



UNIVERSITÉ
CAEN
NORMANDIE



IDENTIFICATION ET DIFFÉRENCIATION
DE L'ESPACE, DE L'ENVIRONNEMENT
ET DES SOCIÉTÉS - CAEN

Mémoire de stage

Construction d'une base de données socio-environnementales spatialisées et rétrospectives :

une première étape pour étudier les inégalités sociales et environnementales de l'asthme en France dans le cadre du projet MultiSocialEGEA



*Licence professionnelle Systèmes d'information géographique, diagnostic et aménagement des territoires
2024-2025*

Mémoire de stage

Construction d'une base de données socio-environnementales spatialisées et rétrospectives :

*une première étape pour étudier les inégalités sociales et environnementales de
l'asthme en France dans le cadre du projet MultiSocialEGEA*

SARR Babacar

Soutenance le 11 Septembre 2025

Remerciements :

Je tiens à remercier Thierry FEUILLET et Sophia TEMAM de m'avoir accordé l'opportunité d'effectuer ce stage au sein du laboratoire Identité et différenciation de l'espace, de l'environnement et des sociétés -IDEES Caen-UMR 6266. Leur encadrement, conseil, et confiance m'ont permis de finaliser avec succès ce stage très enrichissant et de découvrir le monde de la recherche dans le domaine de la géographie de la santé.

Je remercie également les membres de l'équipe pédagogique de la licence professionnelle SIG-DAT pour la qualité de l'enseignement dispensé et les connaissances et compétences que j'ai pu acquérir tout au long de l'année.

Je tiens également à exprimer ma profonde gratitude à mes parents, dont le soutien constant du Sénégal jusqu'en France, a été un véritable moteur de la réussite de mon parcours scolaire et universitaire.

Table des matières

| | |
|--|----|
| Remerciements | 5 |
| Liste des figures..... | 9 |
| Glossaire..... | 13 |
| 1. Introduction..... | 14 |
| 1.1 Contexte | 14 |
| 1.2 Présentation générale et objectifs du stage | 15 |
| 2. Le projet MultiSocialEGEA et la cohorte EGEA : cadre scientifique du stage | 16 |
| 2.1 Objectifs scientifiques..... | 16 |
| 2.2 Historique et structure de la cohorte EGEA..... | 16 |
| 2.3 Enjeu des inégalités d'exposition socio-environnementales dans les maladies respiratoires..... | 18 |
| 3. Construction des indicateurs socio-environnementaux rétrospectifs : matériels, méthodes et résultats..... | 20 |
| 3.1 Indicateurs sociaux..... | 23 |
| 3.1.1 Indice de désavantage social..... | 23 |
| 3.1.1.1 Matériel et méthode | 23 |
| 3.1.1.2 Résultats..... | 26 |
| 3.1.2 Indice d'accessibilité aux services de soins..... | 32 |
| 3.1.2.1 Matériel et méthode | 32 |
| 3.2 Indicateurs de l'environnement physique..... | 37 |
| 3.2.1 Typologie de l'occupation des sols | 37 |
| 3.2.1.1 Matériel et méthode | 37 |
| 3.2.1.2 Résultats..... | 39 |
| 3.2.2 Couverture végétale..... | 41 |
| 3.2.2.1 Matériel et méthode | 41 |
| 3.2.2.2 Résultats..... | 43 |
| 3.3 Indicateur de l'environnement chimique | 44 |
| 3.3.1 Proximité aux sites industriels..... | 44 |
| 3.3.1.1 Matériel et méthode | 44 |
| 3.3.1.2 Résultats..... | 45 |
| 4. Discussion et perspectives | 46 |
| 4.1 Apports pour une lecture territoriale des inégalités de santé respiratoire (asthme) | 46 |
| 4.2 Limites | 47 |
| 4.3 Perspectives dans le cadre du projet MultiSocialEGEA..... | 47 |
| 5. Conclusion..... | 48 |
| Bibliographie | 50 |
| Sitographie..... | 51 |

| | |
|-----------------------------|----|
| Sources de données SIG..... | 52 |
| Annexes..... | 53 |

Liste des figures

| | |
|---|----|
| Figure 1 : Schématisation de quelques facteurs socio-environnementaux des maladies respiratoires..... | 18 |
| Figure 2 : Schéma des différentes composantes de l'exposome..... | 20 |
| Figure 3 : Diagramme des pourcentages de variance expliqués par les dimensions de l'ACP..... | 26 |
| Figure 4 : Cercle de corrélations des dimensions 1 et 2 de l'ACP..... | 26 |
| Figure 5 : Carte Fdep France en 2006 avec le taux de bac | 27 |
| Figure 6 : Carte Fdep du 75, 92, 93, 94 en 2006 avec taux de bachelier | 28 |
| Figure 7 : Carte Fdep France en 2006 avec le taux de bac + 2 en 2006..... | 29 |
| Figure 8 : Carte Fdep du 75, 92, 93, 94 en 2006 avec le taux de bac + 2 ou plus..... | 30 |
| Figure 9 : Comparaison entre le Fdep avec le taux de bachelier et le Fdep avec le taux de bac + 2 en 2006 et 2010 | 31 |
| Figure 10 : Méthode de construction de l'indice de Scale | 33 |
| Figure 11 : Carte de l'indice de scale moyen par iris en France..... | 34 |
| Figure 12 : Carte de la mesure de l'accessibilité aux médecins généralistes et pharmacies selon l'indice de Scale dans la métropole du Grand Paris..... | 35 |
| Figure 13 : Distribution du taux d'urbanisation par iris en France | 39 |
| Figure 14 : Distribution du taux de sites miniers par iris en France | 39 |
| Figure 15 : Distribution du taux d'espaces verts par iris en France..... | 39 |
| Figure 16 : Distribution du taux de zones industrielles par iris en France | 40 |
| Figure 17 : Carte du NDVI moyen par iris en France en 2019 | 43 |
| Figure 18 : Méthodologie de construction de la proximité des adresses aux sites industriels | 44 |
| Figure 19 : Carte de la proximité des adresses aux sites industriels dans la Seine Saint-Denis | 45 |

Liste des tableaux

| | |
|---|----|
| Tableau 1 : Historique de l'enquête EGEA | 17 |
| Tableau 2 : Les variables socio-environnementales | 22 |
| Tableau 3 : Les données de l'indice de défavorisation sociale | 23 |
| Tableau 4 : Exemple d'imputation sur les données | 24 |
| Tableau 5 : Tableau des codes du Corine Land Cover | 37 |
| Tableau 6 : Les différentes bandes de Landsat 8..... | 42 |

Glossaire

ACP : Analyse en Composantes Principales

BD TOPO : Base de données Topographique

CLC : Corine Land Cover

EGEA : Etude Epidémiologique des facteurs Génétiques et Environnementaux de l'Asthme

EPCI : Établissement public de coopération intercommunale

Exposome : désigne l'ensemble des expositions à des facteurs non génétiques (environnement de vie) que subit une personne depuis sa naissance jusqu'à la fin de sa vie.

FDEP : French Deprivation Index

IDEES : Identification et Différenciation de l'Espace, de l'Environnement et des Sociétés

IGN : Institut Géographique National

Scale : Spatial Accessibility MultisALar

INSEE : Institut National de la Statistique et des Études Économiques

IRIS : Ilots Regroupés pour l'Information Statistique

LANDSAT 8 : Satellite américain d'observation de la terre

MGEN : Mutuelle Générale de l'Éducation Nationale

MultiSocialEGEA : c'est un projet de recherche qui évalue la contribution des expositions environnementales multiples aux inégalités sociales dans l'asthme (« justice environnementale »), à partir des données de l'Etude épidémiologique des facteurs Génétiques et Environnementaux de l'Asthme (EGEA)

NDVI : Normalized Difference Vegetation Index

OMS : Organisation Mondiale de la Santé

1. Introduction

1.1 Contexte

Ce stage s'inscrit dans le projet de recherche MultiSocialEGEA financé par l'ANSES avec un budget de 230 000 €. Il est effectué au sein du laboratoire IDEES¹ Caen Geophen qui est un laboratoire pluridisciplinaire en science humaine et sociale. Il regroupe des enseignant.e.s- chercheur.e.s, des chercheur.e.s, des ingénieur.e.s de recherche et/ou d'études et des doctorant.e.s. Il est composé différent chercheur.e.s de spécialités reconnues en France et à l'internationale dans le domaine de la modélisation, l'analyse spatiale, le transport, les risques, la santé, l'environnement etc .

Le projet MultiSocialEGEA fait partie un des projets de recherche phare du laboratoire, il est piloté par Sofia TEMAM, chargée de recherche chez Fondation MGEN² pour la santé publique et trois membres du laboratoire IDEES : Thierry Feuillet, responsable scientifique, Fabien Guillot (IDEES-Caen) et Yoann Doignon (CNRS Rouen).

Le projet propose « une approche novatrice en utilisant des études longitudinales pour examiner la mobilité résidentielle et ses déterminants sociaux, sanitaires et environnementaux. En intégrant le concept d'exposome, il explore l'interaction complexe entre les facteurs sociaux et les expositions environnementales protectrices ou à risque. Cette étude permettra d'éclairer les mécanismes sous-jacents aux inégalités sociales et d'améliorer les stratégies de prévention de l'asthme dans des contextes vulnérables³. »

Les recherches du projet MultiSocialEGEA s'appuient sur la cohorte épidémiologique EGEA(Etude Épidémiologique des facteurs Génétiques et Environnementaux de l'Asthme) lancée en 1991 pour permettre l'identification des facteurs génétiques et environnementaux de l'asthme mais aussi d'expliquer l'hétérogénéité clinique de l'asthme autrement dit les phénotypes variés que peut prendre cette maladie. La cohorte EGEA a été constituée dans 5 villes de France : Paris, Lyon, Grenoble, Montpellier et Marseille, par des équipes de recherche de l'Inserm (Institut National de la Santé et de la Recherche Médicale), elle retrace l'histoire résidentielle de tous les participants afin de fournir des informations supplémentaires de leurs déplacements liés à la santé respiratoire.

¹ Identité et Différenciation de l'Espace, de l'Environnement et des Sociétés

² La Fondation d'Entreprise MGEN pour la Santé Publique a été créée en 2002 par MGEN dans la perspective de contribuer à la recherche en santé publique.

³ <https://idees.unicaen.fr/multisocialegea-inegalites-sociales-et-asthme/>

1.2 Présentation générale et objectifs du stage

L'asthme est une maladie respiratoire chronique qui s'accompagne de symptômes respiratoires (essoufflements, sifflements respiratoires, oppression thoracique, toux persistante etc...). Selon l'OMS⁴, en 2019 on estimait que 262 millions de personnes souffraient d'asthme dans le monde et qu'il entraînait 455 000 décès. Cette maladie aujourd'hui avec le contexte du changement climatique pose un véritable problème de santé publique en France et dans le monde.

En France, l'asthme toucherait près de 4 millions de personnes avec environ 900 décès par an et près de 60 000 hospitalisations (Santé Publique France 2021, AMELI⁵2022). De nombreux facteurs différents sont associés à un risque d'asthme mais il est souvent difficile de trouver une cause directe. Cependant, l'urbanisation, l'exposition à différents allergènes et produits irritants de l'environnement, la pollution atmosphérique pourraient accroître le risque. Elle est une maladie multifactorielle à l'étiologie complexe impliquant à la fois des facteurs de risque génétiques et environnementaux, tels que le lieu de résidence (zone urbaine/ rurale), le mode de vie ou encore le tabagisme (Guillien et al., 2021). Ainsi c'est cette multiplicité de facteurs complexe catalysant l'asthme qui rend difficile d'identifier les causes de celle-ci. C'est pour cela pour chercher les causes, les chercheurs vont au-delà du domaine biomédical afin d'interroger l'environnement des personnes asthmatiques et leur mode de vie. Cette complexité a conduit au développement de l'approche par l'exposome par le chercheur britannique Christopher Wild en 2005, qui vise à prendre en compte l'ensemble des expositions environnementales tout au long de la vie de l'individu. En effet, l'environnement joue un rôle crucial dans les maladies chroniques c'est ce qui a inspiré Wild à inventer le terme d'exposome désignant « la totalité des expositions environnementales humaines depuis la conception, en complément du génome ». Il catégorise trois grands types d'exposition non-génétique qui sont l'exposition interne, l'exposition externe spécifique et l'exposition externe générale. L'exposition interne représente l'ensemble des processus internes à l'organisme, le métabolisme, le vieillissement ou l'activité physique. L'exposition externe spécifique dépend du mode de vie de l'individu, l'alimentation, l'environnement immédiat etc. C'est un ensemble d'élément externe à l'organisme pouvant affecter la santé et exposé les individus ayant une maladie chroniques telle que l'asthme. Et enfin l'exposition externe générale qui combine l'ensemble des influences des déterminants sociaux, psychologiques et économiques (éducation, environnement, milieu de vie etc.).

Ainsi c'est dans ce sillage que ces expositions externes générales vont alimenter nos recherches au cours de ce stage en explorant des indicateurs d'inégalité sociale, de l'environnement à travers l'occupation du sol, le niveau de défaveur sociale, la présence ou absence de végétation, le milieu de vie et l'accessibilité à des services de soins. Donc ce travail explore de manière générale la dimension de la caractéristique territoriale des facteurs d'expositions externes de l'asthme pouvant accroître ou protéger l'individu, d'identifier leur niveau d'exposition mais aussi d'aider à la prise de

⁴ Organisation Mondiale de la Santé

⁵ C'est le portail officiel de l'Assurance Maladie française (gérée par la Caisse Nationale d'Assurance Maladie - CNAM), accessible à l'adresse : www.ameli.fr

décision pour l'aménagement afin atténuer les risques. Ainsi l'objectif général est de construire des bases de données à l'échelle infra-communales (Iris⁶), constituant des indicateurs d'expositions socio-environnementales rétrospectives couvrant une partie des années où les enquêtes de la cohorte l'EGEA ont été établies. Ainsi les indicateurs créés dans les bases de données spatialisées vont servir dans des recherches futures des bases pour corrélérer les informations de l'historique de la cohorte de l'EGEA (adresses, mobilités, déménagements) avec les expositions en fonction des années et de répondre à la question des mécanismes sous-jacents aux inégalités sociales et environnementales qui peuvent être des éléments catalysant ou protégeant de l'asthme.

Ainsi la compréhension des relations entre les caractéristiques individuelles (niveau de vie, revenu médian, niveau d'éducation, etc.), environnement (zone urbaine/rurale, pollution, couverture végétation) et l'asthme permettra d'initier des mesures ciblées de prévention et de nouveaux protocoles de soin.

2. Le projet MultiSocialEGEA et la cohorte EGEA : cadre scientifique du stage

2.1 Objectifs scientifiques

Les recherches sur MultiSocialEGEA ont pour objectif d'évaluer et de mieux comprendre la contribution des expositions environnementales multiples aux inégalités sociales dans l'asthme dans une perspective de justice environnementale. Elles s'appuient sur les données de l'Etude épidémiologique des facteurs Génétiques et Environnementaux de l'Asthme (EGEA) qui a suivi une cohorte pendant 30 ans (<https://cohorte-egea.fr/>). Ces enquêtes de l'EGEA permettent de comprendre les relations entre l'environnement, la génétique et l'asthme pour prendre de nouvelles solutions de prises en charges et d'identifier les causes.

L'étude met en lumière des informations de santé publique afin d'améliorer la santé respiratoire de la population.

2.2 Historique et structure de la cohorte EGEA

Afin d'atteindre la finalité scientifique du projet, 388 enfants et adultes asthmatiques ont été recrutés en consultation de pneumologie ou de pédiatrie dans l'un des huit centres hospitaliers participants (Cochin, Lyon, Marseille, Montpellier, Necker, Bichat, Grenoble et Trousseau). Leurs apparentés du premier degré ont également été inclus (parents et frères/sœurs pour les enfants, conjoints et enfants pour les adultes), ainsi que 415 témoins issus de la population générale. Au total, 2047 participants ont été examinés entre 1992 et 1995. Un premier suivi a eu lieu entre 2003 et 2007, puis un second entre 2011 et 2013. Un troisième suivi a commencé en 2023, faisant de la cohorte EGEA l'une des rares cohortes mondiales à disposer d'un suivi sur 30 ans.

⁶ Ilots regroupés pour l'information statistique

D'après la publication de l'EGEA, la cohorte comprend un total de 2120 participants asthmatiques et non-asthmatiques.

Une première enquête (EGEA1) s'est déroulée entre 1991 et 1995. Une deuxième enquête (EGEA2) s'est déroulée entre 2003 et 2007. Lors de cette seconde enquête, 2% de la population initiale était

décédée, 92,2 % de la population initiale non décédée ont répondu à un questionnaire court et 77,1% à un questionnaire long.

Lors de la troisième enquête (EGEA3) qui a été réalisée entre 2011 et 2013, 1558 participants ont retourné leur questionnaire (taux de réponse = 79,2%).

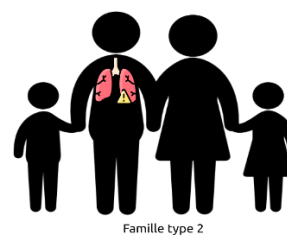
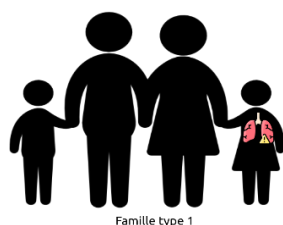
Une quatrième enquête (EGEA4), incluant un examen clinique a débuté en 2023.



Tableau 1 : Historique de l'enquête EGEA

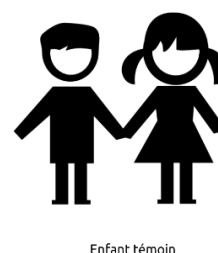
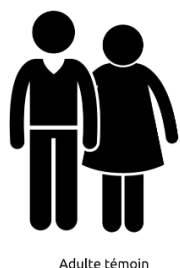
Source : cohorte EGEA

Dans la première phase de l'étude, 388 familles d'asthmatiques et 415 participants venant de la population générale ont été inclus, soit 2047 participants asthmatiques ou non- asthmatiques.



388 familles incluses à partir d'un participant avec le statut d'enfant dans la famille (famille type 1) ou par un participant avec le statut de parent dans la famille (famille type 2), soit au total 1632 participants.

Lors de la seconde phase de l'étude 73 nouveaux membres de familles se sont joints à l'étude. Au total, 2120 participants ont participé à au moins une phase de l'étude.



Ont été également inclus dans l'étude, 415 participants issus de la population générale, qu'ils soient enfants ou adultes.

2.3 Enjeu des inégalités d'exposition socio-environnementales dans les maladies respiratoires

Selon l'Inserm, touche plus de 4 millions de personnes en France. En effet, 10 à 12% de la population ont ou auront l'asthme au cours de la vie, on estime que plus de trois enfants sur quatre respirent un air pollué. Au cours des dernières années, beaucoup d'études épidémiologiques ont mis en avant le rôle de l'environnement dans la morbidité et la mortalité d'origine respiratoire. Parmi d'autres facteurs environnementaux, la pollution atmosphérique d'origine anthropique apparaît comme élément de l'aggravation de pathologies telles que l'asthme et les allergies (Burnett et al 1995, Schwartz et al 1996, Segala et al 1998, Kunzli et Tager 2000) ainsi que de l'augmentation de l'incidence de la broncho-pneumopathie chronique obstructive. Ces pollutions de l'air viennent souvent des activités anthropiques néfastes à l'environnement.

