



HAL
open science

Pour une dose [géographique] réponse en santé environnementale

Sandra Perez

► **To cite this version:**

Sandra Perez. Pour une dose [géographique] réponse en santé environnementale. Géographie. Université Cote d'Azur, 2023. tel-04201635

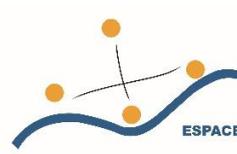
HAL Id: tel-04201635

<https://hal.science/tel-04201635v1>

Submitted on 11 Sep 2023

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

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



HABILITATION A DIRIGER DES RECHERCHES

Sandra PÉREZ
Volume 1

**Pour une dose [géographique] réponse en
santé environnementale**

Soutenue le 10 mars 2023

Membres du jury

- . Virginie Chasles, Professeur de Géographie, Université de Saint Etienne, *Rapportrice*
- . Céline Rozenblat, Professeur de Géographie, Université de Lausanne, *Rapportrice*
- . Francelyne Marano, Professeur émérite de Biologie cellulaire et de Toxicologie, Université de Paris, *Rapportrice*
- . Sébastien Gadal, Professeur de Géographie, Aix-Marseille Université, *Examineur*
- . Philippe Weber, Professeur en Génie informatique, Université de Lorraine, *Examineur*
- . Pascal Staccini, Professeur de Santé Publique, Université Côte d'Azur, *Garant*

À ma fille, Iris

*There is no
universally valid
way to analyze data
and thus no single
solution to the
problems at hand
(Rafi, Greenland,
2020)*

. Sigles

1,2,3,4,6,7,8 HpCDF : 1,2,3,4,6,7,8-Heptachlorodibenzofurane

1,2,3,4,6,7,8,9 OCDD : 1,2,3,4,6,7,8,9-Octachlorodibenzodioxine

1,2,3,4,6,7,8,9 OCDF : 1,2,3,4,6,7,8,9-Octachlorodibenzofurane

1,2,3,6,7,8 HeCDF : Hexachlorodibenzo-*p*-furane

2,3,4,7,8 PeCDF : 2,3,4,7,8-Pentachlorodibenzofurane

2,3,7,8 TeCDF : 2,3,7,8-Tetrachlorodibenzofurane

Ace : Acenaphthène

Acy : Acenaphthylene

Ant : Anthracène

As : Arsenic

BaA : Benzo[a]anthracène

BaP : Benzo[a]pyrène

BbF : Benzo[b]fluoranthène

Benz : Benzène

BkF_p : Benzo[k]fluoranthène phase particulaire (un p après un polluant indique qu'il s'agit de la phase particulaire de ce polluant, et non de sa phase gazeuse)

Bpe : Benzo(ghi)pérylène

C14_PMEN_MENFAMMONO : nombre de personnes des ménages dont la famille principale est une famille monoparentale

C34 tous les types de **cancer des bronches et du poumon**

C64 tous les types de **cancer du rein**

C67 tous les types de **cancer de la vessie**

CB101 : 2,2',4,5,5'-Pentachlorobiphényle

CB105 : 2,3,3',4,4'-Pentachlorobiphényle

CB114 : 2,3,4,4',5-Pentachlorobiphényle
CB118 : 2,3',4,4',5-Pentachlorobiphényle
CB123 : 2,3',4,4',5'-Pentachlorobiphényle
CB126 : 3,3',4,4',5-Pentachlorobiphényle
CB128 : 2,2',3,3',4,4'-Hexachlorobiphényle
CB138 : 2,2',3,4,4',5'-Hexachlorobiphényle
CB141 : 2,2',3,4,5,5'-Hexachlorobiphényle
CB149 : 2,2',3,4',5',6-Hexachlorobiphényle
CB151 : 2,2',3,5,5',6-Hexachlorobiphényle
CB153 : 2,2',4,4',5,5'-Hexachlorobiphényle
CB156 : 2,3,3',4,4',5-Hexachlorobiphényle
CB157 : 2,3,3',4,4',5'-Hexachlorobiphényle
CB169 : 3,3',4,4',5,5'-Hexachlorobiphényle
CB170 : 2,2',3,3',4,4',5-Heptachlorobiphényle
CB18 : 2,2',5-Trichlorobiphényle
CB180 : 2,2',3,4,4',5,5'-Heptachlorobiphényle
CB183 : 2,2',3,4,4',5',6-Heptachlorobiphényle
CB187 : 2,2',3,4',5,5',6-Heptachlorobiphényle
CB194 : 2,2',3,3',4,4',5,5'-Octachlorobiphényle
CB195 : 2,2',3,3',4,4',5,6-Octachlorobiphényle
CB196 : 2,2',3,3',4,4',5,6'-Octachlorobiphényle
CB20 : 2,3,3'-Trichlorobiphényle
CB201 : 2,2',3,3',4,5',6,6'-Octachlorobiphényle
CB28 : 2,4,4'-Trichlorobiphényle

CB31 : 2,4',5-Trichlorobiphényle

CB44 : 2,2',3,5'-Tétrachlorobiphényle

CB52 : 2,2',5,5'-Tétrachlorobiphényle

CB77 : 3,3',4,4'-Tétrachlorobiphényle

CB87 : 2,2',3,4,5'-Pentachlorobiphényle

CB92 : 2,2',3,5,5'-Pentachlorobiphényle

CB95 : 2,2',3,5',6-Pentachlorobiphényle

Cd_p : Cadmium phase particulaire

Cu : Cuivre

Daha : Dibenzo [a,h]anthracène

DCE : Dichloroéthane

DEC_MED14 : revenus médians

E10 à E1415ans tous les types de diabète chez les moins de 15 ans

E10 à E141565ans tous les types de diabète chez les 15-65 ans

E10 à E1465ans tous les types de diabète chez les plus de 65 ans

Fe : Fer

FIA : Fluoranthène

Flu : Fluorène

HAP : Hydrocarbures Aromatiques Polycycliques

HCl : Acide chlorhydrique

HF : Acide fluorhydrique

I00I9915ans : toutes les maladies cardiovasculaires chez les moins de 15 ans

I00I991565ans : toutes les maladies cardiovasculaires chez les 15-65 ans

I00I9965ans : toutes les maladies cardiovasculaires chez les plus de 65 ans

IcdP : Indéno [1,2,3 c, d] pyrène

ICPE : Installations Classées pour la Protection de l'Environnement

IPA : Indice de Pureté Atmosphérique

IRIS : Ilots Regroupés pour l'Information Statistique, découpage infra-communal de l'INSEE

J00 à J19915ans : toutes les maladies de l'appareil respiratoire chez les moins de 15 ans

J00 à J1991565ans : toutes les maladies de l'appareil respiratoire chez les 15-65 ans

J00 à J19965ans : toutes les maladies de l'appareil respiratoire chez les plus de 65 ans

lpy : Indeno(1,2,3-cd)pyrène

Nap : Naphthalene

Ni : Nickel

NO2 : dioxyde d'azote

P14_CHOM15642 : nombre de chômeurs de 15 à 64 ans

P14_NPER_RP_LOC : nombre de personnes en résidences principales occupées par des locataires

P14_NSCOL15P_DIPL0 : nombre de personnes non scolarisées de 15 ans ou plus, titulaires d'aucun diplôme

P14_PMEN_ANEM10P : population des ménages qui a emménagé dans l'IRIS en 2014 depuis plus de 10 ans

P14_POP_IMM1 : nombre de personnes immigrées

Pb : Plomb

Phe : Phenanthrène

PM10 diesel : Poussières de moins de 10 µm de diamètre

PM2,5 diesel : Poussières de moins de 2,5 µm de diamètre

R : Coefficient de corrélation

S02 : Dioxyde de soufre

Se ou Sn : Sélénium

V : Vanadium

VDL : Valeur de Diversité lichénique

Zn : Zinc

. Remerciements

Je tiens à remercier en premier lieu les membres éminents du Jury qui me font l'honneur de leur présence. Merci pour ce très beau jury pluridisciplinaire ! Plus particulièrement mes rapporteurs pour le temps consacré à l'examen de ce Mémoire d'HDR.

Mes plus profonds remerciements vont également à :

Mon garant, Pascal Staccini, Professeur de Santé Publique à l'Université Côte d'Azur.

Mes directeurs d'unité et de laboratoire pour leur soutien sans failles. Mes collègues universitaires de l'UMR ESPACE avec qui les échanges méthodologiques sont toujours passionnants ! Merci pour cette dynamique de Laboratoire ! Les membres de l'Action Espace & Santé de l'UMR ESPACE : Nicole Vernazza-Licht, Daniel Bley, Régis Darques, Eric Caroll.

Plus largement, mes collègues d'UCA qui appartiennent à d'autres disciplines (Sciences ou SHS) pour leurs partages de connaissances enrichissants et forts sympathiques. Mentions amicales à Damienne Provitolo du laboratoire Géoazur, ainsi qu'à l'équipe du projet NR2P2.

Mes collègues et amis de la Géographie de la santé.

Je n'oublie pas non plus mes anciens professeurs, qui m'ont toujours encouragée au fil des décennies, je pense en particulier à Christine Voiron et à Maryse Guigo.

Lionel Jouffe et Stephan Conrady de la société BayesiaLab pour leur sympathie et leur aide à chaque fois que j'en avais besoin. Le personnel d'AtmoSud, et de l'IECP avec qui c'est un véritable plaisir de travailler et d'échanger.

Mes amis Pierre Sicard de la société Argans et Pierre Hébel d'Arches Conseil.

Merci également aux étudiants de Géographie, passés, présents, et futurs qui me donnent l'énergie de développer ces recherches, dans l'espoir de créer, un jour, une « école » de pensée à Nice, sur les questions de santé-environnementale. A mes doctorants pour leur confiance.

A mon amie d'enfance, Garance, pour les quelques jours de « tracances » passés cet été en Provence.

A mes camarades de méditation.

Bien entendu, à mes parents, ma famille, mes amis.

. Table des matières

. Sigles.....	4
. Remerciements	9
. Table des matières	10
. Introduction	12
1. Positionnement de la recherche	16
1.1 Parcours de recherche.....	16
1.2 La géographie de la santé, une géographie en santé	20
1.3 Objet de recherche	25
1.3.1 L'exposome.....	26
1.3.2 Le concept d'espace pathogène	28
1.4 De la nécessaire Causalité.....	35
1.4.1 Les modèles contrefactuels	39
1.4.2. La théorie de la causalité & les réseaux bayésiens	41
1.4.3. Eléments de compréhension de l'expression de la causalité dans un DAG.....	43
1.4.4. Les modèles de type Médiation	45
1.4.5 La création d'un réseau bayésien	54
2. La preuve de concept, les résultats de l'étude Air Santé en Pays de Martigues	57
2.1 Etat de l'art des études menées sur ce territoire	57
2.2 Les données sanitaires.....	60
2.3 Les données de pollution atmosphérique	61
2.4 Les relevés lichéniques : bio-indication et bio-imprégnation	65
2.5 Les données socio-économiques	67
2.6 Méthodologie : la modélisation bayésienne	68
2.7 Les interactions entre les variables : le monitoring	77
2.7.1. Les interactions entre les maladies respiratoires chez les plus de 65 ans & CB169	77
2.7.2 Les interactions entre les maladies respiratoires chez les 15-65 ans & BkF_p	79
2.7.3 Les interactions entre les cancers des bronches et du poumon chez les plus de 65 ans & plusieurs variables.....	80
2.7.4 Les interactions entre les cancers des bronches et du poumon chez les 15-65 ans & l'acide fluorhydrique	82
2.7.5 Les interactions entre les cancers de la vessie chez les plus de 65 ans & plusieurs variables	83
2.7.6. Les interactions entre le diabète chez les plus de 65 ans & le cadmium.....	85
2.7.7 Les interactions entre les maladies cardiovasculaires chez les plus de 65ans & plusieurs variables	86
2.7.8 Les interactions entre les revenus et l'exposition à l'acide chlorhydrique	88
2.7.9 Les interactions entre le fait d'habiter dans un IRIS depuis plus de 10 ans & plusieurs variables	89

2.8	Les analyses de sensibilité par IRIS	92
2.9	Les profils sanitaires des IRIS	100
2.10	La validation de la modélisation bayésienne	101
2.11	L'introduction du spatial dans les réseaux bayésiens	107
3.	Perspectives de recherche	126
3.1	De la causalité au potentiel	126
3.2	Le concept d'eXpace	131
3.3	Du potentiel à l'émergence	140
3.4	La prévention de l'émergence	141
	Conclusion	144
	Références	147
	Annexes	169
	Liste des figures	172
	Liste des cartes.....	173
	Liste des tableaux.....	173
	Résumé.....	174
	Summary	174

. Introduction

Après deux années de pandémie, nul besoin de rappeler l'importance de la santé individuelle et collective, l'interaction entre ces éléments, voire, plus globalement, avec notre environnement, dans une perspective de santé environnementale.

