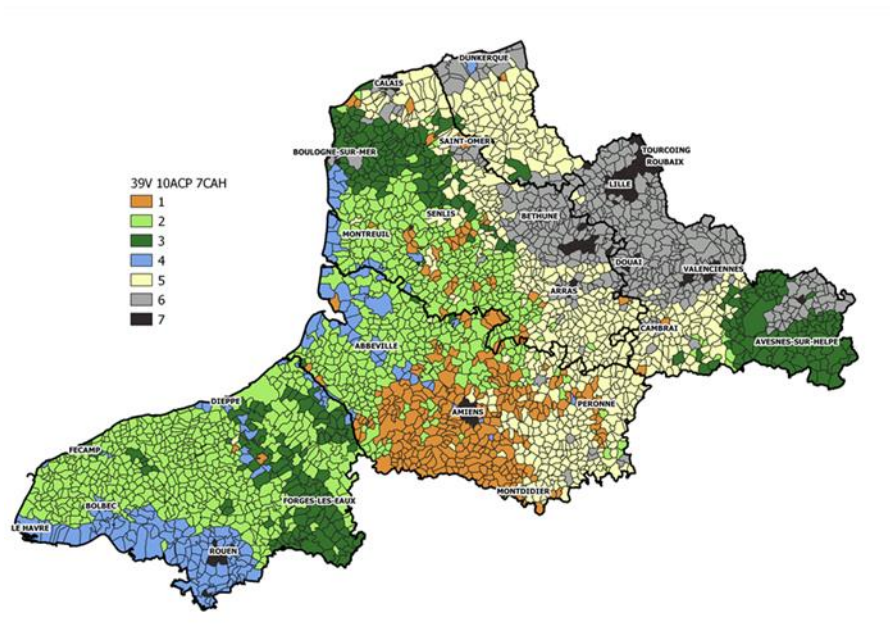


Graphique et Tableau de distribution des SIRs de la MC par quartile croissant de multi-contamination de l'eau (RIMeau)

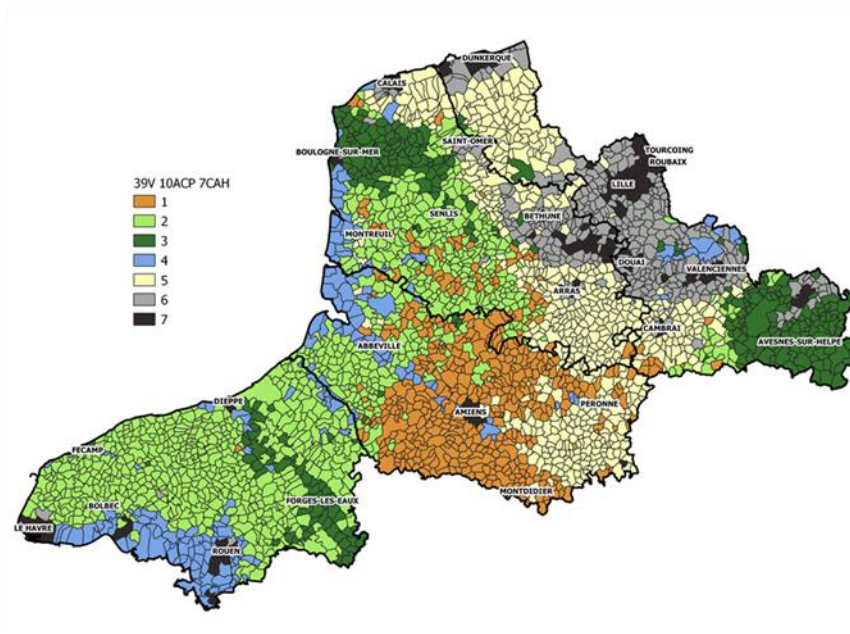
Si deux quartiles partagent une lettre en commun alors ils ne sont pas significativement différents, au contraire si deux quartiles ne partagent aucune lettre en commun alors ils sont significativement différents (Test de Kruskal-Wallis avec comparaisons multiples de Dunn)

ANNEXE 5 : Comparaison entre la classification obtenue avant et après consolidation des k-means.

Classification before k-means consolidation



Final classification after k-means consolidation



ANNEXE 6 : Résultats des régressions écologiques univariées menées sur les indicateurs d'entrée ayant permis la détermination des profils environnementaux.

SUB-DIMENSION	INDICATOR	RR (95% CI)
CONTAMINATION LEVELS	PM2.5_AIR ¹	0.919 [0.871 ; 0.970]
	AS_SOIL ²	1.049 [1.005 ; 1.095]
	HG_SOIL ³	0.982 [0.944 ; 1.021]
	ZN_SOIL ⁴	1.042 [1.008 ; 1.078]
	TL_SOIL ⁵	1.017 [0.985 ; 1.051]
	ENTERO_WATER ⁶	1.000 [0.991 ; 1.009]
	PH_WATER ⁷	1.015 [0.990 ; 1.041]
	CONDUCT_WATER ⁸	0.976 [0.942 ; 1.011]
	TURBIDITY_WATER ⁹	1.002 [0.969 ; 1.036]
	CL2TOT_WATER ¹⁰	0.991 [0.955 ; 1.028]
EMISSION LEVELS	GHGs_EM ¹¹	1.001 [0.995 ; 1.007]
	SO2_EM ¹²	0.987 [0.963 ; 1.011]
	NMVOC_EM ¹³	0.956 [0.924 ; 0.989]
	NH3_EM ¹⁴	1.002 [0.964 ; 1.041]
	AS_EM ¹⁵	0.978 [0.955 ; 1.002]
LOCATION OF EMISSION SOURCES	POLLUT_SITES ¹⁶	0.994 [0.985 ; 1.003]
	CLASSIF_FACILITIES ¹⁷	0.993 [0.98 ; 1.006]
	INDUST_ACCIDENTS ¹⁸	0.992 [0.979 ; 1.004]
	WASTE_PROX ¹⁹	1.001 [0.972 ; 1.031]
	SEWAGE_PROX ²⁰	0.982 [0.951 ; 1.014]