Les facteurs déclenchant les maladies respiratoires se trouvent en majeure partie dans notre environnement tels que les radiations, la pollution de l'air accentuée de nos jours par l'industrialisation dans les milieux urbains et péri-urbains. Ils peuvent aussi dépendre du contexte géographique (zones urbaines/ rurales, artificialisation des sols). Ainsi notre milieu de vie est un déterminant clé du niveau de notre santé respiratoire.

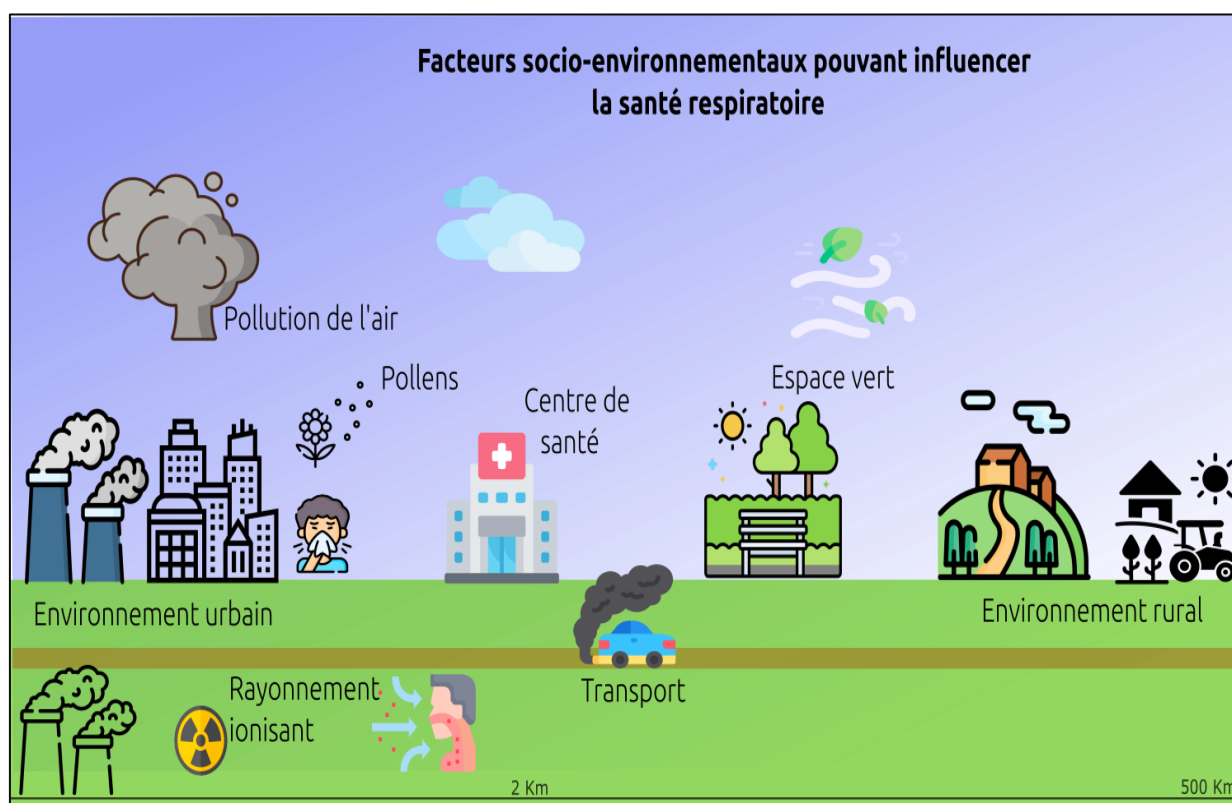


Figure 1 : Schématisation de quelques facteurs socio-environnementaux des maladies respiratoires

L'identification et la connaissance de ces expositions croisées avec les indicateurs sociaux et environnementaux tels que le niveau de vie, l'accessibilité à des soins,

l'occupation du sol, la couverture végétale sont des étapes cruciales afin d'identifier les zones cumulant exposition et défaveur sociale.

Les inégalités d'exposition peuvent aussi résulter par des disparités géographiques :

- Accès limité à des espaces verts ou à un cadre de vie de qualité selon milieu de vie urbain ou rural
- Proximité avec des infrastructures polluants comme les zones d'activités.

L'étude réalisée par Santé publique France et l'Inserm, publiée dans la [revue J Expo Sci Environ Epidemiol](#) montre « la surexposition à la chaleur était plus élevée en 2015-2018 que pour 2000-2014. L'exposition à tous les polluants atmosphériques, à l'exception de l'ozone, a diminué au cours de la période d'étude. En 2018, plus de 79 % des districts de recensement urbains ont dépassé les directives 2021 de l'OMS sur la qualité de l'air. L'évolution de la densité de la végétation entre 2000 et 2018 a été hétérogène dans toute la France continentale⁷ ».

A ces vulnérabilités s'ajoutent les inégalités sociales où les facteurs sociaux comme la pauvreté, l'accès difficile à des services de soin aggrave la santé respiratoire d'une frange de population défavorisée sur le plan économique et géographique. Une étude à montrer qu'à Paris les populations les plus défavorisées risquent trois fois plus de mourir d'un épisode de pollution que les populations les plus aisées.

Ainsi, les déterminants extérieurs (socio-environnementaux) des maladies respiratoires constituent aujourd'hui un enjeu majeur en France et dans le monde. En effet, le contexte du changement climatique, l'hyper-industrialisation des villes et l'augmentation des épisodes de pollution combinés aux inégalités sociales rendent les maladies respiratoires comme l'asthme un des axes de recherches prioritaires par les politiques publiques et les organismes de santé publique en France afin d'identifier ses causes et de trouver des solutions pour mieux protéger la population. Par conséquent, le but de ce travail est de travailler sur les bases, c'est-à-dire d'identifier les variables socio-environnementales, de les calculer sur une partie du période des enquêtes de l'EGEA pour construire des indicateurs rétrospectives et spatialisés.

⁷ Article : <https://www.nature.com/articles/s41370-024-00641-6>

3. Construction des indicateurs socio-environnementaux rétrospectifs : matériels, méthodes et résultats

En matière de santé respiratoire, les expositions socio-environnementales peuvent être diverses et variées. Après des décennies à chercher les facteurs génétiques des pathologies humaines, en 2005 Christopher Wild invente le concept d'exposome dans le cadre de l'épidémiologie moléculaire pour compléter les études des facteurs génétiques dans les maladies chroniques afin de mieux les comprendre. Dans un article publié dans la revue *Cancer Epidemiology, Biomarkers & Prevention*, l'épidémiologiste définissait alors l'exposome comme « la totalité des expositions auxquelles un individu est soumis de la conception à la mort. C'est une représentation complexe et dynamique des expositions auxquelles une personne est sujette tout au long de sa vie, intégrant l'environnement chimique, microbiologique, physique, récréatif, médicamenteux, le style de vie, l'alimentation, ainsi que les infections ». En 2012, il affine plus le concept, selon lui il y'a trois domaines de l'exposome :

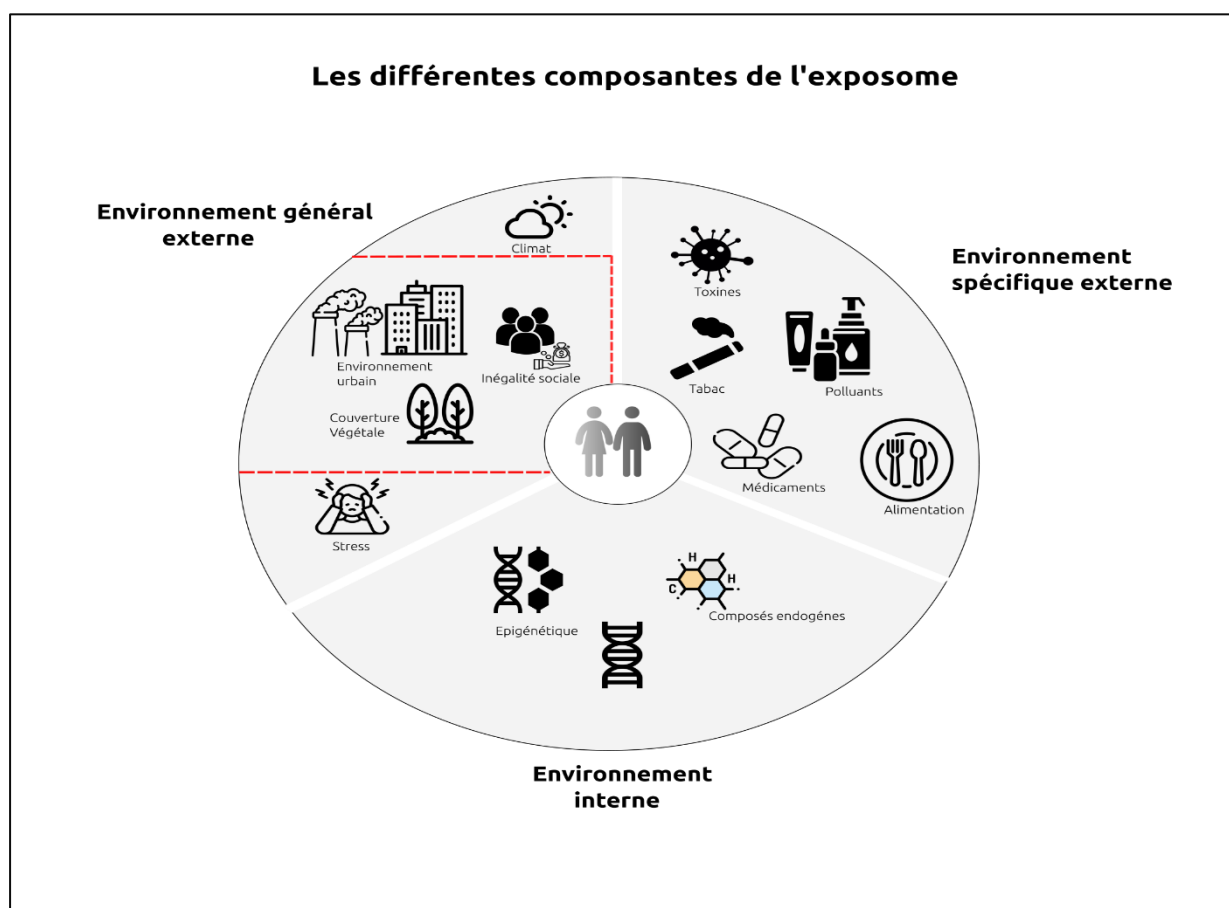


Figure 2 : Schéma des différentes composantes de l'exposome

- Exposome générale interne : correspondant aux facteurs biologiques et physiologiques internes à une personne par exemple les activités hormonales, les métabolites issus de l'alimentation ou de la pollution.
- Exposome spécifique externe : comprend les expositions individuelles et qui être mesurées dans l'environnement immédiat du quotidien. Par exemple les produits de nettoyages, les médicaments, le tabagisme actif ou passif.
- Exposome général externe : rassemble les facteurs contextuels globaux qui influence la santé de manière indirecte. Par exemple la catégorie socio-professionnelle, le niveau d'éducation, l'aménagement du territoire, l'environnement urbain ou rural.

Ainsi c'est l'exposome général externe qui va être le sujet de recherche avec l'exploration de différents indicateurs socio-environnementaux pour mesurer les inégalités d'exposition face à l'asthme selon les territoires. L'échelle de représentation globale des indicateurs est le découpage par iris⁸ de l'INSEE.

Les indicateurs sont divisés en 3 domaines :

- L'environnement social :

Pour évaluer l'exposition sociale deux indicateurs seront calculer :

En France le Fdep⁹ est un indicateur adapté aux études de santé sur la population, c'est pour cela il est un indice de mesure du niveau de défaveur social afin de trouver les quartiers les plus aisés ou défavorisés à l'échelle communale et infra-communale.

L'accessibilité aux services de soins est aussi mesurée à l'aide de l'indice Scale pour voir le niveau d'accessibilité géographique aux services de santé en tenant en compte la répartition des professionnelles de santé, de la population et des distances à parcourir avec l'innovation de l'approche par zone résidentiel.

- L'environnement physique :

Pour mesurer l'environnement physique, le Corine Land Cover permettra d'identifier les types de milieux (urbains, agricole, industriels, naturels) à l'échelle infra-communale afin de modéliser l'exposition environnementale des populations à certains milieux.

- L'environnement chimique :

L'exposition à l'environnement chimique sera mesurée par des zones tampons mesurant la proximité géographique entre les adresses résidentielles et les sites industrielles. Cependant, afin de respecter les règles du RGPD¹⁰ des fausses adresses vont être utilisées pour juste visualiser la proximité de la localisation des adresses par rapport aux zones industrielles.

⁸ C'est un découpage infra-communale de l'INSEE qui signifie « Ilots Regroupés pour l'Information Statistique »

⁹ French Deprivation Index ou Indice de désavantage social

¹⁰ Règlement Général de Protection des données

L'objectif général de ces travaux sur les indicateurs est de développer des scripts et des bases de données reproductibles sur plusieurs années afin d'aider aux chercheurs de faire de future corrélation entre les données de la cohorte de l'EGEA et les résultats des niveaux d'exposition socio-environnementale étudiés.

| Domain | Indicators | Description | Sources | Ressources biblio | Spatial resolution |
|-----------------------------|---------------------------------|---|-----------|---------------------|-----------------------------------|
| Social environment | French Deprivation Index (FDep) | 4 census-derived variables (median household income, % of high school graduates, % of blue-collar workers, and % of unemployment) | INSEE | Rey et al., 2009 | IRIS |
| | Health care services access | Spatial accessibility multi-scalar (SCALE) index | data-gouv | Launay et al., 2019 | IRIS |
| Physical environment | Land use diversity | "Urbanized areas", "industrial or commercial areas and communication networks", "mines, landfills and construction sites" and "artificial green spaces, non-agricultural" | CLC | | IRIS |
| | Greenness level | Normalized Difference Vegetation Index (NDVI): greenness indicator ranging from -1 to +1 with higher values indicating denser vegetation. | LANDSAT | | IRIS |
| Chemical environment | Industrial pollution | Proximity to industrial sites | IGN | | Distance from address coordinates |

Tableau 2 : Les variables socio-environnementales

3.1 Indicateurs sociaux

3.1.1 Indice de désavantage social

L'analyse des inégalités sociales de santé s'appuie sur la mise en place d'indicateurs de référence et reproductibles des situations sociales comme le FDep (French Deprivation Index) qui est un indice synthétique de mesure du niveau de désavantage social utilisé dans le domaine la santé publique en France. Il se base sur le statut individuel (niveau d'étude, catégorie socio-professionnelle, profession etc.) et sur des caractéristiques du contexte de résidence (revenu médian par commune ou iris).

3.1.1.1 Matériel et méthode

L'indice de désavantage social noté Fdep est construit à l'échelle des iris de toute la France par le biais des données des recensements et des revenus fiscaux de la population selon l'INSEE sur plusieurs années : 2006, 2009 et 2010.

Les variables collectées sont :

- Le taux de chômage dans la population active 15 à 64 ans
- Le taux d'ouvriers dans la population active de 15 à 64 ans
- Le taux de bacheliers dans la population non scolarisée des plus de 15 ans
- Le revenu médian par unité de consommation (Iris et commune).

Les données collectées sur INSEE sont nettoyées, ensuite à l'aide de scripts R, on fait l'analyse par composante principale (ACP) de ces quatre variables selon la méthode de Rey et al. (2009) - BMC public health, pondérées par rapport à la population de chaque iris.

| Variables | Source | Année | Échelle | Description |
|----------------|--------|------------------|---------|--|
| <i>txChom</i> | INSEE | 2006, 2009, 2010 | IRIS | Taux de chômage |
| <i>tx_bac</i> | INSEE | 2006, 2009, 2010 | IRIS | Taux de diplômés du baccalauréat ou bac +2 |
| <i>txouvr</i> | INSEE | 2006, 2009, 2010 | IRIS | Taux d'ouvriers |
| <i>revIris</i> | INSEE | 2006, 2009, 2010 | IRIS | Revenu médian par IRIS |
| <i>revCom</i> | INSEE | 2006, 2009, 2010 | Commune | Revenu médian par commune (données de secours) |
| <i>Pop</i> | INSEE | 2006, 2009, 2010 | IRIS | Population totale |

Tableau 3 : Les données de l'indice de défavorisation sociale

Trois fichiers excel varIris avec (*txChom*, *tx_bac*, *tx_ouvrier* et *pop*), *revIris* qui contient les revenus médians par Iris, et *revCom* les revenus médians par communes sont importés sur R. Il faut noter que les données des revenus médians par iris ne sont pas exhaustives. Ainsi pour compléter les données manquantes dans certains iris, lorsque le revenu médian à l'échelle Iris est manquant, la valeur communale est utilisée en substitution. Ensuite le champ obtenu est joint à la table *varIris*.

Une fois toutes les tables disponibles, on crée un tableau final nommé *dtFdep* contenant les quatre variables nécessaires à la construction du fdep :

```
dtFDep <- varIris[,c("IRIS", "Pop06", "tx_bac06", "revMed06", "txChom06", "txouvr06")]
```

Chaque variable de la table *dtFdep* est convertit en numérique. Cependant, malgré le remplissage des valeurs manquantes de revenu à l'iris par celles de la commune, on constate toujours des données manquantes dans la table *dtFDep*.

Ainsi, l'indice de défavorisation sociale FDep va être calculé avec et sans imputation sur les données manquantes, cela nous permettra de quantifier la qualité des résultats obtenus afin de faire une comparaison. Le fdep se calculera aussi avec le taux de bachelier et avec le taux de bac+2 ou plus pour plus affiner la qualité des résultats. L'imputation consiste à remplacer les données manquantes dans le *dtFdep* par des valeurs estimées afin de rendre le traitement exhaustif. C'est une étape importante afin de garantir la robustesse de l'analyse.

Par exemple :

| IRIS | Revenu (revMed06) | Taux de chômage | Taux d'ouvriers | Taux de diplômés |
|------------|-------------------|-----------------|-----------------|------------------|
| IRIS_00001 | 19200 | 11.5% | 27.4% | 30.2% |
| IRIS_00002 | NA | 13.8% | 29.5% | 28.0% |
| IRIS_00003 | 18500 | 10.9% | 26.1% | 32.0% |
| | | | | |
| IRIS | Revenu (estimé) | Taux de chômage | Taux d'ouvriers | Taux de diplômés |
| IRIS_00002 | 17600 | 13.8% | 29.5% | 28.0% |

Tableau 4 : Exemple d'imputation sur les données

Dans l'IRIS_00002 la valeur du revenu médian est manquante, grâce à l'imputation de l'ACP, la valeur est estimée à partir de la corrélation entre les autres variables (taux

de chômage, taux d'ouvriers, taux de diplômé bac). Si un Iris a un fort taux de chômage et peu de proportion de diplômé, il est probable que son revenu médian soit inférieur à la moyenne.

Pour finir l'analyse par composante principale est effectuée sur les quatre variables conformément à la méthode de *Rey et al. (2009)*. L'ACP est centrée réduite, pondérée par la population totale par iris, et limitée à la première composante qui capte l'essentiel de la variance commune entre les variables.

3.1.1.2 Résultats

L'ACP est appliquée sur les quatre variables (taux de chômage, taux d'ouvriers, taux de bachelier et revenu médian), pondérée à la population de chaque Iris pour que les iris les moins peuplés ne pèsent pas autant que ceux très peuplés. L'histogramme de la variance expliquée par chaque dimension montre les dimensions de la variance dans le jeu de données. Dans notre analyse l'axe 1 a été choisi car il capte d'environ 61% de la variance totale des données ce qui explique l'essentiel des données. Ainsi, le score de l'indice de défavorisation sociale est construit à partir de la première dimension représentant donc l'axe principale.