Sur les 102 principales maladies couvertes par un rapport de l'OMS sur la santé mondiale (WHO, 2016), les facteurs de risque environnementaux intervenaient dans plus de 85 d'entre elles. La part des maladies imputable à l'environnement varie bien entendu entre les pathologies, mais aussi entre les pays en développement (25%), et les pays développés (17%). Selon l'étude *Global Burden of Diseases, Injuries, and Risk Factors Study (2015)*, (actualisée par Fuller et al ; en 2022) la pollution est responsable chaque année dans le monde de 9 millions de morts prématurées, et la pollution de l'air, d'1 décès sur 6. Il s'agit là d'une estimation prudente car, dans de nombreux cas, la causalité entre l'environnement et la maladie est, nous le verrons, complexe à détecter. Contrairement à ce que l'on pourrait penser, ces chiffres ne cessent malheureusement de croître, et les tensions énergétiques, économiques, ou bien encore climatiques (émissions de poussières) ne nous rendent pas optimistes pour l'avenir.

Selon l'OMS la « *santé environnementale comprend les aspects de la santé humaine, y compris la qualité de vie, qui sont déterminés par les facteurs physiques, chimiques, biologiques, sociaux, psychosociaux et esthétiques de notre environnement. Elle concerne également la politique et les pratiques de gestion, de résorption, de contrôle et de prévention des facteurs environnementaux susceptibles d'affecter la santé des générations actuelles et futures* ». Elle estime que pour chaque euro investi aujourd'hui dans le domaine de la santé environnementale ce sont 5 euros qui seront économisés demain dans les soins de la population.

Il faut savoir que nous respirons quotidiennement 15 m³ d'air soit 20 kg, à mettre en rapport avec les 1.5 l d'eau que nous buvons et les 2 kg d'aliments environ que nous ingérons. Les pathologies dues à des expositions en lien avec l'environnement sont variées, elles vont des allergies, à des sensibilisations cutanées, des bronchites à répétition qui peuvent déboucher sur de l'asthme, mais parfois cela peut être plus grave, et les personnes exposées peuvent développer des leucémies ou myéloïdes aigües pour une exposition trop fréquente au benzène (émanations des stations-services), ou au formaldéhyde, qui peut être en trop grande quantité dans les locaux, ou les maisons mal ventilés. En fait, nous sommes en permanence soumis à une série de xénobiotiques, à un cocktail de polluants d'origines diverses, estimé d'ailleurs à plus de 200 classes différentes, et qui peuvent se potentialiser.

Mais, nous ne respirons pas tous le même air. De grandes **inégalités écologiques** existent entre non seulement les pays du nord et du sud, mais également au sein des grandes agglomérations (problème multi-scalaire), plus précisément, entre les différents quartiers qui les composent (quartiers anciens, récents, favorisés, moins aisés, mixtes...). Ainsi, l'absence de logements sociaux en quantité suffisante, le prix de l'immobilier et le manque d'espace dans les grandes villes, conduisent de plus en plus de personnes à accepter des logements qui peuvent représenter de par leur situation, leur exposition ou leur ancienneté un risque pour leur santé, et ce, quel que soit le milieu (urbain, périurbain, rural).

Par conséquent, il existerait un lien entre éléments délétères et les disparités sociales rencontrées au sein d'une ville, entre les revenus de certains habitants, et leur acceptation implicite à habiter dans des conditions susceptibles de présenter un risque pour leur santé (proximité d'une autoroute, d'une 4 voies, d'un centre de traitement des déchets, d'une station-service...). Ces **inégalités sociales**, doublées donc d'inégalités environnementales vont impacter la santé des habitants, elles se sont mises en place plus ou moins progressivement au sein des espaces, comme le souligne un rapport de l'Inspection Générale de l'Environnement dès 2005 : « *Les inégalités écologiques des territoires sont celles produites par l'activité humaine, et tout particulièrement celles issues de l'histoire urbaine. Ces activités ne se sont pas développées d'une façon homogène sur le territoire urbain. Depuis toujours, la ville s'est organisée en quartiers et l'histoire en a façonné les différentes composantes, économiques, sociales et environnementales. Leurs transformations successives se sont accompagnées à travers le temps d'une mobilité des différentes catégories de population et des activités. Il existe bien un marquage social et écologique des territoires urbains* » (Diebolt et al., 2005).

Bien souvent, ces personnes vont avoir moins de ressources pour maintenir un bon état de santé, il s'agit par exemple de familles monoparentales, pouvant avoir un accès aux soins plus difficile, notamment à des soins de spécialités (Lucas-Gabrielli et al., 2022). L'ensemble de ces facteurs (environnemental, social, accès aux soins) dessinant alors des disparités spatiales, et des **espaces potentiellement pathogènes**.

De plus, si certaines populations semblent **plus exposées**, d'autres semblent **plus vulnérables**, comme les personnes âgées, qui peuvent souffrir d'insuffisances respiratoires, ou bien encore les enfants dont le système respiratoire immature est particulièrement fragile.

Mais, l'exposition à des polluants de l'air n'est qu'un élément parmi d'autres, qui semblent également intervenir dans l'état de santé des populations, comme l'exposition au bruit qui peut quant à elle générer des problèmes de diabète, de cholestérol, voire d'obésité, tout comme d'ailleurs une alimentation de mauvaise qualité. Dès lors, tout le problème de l'évaluation du risque sanitaire est là : la multiplicité des facteurs, et plusieurs facteurs qui n'ont *a priori* rien en commun (le bruit et l'alimentation) qui semblent concourir à une même pathologie. En outre, ce n'est pas parce que des personnes sont exposées de façon identique aux mêmes polluants qu'elles développeront la même maladie en raison de la susceptibilité individuelle (génétique).

Incertitude et complexité dominant donc dans l'évaluation du risque sanitaire, mais l'absence de preuves ne signifie pas pour autant l'absence de risques (principe de précaution), dès lors, nous devons tenter dans une perspective de santé publique de mieux comprendre l'enchaînement de causes-conséquences qui conduisent à des risques sanitaires.

En effet, l'importance croissante donnée par les individus à leur santé, à leur environnement, à leur qualité de vie, la prise en compte de ces questions par des chercheurs au-delà du contexte médical, et l'intérêt manifeste des politiques et institutions à ces questions, nous amène à proposer un positionnement de recherche qui place l'**Espace** dans ses multiples composantes (physique, humaine) au centre de ses interrogations, et notamment dans son potentiel à influencer sur la santé des populations. En effet, si nous avons beaucoup travaillé en tant que géographes sur la répartition et la variation des phénomènes dans l'espace *via* l'analyse spatiale, que savons-nous au fond de l'espace en lui-même, dans sa capacité à entrer dans notre corps ?

Nous pouvons commencer à poser que l'espace est un ensemble de propriétés au sens philosophique et mathématique du terme, qui le constituent, qui interagissent entre elles avec des temporalités différentes (le temps de la géographie physique n'est pas celui de la géographie humaine), et ce, à de multiples échelles. C'est la **combinaison simultanée** de ces différentes propriétés, leurs arrangements, leur concomitance, et leur impact sur la santé des populations qu'il nous paraît intéressant d'étudier ici.

La géographie de la santé place l'homme, son état de santé, ses comportements au sein de son espace de vie, voire de son espace au cours de la vie (trajectoires) et d'activités. Nous partons du principe que l'espace de vie des individus n'est pas aléatoire au sens de Zusman (2022), il est un construit, parfois sur plusieurs siècles en fonction des potentialités d'un milieu, en vue d'assurer la vie économique des habitants qu'il « enveloppe ».

Nous vivons simultanément dans un *continuum* d'échelles, dans une multitude d'espaces, avec lesquels nous sommes en résonance. Chacun d'eux peut potentiellement avoir un impact sur notre santé, depuis l'espace monde et le changement climatique, en passant par notre pays et l'organisation de son système de soins, la ville où nous habitons et son taux de pollution, l'espace de notre profession, de nos déplacements, de notre maison, sans oublier l'espace de nos cellules et de leurs interactions qui intéressent plus particulièrement les médecins. L'espace est vu ici comme une combinaison de multiples facteurs synergétiques qui dans une certaine configuration (proportion, concomitance) porte en lui des conditions favorables au développement d'une pathologie, ou au maintien d'un bon état de santé.

Mes recherches s'inscrivent dans le champ de la **géographie théorique et quantitative** dont je suis issue. J'essaye d'appréhender le lien qui peut exister entre l'environnement et la santé des individus qui vivent dans un espace particulier. L'environnement étant pris au sens large et recouvrant aussi bien l'exposition à des polluants que le contexte socio-économique. Ce qui m'intéresse, c'est, en tant que géographe, la différenciation spatiale de ce lien (**analyse**

spatiale), mais aussi le potentiel d'un espace à influencer sur la santé des individus vivant au sein de cet espace (**analyse de l'espace**), et c'est ce dernier aspect qui différencie mes travaux de ceux d'un épidémiologiste.

Ainsi, ce qui m'importe ce n'est pas seulement de savoir ce qui influe le plus sur la santé (socio-économie, mode de vie, ou bien environnement), ce qui compte c'est de savoir à quelle hauteur et pour CET espace d'étude chacun de ces déterminants intervient, et comment ils interagissent les uns avec les autres pour dessiner des conjonctions spatiales particulières, voire dans certains cas des situations plus ou moins pathogènes.

Le problème étant que nous devons travailler à une échelle fine, qui rend difficile *de facto* toute généralisation, car les recherches sont dès lors spatio-dépendantes. Néanmoins, nous verrons qu'à l'aide d'un nouveau concept, le concept d'**eXpace** nous allons essayer de dépasser ces difficultés pour pouvoir proposer dans les années à venir un cadre théorique, et être en mesure de fournir une **dose [géographique] réponse** dans le domaine d'une exposition à la pollution de l'air.