LAND USE	ARTIF_LAND ²¹	0.978 [0.962 ; 0.993]
	HUMID_LAND ²²	0.965 [0.937 ; 0.995]
	WATER_LAND ²³	0.996 [0.975 ; 1.018]
	GRASSLAND ²⁴	0.997 [0.956 ; 1.040]
	PERENNIAL_CROPS ²⁵	1.031 [0.996 ; 1.067]
	CEREALS ²⁶	1.050 [1.025 ; 1.076]
	OILSEEDS ²⁷	1.075 [1.034 ; 1.117]
	PROTEIN_CROPS ²⁸	1.018 [0.978 ; 1.059]
	VEGETABLES ²⁹	1.022 [0.987 ; 1.058]
	INDUSTRIAL_CROPS ³⁰	1.079 [1.040 ; 1.119]
	ARTIF_2009/2019 ³¹	1.005 [0.987 ; 1.023]
	AGRICULTURAL PRACTICES	TOTPEST ³²
INSECTICIDE ³³		1.019 [0.994 ; 1.045]
LIVESTOCK ³⁴		1.008 [0.971 ; 1.047]
HNV ³⁵		0.982 [0.956 ; 1.010]
AGRI_PRACTICES ³⁶		1.003 [0.982 ; 1.025]
NATURALNESS	NATURALNESS ³⁷	1.003 [0.981 ; 1.025]
	BIODIV_RESERVOIRS ³⁸	0.970 [0.942 ; 0.998]
	GREEN_SPACES ³⁹	1.005 [0.974 ; 1.037]

¹ PM2.5_AIR - Mean concentration of PM2.5 in ambient air (µg/m³)

² AS_SOIL - Arsenic concentration in soils (mg/kg)

³ HG_SOIL - Mercury concentration in soils (mg/kg)

⁴ ZN_SOIL - Zinc concentration in soils (mg/kg)

⁵ TL_SOIL - Thallium concentration in soils (mg/kg)

⁶ ENTERO_WATER - Enumeration of enterococci (CFU/100mL)

⁷ PH_WATER - Mean pH in drinking water

⁸ CONDUCT_WATER - Mean conductivity in drinking water (µS/cm)

⁹ TURBIDITY_WATER - Mean turbidity in drinking water (NFU)

¹⁰ CL2TOT_WATER - Mean concentration of total chlorine in drinking water (mg/L)

¹¹ GHGs_EM - Greenhouse gas atmospheric emissions (thousand metric tons of CO2 equivalent)

¹² SO2_EM - Sulfur dioxide atmospheric emissions (kg/ha/an)

¹³ NMVOC_EM - Non-methane volatile organic compound atmospheric emissions (kg/ha/an)

¹⁴ NH3_EM - Ammoniac atmospheric emissions (kg/ha/an)

¹⁵ AS_EM - Arsenic atmospheric emissions (kg/ha/an)

¹⁶ POLLUT_SITES - Density of contaminated sites and soils (number per hectare)

¹⁷ CLASSIF_FACILITIES - Density of facilities classified for environmental protection (number per hectare)

¹⁸ INDUST_ACCIDENTS - Density of industrial accidents recorded since 1960 (number per hectare)

¹⁹ WASTE_PROX - Distance to the closest waste management installation (km)

²⁰ SEWAGE_PROX - Distance to the closest waste water treatment plant (km)

²¹ ARTIF_LAND - Percentage of municipal area occupied by artificial lands (%)

²² HUMID_LAND - Percentage of municipal area occupied by wetlands (%)

²³ WATER_LAND - Percentage of municipal area occupied by water surfaces (%)

²⁴ GRASSLAND - Percentage of municipal area occupied by grasslands (%)

²⁵ PERENNIAL_CROPS - Percentage of municipal area occupied by perennial crops (%)

²⁶ CEREALS - Percentage of municipal area occupied by cereals (%)

²⁷ OILSEEDS - Percentage of municipal area occupied by oil crops (%)

²⁸ PROTEIN_CROPS - Percentage of municipal area occupied by protein crops (%)

²⁹ VEGETABLES - Percentage of municipal area occupied by vegetables and flowers (%)

³⁰ INDUSTRIAL_CROPS - Percentage of municipal area occupied by industrial crops (%)

³¹ ARTIF_2009/2019 - Percentage of municipal area artificialized between 2009 and 2019 (%)

³² TOTPEST - Total phytosanitary pressure index (kg/ha/an)

³³ INSECTICIDE - Insecticide pressure index (kg/ha/an)

³⁴ LIVESTOCK - Number of livestock units per hectare of municipal area (livestock units per hectare)

³⁵ HNV - High Natural Value indicator

³⁶ AGRI_PRACTICES - Indicator of extensification of farming practices

³⁷ NATURALNESS - Potential naturalness indicator

³⁸ GREEN_SPACES - Percentage of the population having access to a green space within 400 meters (%)

³⁹ BIODIV_RESERVOIRS - Percentage of municipal area occupied by biodiversity reservoirs (%)