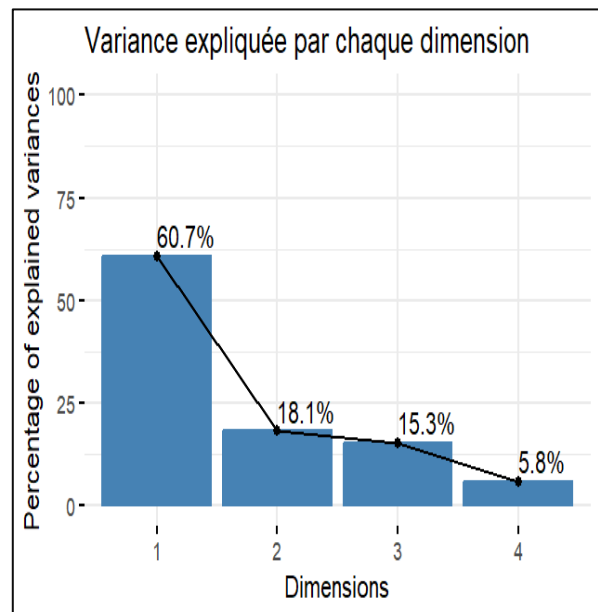


Figure 3 : Diagramme des pourcentages de variance expliqués par les dimensions de l'ACP

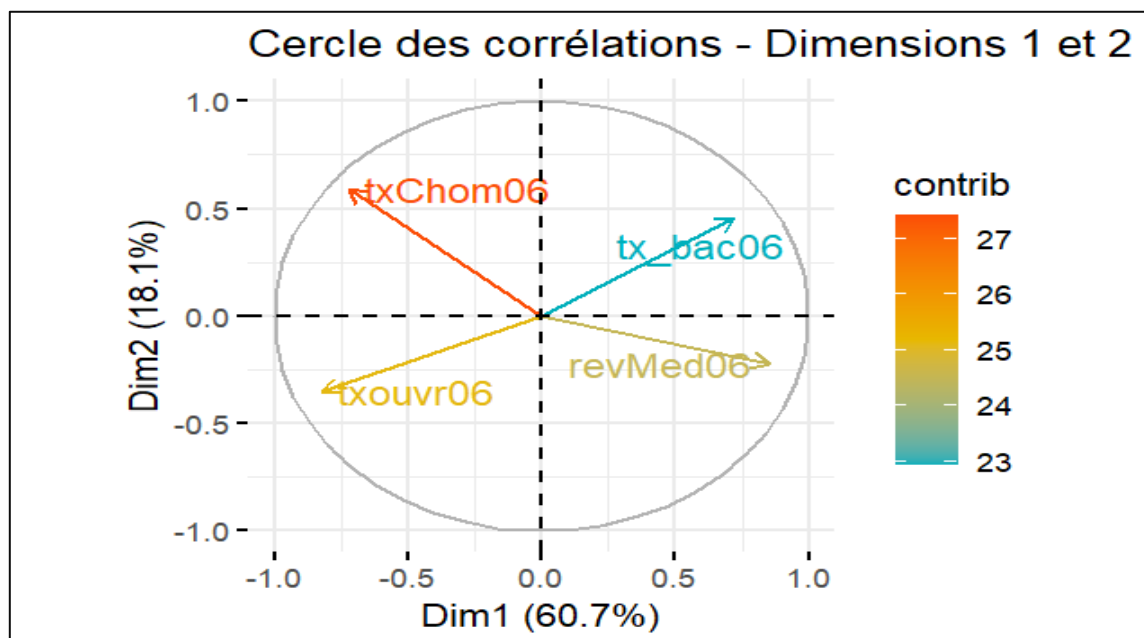


Figure 4 : Cercle de corrélations des dimensions 1 et 2 de l'ACP

Ce diagramme du cercle des corrélations issu de l'ACP (figure:4) entre les dimensions 1 et 2 montre une opposition forte entre les variables du FDep. La première dimension est fortement corrélées positivement au taux de chômage (txchom06) et au

taux d'ouvriers(txouvr06), et négativement au revenu médian (recMed06) et au taux de diplômés au bac (tx_bac06) .

Cette structuration valide l'interprétation de la première composante comme un axe de défavorisation sociale. En effet, le taux de chômage et le taux d'ouvrier positivement projetés contribuent fortement aux iris défavorisés alors que le revenu médian et le taux de bachelier négativement projetés caractérisent des territoires plus favorisés sur le plan social. Ainsi, la dimension 1 oppose la favorisation et la défavorisation des territoires.

- Résultats du fdep avec le taux de bachelier

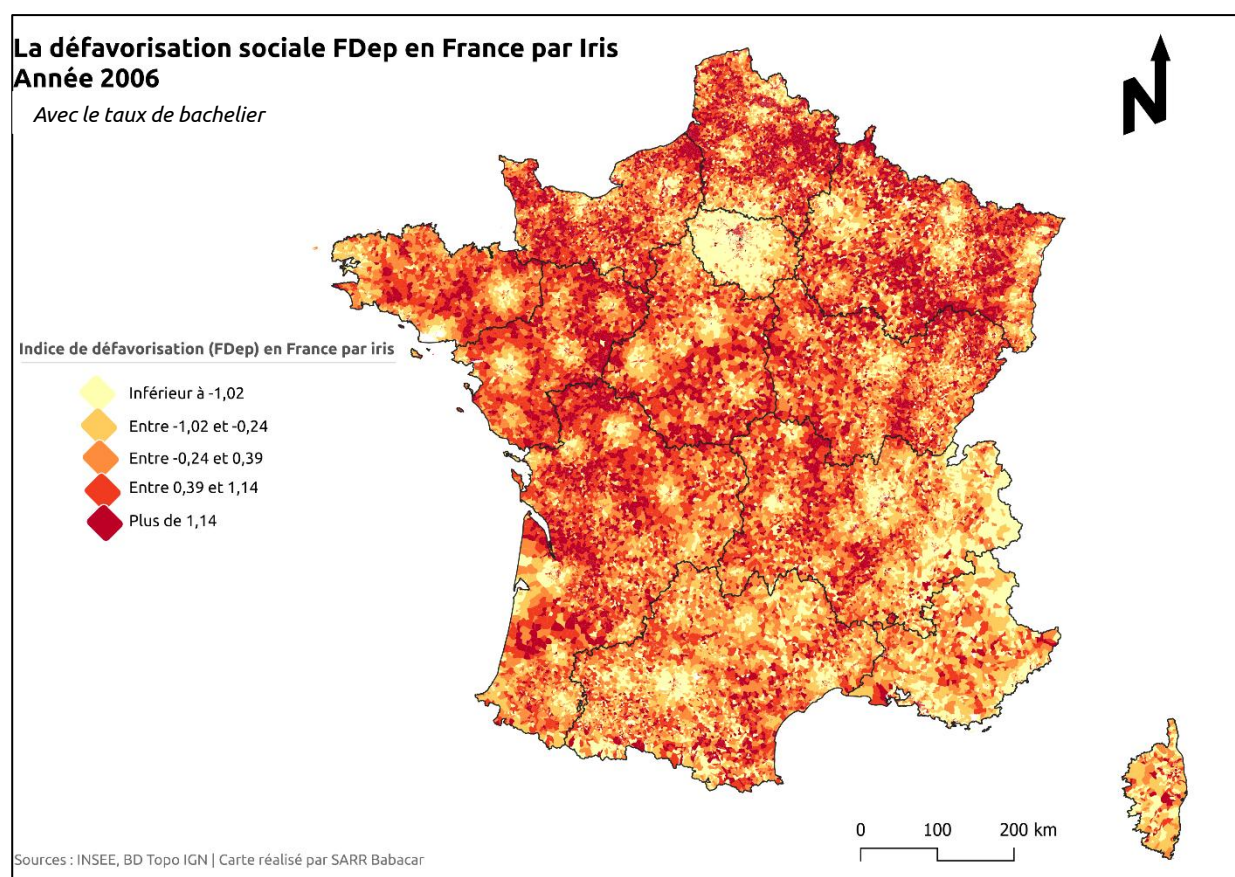
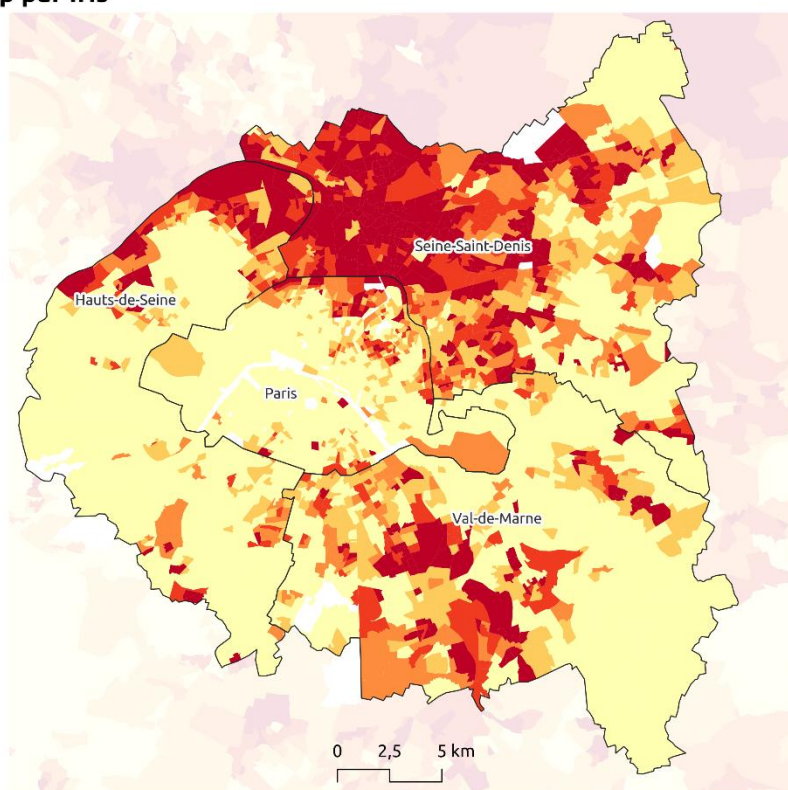


Figure 5 : Carte Fdep France en 2006 avec le taux de bac

La défavorisation sociale FDep par Iris
Zoom sur le 75, 92, 93, 94
Année 2006

Avec le taux de bachelier

Indice de défavorisation (FDep) par iris



Sources : INSEE, BD Topo IGN | Carte réalisée par SARR Babacar

Figure 6 : Carte Fdep du 75, 92, 93, 94 en 200INF avec taux de bachelier

- Résultats du fdep avec le taux de bac 2 ou plus

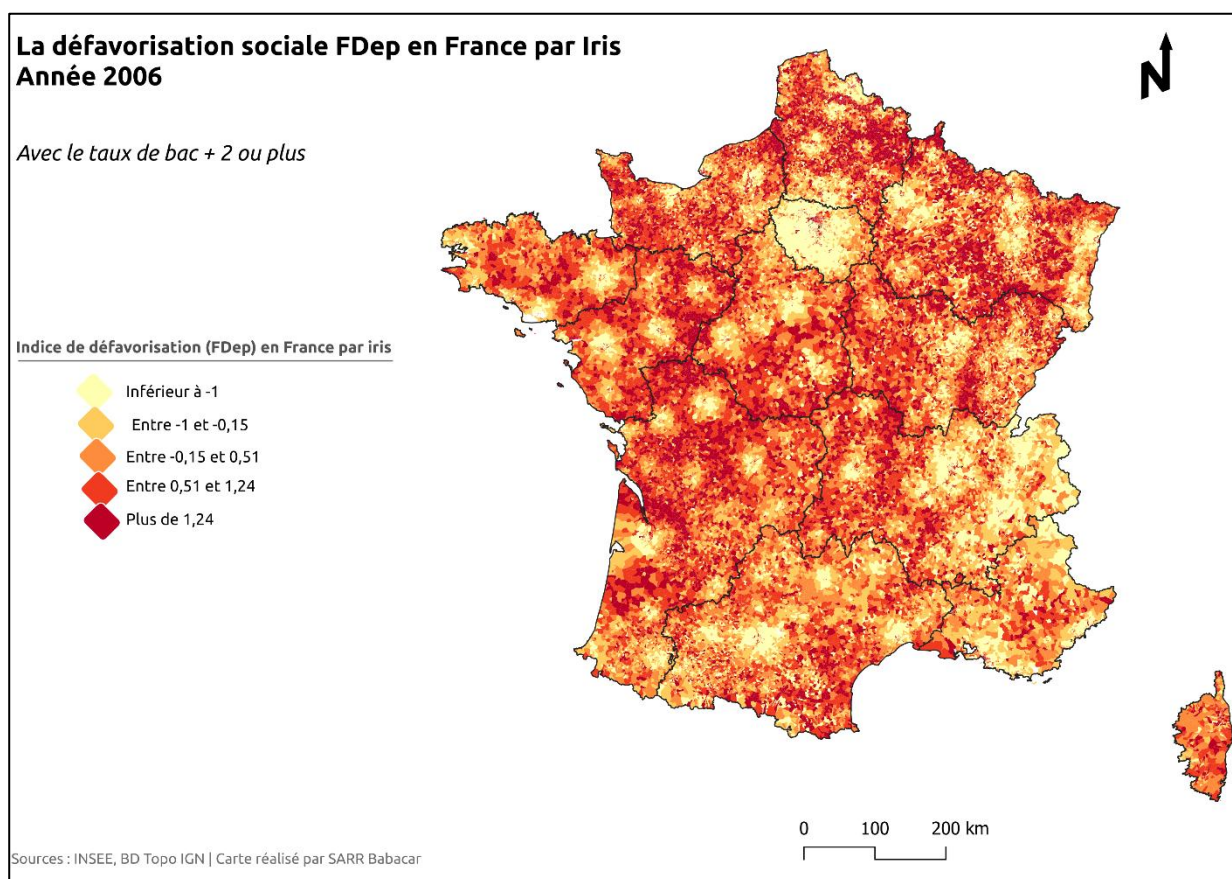


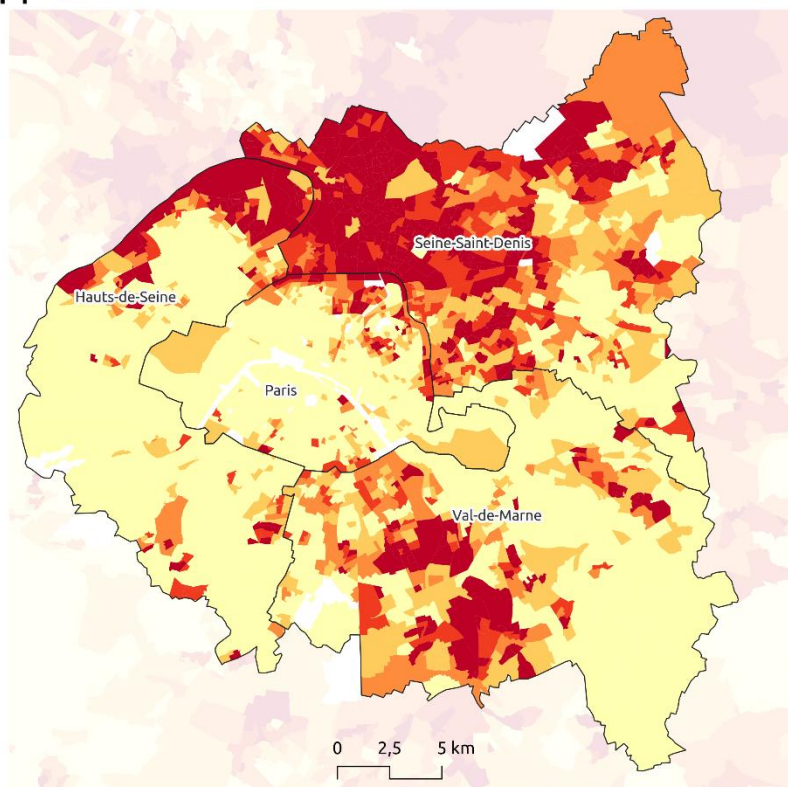
Figure 7 : Carte Fdep France en 2006 avec le taux de bac + 2 en 2006

La défavorisation sociale FDep par Iris
Zoom sur le 75, 92, 93, 94
Année 2006

Avec le taux de bac + 2 ou plus

Indice de défavorisation (FDep) par iris

- Inférieur à -1
- Entre -1, et -0,15
- Entre -0,15 et 0,51
- Entre 0,51 et 1,24
- Plus de 1,24



Sources : INSEE, BD Topo IGN | Carte réalisée par SARR Babacar

Figure 8 : Carte Fdep du 75, 92, 93, 94 en 2006 avec le taux de bac + 2 ou plus

- Comparaison du Fdep avec le taux de bachelier et le taux de bac+2

Le baccalauréat n'est plus un élément discriminant dans la mesure des inégalités sociales, c'est ainsi que la méthodologie de calcul du fdep a évolué. Maintenant le taux de bachelier est remplacé par le taux de bac + 2.

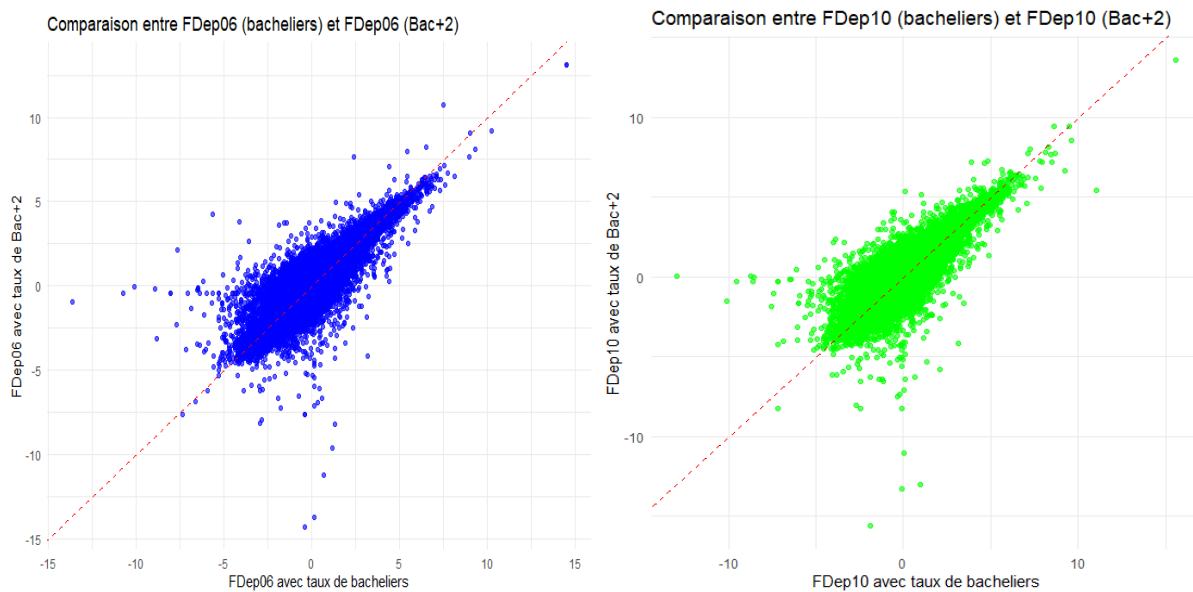


Figure 9 : Comparaison entre le Fdep avec le taux de bachelier et le Fdep avec le taux de bac + 2 en 2006 et 2010

3.1.2 Indice d'accessibilité aux services de soins

L'accessibilité à des services de soins constitue un enjeu majeur dans la santé publique en France. Si de nombreuses recherches démontrent que les inégalités sociales influencent la prise en charge de certaines maladies, la dimension géographique reste aussi un problème à résoudre avec une complexité d'échelle d'accessibilité. Cela s'explique par la variabilité des contextes géographiques et la difficulté d'utiliser des méthodes de mesures comparables et adaptées à tous les territoires. C'est dans cette même ordre d'idée de l'indice de Scale (Spatial aCcessibility multiscALar) développé par Ludivine Launay, Fabien Guillot, David Gaillard, Mohand Medjkane, Thierry Saint-Gérard, Guy Launoy et Olivier Dejardin a été calculé à l'échelle infra-communale donnant une représentation fine des territoires Il permet de visualiser l'accessibilité globale en intégrant l'ensemble des professionnels de la santé. Il définit l'accessibilité comme une notion multifactorielle. L'ensemble des données est disponible sur la plateforme data-gouv¹¹ sur lien suivant : <https://www.data.gouv.fr/datasets/accessibilite-aux-soins-de-premier-recours-en-france-metropolitaine/>. L'objectif est à partir de ces données de calculer l'indice moyen de sca <https://www.data.gouv.fr/datasets/accessibilite-aux-soins-de-premier-recours-en-france-metropolitaine/>. le par iris à l'échelle de la France afin d'avoir un score d'accessibilité pour chaque iris français.