Pour cela, je vais avoir recours à une modélisation particulièrement bien adaptée à l'évaluation du risque sanitaire, à savoir une modélisation fondée sur les réseaux bayésiens, car elle permet de remonter aux causes des pathologies (étiologie), et de déterminer le poids de chacun des facteurs (environnement, social, etc.), et ce, de manière simultanée, en fonction de tous les autres facteurs considérés conjointement. S'il a pu être reproché aux réseaux bayésiens d'être a-spatiaux, nous verrons comment il est possible de retrouver l'espace *via* des analyses de sensibilité.

Les résultats pourront servir à de multiples perspectives de recherche, notamment à la métabolomique qui permettra non seulement de valider les eXpaces trouvés, mais également d'affiner l'inclusion de patients dans ce type d'analyses.

L'organisation du manuscrit, et de la pensée suivent les nouvelles consignes de la 23^{ème} section du CNU à laquelle j'appartiens, soit à la manière d'un article scientifique dans un volume de pages compris entre 80 et 150¹. Ainsi, nous présenterons dans un premier temps le positionnement de notre recherche, puis nous démontrerons grâce à une preuve de concept tout l'intérêt de faire appel aux réseaux bayésiens, avant d'envisager les perspectives de recherche.

Je vous invite à présent à la lecture de mon parcours de recherche, pour comprendre comment je suis arrivée à travailler sur ces questions.

¹ <https://conseil-national-des-universites.fr/cnu/#/entite/entiteName/CNU/idChild/32/idNode/3745-3899>

1. Positionnement de la recherche

1.1 Parcours de recherche

Aussi loin que je me souvienne j'ai toujours voulu comprendre le monde qui m'entourait, comprendre les paysages, les espaces qui s'offraient à ma vue. Très vite, la géographie par son aspect englobant m'est apparue comme étant la discipline me permettant d'approcher cela, même si je sais aujourd'hui que toute autre discipline m'aurait procuré le même sentiment. Ce fut amplifié par un professeur de SVT, Gérard Fujiglando, qui organisait au collège des sorties de terrain, et nous donnait les clefs de compréhension d'un monde dont notre soif de connaissances avait besoin.

C'est la raison pour laquelle je suivis d'abord à Nice un cursus de géographie physique orienté vers la géomorphologie, où j'ai pu bénéficier des cours remarquables sur le sujet du Professeur Maurice Jullian, puis de ceux d'Edward Anthony sur la géomorphologie littorale. Je m'inscrivis par la suite au DESS Gestion intégrée de l'Eau et de l'Environnement, ouvert à la Faculté des Sciences en 1992. Mais, la part humaine finalement me manquait, j'avais également été très attirée dans mon cursus par les cours du Professeur André Dauphiné qui nous parlait d'isopadanes, et je me demandais si finalement tout ne serait alors dû qu'à un effet de poids, l'espace géographique pouvant être représenté comme un espace en Sciences Physiques où des particularités agiraient en le déformant sous l'effet de masses et de pressions... ? Les dés étaient jetés. C'est ainsi que je m'engageais dans le DEA Réseau de géographie théorique et quantitative, puis entrepris une thèse de Géographie qui me permis d'appréhender l'espace sous la forme d'une Géométrie (au sens de Jean-Paul Ferrier, 1986) *via* les discontinuités spatiales.

Après avoir soutenue une thèse de Géographie, en 1999, sur le thème de l'« *Analyse spatiale des régions frontalières et des effets de frontière : application aux espaces frontaliers franco espagnols du Pays Basque et de la Catalogne, et à l'espace franco-italien des Alpes-du-Sud* » sous la Direction du Professeur Christine Voiron je fus définitivement confortée dans le choix de l'analyse spatiale comme support de mes réflexions, mis à profit mes connaissances acquises sur les discontinuités, et rejoins très rapidement le Groupe de recherche intersites « Discontinuités, interfaces et organisation de l'espace », composé de plusieurs membres de l'UMR ESPACE. Divers travaux sur ces questions sont, au fil des ans, publiés.

A partir de 2007, je change complètement de thématique de recherche, celle de l'étude des frontières et des discontinuités me paraissant trop restrictive pour être prolongée sur une carrière. C'est Jean-Pierre Renard, Professeur à l'USTL, Rapporteur de ma thèse qui va en être à l'origine. En effet, il va me conseiller d'aborder cette fois les effets de frontière sous l'angle de la santé, entre par exemple la France et l'Italie. Les deux pays ayant des systèmes de santé complètement différents (de nature bismarkienne pour la France et beveridgienne pour l'Italie) je n'ai pas de mal à mettre en évidence des inégalités d'accès aux soins. Inégalités, qui sont d'autant plus dramatiques que dans certains cas elles peuvent se traduire par des pertes de chance de survie.

Mais, si les inégalités d'accès aux soins peuvent en partie expliquer les inégalités de santé observées, elles ne recouvrent pas cependant totalement ces dernières, dès lors je décide d'inclure dans ma démarche les inégalités sociales, mais aussi environnementales, qui de plus peuvent se combiner, plus rarement se soustraire. Ce qui me conduit à considérer simultanément de multiples variables, de natures différentes, et à me rapprocher dès 2009 d'outils comme les réseaux bayésiens (*deep learning*) pour traiter cette masse d'informations, et analyser les liens entre toutes ces variables. En 2009, j'obtiens un BQR de l'Université pour développer ces questions, sur une proposition de « Modélisation du risque sanitaire en milieu urbain par les réseaux bayésiens : application à la ville de Nice ».

L'actualité me donne malheureusement l'occasion en 2011 de tester la capacité des réseaux bayésiens à calculer les niveaux de probabilités spatialisés des enfants japonais à développer des cancers, 60 ans après leur exposition aux radiations émises par la Centrale nucléaire de Fukushima-Daiichi. A cette époque, je m'intéresse également aux effets de la radioactivité naturelle de type radon dans la région Sud. Ce qui m'amène à être associée au projet Idex UCA JEDI NR2P2 (Nuclear Risk, from Radioisotope (bio)chemistry to Public Perception), à travailler au sein d'une équipe pluridisciplinaire composée de radio-chimistes, de biologistes, de spécialistes de la gestion de crise et du risque, du droit de l'environnement, et à réfléchir à la question de la perception du risque nucléaire, aux différences qui semblent exister entre une population experte (sapeurs-pompiers) et une population non experte (composée d'étudiants de mon Université).

En 2015, suite à l'exposition des résultats de Fukushima dans une Conférence organisée par le Cancéropôle PACA, je suis contactée par le service de santé de la mairie de Martigues qui me commande une étude de type Air-Santé afin d'établir un diagnostic territorial à fine échelle entre pathologies d'intérêt et pollution de l'air. Je m'y consacre pendant plusieurs années (2015-2019), développe une méthodologie inédite, qui donne lieu à des résultats pertinents (plusieurs relations trouvées parmi les 178 variables mobilisées). Ces résultats font l'objet d'une validation par les pairs (Pérez, 2021), ainsi que d'une audition par la Commission des Affaires Sociales de l'Assemblée Nationale, le 4 octobre 2021.

J'intègre le GDR O3 (Odorants, Odeurs, Olfaction) également en 2015, et entame avec un doctorant, Jamel Ben Hassine, un travail de recherche sur les odeurs, plus précisément sur les nuisances olfactives, ce qui me permet d'affiner les niveaux de pollution auxquels la population du pays de Martigues est soumise (derrière les odorants, se cachent parfois des polluants), puisque les plaintes déposées sur le site de surveillance régionale des odeurs d'AtmoSud sont cette fois aux adresses.

J'obtiens en 2016 un CRCT d'un semestre au titre de l'établissement (Septembre 2015 - Février 2016), qui me permet de réaliser une *meta review* des travaux menés entre 2002 à 2015 en Géographie de la Santé, sur la base de 360 articles publiés dans une des revues de référence : *International Journal of Health Geographics*.

En 2020, suite à la pandémie, je présente le projet MoDeus (MoDelisation dans l'espace du Coronavirus) à l'ANR *Reacting* qui n'est malheureusement pas retenu. Je réalise avec mon doctorant 2 vidéos sur la diffusion spatiale du virus en Russie, et publie un article en 2021 sur le décalage entre les mesures prises sur la base du principe de précaution par le gouvernement dans la gestion des décès lors de la 1^{ère} vague, et la perception de ces mêmes mesures par les familles en grande souffrance, car ils n'ont pu dire au revoir à leurs proches, ni pratiquer certains rites funéraires.

A l'heure actuelle, mes recherches sont en lien avec la notion d'exposome, car j'ai de plus en plus le sentiment, comme d'autres, que la pollution doit être analysée de manière globale, touchant simultanément plusieurs compartiments (air, eau, sol) comme j'ai pu m'en rendre compte dans l'Etude Air-Santé du pays de Martigues. En 2021, suite à la demande d'encadrement en M2 d'Anaïs Squarta, étudiante du Master Humanités Médicales, Parcours "Anthropologie Evolutive : os, gènes, cultures" hébergé à La Timone, je travaille sur les agénésies transverses des membres, pathologies particulièrement invalidantes qui nécessitent dans la recherche de leur étiologie de raisonner en termes d'exposome. Une demande de bourse régionale a été présentée à l'ED SHAL, (classée seconde) afin de poursuivre ces recherches. J'ai également été membre de l'ANR Expostroke (Exposome and Stroke) déposée dans le cadre de l'AAPG 2022 de l'ANR, par mon collègue Thierry Pourcher du laboratoire Tiro-Matos, qui combinait approches métabolomiques et protéomiques. Ce projet visait à apporter de nouvelles connaissances sur la relation entre l'exposition aux polluants atmosphériques et la survenue d'AVC, au niveau moléculaire, en identifiant des biomarqueurs d'exposition. Nous verrons dans la partie concernant les perspectives de recherches comment la métabolomique permettrait de lever certains verrous scientifiques, et le bénéfice qu'elle pourrait également tirer de la méthodologie que nous allons déployer dans les années à venir.

Dans la continuité de l'Étude Air-Santé un projet de recherche a également été déposé dans le cadre du Programme National de Recherche Environnement-Santé-Travail (PNR EST), « *Analyse par intelligence artificielle du lien entre polluants atmosphériques et diabète de type 1 : cas de la région industrielle Fos-Berre* ». L'objectif est d'étudier sur la base de l'étude Air-Santé, plus précisément, la relation entre polluants diabétogènes et diabète de type 1, entre 2013 et 2020. Une Attachée de Recherche Clinique (ARC) qui travaille au sein du service de Diabétologie Endocrinologie Nutrition du Dr Sarde au Centre Hospitalier de Martigues, Clémentine Kerdat, travaillera sur ces questions dans le cadre d'une thèse, à partir de la rentrée 2022-2023.

Mes travaux scientifiques sont par essence pluridisciplinaires, ils sont à l'intersection de la géographie, de la médecine, et des sciences de l'environnement, d'où la nécessaire collaboration avec des médecins en charge de certaines pathologies (Pascal Staccini, Pierre Hébel (ARC), Thierry Pourcher, Laurent Suissa, Barbara Seitz-Polski, Laurence Pascal, Patrick Courtin, Elisa Sarde, Annie Lévy-Mozziconacci).

Parfois, je suis conduite à travailler sur un objet transdisciplinaire, comme l'odeur. En effet, certaines associations agréées de surveillance de la qualité de l'air consignent depuis quelques années sous la forme d'un *corpus* les plaintes de riverains qui sont invités à décrire les nuisances olfactives dont ils sont l'objet (description de l'odeur, effets physiques associés comme les maux de têtes, voire la nausée, l'origine supposée de l'odeur etc.). Ces informations intéressent bien sûr les géographes de la santé à la fois pour la description des symptômes évoqués, mais aussi, nous l'avons vu, pour une évaluation plus fine de l'exposition des riverains à certains polluants. Mais, ces données peuvent également intéresser les linguistes pour l'analyse lexicale des termes employés pour décrire l'odeur, les anthropologues qui travaillent sur la sensorialité (Joël Candau, 2016), les psychologues pour approcher ce qui se joue derrière ces plaintes, et les neurosciences pour comprendre comment s'effectue la reconnaissance des odeurs. Sans oublier les chimistes concernés par les analyses physico-chimiques *in situ* afin de mesurer objectivement les molécules présentes dans l'air.

La santé environnementale m'a en fait, permis de réconcilier les deux parcours empruntés au cours de mon cursus : la géographie physique et la géographie humaine, mes recherches se situant à l'interface entre les deux.

Mais, avant de présenter l'objet de mes recherches à proprement parler, et son positionnement en géographie de la santé, voyons en quoi consiste plus précisément cette dernière.

1.2 La Géographie de la santé, une géographie en santé

La Géographie de la santé est un champ relativement nouveau de notre discipline. Elle traite de la façon dont l'espace (géographique) interagit avec les individus, leurs corps, pour produire des effets sur leur santé. Ces effets sont généralement étudiés du point de vue négatif. L'espace peut entrer en relation par le biais de la distance (rugosité, frein) engendrant par exemple une plus grande difficulté d'accès aux soins, ou bien par le voisinage (contamination dans le cas des virus, proximité avec des axes routiers, un aéroport, des industries polluantes).

Les recherches menées sont florissantes et concernent des thèmes aussi divers que la propagation de maladies infectieuses ou vectorielles, l'accessibilité aux soins, ou bien encore la recherche de déterminants environnementaux dans l'émergence d'une pathologie (thèmes non-exhaustifs). L'analyse spatiale à travers l'étude du voisinage, de la diffusion, de l'interaction spatiale, de l'autocorrélation, de l'interpolation, des cheminements, de l'accessibilité, est tout naturellement au cœur de ces recherches.

Il y a quelques années pour rendre compte de cette diversité et de l'évolution des recherches menées en Géographie de la Santé, nous avons entrepris une *meta review* en nous fondant sur les articles publiés entre 2002 et 2015 dans la revue de référence : *International Journal of Health Geographics*, sur la base d'une sélection de 360 articles qui avaient employé le terme « *space* » soit dans leur titre, leur résumé, ou leur texte (Pérez et al., 2016).

Ainsi, au début des années 2000, ce sont les questions de cartographie, de visualisation des données, d'usage des SIG qui ont surtout concentré l'attention des chercheurs. Cela allait de pair avec le géoréférencement des données, notamment des données de cancer rendu possible aux USA, et le développement d'outils d'analyse spatiale. L'analyse spatiale est durant cette période essentiellement utilisée dans le but de détecter des clusters "*we were able to pinpoint geographic areas with higher risk through exploratory spatial analyses, and to assess temporal variability of the risk surfaces, thus providing a working hypothesis on breast cancer and environmental exposures*" (Han et al., 2005), (Pickel et al., 2005).

Les auteurs s'interrogent alors sur les meilleures méthodes statistiques pour détecter ces *hot spots*. Ainsi, par exemple, dans l'article *A flexibly shaped scan statistic for detecting clusters*, Tango et Takahashi remettent en question la statistique de Kulldorff qui utilise une fenêtre circulaire, et proposent une autre méthode pour approcher les clusters de forme linéaire (le long d'une route ou d'une rivière, par exemples), (Tango et al., 2005). Les chercheurs sont également dans la comparaison de méthodes, de modèles statistiques *Comparison of spatial scan statistic and spatial filtering in estimating low birth weight* (Ozdenerol et al., 2005), ou plutôt dans leur combinaison pour produire des résultats plus fiables (Griffith, 2005). Les SIG sont alors perçus comme pouvant jouer un grand rôle dans le suivi et la compréhension de la diffusion d'une épidémie de type SRAS en 2003. Ils permettent d'aller très loin, jusqu'à l'identification des bâtiments infectés qu'il conviendrait d'éviter (Boulos et al., 2004).

Les clusters spatiaux sont recherchés, mais également leur inertie temporelle. Ainsi, dans *Detecting spatio temporal clusters of accidental poisoning among Texas counties, US, 1980-2001* Nikhoma et al., examinent si les clusters varient spatio-temporellement lorsque le genre et l'ethnie sont pris en compte. Les résultats révèlent la persistance à travers le temps de la population noire comme étant plus à risque d'accident par empoisonnement, en comparaisons d'autres groupes ethniques, et ce, d'autant plus lorsque la population masculine est analysée.

Parallèlement, des mises en garde contre l'usage systématique de l'analyse spatiale, ou tout du moins ses limites apparaissent dès 2004 (Jacquez et al., 2004), avec le problème de l'agrégation spatiale (MAUP), la subjectivité contenue dans les modèles spatiaux, la recherche d'associations de variables qui n'impliquent pas forcément la causalité entre elles, le fait qu'il n'existe pas un modèle absolu, et que chacun possède ses propres intérêts et limites, le problème de l'illusion écologique etc. Le terme de "cluster" est lui-même remis en cause, ou du moins son manque de précision.

Entre 2006 et 2010, les questions de clusters mobilisent encore l'énergie des chercheurs, mais le lien entre géographie de la santé et les progrès technologiques est de plus en plus manifeste, à travers les possibilités de géolocalisation (Rainham et al., 2008), (Wiehe et al., 2008). Parallèlement, les questions éthiques, et notamment de droit à la vie privée qui vont avec font leur apparition dans certains articles (Boulos et al., 2009). La désagrégation des données géographiques rendant désormais possible une réflexion à l'échelle des individus.

Plus récemment, entre 2011 et 2015, apparaît la question de la potentialité de l'espace lui-même à influencer sur la santé, à travers par exemple sa capacité à favoriser la marche (*walkability*) (Mayne et al., 2013), ou inversement des environnements peu propices à cette activité, voire même favorables aux chutes (Lai et al., 2011). Ainsi, dans ce dernier article ont été identifiés, dans un quartier particulièrement dense de Hong Kong (Mong Kok) des lieux propices à la chute, comme le voisinage des stations de métro, la jonction de rues "embouteillées », des feux avec un signal de passage trop court pour les piétons, des bordures de rue mal adaptées à la fluidité de la marche, ou bien encore un changement soudain et abrupt de bordure de trottoir. Dans *Geospatial examination of lithium in drinking water and suicide mortality* (Helbich et al., 2012) il semblerait même, que dans certains espaces, le taux de suicide soit inférieur à la moyenne nationale en raison de la présence naturelle dans l'eau de distribution d'une certaine concentration de lithium, qui est, on le sait, un régulateur d'émotions.

Enfin, dans les articles plus récents (2015-2022) les auteurs parlent d'espace intelligent dans le sens par exemple de la "*smart city*", (Gong et al., 2015). L'espace y est alors appréhendé à travers un réseau de capteurs intelligents et communicants qui seraient capables de produire des données en temps réel (couplage SIG et plateformes web), et de nous alerter lorsqu'un « événement » se produit. Il faut entendre ici le terme événement comme un phénomène géographique qui se produit en un point spécifique à partir d'un certain seuil de déclenchement, par exemple, après un pic de pollution de l'air. Tandis que, l'espace-vert, ou du moins les espaces-verts contribueraient, sous certaines conditions, non seulement à notre santé, mais aussi à notre bien-être général, voire à notre santé mentale. Néanmoins, Wheeler et al., soulignent l'importance d'études approfondies, car cette relation, a priori positive, ne serait pas aussi linéaire que l'on voudrait bien le supposer. La corrélation entre espace-vert et bonne santé ne semble pas aussi évidente.

Nous venons de voir, l'évolution au fil du temps des articles publiés dans cette revue de référence. Cependant, même si des tendances peuvent être décelées, les articles sont très diversifiés, ce qui révèle la richesse de ce domaine de recherche. De plus, la façon dont l'espace est abordé peut-être très différent. Pour nous en convaincre nous avons sélectionné des articles qui concernent cette fois la même thématique de recherche, mais où l'espace n'est pas abordé de la même manière. Le rapport à l'espace des différents auteurs peut aller de l'espace considéré comme un simple support des variables qui s'y déploient, à un espace reconnu comme un véritable acteur, porteur d'attributs spécifiques qui vont influencer dans un sens soit positif (amplification, facilitation...), ou au contraire négatif (barrière, contrainte, atténuation..), le phénomène, ou la pathologie étudiée.

Regardons à présent, comment l'espace est considéré à travers deux domaines qui ont concentré l'énergie de beaucoup de chercheurs en Géographie de la santé, à savoir les maladies vectorielles, et l'accessibilité aux soins. Dans le cas d'articles portant sur les maladies vectorielles, l'espace, est à la fois acteur, à travers les multiples déterminants (environnementaux, socio-économiques, climatiques) qui influent sur la diffusion des virus, que support. Ainsi, in *Environmental predictors of West Nile fever risk in Europe* (Tran et al., 2014) l'accent est mis sur le rôle des déterminants environnementaux dans la diffusion spatiale du virus du *West Nile*, maladie vectorielle transmise à l'homme par la piqûre d'un moustique infecté. Les déterminants environnementaux désignent ici les caractéristiques physiques et climatiques des espaces étudiés. Afin de comprendre la place de ces facteurs environnementaux, un modèle de régression logistique est utilisé, avec comme variable dépendante le statut à l'infection du virus (infecté/non-infecté), et comme variables explicatives des déterminants environnementaux issus de données de télédétection, comme par exemple, la présence de zones humides. Le modèle permet de souligner la part de l'environnement qui explique le risque d'infection au virus. Ce modèle peut par la suite être utilisé comme modèle prédictif afin d'estimer la probabilité de futurs cas. Ici, l'espace est donc perçu comme un acteur présentant des caractéristiques physiques et climatiques variables dans le temps et l'espace, ce qui joue un rôle déterminant dans la transmission de la maladie vectorielle, et explique son inégale distribution (spatiale et temporelle). Grâce à l'étude des caractéristiques environnementales de l'espace étudié, il est ainsi possible de mieux comprendre et de prédire la variabilité spatiale du risque de transmission.

In Agricultural landscape and spatial distribution of Toxoplasma gondii in rural environment: an agent-based model (Gotteland et al., 2014) les auteurs considèrent cette fois l'espace essentiellement comme un espace support visant par l'intermédiaire d'un modèle multi-agents à prédire la distribution spatiale d'un pathogène (difficile à appréhender dans la réalité), via un certain nombre de variables qui semblent influencer sur sa répartition. Le parasite en question est le *Toxoplasma gondii*, responsable de la toxoplasmose, excrété dans l'environnement par les chats infectés (hôte définitif), et les rongeurs (hôte intermédiaire). La répartition spatiale des fermes (en tant que lieu de refuge pour les chats), ainsi que la distance à la ferme la plus proche serviraient entre autres de supports à la diffusion du parasite, et expliqueraient de hauts niveaux de contamination. Le risque de contamination serait plus important si les fermes sont à l'intérieur d'un village que si elles sont plus dispersées. Les entités du modèle sont les chats, les rongeurs, les bâtiments de fermes et les cellules de l'environnement, chaque agent et cellule sont caractérisés par un état de contamination et des règles qui leur sont propres (dynamique de population, activité, mobilité, par exemples, pour les hôtes).