3.1.2.1 Matériel et méthode

Les bâtiments résidentiels sont les points de références de cette méthodologie au lieu d'utiliser l'échelle administrative. La population est localisée au centre des zones résidentielles définies par agrégation du bâti à partir du BD topo IGN. Les équipements de santé comprennent l'essentiel des soins de premiers secours : les masseurs-kinésithérapeutes, les médecins-généralistes, les chirurgiens-dentistes, les infirmiers, les pharmacies, les établissements de santé court-séjour, les spécialistes en pédiatrie, en gynécologie médicale ou obstétrique et les maternités, les urgences de la BPE de l'INSEE en 2013.

¹¹ C'est le portail officiel français des données ouvertes. Il rend accessible à toutes les données publiques produites par les collectivités, administrations, services publics etc.

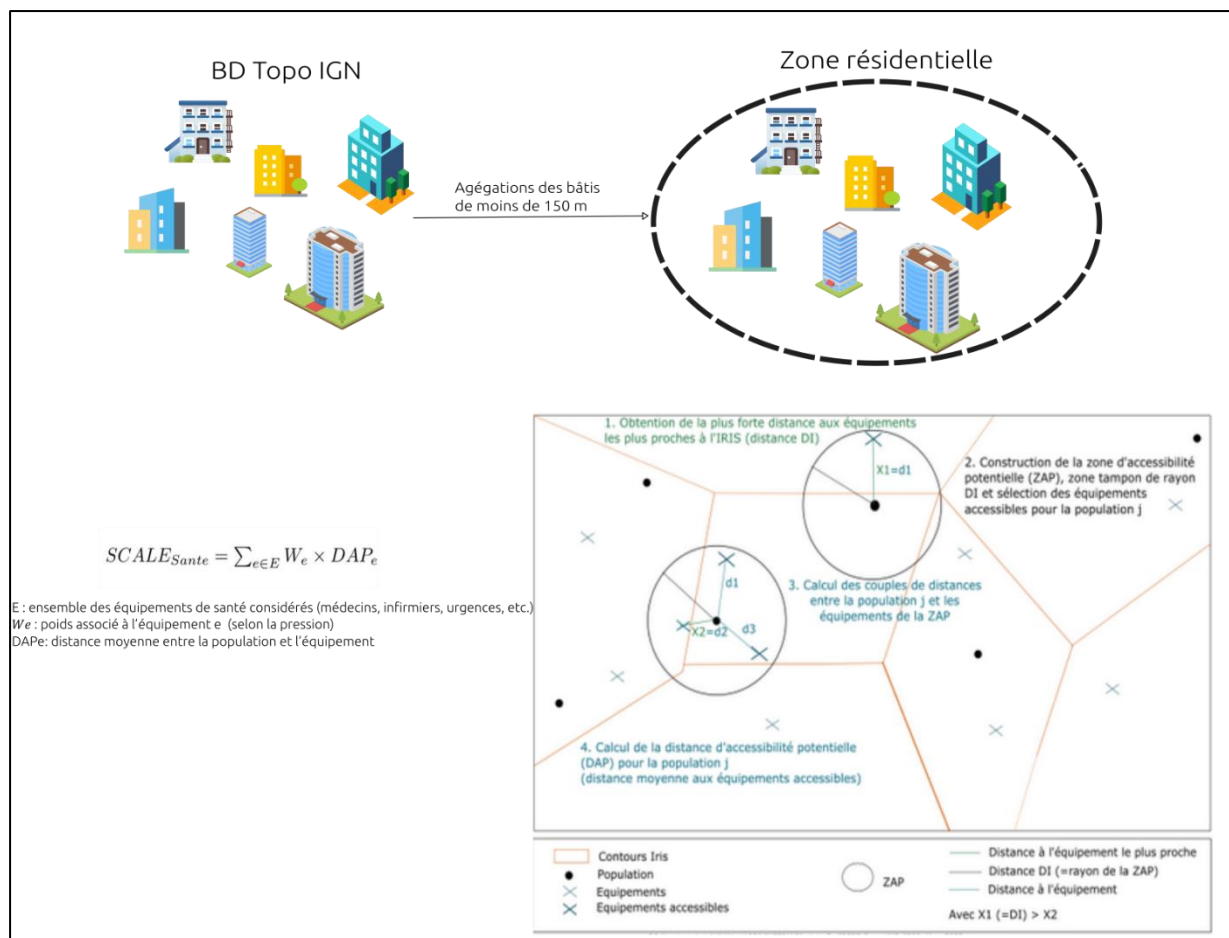


Figure 10 : Méthode de construction de l'indice de Scale

Cette méthodologie de scale est aussi appliquée dans le métropole du Grand Paris avec seulement l'accès aux médecins généralistes et pharmacies.

3.1.2.2 Résultats

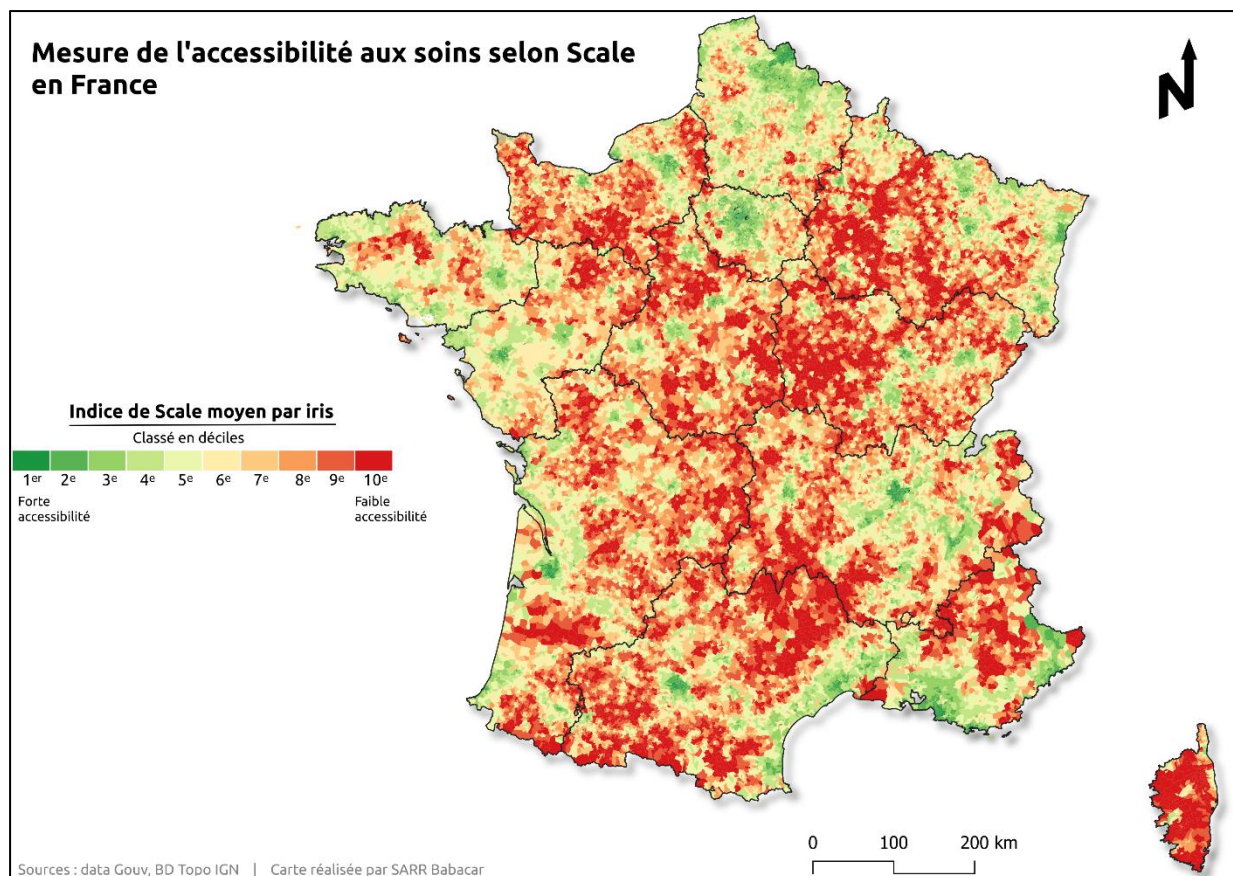


Figure 11 : Carte de l'indice de scale moyen par iris en France

La cartographie à l'échelle de la France métropolitaine démontre une disparité de l'accessibilité aux équipements de santé. De manière générale, les iris des pôles urbains et métropoles (métropole du Grand Paris, métropole européenne de Lille, métropole du Grand Lyon etc.) offre une bonne accessibilité aux équipements sanitaires, cependant, plusieurs iris présentent l'effet contraire surtout dans les zones plus reculées (massif central, sud du calvados etc.) Cela montre des disparités d'accessibilité aux soins à l'échelle des iris en France. Parmi les zones avec aussi une faible accessibilité aux soins on peut noter l'axe du Contentin à au centre val de Loire et une partie du champagne ardenne. Une seconde zone est qui se trouve dans l'ensemble alpin et dans les Cévennes présente aussi des incs d'accesibilité faible à l'échelle des iris.

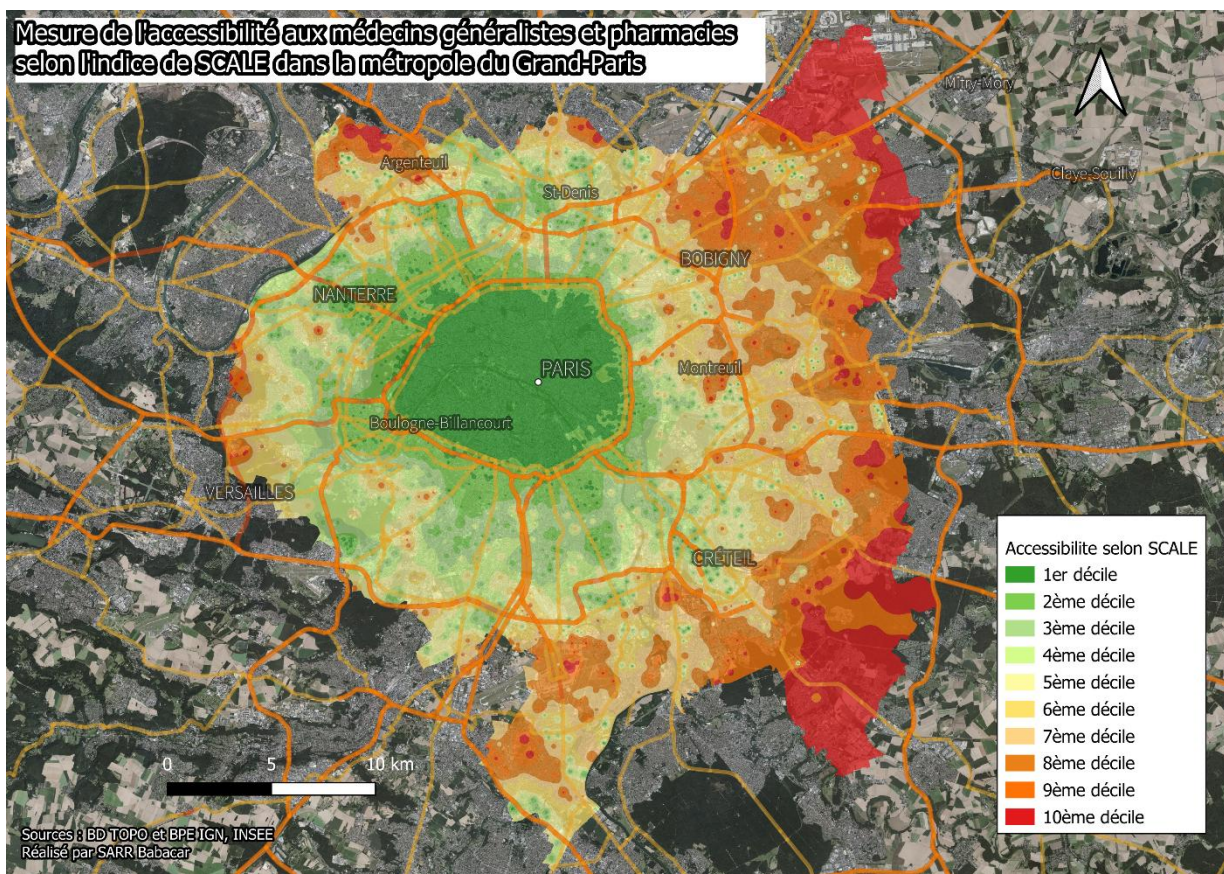


Figure 12 : Carte de la mesure de l'accessibilité aux médecins généralistes et pharmacies selon l'indice de Scale dans la métropole du Grand Paris

Une autre approche cartographique est réalisée à travers cette méthodologie, c'est l'approche par EPCI¹² en tenant en compte que deux attributs : médecin généraliste et pharmacie à l'échelle de la métropole du Grand Paris. Cette carte montre une disparité d'accessibilité à des médecins généralistes et pharmacies. Le centre de Paris est plus accessible, et de manière générale l'accessibilité est concentrique par rapport à Paris plus on s'éloigne du centre plus l'accessibilité diminue. La faible de l'accessibilité est plus marquée vers le sud et nord-est du métropole mais elle reste néanmoins accessible à grande échelle.

¹² Etablissement public de coopération intercommunale

3.2 Indicateurs de l'environnement physique

Pour mesurer le niveau d'exposition à l'environnement physique impactant la santé respiratoire, elle est nécessaire de quantifier la végétation autour de ces lieux c'est à la couverture végétale et des indicateurs d'occupation du sol pour mesurer les typologies d'occupation selon les iris afin de savoir la fonction des proportions de l'utilisation du sol. C'est un élément essentiel car le milieu de vie détermine beaucoup l'influence sur la santé surtout celle liée à la respiration.

3.2.1 Typologie de l'occupation des sols

3.2.1.1 Matériel et méthode

Les données de l'occupation du sol sont issues du corine land cover diffusé par le programme européen Copernicus. C'est une base disponible sous forme de polygones de l'utilisation et l'occupation du sol selon des nomenclatures. Pour les besoins de ce travail nous avons besoin uniquement certaines classes correspondant aux espaces artificialisés et aux espaces verts urbains. Les codes dans le tableau 5 suivant ont été sélectionnés :

Tableau 5 : Tableau des codes du Corine Land Cover

| Codes Corine Land Cover | Libellé officiel | Catégorie regroupée |
|-------------------------|--|---------------------------|
| 111, 112 | Tissus urbains continus et discontinus | Urbanisation |
| 121, 122 | Zones industrielles et commerciales ; Réseaux routiers et ferroviaires et espaces associés | Industriel |
| 131, 132, 133 | Extraction de matériaux, Décharges, Chantiers | Mines et chantiers |
| 141, 142 | Espaces verts urbains ; Équipements sportifs et de loisirs | Espaces verts |

Afin d'estimation la proportion d'occupation du sol par type au sein des iris en France, l'intersection spatiale est réalisée entre les polygones des iris et du corine land cover filtrées par types. Ce traitement permet de découper les polygones du corine land cover à l'intérieur des iris et de calculer en même temps les surfaces artificialisées correspondantes. Les résultats de chaque surface est exprimé en hectare. Ensuite pour chaque iris, les surfaces intersectées issues du CLC sont agrégées par catégories

(urbanisation, industriel, mines, et espaces verts) comme dans la catégorie regroupée dans le tableau 5. Ces surfaces sont rapportées à la surface totale des iris afin d'obtenir le taux d'occupation du sol pour chaque catégorie en pourcentage. Ainsi dans chaque table on aura la part en hectare de la surface de l'iris, la surface artérialisée selon la catégorie (urbanisation, mines etc.) en hectares et la proportion d'artificialisation de chaque catégories par iris en pourcentage en France selon le corine land cover.

3.2.1.2 Résultats

Les quatre graphiques suivantes (figures 13, 14, 15, 16) démontrent que l'occupation du sol en France est globalement très inégale selon les iris. La majorité des iris sont très faiblement occupées par les activités minières et industrielles alors que la répartition entre espaces verts et zones urbanisées dessine une géographie à deux vitesses.