Dans l'article de Kienberger et al., *Spatial-explicit modeling of social vulnerability to malaria in East Africa* (2014) l'espace est à la fois perçu comme support du phénomène étudié et acteur dans la représentation de celui-ci. Cet article a pour intérêt de proposer une méthode originale d'analyse du risque au paludisme, qui peut servir d'appui aux mesures d'intervention. La méthode est basée sur un processus de régionalisation. Dans cet article, les résultats (un indice de vulnérabilité au paludisme dans la région Est de l'Afrique subsaharienne) sont visualisés par une unité géographique construite appelée "Geon". L'idée est de proposer une alternative au MAUP par une méthode d'agrégation des données qui s'affranchit des limites administratives, ou des cellules d'une grille lorsque les données sont carroyées. Le choix même de l'agrégation des données et de leur régionalisation étant nous le savons un processus qui influe directement sur la modélisation et la représentation de la variabilité du phénomène étudié.

A la différence des maladies vectorielles qui ne se transmettent donc pas directement d'homme à homme, les interactions de type contact plus ou moins rapproché entre individus sains et infectieux sont déterminantes dans les maladies infectieuses contagieuses (comme nous avons pu le constater après 2 ans de pandémie due au Covid) (Qi et al., 2013). Le suivi détaillé des comportements individuels (leur mobilité dans l'espace-temps, leur exposition au facteur de risque en résultant) est rendu possible grâce aux outils de géo-localisation. Ce sont alors les individus qui par leurs interactions dans leurs espaces d'activité propageraient la maladie et créeraient par la même des espaces plus ou moins à risque d'infection.

Nous allons voir à présent comment l'espace a été pris en compte dans les recherches concernant l'accessibilité à des éléments favorisant la santé (établissements de soins, magasins alimentaires de qualité). Nous nous sommes focalisés sur 3 articles (Larsen et al., 2008, Paez et al., 2010, Salzé et al., 2011), mais nous aurions pu en choisir bien plus tant cette thématique a mobilisé les recherches des géographes de la santé pendant au moins 2 décennies. Nous allons voir que pour un même objectif (accessibilité) l'espace a été analysé de manières différentes.

Les facteurs spatiaux ont été utilisés pour examiner comment les mouvements habituels des personnes interagissent avec leur environnement. L'accès est influencé par la forme et la superficie de l'espace de mobilité d'un individu, la distribution spatiale des opportunités d'accès, et par les structures spatiales qui contraignent et dirigent le mouvement dans l'espace. La portée de l'espace de mobilité dépend de la façon dont il est conceptualisé et mesuré. Alors que Salzé et Larsen prennent en compte les facteurs spatiaux, Paez et al. prennent en compte à la fois les facteurs individuels (capacité de déplacements des personnes âgés) et spatiaux.

Diverses méthodes ont été mobilisées pour estimer l'accessibilité spatiale. Les SIG ont été utilisés pour mesurer la distance entre l'habitat et les services de soin, ou les magasins alimentaires. Pour ces mesures, les auteurs ont considéré la distance à pied, en voiture, en transports en commun et dans les rues, *via* les calculs de la distance moyenne, la distance en ligne droite, l'approche basée sur le réseau (Larsen et al.), ou bien encore des indicateurs comme l'accessibilité relative (Paez et al.).

Nous venons de passer en revue les principaux sujets de recherche en Géographie de la santé, qui nous l'avons vu sont variés, aussi bien dans les thématiques que les méthodologies employées. Mais, comment s'inscrivent mes recherches dans le champ de la Géographie de la santé ?

1.3 Objet de recherche

Elles s'inscrivent depuis plus de 15 dans le cadre de la santé environnementale, plus précisément dans celui des études écologiques. Ce qui m'intéresse particulièrement, ce sont les liens entre l'influence des individus sur l'environnement (pollution), l'influence de ce même environnement sur les individus (santé), et les multiples interactions entre ces deux éléments : pollution & santé. Je suis persuadée que la prise en compte de l'environnement passe par une prise de conscience de ses effets sur notre santé, lorsque celui-ci est dégradé.

L'environnement est compris ici dans son acception la plus large, c'est-à-dire incluant à la fois par exemple des mesures de pollution, mais également le contexte socio-économique de vie des habitants. Dans une perspective d'approche de l'exposome, l'environnement, la socio-économie sont donc considérés simultanément. La socio-économie nous le savons peut parfois compenser certaines expositions environnementales délétères, mais la plupart du temps du fait de la rente foncière, les deux tendent à évoluer en sens inverse (l'exposition augmente au fur et à mesure que la socio-économie diminue), dessinant alors des **configurations spatiales pathogènes** et produisant *de facto* des inégalités socio-environnementales de santé (Chasles, 2019).

Un effet en santé environnementale dépend de plusieurs facteurs, il dépend de la durée de l'exposition, de l'intensité de celle-ci (concentration, toxicité du polluant), de la vulnérabilité de la personne exposée (enfants, femmes, personnes âgées), de sa capacité à atténuer/accentuer son exposition (accès aux soins, alimentation, hygiène de vie, type de travail), et enfin de sa susceptibilité à être impactée qui peut être fonction de son patrimoine génétique.

Pour reprendre une terminologie empruntée à l'écotoxicologie, **un effet peut être de différentes natures, il peut être :**

. **additif**, quand la toxicité d'un mélange est égale à celle prédite par l'addition de la toxicité de chaque composant (polluant) du mélange;

. **supra-additif** (synergie ou potentialisation), quand la toxicité induite par le mélange est plus élevée que la somme de celles qui seraient induites par chaque composante du mélange, on parle alors d'un effet **synergique**; quand un mélange de plusieurs toxiques produit une toxicité supérieure à la somme de celles de chacun des toxiques (p.ex. $2 + 3 = 10$), ou bien d'effet **potentialisé**; quand un composant non toxique seul, augmente la toxicité d'un autre produit (ou de plusieurs autres), (c'est aussi un effet supra-additif, de type « $3 + 0 = 7$ »);

. et enfin, **infra-additif** (antagonisme), quand le mélange est moins toxique que l'addition de la toxicité de tous ces composants.

Nous voyons déjà à ce niveau que l'étude de ces liens, va demander de travailler avec de multiples variables, et que cela va donner lieu à des études complexes pour identifier la part de chacune dans les pathologies observées.

Il convient à présent de préciser comment mes recherches se positionnent par rapport à celles menées sur l'exposome, de même, que par rapport au concept d'espace pathogène.

1.3.1 L'exposome

Dans l'article de référence de Christopher Wild au sujet de l'exposome (2005) celui-ci plaide pour un développement de mesures visant à évaluer, quantifier l'environnement des patients à la même hauteur que les études déployées alors pour mesurer le génome. « *There is a desperate need to develop methods with the same precision for an individual's environmental exposure as we have for the individual's genome. I would like to suggest that there is a need for an "exposome" to match the "genome."* (Wild, 2005).

De nombreuses études en Europe et dans le monde se concentrent à l'heure actuelle sur l'exposome qui est inclus dans le Plan National (français) de Santé Environnement depuis le PNSE² grâce à des personnes comme Francelyne Marano qui ont œuvré pour que cela le soit.

²<https://solidarites-sante.gouv.fr/sante-et-environnement/les-plans-nationaux-sante-environnement/article/plan-national-sante-environnement-pnse3-2015-2019>

Il s'agit d'études complexes, très fines, faisant appel aux omiques (métabolomique, protéomique, adductomique, transductomique...), elles sont menées la plupart du temps par des épidémiologistes, ou bien des (éco)toxicologues (Slama, 2015). Leur objectif est d'avancer dans la connaissance des doses-réponses, en étudiant finement l'exposition de groupes d'individus suffisamment homogènes pour distinguer la variabilité interindividuelle (même sexe et âge par exemples), et suffisamment hétérogènes (exposés *versus* non exposés) pour percevoir les effets de la variabilité d'exposition.

Dans la version idéalisée de l'exposome celui-ci devrait pouvoir couvrir l'ensemble des expositions environnementales d'un individu tout au long de sa vie. Nous voyons là toute la difficulté à réaliser cela à la fois pour des raisons éthiques, de vie de la recherche au sein des laboratoires (conduire une même étude qui mobiliserait toutes les énergies pendant plus de 80 ans), que de traitements et d'analyses de données. Cependant, certaines cohortes (Elfe en France³, Sapaldia en Suisse⁴) se sont engagées sur ce chemin. Et le plus important n'est peut-être pas d'atteindre l'objectif visé que de cheminer, de voir comment on peut « attaquer » cet objet de la façon la plus efficace possible, car nous ne pouvons pas pour l'instant étudier sérieusement l'exposome dans son entièreté (hors métabolomique), la tâche est titanesque. De plus, ne serait-ce que dans le domaine de la pollution de l'air nous ne connaissons pas encore tous les mécanismes d'interactions entre les multiples polluants auxquels les individus sont exposés (effets cocktails), et pour certains polluants les valeurs toxiques de référence ne sont même pas établies. Si des incertitudes subsistent à ce niveau il ne sera dès lors pas possible de projeter dans le temps (prospectif) les relations trouvées.

Mais, ce concept a le mérite de faire sauter les silos entre les spécialistes de la pollution de l'air, du sol, des eaux, du biotope, et de poser le problème de la multi-exposition, qui plus est cumulative. Ce faisant ces spécialistes doivent travailler ensemble et mettre leurs connaissances en commun afin de voir dans un premier temps **ce qui se passe aux interfaces**, comment se font les transferts de contamination air-sol, sol-air, sol-eau, suivant quels mécanismes ? Et comment les populations peuvent être exposées (voies de contamination).

Une adaptation raisonnée de ce concept à la réalité est également de porter l'attention à des fenêtres d'exposition critiques à certains moments de la vie des individus (qui pourraient avoir des effets sur le long terme), ainsi qu'aux périodes sensibles (susceptibles d'effets importants) (Ben-Shlomo et Kuh, 2002). Les deux (critiques et sensibles) se succédant au fil de la vie, et constituent ce que l'on appelle des chaînes de risque (Vineis et al., 2017).

³ <https://www.elfe-france.fr/> (de la naissance jusqu'au 20 ans des enfants inclus dans la cohorte)

⁴ <https://www.swisstph.ch/fr/topics/non-communicable-diseases/human-biomonitoring/sapaldia-fr/>

Néanmoins, si ce n'est pas encore un concept tout à fait opérationnel, le concept d'exposome a le mérite d'organiser la pensée, les réflexions, de montrer le cap, de fédérer les énergies autour d'un but commun.

Nous nous plaçons ici dans une autre perspective que le concept englobant d'exposome. En effet, si ce concept est performant pour passer en revue tous les facteurs auxquels un individu doit faire face au cours de sa vie, ou bien lors de fenêtres d'exposition critique (pré ou post-natales), il n'a pas pour objet de catégoriser différents espaces de vie entre eux, la mission fixée est déjà énorme. Si ces études mettent bien en relation des niveaux d'exposition différents et des effets proportionnés, elles investiguent moins le « **pourquoi là ?** » si cher aux géographes. L'espace est dans ces recherches un facteur parmi les autres, mais ce n'est pas LE facteur déterminant. En effet, si nous prenons les publications de l'étude européenne « *Health and Environment-wide Associations based on Large population Surveys, HEALS* »⁵ elles ont trait à des polluants particuliers, dont les impacts ont été mesurés sur des populations cibles, en des points particuliers, mais sans qu'aucune règle généralisable au niveau des espaces n'en soit pour l'instant sortie, ce qui est bien entendu normal compte tenu de la complexité de ces recherches, et de la spécificité de chacun des espaces d'étude, nécessaires pour évaluer la diversité des doses réponses potentielles.

Or, en tant que géographe ce qui m'intéresse c'est de comprendre la **capacité de l'espace géographique à induire des pathologies** en raison de caractéristiques intrinsèques qu'il posséderait *in se, per se*. Cela en vue de catégoriser les espaces entre eux, dans l'espoir d'en dégager une théorie, nous développerons ces points dans la dernière partie de cette HDR. Je fais donc l'hypothèse que certains **lieux** (espaces pathogènes) possèdent des caractéristiques suffisamment fortes pour entrer dans les corps, et être alors en capacité de générer des pathologies. Par caractéristiques fortes, il est par exemple question d'intensité, de temporalité (pollution chronique), de dangerosité, de multiplicité des polluants, mais aussi de topographie, de vents etc.

1.3.2 Le concept d'espace pathogène

Hippocrate déjà (V^e siècle av. J.-C) dans son « *Traité des airs, des eaux et des lieux* » pressent le rôle que peuvent exercer les lieux sur la plus ou moins bonne santé de leurs habitants. Henri Picheral, géographe français, fut l'un des premiers à conceptualiser l'effet pathogène d'un espace géographique (Picheral, 1983) à travers le socio-système, le géon pathogène (composé aussi bien d'éléments appartenant à la géographie physique, à l'environnement, qu'à la géographie humaine, et à la part sociale des lieux habités). Ainsi, à propos de la Géographie de la santé, il écrit : « *S'il faut à tout prix proposer une énième définition, peut-être pourrions-nous dire simplement qu'il s'agit de l'étude spatiale de la*

⁵ <https://cordis.europa.eu/project/id/603946/results/fr>

qualité de la santé dans ses relations avec l'environnement physique, biologique, socio-économique, comportemental et culturel » (Picheral, 1982). Picheral parle même de « climat épidémiologique » en faisant référence aux nombreux facteurs qui peuvent être associés à une pathologie : « *On ignore encore tout de l'agent pathogène de bon nombre de maladies. Par contre, on cerne peu à peu leur « climat épidémiologique » par la découverte de multiples facteurs qui leur sont associés* », et un peu plus loin « *Leur analyse réclame des moyens considérables de traitement des données afin d'évaluer le poids relatif de chacune des composantes étiologiques, et de mesurer leurs interactions* » (Picheral, 1982).

Picheral, reprend en fait les travaux de Max Sorre avant lui (1926, 1933, 1951) sur les complexes (éco-)pathogènes. Ces complexes énoncés par Sorre dans son article intitulé « *Complexes pathogènes et géographie médicale* » consistent en une série d'éléments de type parasites, vecteurs, hôtes qui vont interagir entre eux pour produire des maladies. Au temps de Sorre, ce sont les maladies infectieuses qui retiennent alors tout l'intérêt des géographes tropicalistes, de là à présupposer que Max Sorre se serait de nos jours penché sur les pathologies dérivées de la pollution ? Il y n'a qu'un pas, que j'ose franchir.

Simon (2021), dans sa remarquable biographie consacrée à Max Sorre, écrit que pour Sorre « *La maladie est appréhendée comme le résultat d'interactions entre différents agents, des combinaisons vivantes qui sont d'abord énoncées de manière relativement simple et linéaire, avec deux ou trois termes ; le germe infectieux, diverses espèces d'insecte (servant de vecteurs), et l'homme. Toutefois, le géographe opère progressivement une complexification de ces combinaisons qui se multiplient et s'interpénètrent à l'intérieur du complexe* ». A travers ces deux auteurs, tout est déjà là, **les lieux, les liens, les interactions.**

Mais, ce concept a surtout été développé dans les travaux sur les maladies vectorielles, ou les épizooties, l'école soviétique reprenant alors le concept d'écologie du paysage sous l'appellation « *landscape epidemiology* » (Pavlosky, 1966) avec l'emploi massif des SIG et de la télédétection, ou bien encore « *pathogenic landscape* » pour les anglo-saxons (Lambin et al., 2010).

Plus récemment, Handshumacher et Hervouët (2004) dans leur article intitulé « *Des systèmes pathogènes à la santé publique : une nouvelle dimension pour la géographie de la santé tropicale* » décrivent très bien l'investigation menée en tant que géographes de la santé, dans une optique toutefois anthro-écologique afin de comprendre les cycles d'occupation/abandon des vallées de la Volta soumises à l'onchocercose, maladie transmise par un vers parasite (*Onchocerca volvulus*) qui peut aller jusqu'à provoquer la cécité en cas de contamination sévère.

Enfin, dernièrement, ce concept est mobilisé par Eric Daudé (Daudé, 2017) *via* ce qu'il appelle un système pathogène dit complexe, car il s'agit d'un système ouvert, sensible aux conditions initiales, composé de multiples composants en équilibre instable, évolutifs, qui ont leur dynamique propre, qui parfois peut être non linéaire, capables d'auto-organisation, avec des agents qui interagissent mutuellement avec leur environnement. Ce

système pathogène complexe a un objectif à atteindre, et il peut être appréhendé par une modélisation à base d'agents.

Mais, ce qui m'importe ce n'est pas de tant de connaître la dynamique d'un système, car je ne travaille pas sur des maladies infectieuses ou vectorielles pour lesquelles la modélisation à base d'agents est particulièrement bien adaptée, mais plutôt sur des maladies chroniques dont je recherche les causalités. Ce qui m'intéresse alors ce sont les interactions entre les variables, plutôt qu'entre les agents. Cependant, connaissant des causalités bien établies dans l'espace, il serait intéressant de les réinjecter dans une simulation à base d'agents afin d'élaborer une modélisation fondée sur une réalité éprouvée, ou bien de transformer les variables en agents, et voir comment elles évoluent simultanément au sein de différents espaces.

Nous reprenons le concept d'espace pathogène à notre compte pour caractériser des espaces suivant un certain nombre de critères que nous devons découvrir.

Les espaces étudiés dans mes recherches peuvent concerner des milieux très divers : urbain (pollution de l'air due au trafic de véhicules), périurbain (pollution de l'air due à la proximité d'industries de type ICPE), rural (pesticides), voire naturel (mais soumis à une forte pollution, comme c'est le cas par exemple de la Camargue).

Aborder la santé de manière holistique, globale en s'approchant le plus possible de l'espace de vie, nécessite de travailler à **fine échelle** (infra-communale : IRIS). En effet, il existe une grande variabilité des polluants d'une rue à l'autre, tout comme la socio-économie, d'un quartier à l'autre - un quartier plus aisé bordant parfois un autre qui l'est beaucoup moins. Les variables doivent être les plus désagrégées possibles afin de savoir précisément celles qui sont les plus influentes, celles sur lesquelles les pouvoirs publics doivent agir en priorité (préconisations), et de multiples natures.

La nature des liens entre les variables (déterminants de santé) est également à investiguer, de même que les seuils à partir desquels leur influence sur la santé est observée, car si nous connaissons bien l'effet d'un xénobiotique sur la santé (valeurs guides de l'OMS), tout reste à faire concernant l'effet cocktail auquel nous sommes réellement exposés. En effet, en matière d'interactions entre de multiples polluants nous devons progresser, sachant que de plus certains d'entre eux peuvent se potentialiser. Or, il existe encore trop peu d'études qui prennent en compte un ensemble de polluants, et pas seulement quelques-uns sur un espace donné. **L'idée est non pas de diagnostiquer les individus, domaine réservé à la médecine, mais de diagnostiquer l'espace dans ses multiples dimensions, au regard de la santé des individus qui y habitent.**

La pollution de l'air extérieur est un enjeu majeur de santé publique dans le monde (Lelieveld et al., 2015 ; Cohen et al., 2017 ; Khaniabadi et al., 2018 ; Sicard et al., 2019 ; Rovira et al., 2020) entraînant un demi-million de décès prématurés dans l'Union européenne en 2015 (EEA, 2018).

Les activités industrielles, les émissions des véhicules à moteur, les transports et le chauffage domestique sont autant de différentes sources de pollution de l'air dans les zones urbaines (EEA, 2019), libérant un large éventail de polluants atmosphériques tels que les oxydes d'azote (NOx), les composés organiques volatils (COV), les oxydes de soufre (SOx), de carbone noir, des hydrocarbures aromatiques polycycliques (HAP), et des particules (PM) (EEA, 2019). Des études antérieures ont rapporté une association entre l'exposition à la pollution de l'air ambiant, la mortalité et des morbidités *via* les hospitalisations pour maladies respiratoires et cardiovasculaires (Langrish et al., 2012 ; Suissa et al., 2013 ; Pascal et al., 2014 ; Atkinson et al., 2015 ; Cai et al., 2016 ; Cohen et al., 2017 ; Guo et al., 2018 ; Mannucci et al., 2019 ; Yang et al., 2020 ; Park et al., 2021). L'Union européenne et l'Organisation Mondiale de la Santé réglementent les concentrations de polluants atmosphériques en fixant des valeurs limites et cibles pour la protection de la santé humaine (directive 2008/50/CE, et les lignes directrices de l'OMS sur la qualité de l'air revues à la baisse en 2021).

Les enfants sont particulièrement vulnérables à la pollution de l'air⁶. En effet, il ne faut pas croire que ce sont de petits adultes « *they are not little adults, their organs and metabolism are under delicate formation and their capacity to face pollutant stress is different than those in adults* » (Grimalt et al., 2018). La trachée est plus courte chez l'enfant que chez l'adulte, le conditionnement de l'air, le rôle d'épuration et d'élimination des polluants associés moins efficaces, les bronches étant plus petites⁷, leur obstruction est plus fréquente. De plus, les enfants consomment plus d'oxygène, du fait d'une fréquence respiratoire plus élevée. Autant de raisons qui expliquent que les enfants soient plus sensibles que les adultes à la pollution de l'air. Cette sensibilité peut se traduire dans un premier temps chez l'enfant par une allergie. En effet, notre organisme est protégé de l'extérieur par deux interfaces : la peau et les muqueuses. Chez un sujet allergique, l'exposition, la mise en contact de l'organisme avec un pneumallergène potentiel va déclencher une défense de l'organisme disproportionnée face à ce qui est considéré comme une attaque, s'ensuit alors une hyperactivité nasale sous la forme de rhinite allergique, de rhume des foins, de rhino-conjonctivite, voire quand il s'agit de la peau de dermatite. Ces allergies risquant progressivement d'évoluer avec le temps, pour se transformer en asthme.