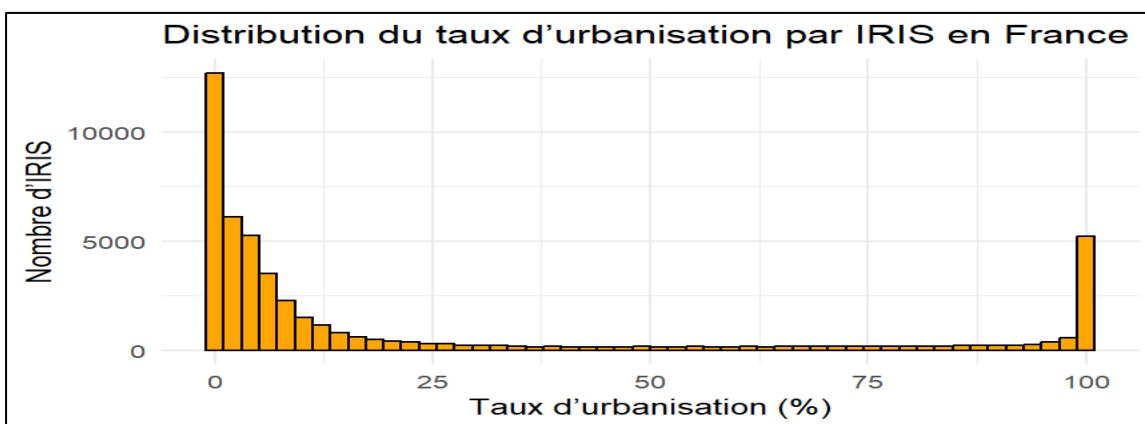


Figure 13 : Distribution du taux d'urbanisation par iris en France

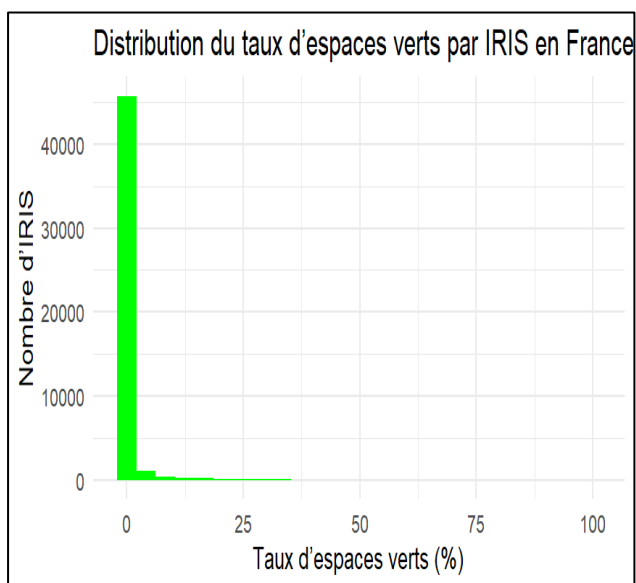


Figure 15 : Distribution du taux d'espaces verts par iris en France

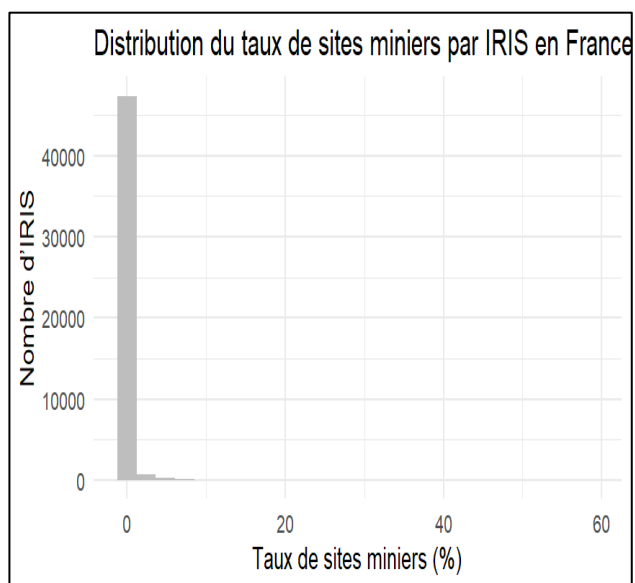


Figure 14 : Distribution du taux de sites miniers par iris en France

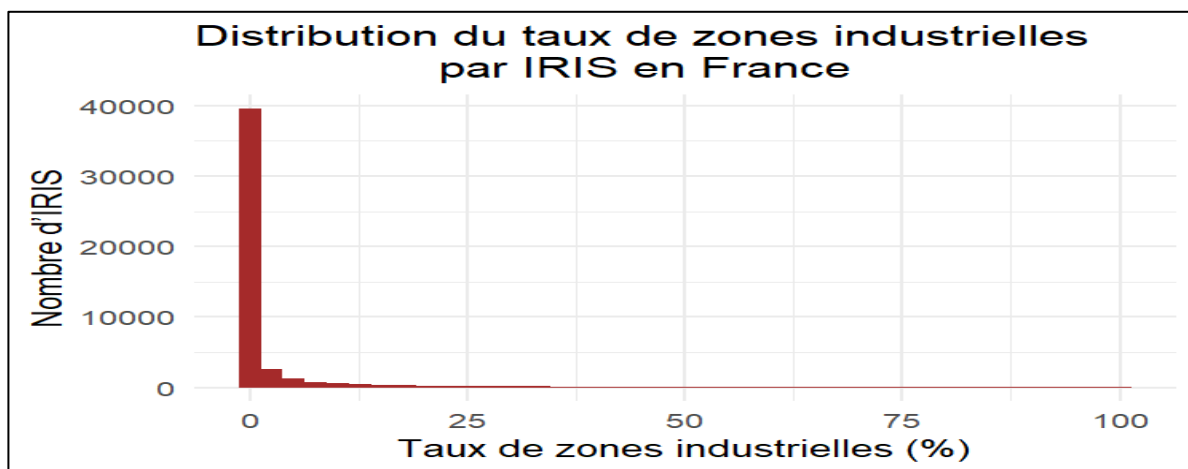


Figure 16 : Distribution du taux de zones industrielles par iris en France

3.2.2 Couverture végétale

Pour mesurer la couverture végétale, les espaces verts et la santé végétale sont des éléments à tenir en compte pour quantifier cette couverture à l'échelle des iris. C'est pour cela le NDVI¹³ est un outil de mesure efficace en télédétection afin de classer la santé et la couverture de la végétation à grande échelle. Il peut être utilisé pour la surveillance de la végétation, la différenciation des types de végétation,

3.2.2.1 Matériel et méthode

Cette méthodologie se repose sur l'outil Google Earth Engine pour calculer et analyser l'indice de la végétation directement avec ses bibliothèques de photographie aérienne sur plusieurs années. Ceci permettra de faire des mises à jour sur l'année ou la période souhaitée automatiquement à l'échelle de chaque iris en France métropolitaine. Le NDVI est un indice de végétation utilisé pour analyser la vigueur, d'évaluer la densité et la santé de la végétation à partir d'images satellitaires. Sa formule

Les images Landsat 8 de la période choisie (2019) sont chargées sur GEE¹⁴. Elles sont ensuite filtrées selon la date et la zone géographique, puis nettoyées pour supprimer les nuages et les ombrages afin d'avoir des images aériennes propres. Les valeurs spectrales sont ensuite converties en réflectance normalisée en les divisant par 10000. Cette procédure permet de travailler avec des données homogènes.

Le NDVI calculé ici à partir de LANDSAT 8, la formule de calcul est la suivante :

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

NIR = la bande du proche infrarouge

RED = la bande du spectre du visible

¹³ Normalized Difference Vegetation Index

¹⁴ Google Earth Engine

| Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) Launched February 11, 2013 | Bands | Wavelength (micrometers) | Resolution (meters) |
|---|-------------------------------------|-------------------------------------|--------------------------------|
| | Band 1 - Coastal aerosol | 0.43 - 0.45 | 30 |
| | Band 2 - Blue | 0.45 - 0.51 | 30 |
| | Band 3 - Green | 0.53 - 0.59 | 30 |
| | Band 4 - Red | 0.64 - 0.67 | 30 |
| | Band 5 - Near Infrared (NIR) | 0.85 - 0.88 | 30 |
| | Band 6 - SWIR 1 | 1.57 - 1.65 | 30 |
| | Band 7 - SWIR 2 | 2.11 - 2.29 | 30 |
| | Band 8 - Panchromatic | 0.50 - 0.68 | 15 |
| | Band 9 - Cirrus | 1.36 - 1.38 | 30 |
| | Band 10 - Thermal Infrared (TIRS) 1 | 10.60 - 11.19 | 100 |
| | Band 11 - Thermal Infrared (TIRS) 2 | 11.50 - 12.51 | 100 |

Tableau 6 : Les différentes bandes de Landsat 8

Pour faire une analyse infracommunale, plus précisément à l'échelle des iris, une analyse statistique zonale est réalisée, chaque polygone des iris sert de zone d'agrégation où la valeur moyenne du NDVI est calculée avec une résolution spatiale de 30m comme la résolution du Landsat 8. Ainsi, les résultats sont enregistrées sous une nouvelle couche contenant la valeur moyenne du NDVI de chaque iris en France.

3.2.2.2 Résultats

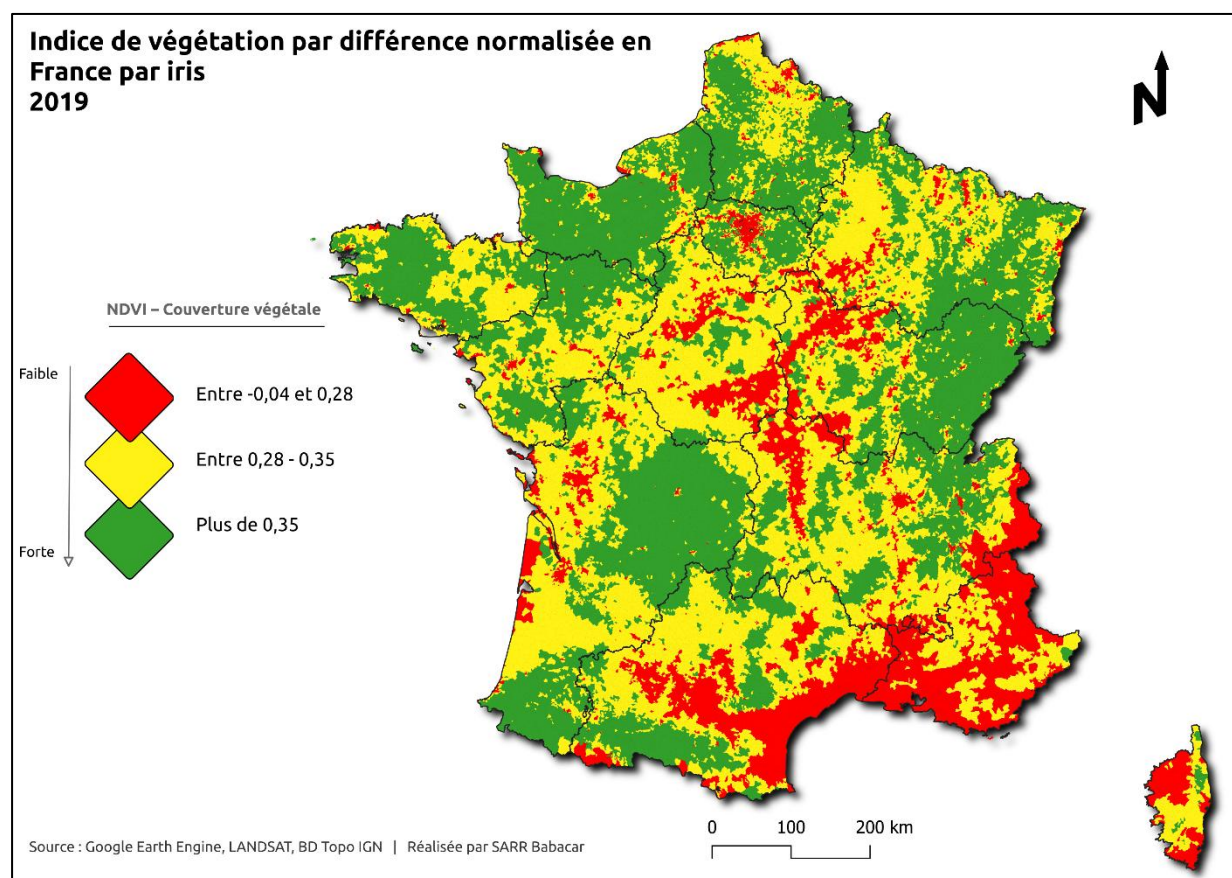


Figure 17 : Carte du NDVI moyen par iris en France en 2019

Cette cartographie de l'indice de végétation par différence normalisée (NDVI) représente le NDVI moyen par iris en France à l'été 2019. L'indice varie de -1 à 1, les valeurs proches de 1 indiquent une végétation saine et dense et les valeurs -1 indiquent une végétation moins dense, peu saine ou inexistante. En s'appuyant sur la cartographie, on peut trouver trois niveaux de densité du NDVI moyen par iris. Les valeurs des zones les plus denses et saines sont supérieures à 0,35 comme une partie du massif central, le pays d'auge en Normandie, le Sud-Ouest de la France. On note les indices les plus faibles dans les agglomérations denses avec beaucoup d'artificialisations des sols, comme Paris, Lyon, les métropoles de Marseille, Montpellier et Toulouse et aussi les zones des cours d'eau et des montagnes de glace comme les Alpes.

3.3 Indicateur de l'environnement chimique

3.3.1 Proximité aux sites industriels

La théorie e de l'exposome de Wild décrit les expositions extérieures de l'individu par rapport à son environnement de vie. C'est ainsi que le milieu de vie constitue un élément déterminant dans la santé respiratoire.

3.3.1.1 Matériel et méthode

Cette méthodologie vise à faire des zones tampons autour des adresses afin de créer une base de données comportant la proximité de chaque adresse par rapport aux zones industrielles en faisant des zones tampons de 500 et 1000 m.

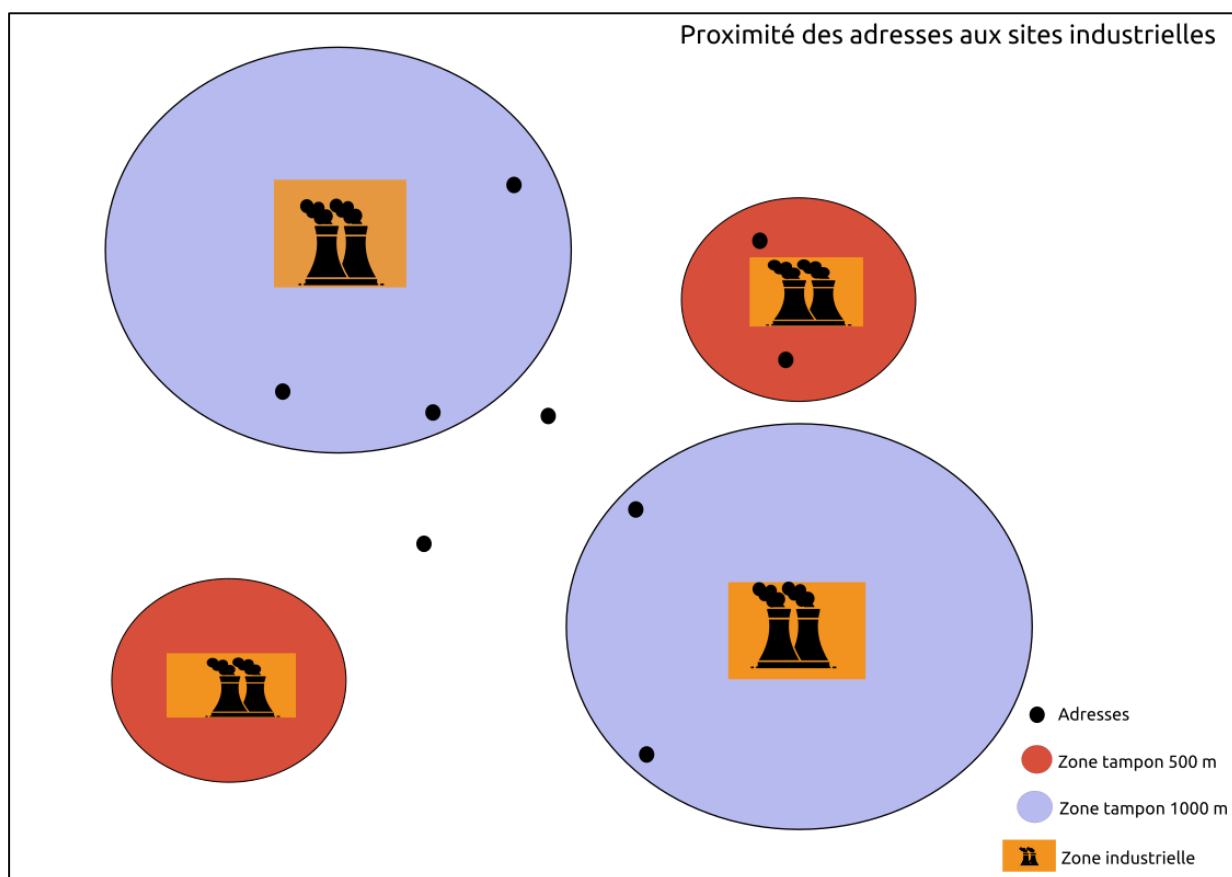


Figure 18 : Méthodologie de construction de la proximité des adresses aux sites industriels

La première étape consiste à charger les couches nécessaires à l'étude : les données vectorielles des limites des zones industrielles, le département d'étude et la base des adresses cadastrales de la zone d'étude ici considérée comme des exposés par rapport aux sites industriels. Les données ont été ensuite harmonisées dans la même projection cartographique (Lambert 93 EPSG : 2154) ce qui permet d'exprimer les distances en mètres pour correctement réaliser les calculs.

Ensuite la deuxième étape constitue à créer des zones tampons (buffers) de 500 et 1000 m autour des zones industrielles correspondant à une proximité aux sources de pollutions. Quand les zones tampons de plusieurs sites se chevauchent, les buffers sont fusionnés afin de constituer des limites d'enveloppe unique. Ensuite, la jointure spatiale est réalisée entre les adresses et les zones tampons, ce géotraitement permet d'attribuer chaque adresse deux variables booléens (proche 500m et proche 1km) indiquant si l'adresse est à l'intérieur des ou no de la zone 500m ou 1000m. Ainsi chaque adresse est caractérisée par son de proximité à une zone industrielle le plus proche.

3.3.1.2 Résultats

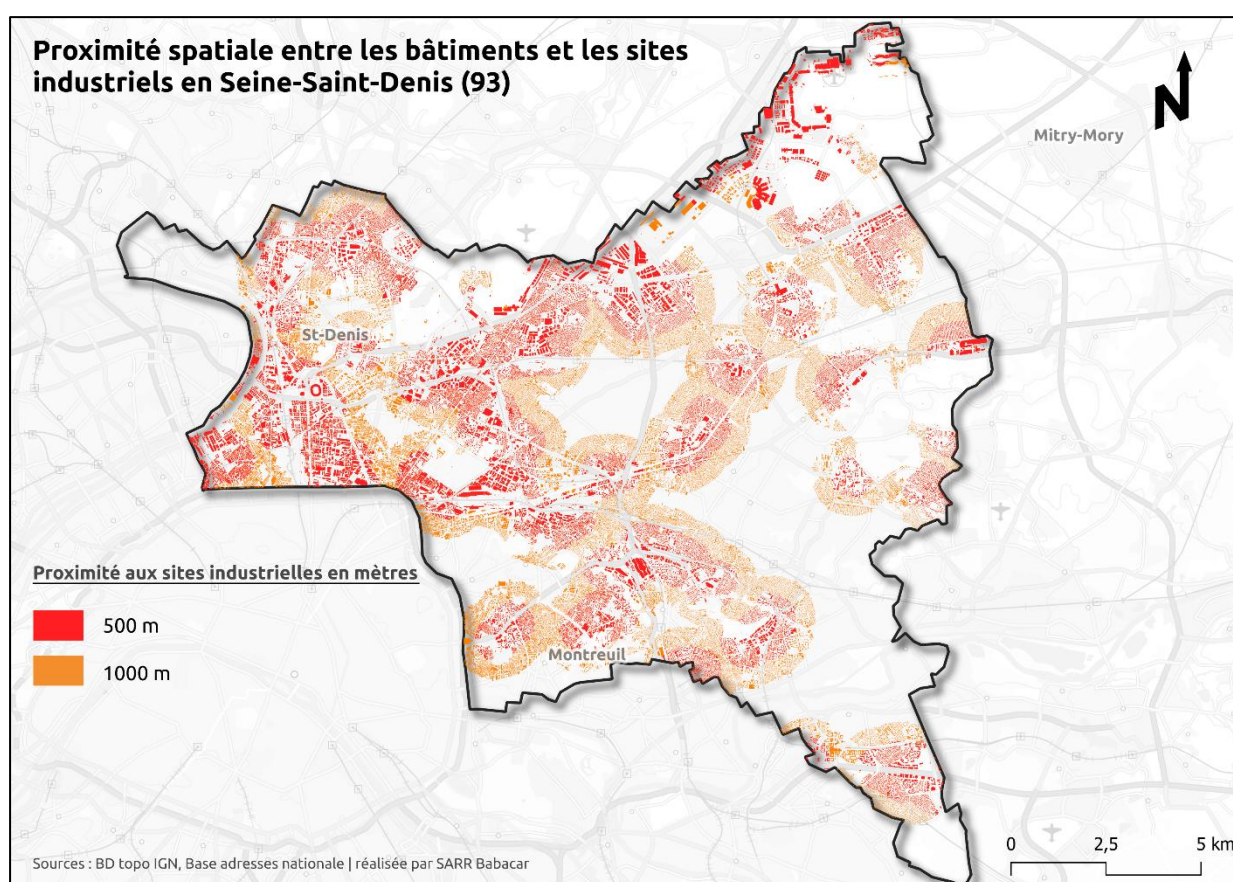


Figure 19 : Carte de la proximité des adresses aux sites industriels dans la Seine Saint-Denis

Cette carte matérialise la méthodologie de construction de la proximité des adresses aux sites industrielles. On constate ici dans la Seine-Saint Denis que beaucoup d'adresses sont exposées à de potentielles pollution venant de ses zones industrielles. Ainsi cette exposition par rapport au milieu de vie des individus détermine fortement la santé respiratoire.

La construction de cette méthodologie adaptable dans plusieurs départements en France va permettre à l'EGEA à travers les adresses de l'histoire résidentiel des asthmatiques de faire de potentiels corrélation entre le milieu de vie des patients et leur localisation.

4. Discussion et perspectives

4.1 Apports pour une lecture territoriale des inégalités de santé respiratoire (asthme)

L'objectif de ce stage était de construire des bases de données spatialisées avec des indicateurs sur les inégalités d'exposition socio-environnemental liées à l'asthme. Afin de venir en appui le projet MultiSocialEGEA, l'approche spatiale mobilisée a permis de mettre en évidence des contrastes territoriaux sur les niveaux d'exposition potentielles des populations aux facteurs socio- environnementaux tels que la pauvreté, les pollutions, l'accessibilité, le type d'occupation des sols et la couverture végétale susceptibles d'affecter la santé respiratoire. Une étude sur les inégalités sociales a été réalisée par le biais de l'indice de défavorisation social (Fdep) sur plusieurs années ce qui permettra à d'étudier la corrélation à travers l'histoire résidentielles des individus du cohorte EGEA. Cet indicateur construit à travers l'analyse par composante principale du taux de chômage, d'ouvrier, de bachelier ou bac + 2 ou supérieur et le revenu médian permet de savoir le niveau de défavorisation sociale selon chaque iris en France à travers plusieurs années. Ensuite l'accessibilité aux infrastructures est aussi mesurée à travers le scale, les résultats montrent à travers la France le maillage territorial de l'accessibilité à des services de soins de premiers recours et les inégalités d'accessibilité à travers les iris. Les résultats sur la proximité aux zones industrielles et l'analyse d'indicateurs environnementaux tels que le NDVI, contribuent à affiner la compréhension des inégalités environnementales sur le type de milieu de vie immédiat et la couverture végétale. Ainsi, tous ces résultats permettent de repérer de manière générale les zones à risques. En effet, l'établissement de la corrélation entre les inégalités socio-environnementales et la santé respiratoire, en tenant compte des niveaux d'exposition selon les zones (urbain, rural, périurbain) peuvent permettre de mieux identifier les zones les plus vulnérables. L'analyse territoriale met en évidence une superposition de vulnérabilités socio-environnementales des territoires qui cumulent plusieurs facteurs de risques. En plus, elle apporte une compréhension plus fine des déterminants sociaux et environnementaux en mettant en évidence les effets de localisation de l'espace sur la santé respiratoire notamment l'asthme. Ainsi, elle constitue un outil primordial pour orienter les politiques publiques dans l'aménagement du territoire à travers la distribution des infrastructures sanitaires au niveau national, la réduction des inégalités entre le milieu urbain et le milieu rural, de renforcer la végétation et les espaces verts dans les grandes villes majoritairement très artificialisées.

4.2 Limites

Les enquêtes de la cohorte de l'EGEA ont commencé de 1992 EGEA 1 à EGEA 4 qui est en cours depuis 2023. Cependant certains indicateurs comme l'indice de défavorisation sociale (Fdep) ne couvre pas la totalité des années de l'enquête (donnée disponible à partir de 2006 sur INSEE). Ainsi, pour les données de 1992 à 2006 pour calculer le Fdep, nous avons fait la commande des données anciennes de l'INSEE sur Progedo ce qui va être disponible dans les prochains mois pour compléter ce travail afin de couvrir toutes les années de l'enquête.

4.3 Perspectives dans le cadre du projet MultiSocialEGEA

Ce travail permet de faire avancer le projet de recherche de MultiSocialEGEA qui par une approche interdisciplinaire croise les données médicales, sociales et environnementales. Les recherches du projet MultiSocialEGEA vont faire l'objet de publications scientifiques et de participation à des programmes européens et l'intégration dans des projets de transition écologique et de justice sociale et environnementale. Ainsi ce travail de recherche pose les bases d'une analyse spatiale afin de faire des corrélations à travers la cohorte de MultiSocialEGEA entre l'asthme et les déterminants socio-environnementaux.

5. Conclusion

En somme, ce stage s'inscrit dans la dynamique de recherche interdisciplinaire qui a pour objectif de mieux comprendre les liens entre les facteurs sociaux, environnementaux et les inégalités de santé respiratoire notamment l'asthme. En mobilisant le cadre scientifique du projet MultisocialEGEA, il s'agit ici de dépasser la dimension purement biomédicale de l'asthme pour intégrer la dimension de la lecture territoriale des expositions qui pèsent sur les individus. Cette approche repose sur l'idée que les conditions de vie, le lieu de résidence, la qualité de l'environnement ou encore l'accessibilité au soin sont autant des déterminants qui catalysent et favorisent le développement des maladies respiratoires. Dans ce contexte, la construction d'indicateurs socio-environnementaux rétrospectifs à l'échelle des iris couvrant plusieurs années des enquêtes de l'EGEA constitue l'élément central de ce travail. Ces indicateurs sociaux (la défavorisation sociale, l'accessibilité aux services de santé) environnementaux (occupation du sol, couverture végétale) ou encore liés à des sources de pollution (proximité aux sites industriels), permettront d'alimenter une base de données spatiale afin de faire une lecture territoriale des inégalités d'exposition. Ainsi, elle vise à documenter l'exposition différenciée des individus en fonction de leur milieu de vie et de leur trajectoire résidentielle.

Au-delà de leur apport scientifique, ces indicateurs présentent des leviers opérationnels pour les politiques de l'aménagement du territoire. Ils permettent d'orienter les politiques publiques de prévention en identifiant les contextes territoriaux des territoires vulnérables, et en développant des stratégies d'adaptation afin de réduire les inégalités sociales et environnementales face à l'asthme. Bien que certaines limites demeurent dans ce travail notamment lié à la disponibilité des données, le travail de recherche effectué constitue une première étape vers une compréhension plus fine des relations entre territoire et santé.

Enfin, ce mémoire contribue à alimenter les recherches menées dans le cadre du projet MultiSocialEGEA en posant les bases d'analyses futures. Les perspectives de ce travail sont le perfectionnement des indicateurs construits et la couverture de tous les indicateurs à toutes les phases des études de la cohorte de l'EGEA aussi l'intégration de nouvelles dimensions de l'exposition comme par exemple l'EDI (**European Deprivation Index**). À terme, cette approche intégrée pourrait jouer un rôle majeur dans l'élaboration des politiques de santé publique plus équitables et surtout plus efficaces en tenant en compte les facteurs socio-environnementaux qui déterminent la santé respiratoire des populations.

Bibliographie

1. E. Bouzigon, R. Nadif, N. Le Moual, M.-H. Dizier, H. Aschard, A. Boudier, J. Bousquet, S. Chanoine, C. Donnay, O. Dumas, F. Gormand, B. Jacquemin, J. Just, P. Margaritte-Jeannin, R. Matran, C. Pison, E. Rage, M. Rava, C. Sarnowski, L.A.M. Smit, S. Temam, R. Varraso, L. Vignoud, M. Lathrop, I. Pin, F. Demenais, F. Kauffmann, V. Siroux. (2015). Facteurs génétiques et environnementaux de l'asthme et de l'allergie : synthèse des résultats de l'étude EGEA. *Revue des Maladies Respiratoires*, <http://dx.doi.org/10.1016/j.rmr.2014.12.005>
2. A. Guillien, A. Bédard, O. Dumas, J. Allegre, N. Arnault, A. Bochaton, N. Druesne-Pecollo, D. Dumay, L. K. Fezeu, S. Hercberg, N. Le Moual, H. Pilkington, S. Rican, G. Sit, F. Szabo de Edelenyi, M. Touvier, P. Galan, T. Feuillet, R. Varraso, V. Siroux. *Exposome Profiles and Asthma among French adults*.
3. Guillien, A., Nadif, R., Kauffmann, F., & Bouzigon, E. (2020). Exposome and asthma: Results from the EGEA study. *Environmental Research*, 191, 110055. <https://doi.org/10.1016/j.envres.2020.110055>
4. France Stratégie. (2022, mars). *Améliorer la santé et le bien-être au travail : Rapport final*. France Stratégie.
5. Launay, L., et al. (2024). Profile of exposures and lung function in adults with asthma: An exposome approach in the EGEA study. *International Journal of Environmental Research and Public Health*.
6. Launay, L., et al. (2019). SCALE : A spatial accessibility index for primary health care in France. *PLOS ONE*, 14(9), e0221417. <https://doi.org/10.1371/journal.pone.0221417>
7. Rican, S., et al. (2014). Inégalités géographiques de santé en France. *Les Tribunes de la santé*, 43(1), 43–52.
8. Mbath. O Géographie de l'asthme en France : distribution spatiale des facteurs de risque de l'asthme d'après l'étude Nutrinet-Santé. (2016).

Sitographie

Le Temps : <https://www.letemps.ch/societe/mode-vie-occidental-rime-souvent-crisis-dasthme>

Santé publique France : <https://www.santepubliquefrance.fr/revues/articles-du-mois/2024/inequites-sociales-et-environnementales-en-france-continentale-une-analyse-de-l-exposition-a-la-chaaleur-a-la-pollution-de-l-air-et-au-manque-de>

Unicaen :
<https://unicaen.maps.arcgis.com/apps/MapJournal/index.html?appid=aeab64aec57442f0b37ec61a47a43fa6>

EGEA : <https://cohorte-egea.fr/fr/la-cohorte/objectifs-et-suivis>

Labiratoire IDEES Caen : <https://idees.unicaen.fr/multisocialegea-inegalites-sociales-et-asthme/>

Hal Open Science : <https://hal.science/hal-03003960/file/exposome.pdf>

Sources de données SIG

BD TOPO IGN : <https://geoservices.ign.fr/bdtopo>

INSEE : <https://www.insee.fr/fr/accueil>

Google Earth Engine : <https://earthengine.google.com/>

Corine Land Cover : <https://land.copernicus.eu/en/products/corine-land-cover>

Annexes

Annexe 1 : Scripts

Annexe 1.1 Code R de création de l'indice de défavorisation sociale (Fdep)

- Avec le taux de bachelier en 2006 (imputation des données manquantes)

```
# ---- Calcul du French deprivation index (FDep06) ----#
#-----Rey et al. (2009) - BMC public health-----#
# installation des packages
install.packages("FactoMineR", dependencies = TRUE)
install.packages("readxl", , dependencies = TRUE)
install.packages("writextl")
install.packages("factoextra", , dependencies = TRUE)

# package
library(FactoMineR)
library(readxl)
library(missMDA)
library(factoextra)

# data
varIris <- read_excel("Data/varIris06.xlsx")
revCom <- read_excel("Data/revCom06.xlsx")
revIris <- read_excel("Data/revenu_mediane_2006.xlsx")

# Calcul du FDep 2006
# Jointures
t <- merge(x = varIris, y = revIris, by="IRIS", all.x=T)
varIris$revIris <- t$med_iris

t <- merge(x = varIris, y = revCom, by="COM", all.x=T)
varIris$revCom <- t$mediane

varIris$revMed06 <- varIris$revIris
```

```
varIris$revMed06 <-
ifelse(is.na(varIris$revMed06), varIris$revCom, varIris$revIris)
# revenu des communes quand no data à l'IRIS

# Sélection des variables pour le FDep
dtFDep <-
varIris[,c("IRIS", "Pop06", "tx_bac06", "revMed06", "txChom06", "txouvr06")]

# Conversion en numérique
dtFDep$tx_bac06 <-
as.numeric(as.character(dtFDep$tx_bac06))

dtFDep$revMed06 <-
as.numeric(as.character(dtFDep$revMed06))

dtFDep$txChom06 <-
as.numeric(as.character(dtFDep$txChom06))

dtFDep$txouvr06 <-
as.numeric(as.character(dtFDep$txouvr06))

X <- dtFDep[, 3:6]
X <- data.frame(lapply(X, function(x)
as.numeric(as.character(x))))) # conversion propre

# Imputation des valeurs manquantes
res.impute <- imputePCA(X, scale = TRUE) # Imputation
PCA, bien meilleure
X.imputed <- res.impute$completeObs

# Calcul du FDep via ACP
dtFDep$Pop06 <-
as.numeric(as.character(dtFDep$Pop06))

res.pca <- PCA(X = X.imputed, scale.unit = TRUE, ncp = 1,
row.w = dtFDep$Pop06)

dtFDep$FDep06 <- -res.pca$ind$coord[, 1]

varIris$FDep06 <- dtFDep$FDep06
```

- Avec le taux de bachelier en 2006 (sans imputation des données manquantes)

```
# ---- Calcul du French deprivation index (FDep06) ----
# Rey et al. (2009) - BMC Public Health

# Installation des packages (à faire une seule fois si pas
encore installés)
install.packages("FactoMineR", dependencies = TRUE)
install.packages("readxl", dependencies = TRUE)
install.packages("writexl")
install.packages("factoextra", dependencies = TRUE)

# Chargement des bibliothèques
library(FactoMineR)
library(readxl)
library(factoextra)

# Chargement des données
varIris <- read_excel("Data/varIris06.xlsx")
revCom <- read_excel("Data/revCom06.xlsx")
revIris <- read_excel("Data/revenu_mediane_2006.xlsx")

# Jointures
t <- merge(x = varIris, y = revIris, by = "IRIS", all.x = TRUE)
varIris$revIris <- t$med_iris

t <- merge(x = varIris, y = revCom, by = "COM", all.x =
TRUE)
varIris$revCom <- t$mediane

# Construction de la variable de revenu médian
varIris$revMed06 <- varIris$revIris # on garde
uniquement revIris, sans fallback revCom
varIris$revMed06 <-
ifelse(is.na(varIris$revMed06), varIris$revCom, varIris$revIri
s) # revenu des communes quand no data Ã l'IRIS

# Création de la table FDep
dtFDep <- varIris[, c("IRIS", "Pop06", "tx_bac06",
"revMed06", "txChom06", "txouvr06")]

# Conversion des variables en numérique
dtFDep$tx_bac06 <-
as.numeric(as.character(dtFDep$tx_bac06))
dtFDep$revMed06 <-
as.numeric(as.character(dtFDep$revMed06))
dtFDep$txChom06 <-
```

```
as.numeric(as.character(dtFDep$txChom06))
dtFDep$txouvr06 <-
as.numeric(as.character(dtFDep$txouvr06))
dtFDep$Pop06 <-
as.numeric(as.character(dtFDep$Pop06))

# Résumé des valeurs manquantes
missing_counts <- sapply(dtFDep[, 3:6], function(x)
sum(is.na(x)))
missing_percents <- sapply(dtFDep[, 3:6], function(x)
mean(is.na(x)) * 100)

missing_summary <- data.frame(
  Indicateur = names(missing_counts),
  N_manquants = missing_counts,
  Pourcentage = round(missing_percents, 2)
)

print("Résumé des données manquantes par indicateur :")
print(missing_summary)

cat("Nombre d'IRIS initiaux :", nrow(dtFDep), "\n")

# Suppression des lignes incomplètes (sans imputation)
dtFDep_clean <- dtFDep[complete.cases(dtFDep[, 3:6]), ]

cat("Nombre d'IRIS avec données complètes :",
nrow(dtFDep_clean), "\n")

# PCA sur les données complètes
X <- dtFDep_clean[, 3:6]
res.pca <- PCA(X = X, scale.unit = TRUE, ncp = 1, row.w =
dtFDep_clean$Pop06)

# Attribution du score FDep à la table propre
dtFDep_clean$FDep06 <- res.pca$ind$coord[, 1]

# Jointure des scores FDep à varIris (en conservant
uniquement les IRIS présents dans les données complètes)
varIris <- merge(varIris, dtFDep_clean[, c("IRIS",
"FDep06")], by = "IRIS", all.x = TRUE)

writexl::write_xlsx(varIris, "Fdep06_sans_imputation.xlsx")
```

- Avec le taux de bachelier en 2010 (imputation des données manquantes)

```
# ---- Calcul du French deprivation index (FDep10) ----#
```

```
#-----Rey et al. (2009) - BMC public health-----#
```

```
# Installation des packages (à faire une seule fois)
```

```
install.packages("FactoMineR", dependencies = TRUE)
```

```
install.packages("readxl", dependencies = TRUE)
```

```
install.packages("writexl", dependencies = TRUE)
```

```
install.packages("missMDA", dependencies = TRUE)
```

```
# Chargement des librairies
```

```
library(FactoMineR)
```

```
library(readxl)
```

```
library(writexl)
```

```
library(missMDA)
```

```
# Import des données
```

```
varIris <- read_excel("data/VarIris.xlsx")
```

```
revCom <-
```

```
read_excel("data/revenu_median_par_com.xlsx")
```

```
revIris <- read_excel("data/revenu_median_iris.xlsx")
```

```
names(varIris)
```

```
names(revCom)
```

```
names(revIris)
```

```
# Jointures
```

```
t <- merge(varIris, revIris[, c("IRIS", "revenu median")], by = "IRIS", all.x = TRUE)
```

```
t <- merge(t, revCom[, c("COM", "mediane")], by = "COM", all.x = TRUE)
```

```
varIris$revIris <- t$'revenu median'
```

```
varIris$revCom <- t$mediane
```

```
# Substitution du revenu médian IRIS manquant par celui de la commune
```

```
varIris$revMed10 <- ifelse(is.na(varIris$revIris),
```

```
varIris$revCom, varIris$revIris)
```

```
##Renommer les colonnes
```

```
colnames(varIris)[colnames(varIris) == "Pop 15-64 ans en 2010"] <- "pop15_64"
```

```
colnames(varIris)[colnames(varIris) == "tx_chomage"] <- "tx_chomage"
```

```
colnames(varIris)[colnames(varIris) == "tx_ouvrier"] <- "tx_ouvrier"
```

```
# Sélection des variables pour le FDep
```

```
dtFDep <- varIris[, c("IRIS", "pop15_64", "tx_bac10", "revMed10", "tx_chomage", "tx_ouvrier")]
```

```
# Conversion des variables en numérique
```

```
dtFDep$tx_bac10 <- as.numeric(as.character(dtFDep$tx_bac10))
```

```
dtFDep$revMed10 <- as.numeric(as.character(dtFDep$revMed10))
```

```
dtFDep$tx_chomage <- as.numeric(as.character(dtFDep$tx_chomage))
```

```
dtFDep$tx_ouvrier <- as.numeric(as.character(dtFDep$tx_ouvrier))
```

```
# On sélectionne les variables à inclure dans la PCA
```

```
X <- dtFDep[, 3:6]
```

```
# Imputation des valeurs manquantes
```

```
res.impute <- imputePCA(X, scale = TRUE)
```

```
X.imputed <- res.impute$completeObs
```

```
dtFDep$pop15_64 <-
```

```
as.numeric(as.character(dtFDep$pop15_64))
```

```
res.pca <- PCA(X = X.imputed, scale.unit = TRUE, ncp = 1, row.w = dtFDep$pop15_64)
```

```
dtFDep$FDep10 <- -res.pca$ind$coord[, 1]
```

```
varIris$FDep10 <- dtFDep$FDep10
```