La prévalence de l'asthme en France est de 6 % chez l'adulte, contre plus de 10 % chez l'enfant. La prévalence chez l'enfant a doublé en 20 ans (passant de 5 %, à 10-12 % actuellement). De sorte que de nos jours, l'asthme est la maladie chronique infantile la plus fréquente. Il s'agit d'une maladie inflammatoire chronique des bronches, où les petites bronches sont atteintes en premier avec, au fur et à mesure de l'évolution de la maladie, une extension ascendante. L'obstruction des bronches s'accompagne de difficultés à respirer,

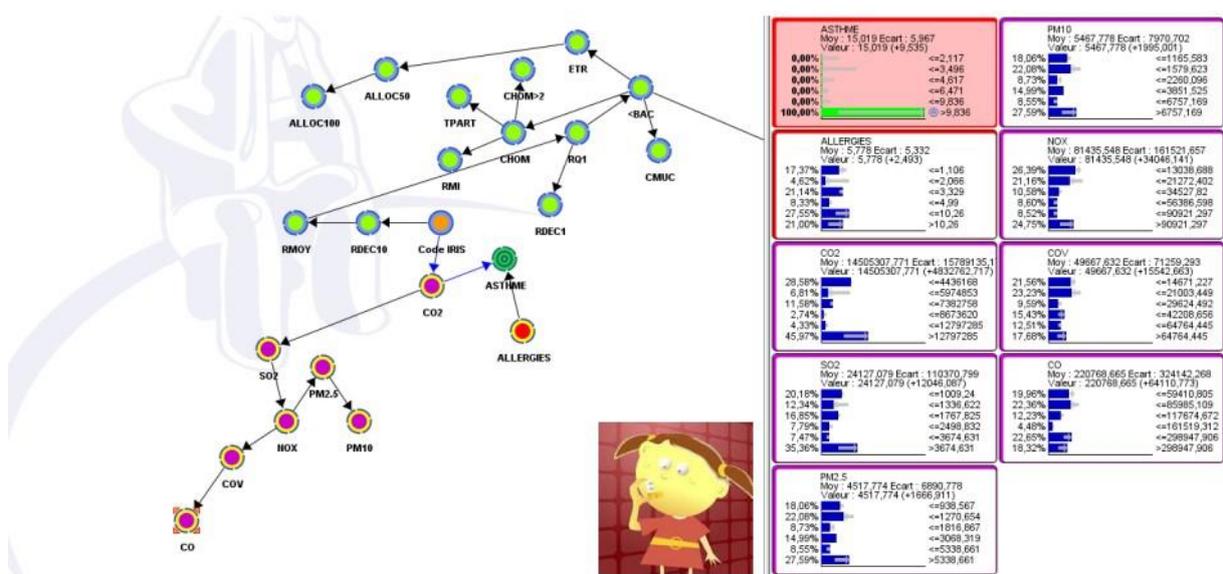
⁶ C'est déjà en soi une bonne raison de travailler sur les enfants, mais de plus, du fait de leur âge ils rendent « en général » bien compte de l'exposition en un seul lieu, et le poids des déterminants individuels est moindre que pour les adultes

⁷ 4 mm chez l'enfant, contre 8 mm pour l'adulte

d'une toux nocturne, de bronchites à répétition sans fièvre, d'un sentiment d'oppression dans la cage thoracique.

Les facteurs favorisant son expression sont complexes et interagissent entre eux, pour induire une susceptibilité à l'asthme. Cette complexité est due à l'interaction entre des éléments de natures différentes liés à l'hérédité, mais aussi aux conditions socio-économiques, environnementales, autant de facteurs qui de plus, peuvent se potentialiser. Ainsi, selon une étude menée par l'IRDES (Afrite et al., 2011) toutes choses étant égales par ailleurs, être obèse, vivre dans un ménage à faibles revenus ou de structure monoparentale augmente, non seulement le risque d'avoir un asthme (prévalence plus élevée), mais également d'avoir un asthme moins contrôlé (problème de l'accessibilité aux soins). « En effet, la prévalence de l'asthme est de 9,7 % parmi les personnes dont le revenu mensuel du ménage est inférieur à 550 € par unité de consommation (UC) contre 5,8 % lorsqu'il est supérieur ou égal à 1 300 €/UC ». De plus, les asthmatiques insuffisamment contrôlés sont plus nombreux au sein des ménages les plus modestes (15 %). Ainsi, le risque d'être insuffisamment contrôlé augmente clairement lorsque le niveau de ressources baisse : il est maximal parmi les bas revenus (< 550 €/UC : OR=2,15), et reste important lorsque les revenus sont compris entre 550 € et 840€/UC (OR=1,52 » (Afrite et al., 2011).

Une étude réalisée en 2010, ayant pour objet l'effet de la pollution de l'air sur les écoliers niçois, à l'échelle des IRIS (Pérez, 2010) a révélé par exemple que là où les enfants asthmatiques étaient nombreux, nous avons également plus de 48% d'enfants allergiques (lien connu entre les 2 variables), mais cela correspondait surtout aux taux les plus élevés des différents polluants (25% des NOX, 35% du SO2, 27% des PM2.5), et notamment pour plus de 45% du taux le plus élevé de CO2.



De plus, les variables de nature socio-économique prises en compte étaient mêlées à celles des polluants. Elles dessinaient alors une configuration particulièrement néfaste à une bonne santé des enfants (pathogène), car à l'exposition aux polluants s'ajoutent un accès moins facile aux soins (CMUC), plus de probabilités d'avoir des parents en situation économique dégradée (chômage depuis 2 ans, RSA etc..), qui fument (tabagisme passif), habitent dans des HLM ou de l'habitat collectif, et où la qualité de l'air intérieur peut aggraver l'asthme.

Travailler dans le cadre de ses recherches sur une population cible comme les enfants est un sujet sensible que ce soit sur une problématique de pollution de fond comme nous venons de l'évoquer, et bien entendu encore plus lorsqu'il s'agit de malformations congénitales, c'est-à-dire des pathologies qui ne seraient pas dues à des anomalies de la séquence ADN (acide désoxyribonucléique), mais liées à des modifications de son activité sous l'influence de substances chimiques extérieures suffisamment délétères pour agir au niveau du fœtus (éléments tératogènes).

Ces modifications peuvent parfois engendrer des malformations, on parle alors de malformation « congénitale », et de cas « *de novo* », car elles ne sont pas dues à l'hérédité mais à l'environnement, elles sont de nature épigénétique (Amar, 2017). L'environnement fait ici référence aux multiples substances chimiques étrangères à celles habituellement présentes dans le corps (xénobiotiques). Ces substances peuvent, par différentes voies, modifier chimiquement la structure de l'ADN et/ou des protéines associées, dont la méthylation ou non de la chromatine, qui permet de rendre lisible ou non un gène, de lui faire produire ou non une protéine, et de déclencher ou empêcher la cascade d'actions et de fonctions qui lui est propre. Dès lors, **l'environnement extérieur** impacte **l'environnement intérieur** d'un individu au niveau cellulaire.

Une méthylation influencée par un polluant étranger peut conduire à une plus grande sensibilité à certaines maladies, voire à des dérèglements métaboliques ou comportementaux (Latham et al., 2012). Plus précisément, la méthylation de l'ADN est opérée lors du développement du fœtus par des enzymes spécialisées, qui vont poser un groupement méthyle sur la double hélice. Leur action est compensée par d'autres enzymes dont le rôle est d'enlever ce groupement. Le mécanisme de ces deux enzymes sert à réguler la position (ouverte ou fermée à) la lecture des gènes et donc d'appliquer ou non leur fonction métabolique. Chaque nucléosome (segment d'enroulement du chromosome) est en effet espacé de manière très précise afin de favoriser ou réprimer l'accès aux gènes impliqués dans le développement, la prolifération, et la survie cellulaire. Ainsi, le développement fœtal est particulièrement sensible à toute modification de nature épigénétique, la plupart des malformations congénitales se produisant avant la 8ème semaine d'aménorrhée, soit lorsque la mère ne se sait pas encore enceinte.

Ces malformations congénitales dues à un problème au cours du développement fœtal, peuvent prendre la forme d'**agénésies**, qui se traduisent par l'absence partielle ou totale, d'un tissu interne (organe) ou externe de manière assez conséquente. Le terme d'aplasie est utilisé quand l'atteinte est moins grave. Dans le cas d'agénésies transverses des membres, une des

hypothèses est la création de brides amniotiques, sorte de filins qui vont former un nœud et bloquer ainsi la croissance d'un membre ou d'un tissu, voire l'amputer. Cette absence de membre peut se déceler à l'échographie, mais elle est encore trop souvent confirmée ou découverte à la naissance.

On estime à environ 1 cas sur 2 350 naissances (ou 4.3 cas sur 10 000) par an touché par une agénésie en France, bien que selon les sources, ce ratio varie. Ainsi, selon l'ASSEDEA⁸, l'association qui soutient les familles des enfants touchés par ces malformations, il y aurait plus de 345 enfants, qui naissent, en France, chaque année, avec une agénésie.

En France, **plusieurs clusters** d'agénésie transverse des membres supérieurs (ATMS) ont été détectés (Pedrot, 2021) : dans l'Ain (Amar, 2016, Gnansia, 2021), le Morbihan (Gagnère, 2018), et en Loire-Atlantique (Benachi, 2020). Récemment, la presse est revenue sur les cas des bébés nés sans bras, en 2016, dans les Bouches du Rhône (Chaud et al., 2021).

Dans le cadre d'un Mémoire de Master 2 (Squarta, 2021) nous avons géoréférencé l'ensemble des cas d'agénésies recensés dans la région Sud entre 2014 et 2019 (27 cas) avec l'aide du Centre pluridisciplinaire de diagnostic prénatal (CPDPN) de l'Hôpital Nord, à Marseille, par le biais du Dr Levy-Moziconnacci, de l'Unité Fonctionnelle de Génétique prénatale. Un regroupement de cas significatif est apparu grâce à une statistique de scan dans les départements des Alpes de Haute Provence et des Hautes Alpes.

Il convient d'**étudier très finement l'environnement de tous ces cas d'ATM, dans l'ensemble de la région Sud**. En effet, si l'attention est toujours portée sur les cas formant un agrégat spatial (cluster), il nous paraît primordial de considérer de la même manière (approfondie) tous les cas d'apparition, c'est à dire même ceux ne formant pas un cluster, **car les causes peuvent être multiples et non pas communes**. De plus, en raison du caractère très impactant de cette malformation sur l'enfant et sa famille la recherche des causes est le moins que la puissance publique puisse faire, et ce, dans une optique de prévention.

*En raison de la sensibilité des recherches menées en santé environnementale qui pointent des agents responsables ou le plus souvent co-responsables de pathologies qui vont impacter la qualité de vie, voire l'intégrité des individus, la **recherche des causes** s'avère indispensable.*

⁸ <http://www.assedea.fr/solution/lagenesie/>

1.4 De la nécessaire Causalité

Ce chapitre sur la causalité a pour but de répondre aux critiques qui sont souvent faites à l'encontre d'études en santé environnementale avec le sempiternel mantra « *Corrélation ne vaut pas causation, (causalité)* » (*cum hoc sed non propter hoc*). En effet, si nous restons sous le paradigme de la statistique classique dire que 2 évènements *A* et *B* sont liés par une causalité alors qu'ils sont simplement corrélés conduit à ce que les anglo-saxons appellent *a spurious correlation*.

Je ne peux pas déduire que *A* est la cause de *B* simplement parceque *A* et *B* sont corrélés, je dois en effet passer sous un autre paradigme statistique pour cela. Mais, déjà, si je prends un peu de distance, si je fais un pas sur le côté, je peux penser que *B* est peut-être la cause de *A*, voire qu'une variable latente peut être une cause commune de *A* et de *B*, ou bien alors qu'il s'agit d'une simple coïncidence.