- Avec le taux de bachelier en 2010 (sans imputation des données manquantes)

```
# ---- Calcul du French deprivation index (FDep10 sans
imputation) ----

#-----Rey et al. (2009) - BMC Public Health-----#

# Installation des packages (à faire une seule fois si pas
encore installés)
install.packages("FactoMineR", dependencies = TRUE)
install.packages("readxl", dependencies = TRUE)
install.packages("writexl", dependencies = TRUE)

# Chargement des librairies
library(FactoMineR)
library(readxl)
library(writexl)

# Import des données
varIris <- read_excel("data/VarIris.xlsx")
revCom <-
read_excel("data/revenu_median_par_com.xlsx")
revIris <- read_excel("data/revenu_median_iris.xlsx")

# Jointures
t <- merge(varIris, revIris[, c("IRIS", "revenu median")], by
= "IRIS", all.x = TRUE)
t <- merge(t, revCom[, c("COM", "mediane")], by = "COM",
all.x = TRUE)

varIris$revIris <- t$'revenu median'
varIris$revCom <- t$mediane

# Substitution du revenu médian IRIS manquant par celui
de la commune
varIris$revMed10 <- ifelse(is.na(varIris$revIris),
varIris$revCom, varIris$revIris)

# Renommer les colonnes
colnames(varIris)[colnames(varIris) == "Pop 15-64 ans en
2010"] <- "pop15_64"
colnames(varIris)[colnames(varIris) == "tx_chomage"] <-
"tx_chomage"
colnames(varIris)[colnames(varIris) == "tx_ouvrier"] <-
"tx_ouvrier"

# Sélection des variables pour le FDep
dtFDep <- varIris[, c("IRIS", "pop15_64", "tx_bac10",
"revMed10", "tx_chomage", "tx_ouvrier")]
```

```
# Conversion des variables en numérique
dtFDep$tx_bac10 <-
as.numeric(as.character(dtFDep$tx_bac10))
dtFDep$revMed10 <-
as.numeric(as.character(dtFDep$revMed10))
dtFDep$tx_chomage <-
as.numeric(as.character(dtFDep$tx_chomage))
dtFDep$tx_ouvrier <-
as.numeric(as.character(dtFDep$tx_ouvrier))
dtFDep$pop15_64 <-
as.numeric(as.character(dtFDep$pop15_64))

# Vérification des données manquantes
missing_counts <- sapply(dtFDep[, 3:6], function(x)
sum(is.na(x)))
missing_percents <- sapply(dtFDep[, 3:6], function(x)
mean(is.na(x)) * 100)

missing_summary <- data.frame(
  Indicateur = names(missing_counts),
  N_manquants = missing_counts,
  Pourcentage = round(missing_percents, 2)
)

print("Résumé des données manquantes par indicateur :")
print(missing_summary)

cat("Nombre d'IRIS initiaux :", nrow(dtFDep), "\n")

# Suppression des lignes avec NA (sans imputation)
dtFDep_clean <- dtFDep[complete.cases(dtFDep[, 3:6]), ]

cat("Nombre d'IRIS avec données complètes :",
nrow(dtFDep_clean), "\n")

# PCA sans imputation
X <- dtFDep_clean[, 3:6]
res.pca <- PCA(X = X, scale.unit = TRUE, ncp = 1, row.w =
dtFDep_clean$pop15_64)

# Ajout du score FDep10
dtFDep_clean$FDep10 <- res.pca$ind$coord[, 1]

# Rattacher FDep10 à la table principale varIris
varIris <- merge(varIris, dtFDep_clean[, c("IRIS",
```



```
"FDep10"]], by = "IRIS", all.x = TRUE)
```

```
writexl::write_xlsx(varIris,  
"Fdep2010_sansimputation.xlsx")
```

- Avec le taux de bac+2 en 2006

```
# ---- Calcul du French deprivation index (FDep06) ----#
#-----Rey et al. (2009) - BMC public health-----#
# installation des packages
install.packages("FactoMineR", dependencies = TRUE)
install.packages("readxl", , dependencies = TRUE)
install.packages("writexl")
install.packages("factoextra", , dependencies = TRUE)

# package
library(FactoMineR)
library(readxl)
library(missMDA)
library(factoextra)

# data
varlris_bac_2 <-
read_excel("Data_fdep_bac_2/varlris06_bac_2.xlsx")
revCom <- read_excel("Data_fdep_bac_2/revCom06.xlsx")
revlris <-
read_excel("Data_fdep_bac_2/revenu_mediane_2006.xlsx")

# Calcul du FDep 2006 avec bac+2
# Jointures
t <- merge(x = varlris_bac_2, y = revlris, by="IRIS", all.x=T)
varlris_bac_2$revlris <- t$med_iris

varlris_bac_2$revMed06 <- varlris_bac_2$revlris
```

```
varlris_bac_2$revMed06 <-
ifelse(is.na(varlris_bac_2$revMed06),varlris_bac_2$revCom,
varlris_bac_2$revlris) # revenu des communes quand
no data Ã l'IRIS

dtFDep_2 <-
varlris_bac_2[,c("IRIS","Pop06","tx_bac_2_06","revMed06",
"txChom06","txouvr06")]

dtFDep_2$tx_bac_2_06 <-
as.numeric(as.character(dtFDep_2$tx_bac_2_06))

dtFDep_2$revMed06 <-
as.numeric(as.character(dtFDep_2$revMed06))

dtFDep_2$txChom06 <-
as.numeric(as.character(dtFDep_2$txChom06))

dtFDep_2$txouvr06 <-
as.numeric(as.character(dtFDep_2$txouvr06))

X <- dtFDep_2[, 3:6]
X <- data.frame(lapply(X, function(x)
as.numeric(as.character(x)))) # conversion propre
res.impute <- imputePCA(X, scale = TRUE) # Imputation
PCA, bien meilleure
X.imputed <- res.impute$completeObs

dtFDep_2$Pop06 <-
as.numeric(as.character(dtFDep_2$Pop06))

res.pca <- PCA(X = X.imputed, scale.unit = TRUE, ncp = 1,
row.w = dtFDep_2$Pop06)

dtFDep_2$FDep_2_06 <- res.pca$ind$coord[, 1]
varlris_bac_2$FDep_2_06 <- dtFDep_2$FDep_2_06

writexl::write_xlsx(dtFDep_2, "resultat_fdep_2_06.xlsx")
```

- Avec le taux de bac+2 en 2010

```
# --- Calcul du French deprivation index (FDep10) ---
-#
#-----Rey et al. (2009) - BMC public health-----#

# Installation des packages (à faire une seule fois)
install.packages("FactoMineR", dependencies = TRUE)
install.packages("readxl", dependencies = TRUE)
install.packages("writexl", dependencies = TRUE)
install.packages("missMDA", dependencies = TRUE)

# Chargement des librairies
library(FactoMineR)
library(readxl)
library(writexl)
library(missMDA)

# Import des données
varIris <- read_excel("data/VarIris.xlsx")
revCom <- read_excel("data/revenu_median_par_com.xlsx")
revIris <- read_excel("data/revenu_median_iris.xlsx")

names(varIris)
names(revCom)
names(revIris)

# Jointures
t <- merge(varIris, revIris[, c("IRIS", "revenu median")], by = "IRIS", all.x = TRUE)
t <- merge(t, revCom[, c("COM", "mediane")], by = "COM", all.x = TRUE)

varIris$revIris <- t$'revenu median'
varIris$revCom <- t$mediane

# Substitution du revenu médian IRIS manquant par celui de la commune
varIris$revMed10 <- ifelse(is.na(varIris$revIris),
```

```
varIris$revCom, varIris$revIris)

##Renommer les colonnes
colnames(varIris)[colnames(varIris) == "Pop 15-64 ans en 2010"] <- "pop15_64"
colnames(varIris)[colnames(varIris) == "tx_chomage"] <- "tx_chomage"
colnames(varIris)[colnames(varIris) == "tx_ouvrier"] <- "tx_ouvrier"

# Sélection des variables pour le FDep
dtFDep_2_10 <- varIris[, c("IRIS", "pop15_64", "tx_bac_2_10", "revMed10", "tx_chomage", "tx_ouvrier")]

# Conversion des variables en numérique
dtFDep_2_10$tx_bac_2_10 <- as.numeric(as.character(dtFDep_2_10$tx_bac_2_10))
dtFDep_2_10$revMed10 <- as.numeric(as.character(dtFDep_2_10$revMed10))
dtFDep_2_10$tx_chomage <- as.numeric(as.character(dtFDep_2_10$tx_chomage))
dtFDep_2_10$tx_ouvrier <- as.numeric(as.character(dtFDep_2_10$tx_ouvrier))

# On sélectionne les variables à inclure dans la PCA
X <- dtFDep_2_10[, 3:6]

# Imputation des valeurs manquantes
res.impute <- imputePCA(X, scale = TRUE)
X.imputed <- res.impute$completeObs

dtFDep_2_10$pop15_64 <- as.numeric(as.character(dtFDep_2_10$pop15_64))
res.pca <- PCA(X = X.imputed, scale.unit = TRUE, ncp = 1, row.w = dtFDep_2_10$pop15_64)
dtFDep_2_10$FDep10 <- res.pca$ind$coord[, 1]
varIris$FDep10 <- dtFDep_2_10$FDep10
```

Annexe 1.2 Code R de calcul de l'accessibilité de scale moyen par iris en France

```
install.packages('sf')
library(sf)
library(dplyr)

# Chargement shapefile SCALE
scale_sf <-
st_read("C:/Data/Stage/Scale/Data/shp/shp/scale_index_
health.shp")

# Calculer la moyenne de scale par IRIS
scale_moyen_iris <- scale_sf %>%
  st_drop_geometry() %>% # On enlève la géométrie pour
un calcul plus rapide
  group_by(DCOMIRIS) %>%
  summarise(scale_moyen = mean(score_sant, na.rm =
TRUE))

## Chargement contours iris
iris_sf <- st_read("C:/Data/Stage/Scale/Data/CONTOURS-
IRIS_2-1__SHP__FRA_2021-01-01/CONTOURS-
IRIS/1_DONNEES_LIVRAISON_2021-06-00217/CONTOURS-
IRIS_2-1_SHP_LAMB93_FXX-2021/CONTOURS-IRIS.shp")

iris_sf <- iris_sf %>%
  rename(DCOMIRIS = CODE_IRIS)

# Joindre les résultats

iris_scale <- iris_sf %>%
  left_join(scale_moyen_iris, by = "DCOMIRIS")
# export

st_write(iris_scale,
"C:/Data/Stage/Scale/Data/iris_scale.shp", delete_layer =
TRUE)
```

Annexe 1.3 Code SQL de calcul l'accessibilité aux médecins traitants et pharmacie du métropole Grand Paris

```
--Filtrer les bâtiments
create table IF NOT EXISTS accessibilite.zr_01_bati_filtre
as
select b.id, b.geom, b."HAUTEUR", b."NB_ETAGES"
FROM accessibilite.bati_gp b
WHERE b."USAGE1" in ('Indifférencié', 'Résidentiel')
or b."USAGE2" in ('Indifférencié', 'Résidentiel');

with b as (select b.*
from accessibilite.bati_gp b,
accessibilite."ZONE_D_ACTIVITE_OU_D_INTERET" za
where st_intersects (b.geom, za.geom))
delete from accessibilite.zr_01_bati_filtre
where id in (select id from b);

-- Découpage par IRIS et calcul des étages
select ROW_NUMBER() OVER () as id_bati,
st_intersection(b.geom, i.geom) as geom, b.id,
b."HAUTEUR", b."NB_ETAGES", i."CODE_IRIS"
from accessibilite.zr_01_bati_filtre b,
accessibilite."CONTOURS_IRIS_2154" i
where st_intersects(b.geom, i.geom);
-- Une colonne nb_floors est ajoutée pour estimer le
nombre d'étages
alter table accessibilite.zr_02_bati_decoup
add column nb_floors real;

update accessibilite.zr_02_bati_decoup
set nb_floors =
CASE
WHEN "NB_ETAGES" IS NOT NULL THEN "NB_ETAGES" --
prend la valeur du champ nb_etages, si il est non nul
WHEN "NB_ETAGES" IS NULL AND "HAUTEUR" < 5 THEN 1
-- si nb_etages est nul, et que la batiment fait moins de
5mètres de haut, on considère qu'il n'y a qu'un étage
WHEN "NB_ETAGES" IS NULL AND "HAUTEUR" >= 5 THEN
"HAUTEUR" / 2.5 -- si la hauteur est supérieur ou égale à
5m, on compte le nombre d'étages de 2,5m de hauteur
else null -- mettre null sinon ? ou 1 ?
end;

-- Ajout des données de population et calcul des surfaces
create table IF NOT EXISTS accessibilite.zr_03_bati_pop as
select b.*, ip."P21_POP"
from accessibilite.zr_02_bati_decoup b
```

```
join accessibilite."iris-pop-2021" ip on b."CODE_IRIS" =
ip."IRIS";

-- La surface au sol (surface) est calculée avec ST_Area,
une fonction qui mesure l'aire d'un polygone.
alter table accessibilite.zr_03_bati_pop
add column surface real;
update accessibilite.zr_03_bati_pop
set surface =
st_area(geom) ;

--Puis la surface habitable (surface_habitable) est obtenue
en multipliant
--la surface par le nombre d'étages.

alter table accessibilite.zr_03_bati_pop
add column surface_habitable real;
update accessibilite.zr_03_bati_pop
set surface_habitable =
surface*nb_floors;

-- La surface habitable totale par IRIS
(surface_habitable_iris) est ensuite calculée par
agrégation.

alter table accessibilite.zr_03_bati_pop
add column surface_habitable_iris real;
UPDATE accessibilite.zr_03_bati_pop
SET surface_habitable_iris = sub.somme_surf_hab
FROM ( SELECT code_iris, SUM(surface_habitable) AS
somme_surf_hab
FROM accessibilite.zr_03_bati_pop
GROUP BY "CODE_IRIS"
) sub
WHERE accessibilite.zr_03_bati_pop."CODE_IRIS" =
sub."CODE_IRIS";

-- Estimation de la population par bâtiment

alter table accessibilite.zr_03_bati_pop
add column pop_batiment real;
update accessibilite.zr_03_bati_pop
set pop_totale =
(surface_habitable / surface_habitable_iris) * "P21_POP";
```

-- Création des zones résidentielles (ZR)

```
CREATE TABLE IF NOT EXISTS accessibilite.zr_04_tampon
AS
SELECT ROW_NUMBER() OVER () as id_tampon,
(st_dump(st_union(ST_Buffer(bc.geom, 75)))) AS geom
FROM accessibilite.zr_03_bati_pop bc ;
```

-- Ces zones sont découpées par les contours IRIS

```
create table IF NOT EXISTS
accessibilite.zr_04_tampon_iris as
select ROW_NUMBER() OVER () as id_tampon,
st_intersection(t.geom, ci.geom) as geom
from accessibilite.zr_04_tampon t,
accessibilite."CONTOURS_IRIS_2154" ci
where st_intersects(t.geom, ci.geom);
```

```
create table IF NOT EXISTS
accessibilite.zr_05_batiment_tampon as
select bc.*, ti.id_tampon
from accessibilite.zr_03_bati_pop bc
join accessibilite.zr_04_tampon_iris ti on
st_intersects(bc.geom, ti.geom)
where st_intersects(bc.geom, ti.geom);
```

-- La population est agrégée par zone

```
create table IF NOT EXISTS
accessibilite.zr_05_pop_par_tampon as
select zb.id_tampon,
zb."CODE_IRIS",
sum(zb.pop_batiment) as pop_zone
from accessibilite.zr_05_batiment_tampon zb
group by zb.id_tampon, zb."CODE_IRIS";
```

-- Génération des points dans la surface des zones résidentielles

```
CREATE TABLE IF NOT EXISTS accessibilite.centre_zr AS
select ti.id_tampon,
zp."CODE_IRIS", ST_PointOnSurface(geom) AS geom,
zp.pop_zone
FROM accessibilite.zr_04_tampon_iris ti
join accessibilite.zr_05_pop_par_tampon zp
on ti.id_tampon = zp.id_tampon;
```

-- Estimer la pression théorique exercée sur les établissements santé

-- Ajout des données de population des différents niveaux administratifs

```
create table IF NOT EXISTS accessibilite.pression_epci_pop
as
select ep.*, p.pop_totale::REAL as pop_epci
from accessibilite.contours_epci ep
join accessibilite.pop_pressions_epci p on
ep."CODE_SIREN" = p."INSEE"
```

--

```
create table IF NOT EXISTS
accessibilite.pression_etablissement_sante2 as
select e.*, ep.pop_epci
from accessibilite.pression_epci_pop ep,
accessibilite.bpe_gp e
where st_intersects(ep.geom, e.geom) AND "Catégorie"
IN ( 'CHIRURGIEN DENTISTE' , 'INFIRMIER' ,
'MASSEUR KINÉSITHÉRAPEUTE' , 'MATERNITÉ' , 'MÉDECIN
GÉNÉRALISTE' , 'PHARMACIE' ,
'SPÉCIALISTE EN GYNÉCOLOGIE (MÉDICALE ET/OU
OBSTÉTRIQUE)' , 'SPÉCIALISTE EN OPHTALMOLOGIE'
, 'SPÉCIALISTE EN PÉDIATRIE' , 'URGENCES' ,
'ÉTABLISSEMENT SANTÉ COURT SÉJOUR' );
```

-- Identifier l'aire d'influence de chaque établissement de santé : les polygones de Voronoï

-- Médecins

```
create table IF NOT EXISTS accessibilite.voronoi_medecins
as
WITH
voronoi as
(SELECT
(st_dump(st_VoronoiPolygons(st_collect(ee.geom),
0.0))) AS geom
FROM accessibilite.pression_etablissement_sante2 ee
WHERE ee."Catégorie" = 'MÉDECIN GÉNÉRALISTE'),
medecin as
(SELECT ee.id, ee.geom
FROM accessibilite.pression_etablissement_sante2 ee
WHERE ee."Catégorie" = 'MÉDECIN GÉNÉRALISTE')
```

```
SELECT s.id, v.geom
FROM voronoi v JOIN medecin s ON st_contains(v.geom,
s.geom);
```

-- Pharmacie

create table IF NOT EXISTS

```

accessibilite.voronoi_pharmacie as
WITH
voronoi as
(SELECT
(st_dump(st_VoronoiPolygons(st_collect(ee.geom),
0.0))).geom AS geom
FROM accessibilite.pression_etablissement_sante2 ee
WHERE ee."Catégorie" = 'PHARMACIE'),
pharmacie as
(SELECT ee.id, ee.geom
FROM accessibilite.pression_etablissement_sante2 ee
WHERE ee."Catégorie" = 'PHARMACIE')

SELECT s.id, v.geom
FROM voronoi v JOIN pharmacie s ON st_contains(v.geom,
s.geom);

-- Estimation de la population présente dans l'aire
d'influence
-- Médecins

create table IF NOT EXISTS
accessibilite.pression_theorique_medecins as
select vm.id,
(sum(cz.pop_zone)) as pop_theorique_medecins
from accessibilite.voronoi_medecins vm,
accessibilite.centre_zr cz
where st_intersects(vm.geom, cz.geom)
group by vm.id;

-- Pharmacie

create table IF NOT EXISTS
accessibilite.pression_theorique_pharmacie as
select vm.id,
(sum(cz.pop_zone)) as pop_theorique_pharma
from accessibilite.voronoi_pharmacie vm,
accessibilite.centre_zr cz
where st_intersects(vm.geom, cz.geom)
group by vm.id;

--Encore au moyen d'une jointure, ce nombre de
médecins/Pharmacie théorique est ajouté à la table
--des établissements de santé, en liant les deux tables par
l'id.

--Médecins

ALTER TABLE accessibilite.pression_etablissement_sante2

```

```

ADD COLUMN pop_theorique_medecins REAL;

UPDATE accessibilite.pression_etablissement_sante2 ee
SET pop_theorique_medecins =
p.pop_theorique_medecins
FROM accessibilite.pression_theorique_medecins p
WHERE ee.id = p.id AND "Catégorie" = 'MÉDECIN
GÉNÉRALISTE';

-- Pharmacie

ALTER TABLE accessibilite.pression_etablissement_sante2
ADD COLUMN pop_theorique_pharma REAL;

UPDATE accessibilite.pression_etablissement_sante2 ee
SET pop_theorique_pharma = p.pop_theorique_pharma
FROM accessibilite.pression_theorique_pharmacie p
WHERE ee.id = p.id AND "Catégorie" = 'PHARMACIE';

-- Taux de pression
-- Médecins

ALTER TABLE accessibilite.pression_etablissement_sante2
ADD COLUMN IF NOT EXISTS pression_medecins REAL;
UPDATE accessibilite.pression_etablissement_sante2
SET pression_medecins = (pop_theorique_medecins /
pop_epci * 100);

-- Taux de pression
-- Pharmacie

ALTER TABLE accessibilite.pression_etablissement_sante2
ADD COLUMN IF NOT EXISTS pression_pharmacie REAL;
UPDATE accessibilite.pression_etablissement_sante2
SET pression_pharmacie = (pop_theorique_pharma /
pop_epci * 100);

--Calcul de l'accessibilité aux équipements
d'enseignement, en fonction de la pression théorique
--et de la fréquence des équipements

-- Distance à l'établissement d'enseignement le plus
proche

-- Médecin euclidienne
create table IF NOT EXISTS accessibilite.medecin_eucl_da
as
SELECT cz."CODE_IRIS",cz.id_tampon, ee.id,

```

```

ST_Distance(cz.geom, ee.geom) AS distance
FROM accessibilite.centre_zr cz
JOIN accessibilite.pression_etablissement_sante2 ee
ON ee.id =
(SELECT ee2.id
FROM accessibilite.pression_etablissement_sante2 ee2
where ee2."Catégorie" = 'MÉDECIN GÉNÉRALISTE'
ORDER BY ST_Distance(cz.geom, ee2.geom) ASC
LIMIT 1);
-- Pharmacie euclidienne
create table IF NOT EXISTS accessibilite.pharma_eucl_da
as
SELECT cz."CODE_IRIS", cz.id_tampon, ee.id,
ST_Distance(cz.geom, ee.geom) AS distance
FROM accessibilite.centre_zr cz
JOIN accessibilite.pression_etablissement_sante2 ee
ON ee.id =
(SELECT ee2.id
FROM accessibilite.pression_etablissement_sante2 ee2
where ee2."Catégorie" = 'PHARMACIE'
ORDER BY ST_Distance(cz.geom, ee2.geom) ASC
LIMIT 1);
-- Médecin Minkowski
select cz."CODE_IRIS", cz.id_tampon, ee.id,
(((ABS(ST_X(cz.geom) - ST_X(ee.geom))) +
(ABS(ST_Y(cz.geom) - ST_Y(ee.geom)))^1.25) ^ (1/1.25))
AS distance_minkowski
FROM accessibilite.centre_zr cz
JOIN accessibilite.pression_etablissement_sante2 ee
ON ee.id =
(SELECT ee2.id
FROM accessibilite.pression_etablissement_sante2 ee2
where ee2."Catégorie" = 'MÉDECIN GÉNÉRALISTE'
ORDER BY (((ABS(ST_X(cz.geom) - ST_X(ee2.geom)))^1.25
+ (ABS(ST_Y(cz.geom) - ST_Y(ee2.geom)))^1.25) ^
(1/1.25))
LIMIT 1);

-- Pharmacie Minkowski

select cz."CODE_IRIS", cz.id_tampon, ee.id,
(((ABS(ST_X(cz.geom) - ST_X(ee.geom))) +
(ABS(ST_Y(cz.geom) - ST_Y(ee.geom)))^1.25) ^ (1/1.25))
AS distance_minkowski
FROM accessibilite.centre_zr cz
JOIN accessibilite.pression_etablissement_sante2 ee
ON ee.id =
(SELECT ee2.id

```

```

FROM accessibilite.pression_etablissement_sante2 ee2
where ee2."Catégorie" = 'PHARMACIE'
ORDER BY (((ABS(ST_X(cz.geom) - ST_X(ee2.geom)))^1.25
+ (ABS(ST_Y(cz.geom) - ST_Y(ee2.geom)))^1.25) ^
(1/1.25))
LIMIT 1);

-- Calcul de la zone d'accessibilité potentielle par IRIS

-- Medecins
create table IF NOT EXISTS
accessibilite.medecin_eucl_da_max as
select "CODE_IRIS", max(distance) as dap_medecin
FROM accessibilite.medecin_eucl_da
group by "CODE_IRIS" ;
-- La DAP est ensuite jointe à la table des zones
résidentielles
create table IF NOT EXISTS
accessibilite.medecin_eucl_rayon as
select cz.*, i.dap_medecin as rayon_zap_m_p
from accessibilite.centre_zr cz
join accessibilite.medecin_eucl_da_max i on
cz."CODE_IRIS" = i."CODE_IRIS";

-- Pharmacie
create table IF NOT EXISTS
accessibilite.pharma_eucl_da_max as
select "CODE_IRIS", max(distance) as dap_pharmacie
FROM accessibilite.pharma_eucl_da
group by "CODE_IRIS" ;
-- La DAP est ensuite jointe à la table des zones
résidentielles
create table IF NOT EXISTS
accessibilite.pharmacie_eucl_rayon as
select cz.*, i.dap_pharmacie as rayon_zap_p
from accessibilite.centre_zr cz
join accessibilite.pharma_eucl_da_max i on
cz."CODE_IRIS" = i."CODE_IRIS";

-- Distance moyenne aux équipements dans la ZAP,
pondérée par la pression théorique
-- sur chaque établissement de santé présent dans la ZAP

--Médecin euclidienne
create table IF NOT EXISTS accessibilite.medecin_eucl_zap
as
select z.id_tampon, count(*) as
nombre_etablissements_dans_zap,
(avg((st_distance(z.geom,
e.geom))*e.pression_medecins)) as dap_medecin

```



```

from accessibilite.medecin_eucl_rayon z,
accessibilite.pression_etablissement_sante2 e
where st_Dwithin(z.geom, e.geom, (z.rayon_zap_m_p +
1))
and e."Catégorie" = 'MÉDECIN GÉNÉRALISTE'
group by z.id_tampon;

```

-- Pharmacie euclidienne

```

create table IF NOT EXISTS
accessibilite.pharmacie_eucl_zap as
select z.id_tampon, count(*) as
nombre_etablissements_dans_zap,
(avg((st_distance(z.geom,
e.geom))*e.pression_pharmacie)) as dap_pharmacie
from accessibilite.pharmacie_eucl_rayon z,
accessibilite.pression_etablissement_sante2 e
where st_Dwithin(z.geom, e.geom, (z.rayon_zap_p + 1))
and e."Catégorie" = 'PHARMACIE'
group by z.id_tampon;

```

--Médecin Minkowski

```

create table IF NOT EXISTS
accessibilite.medecin_mink_dap_max as
select z.id_tampon, count(*) as
nombre_etablissements_dans_zap,
avg((((ABS(ST_X(z.geom) - ST_X(e.geom)))^1.25 +
(ABS(ST_Y(z.geom) - ST_Y(e.geom)))^1.25) ^ (1/1.25)) *
e.pression_medecins)) as dap_medecin_mink
from accessibilite.medecin_eucl_rayon z,
accessibilite.pression_etablissement_sante2 e
where st_Dwithin(z.geom, e.geom, (z.rayon_zap_m_p +
1)) and e."Catégorie" = 'MÉDECIN GÉNÉRALISTE'
group by z.id_tampon;

```

--Pharmacie Minkowski

```

create table IF NOT EXISTS
accessibilite.pharmacie_mink_dap_max as
select z.id_tampon, count(*) as
nombre_etablissements_dans_zap,
avg((((ABS(ST_X(z.geom) - ST_X(e.geom)))^1.25 +
(ABS(ST_Y(z.geom) - ST_Y(e.geom)))^1.25) ^ (1/1.25)) *
e.pression_pharmacie)) as dap_pharmacie_mink
from accessibilite.pharmacie_eucl_rayon z,
accessibilite.pression_etablissement_sante2 e
where st_Dwithin(z.geom, e.geom, (z.rayon_zap_p + 1))
and e."Catégorie" = 'PHARMACIE'
group by z.id_tampon;

```

-- Ces valeurs sont ensuite jointes à la table originale des zones résidentielles, en créant un nouveau champ.

-- Medecin

```

ALTER TABLE accessibilite.centre_zr
ADD COLUMN dap_medecin_eucl double precision;

```

```

UPDATE accessibilite.centre_zr cz
SET dap_medecin_eucl = e.dap_medecin
FROM accessibilite.medecin_eucl_zap e
WHERE cz.id_tampon = e.id_tampon;

```

-- pharmacie

```

ALTER TABLE accessibilite.centre_zr
ADD COLUMN dap_pharmacie_eucl double precision;

```

```

UPDATE accessibilite.centre_zr cz
SET dap_pharmacie_eucl = e.dap_pharma
FROM accessibilite.pharmacie_eucl_zap e
WHERE cz.id_tampon = e.id_tampon;

```

-- Calcul de la fréquence des équipements

--Médecin

```

select count(*)
from accessibilite.bpe_gp
WHERE "Catégorie" = 'MÉDECIN GÉNÉRALISTE'
-- 3535

```

--Pharmacie

```

select count(*)
from accessibilite.bpe_gp
WHERE "Catégorie" = 'PHARMACIE'
-- 2236

```

-- Calcul final de l'indice SCALE pour la santé

```

ALTER TABLE accessibilite.centre_zr
ADD COLUMN dap_eucl_finale double precision;

```

```

UPDATE accessibilite.centre_zr
SET dap_eucl_finale = (dap_medecin_eucl * 1) +
(dap_pharmacie_eucl * 0.632

```

Annexe 1.4 Code R de calcul de la typologie de l'occupation du sol du corine land cover

```
# Installer les packages si nécessaire
#install.packages(c("sf", "dplyr", "ggplot2"))
#install.packages("tmap")
#install.packages("tidyr")

library(sf)
library(dplyr)
library(tmap)
library(tidyr) # pour replace_na()

# Données IRIS (shapefile INSEE)
iris <- st_read("data/CONTOURS_IRIS_2154.shp")

# Données d'occupation du sol ( Corine Land Cover )
clc <- st_read("data/CLC_decoup_fr.gpkg")

iris <- st_transform(iris, 2154)
clc <- st_transform(clc, 2154)

## Filtrer les types d'artificialisation qu'on a besoin
## Nomenclatures
# 111 Tissu urbain continu
# 112 Tissu urbain discontinu
# 121 Zones industrielles et commerciales
# 122 Réseaux routier et ferroviaire et espaces associés
# 131 Extraction de matériaux
# 132 Décharges
# 133 Chantiers
# 141 Espaces verts urbains
# 142 Équipements sportifs et de loisirs

urbanized <- clc %>% filter(code_18 %in% c('111', '112'))
industrial <- clc %>% filter(code_18 %in% c('121', '122'))
mines <- clc %>% filter(code_18 %in% c('131', '132', '133'))
green <- clc %>% filter(code_18 %in% c('141', '142'))

## Intersection avec les IRIS et calcul des surfaces

# Fonction pour intersecter et calculer les surfaces
intersect_artif <- function(artif_layer, iris_layer) {
  inter <- st_intersection(iris_layer, artif_layer)
  inter <- inter %>% mutate(surface_ha = st_area(.) /
    10000) # en hectares
  return(inter)
}

urban_inter <- intersect_artif(urbanized, iris)
indus_inter <- intersect_artif(industrial, iris)
mines_inter <- intersect_artif(mines, iris)
green_inter <- intersect_artif(green, iris)

# Surface totale des IRIS
iris <- iris %>% mutate(surface_totale_ha = st_area(.) /
  10000)

# Fonction d'agrégation par IRIS
library(tidyr) # pour replace_na()
library(units) # pour set_units si besoin

aggregate_taux <- function(inter_layer, iris_layer,
  varname) {
  # Agréger les surfaces par IRIS
  agg <- inter_layer %>%
    group_by(CODE_IRIS) %>%
    summarise(surface_artif_ha = sum(surface_ha, na.rm =
      TRUE)) %>%
    st_drop_geometry()

  # Jointure et calcul du taux avec gestion des unités
  taux <- iris_layer %>%
    left_join(agg, by = "CODE_IRIS") %>%
    mutate(
      surface_artif_ha = replace_na(surface_artif_ha,
        units::set_units(0, "ha")),
      surface_artif_ha = as.numeric(surface_artif_ha),
      surface_totale_ha = as.numeric(surface_totale_ha),
      !varname := (surface_artif_ha / surface_totale_ha) *
        100
    )

  return(taux)
}

iris_urban <- aggregate_taux(urban_inter, iris,
  "taux_urban")
```

| | |
|--|---|
| <pre>iris_indus <- aggregate_taux(indus_inter, iris, "taux_indus") iris_mines <- aggregate_taux(mines_inter, iris, # Export des résultats en GeoPackage st_write(iris_urban, "data/taux_urban.gpkg", delete_dsn = TRUE) st_write(iris_indus, "data/taux_indus.gpkg", delete_dsn =</pre> | <pre>"taux_mines") iris_green <- aggregate_taux(green_inter, iris, "taux_green") TRUE) st_write(iris_mines, "data/taux_mines.gpkg", delete_dsn = TRUE) st_write(iris_green, "data/taux_green.gpkg", delete_dsn = TRUE)</pre> |
|--|---|

Annexe 1.5 Code Javascript du calcul du NDVI moyen par iris en France

```
// -----
// 1. CHARGER LES DONNÉES
// -----

// a. Frontière France
var france =
ee.FeatureCollection("FAO/GAUL/2015/level0")
    .filter(ee.Filter.eq("ADM0_NAME", "France"));

// b. IRIS (à importer comme asset ou depuis Drive)
var iris = ee.FeatureCollection("projects/focal-elf-464111-
i4/assets/CONTOURS_IRIS_2154");

// c. Période d'étude
var startDate = "2019-04-01";
var endDate = "2019-08-31";

// -----
// 2. PRÉTRAITEMENT IMAGES LANDSAT 8
// -----

var L8 = ee.ImageCollection("LANDSAT/LC08/C02/T1_L2")
    .filterDate(startDate, endDate)
    .filterBounds(france);

// Masque nuages/ombres
function maskL8sr(image) {
  var qa = image.select('QA_PIXEL');
  var cloudShadowBitMask = 1 << 3;
  var cloudsBitMask = 1 << 1;
  var mask = qa.bitwiseAnd(cloudShadowBitMask).eq(0)
    .and(qa.bitwiseAnd(cloudsBitMask).eq(0));
  return image.updateMask(mask).divide(10000); //
  réflectance corrigée
}

var L8_clean = L8.map(maskL8sr);

// -----
// 3. CALCUL DU NDVI
// -----

var composite = L8_clean.median().clip(france);
var ndvi = composite.normalizedDifference(["SR_B5",
"SR_B4"]).rename("NDVI");

// -----
// 4. STATISTIQUES PAR IRIS
// -----

var zonalStats = ndvi.reduceRegions({
  collection: iris,
  reducer: ee.Reducer.mean(),
  scale: 30,
  crs: 'EPSG:4326'
});

// Ajouter le champ NDVI moyen et garder code IRIS
zonalStats = zonalStats.map(function(f) {
  return ee.Feature(f.geometry()).set({
    'CODE_IRIS': f.get('CODE_IRIS'),
    'NDVI_moyen': f.get('mean')
  });
});

// -----
// 5. EXPORT DES RÉSULTATS
// -----

Export.table.toDrive({
  collection: zonalStats,
  description: 'NDVI_MOYEN_IRIS_FRANCE_2019',
  fileFormat: 'GeoJSON', // ou 'CSV' si tu ne veux pas la
  géométrie
  selectors: ['CODE_IRIS', 'NDVI_moyen']
});

// -----
// 6. AFFICHAGE (facultatif)
// -----

Map.centerObject(france, 5);
Map.addLayer(ndvi, {min: 0, max: 1, palette: ["red",
"yellow", "green"]}, "NDVI France Été 2019");
```

Annexe 1.6 Code R de calcul de la proximité des adresses aux sites industriels

```
# Chargement des bibliothèques nécessaires
```

```
library(sf)
```

```
library(dplyr)
```

```
# 1. Charger les couches SIG
```

```
zone_industrielle <-
```

```
st_read("C:/Data/Stage/proximité_sites_industrielles/DAT  
A/zone_industrielle_93.shp")
```

```
departement <-
```

```
st_read("C:/Data/Stage/proximité_sites_industrielles/DAT  
A/departement_93.shp")
```

```
adresses <-
```

```
st_read("C:/Data/Stage/proximité_sites_industrielles/DAT  
A/adresses-cadastre-93.geojson/adresses-cadastre-  
93.geojson")
```

```
# 2. Vérifier et harmoniser les projections
```

```
# Utilisation de la projection Lambert 93 (EPSG:2154) pour  
les distances en mètres
```

```
zone_industrielle <- st_transform(zone_industrielle, 2154)
```

```
departement <- st_transform(departement, 2154)
```

```
adresses <- st_transform(adresses, 2154)
```

```
# 3. Création des buffers autour des zones industrielles
```

```
buffer_500m <- st_buffer(zone_industrielle, dist = 500)
```

```
buffer_1km <- st_buffer(zone_industrielle, dist = 1000)
```

```
# Union des buffers si plusieurs zones
```

```
buffer_500m <- st_union(buffer_500m)
```

```
buffer_1km <- st_union(buffer_1km)
```

```
# 4. Intersections (proximité)
```

```
# Ajouter une variable "proche_500m" et "proche_1km"  
aux adresses
```

```
adresses <- adresses %>%
```

```
mutate(
```

```
  proche_500m = st_intersects(geometry, buffer_500m,  
    sparse = FALSE)[, 1],
```

```
  proche_1km = st_intersects(geometry, buffer_1km,  
    sparse = FALSE)[, 1]
```

```
)
```

```
# 5. Sauvegarde de la couche résultante
```

```
st_write(adresses,
```

```
"C:/Data/Stage/proximité_sites_industrielles/DATA/Proxi  
mity to industrial sites.shp")
```

```
# 6. Statistiques simples
```

```
table(adresses$proche_500m)
```

```
table(adresses$proche_1km)
```

```
# les adresses ont maintenant deux colonnes :  
TRUE/FALSE selon leur proximité
```