Ce mantra serait tellement puissant qu'il serait susceptible à lui seul de balayer d'un revers de la main les efforts entrepris par les chercheurs pour rendre compte de l'influence de l'espace géographique à agir sur la santé des individus qui y habitent. Ce qui est bien dommage tant la Géographie de la Santé semble renouveler la discipline. Il est bien entendu difficile de remonter sans tomber dans une certaine forme de métaphysique aux vraies causes, aux causes véritables, premières, (Pearl, *The Art and Science of cause and effect In Causality*, 2009). Les causes d'un phénomène sont souvent multiples, diverses, plus ou moins directes. La géographie n'échappe pas à cette règle. Si elle se pose la question du *pourquoi ici et pas ailleurs* c'est bien que des causes ont prévalu à l'hétérogénéité de l'espace observé. Ce qui élève *de facto* la Géographie au rang de Science, penser le contraire par crainte d'un certain déterminisme qui contraindrait notre sacro-saint libre arbitre, qui ne saurait se limiter à un enchaînement de causes et de conséquences me semble extrêmement délétère pour notre discipline.

Cette notion de causalité en santé est fondamentale, car derrière, la responsabilité juridique peut être engagée, à travers notamment *l'imputabilité* qui peut donner droit à des indemnisations pour les victimes.

Par contre, l'imputabilité, ne souffre pas d'incertitudes : « *Que cette causalité soit directe ou indirecte, l'essentiel reste qu'elle soit certaine, parce que ce principe ne peut connaître de dérogation : pas de condamnation possible sans démonstration d'un lien de causalité certain entre le comportement dommageable et le préjudice constaté* » (Académie Nationale de Médecine, 2005). En droit, pour prouver le lien de causalité, il suffit d'une probabilité suffisante de causalité, d'un faisceau d'indices graves, précis et concordants. Deux approches peuvent être distinguées, la théorie de l'équivalence des conditions et la théorie de la causalité adéquate (Le Tourneau, 2003). Dans la théorie de l'équivalence des conditions il s'agit de considérer que chaque élément qui a concouru à la réalisation d'un dommage en est une cause. Tandis que pour la théorie de la causalité adéquate il s'agit de ne retenir qu'une seule cause génératrice du dommage, celle qui a mené directement à la réalisation de ce dernier (effet direct).

De plus, sans causalité établie la puissance publique en charge de ces questions ne peut savoir sur quel aspect intervenir : sur quel polluant, sur quel pan de la socio-économie, sur les 2 simultanément ? Ce qui est quand même problématique, et peut trop souvent conduire à des décisions à l'emporte-pièces, car non éclairées.

En épidémiologie, les recherches observationnelles sont jugées moins robustes que les recherches expérimentales pour établir un lien causal entre un état de santé et un facteur de risque, et la notion de causalité serait plus de « *nature probabiliste que déterministe du fait de la variabilité interindividuelle* » « *chaque être vivant [étant] unique, il réagit différemment dans la même situation* » (Flahault, 2011).

Mais, **l'espace ne peut pas être expérimenté dans le sens scientifique du terme**, il ne peut pas bénéficier d'essais contrôlés randomisés (ECR), il est difficile d'y introduire des éléments dont on voudrait mesurer directement les effets. Or, la causalité scientifique repose sur l'expérimentation, où les conditions initiales sont contrôlées, randomisées, et au sein de laquelle nous pouvons mesurer l'effet d'un nouvel élément, qui viendrait modifier ces conditions initiales. C'est la grande force nous le verrons du formalisme bayésien, qui peut rechercher des causes dans une perspective étiologique à partir de données non expérimentales, voire parfois en l'absence totale de données sur la base de l'élicitation ou de dire d'experts, et sous un paradigme probabiliste particulièrement bien adapté aux questions sanitaires.

Par ailleurs, la science recherche des réponses univoques du type « *telle cause produit tel effet* », et si possible linéaires afin d'en assurer la prédiction. Or, en santé environnementale, les réponses sont la plupart du temps équivoques, car le résultat peut venir de multiples causes (multifactoriel). Dès lors, la complexité, la méconnaissance rend (pour l'instant) difficile, voire impossible toute projection sérieuse dans le futur. Par contre, les investigations présentes peuvent être vérifiées sur le passé dans le cadre d'études rétrospectives, où nous observons des pathologies actuelles, fruits d'une exposition plus ancienne (abduction).

Ainsi, à partir de données non expérimentales, observationnelles, la plupart du temps sur des populations, nous allons chercher à remonter aux causes, par exemple d'exposition, qui ont conduit à une pathologie d'intérêt donnée. Le cadre est donc celui d'un raisonnement hypothético-déductif sur lequel se base le raisonnement scientifique depuis Descartes (1637) : soit des pathologies avérées, vraies (prémises) supposées être le résultat d'expositions (assertions) qui en seraient les causes. La recherche des causes, l'étude de la causalité est cruciale en santé environnementale.

Hill, en tant qu'épidémiologiste recense en 1965, (Hill, 1965) 9 caractéristiques qui doivent se retrouver pour établir un lien de causalité entre 2 variables X et Y , ces 9 caractéristiques sont connues sous le nom de *critères de Hill pour la causalité* :

1. La force de l'association (qui peut être évaluée selon plusieurs méthodes statistiques mais qui doit dans tous les cas être retrouvée) ;
2. La consistance (*consistency*) de l'association reproductible dans le temps, dans d'autres lieux, toutes choses étant égales par ailleurs ;
3. La spécificité de l'association : une cause suspectée ne doit conduire qu'à la seule conséquence suspectée ;
4. La temporalité : la cause suspectée doit impérativement précéder l'apparition de la maladie étudiée (antécédence de la cause) ; de la même manière, si la cause suspectée disparaît, la morbidité qui lui est imputée doit décroître proportionnellement ;
5. La dose réponse (*biological gradient*) l'effet doit être proportionnel à la dose ;
6. La plausibilité : plus l'association entre l'agent causal et la maladie est compatible avec les données de la science (plausibilité biologique), plus fort sera ce critère, et inversement ;
7. La cohérence avec les connaissances (une modification de X entraîne *de facto* une modification de Y) ;
8. L'expérimentation (à envisager dès que possible) ;
9. L'analogie.

Bien entendu, ces critères ne doivent pas tous être remplis pour définir une causalité entre 2 variables. Inversement, aucun critère à lui seul (excepté celui de l'antécédence de la cause) ne peut établir une causalité (Rothman, Greenland, 1998). Ainsi, et à minima « *We may define a cause to be an object followed by another, where if the first object had not been, the second never had existed*'' (Hume, 1748).

Si X a un effet sur Y peut s'écrire selon le théoricien de la causalité Judea Pearl (Pearl, 2000a) ainsi :

Soit 2 objets,

$$X(U) = x$$

$$Y(U) = y$$

Sous l'action de X , Y est différent de y (de son état antérieur) :

$$Y(x')(U) \neq y \text{ for } x' \neq x$$

Pour qu'une action de X sur Y se produise il est nécessaire que les 2 objets soient en présence, on pourrait parler aussi de *concomitance*, Pearl le définit comme le concept de *nécessité* (2000a). Or, en santé, un différé temporel peut exister entre, une cause et son effet. De même, la « présence » entre 2 évènements comme par exemple celui de l'émission d'une substance polluante, et sa réception exprimée plus généralement par le terme exposition n'est pas dépendante de la distance, l'effet d'un polluant n'étant pas forcément le plus fort à proximité immédiate de l'émetteur (effet protecteur en cloche). Souvent, il se fait plus sentir à une certaine distance, dans une certaine direction par rapport à ce dernier, en fonction essentiellement du relief, et surtout de la direction dominante et de la vitesse du vent.

Pearl nuance ce concept de nécessité en introduisant le concept de *suffisance* (Tian, Pearl, 2000) qui implique l'idée de seuil, soit à partir de quel seuil X produit un effet sur Y , ce qui correspond au concept opératoire de « dose-réponse » bien connu en écotoxicologie et en santé environnementale. En effet, nous savons depuis Paracelse que c'est la dose qui fait le poison « *Tout est poison et rien n'est sans poison, la dose seule fait que quelque chose n'est pas un poison* » (1526), et nous disposons de tables, de guides (Niosh, Ineris) qui nous fournissent la dose-réponse suivant les modes de contamination.

Hume (1748), reconnaît quant à lui 3 critères de base pour établir une causalité : (a) la *contiguité spatio/temporelle*, (b) la *succession dans le temps* et (c) la *constante conjonction* (association) des 2 phénomènes. Le premier critère (la contiguité), nous l'avons vu, est parfois difficile à respecter en santé environnementale, et ce, aussi bien dans le temps (différé), que dans l'espace (distance). Le second (succession temporelle) l'est beaucoup plus, sans pour autant que nous puissions le définir précisément, mais nous partons du principe que l'effet, la pathologie est toujours postérieure à l'exposition. Enfin, l'association est rendue compte par les tables de « dose-réponse » que nous venons de citer, qui sont bien documentées, car issues de multiples expérimentations (mais a-spatiales), même si les effets cocktails le sont beaucoup moins, nous y reviendrons plus loin. « *In the analysis of the idea that A causes B this means that (a) A and B are contiguous in space and time, (b) A precedes B in time, and (c) A and B always occur (or do not occur) together* » (Holland, 1986).

Pour Holland (1986) : « *Correlation does not imply causation, and yet causal conclusions drawn from a carefully designed experiment are often valid* ». Il formalise ainsi l'effet causal : lorsque le "phénomène étudié" se produit $Y = 1$; lorsqu'il ne se produit pas $Y = 0$; "la circonstance dans laquelle les instances diffèrent" lorsqu'elles sont présentes = t , lorsqu'elles sont absentes = c . Alors, $Y_t(u) = 1$ dénote le fait que lorsque la circonstance est présente le phénomène se produit, et $Y_c(u) = 0$ dénote le fait que lorsque la circonstance est absente le phénomène ne se produit pas. L'égalité de toutes les autres circonstances est modélisée en considérant la même unité. Ainsi $Y_t(u) - Y_c(u) = 1$, donc l'effet causal de la circonstance sur l'unité est 1, et correspond à l'affirmation de Hill (1965) selon laquelle la circonstance est "la cause ou une partie indispensable de la cause du phénomène". Autrement-dit, **lorsque la ou les circonstances sont réunies les conséquences le sont tout autant.**

1.4.1 Les modèles contrefactuels

Dans les domaines des sciences sociales et de la biostatistique, les modèles contrefactuels sont largement acceptés pour étudier les effets causaux (on les retrouve dans la littérature sous les terminologies de cadre des résultats potentiels (*framework potential outcomes*, modèle de Rubin, ou modèle Neyman-Rubin).

Le modèle contrefactuel est relatif à ce qui aurait pu se produire si on avait modifié les causes, il permet de vérifier que les causes sont bien les bonnes, par exemple en cas de moindre exposition le nombre de cas devrait diminuer, cela peut être rapproché de l'échantillon témoin dans les études de types « cas-témoin » bien connues en épidémiologie, et par lesquelles la causalité est recherchée.

Dans ce cas, pour un individu i , deux résultats sont alors possibles pour une exposition binaire de type oui/non (exposition/non exposition) :

$Y_{x=0}(i)$ représente le résultat que l'on observerait si l'individu n'est pas exposé, autrement dit si $X = 0$

A contrario,

$Y_{x=1}(i)$ représente le résultat que l'on observerait si l'individu est exposé, autrement dit si $X = 1$

Si nous extrapolons à présent la procédure à plusieurs individus, l'effet causal individuel (*individual-level causal effect*, ICE), serait défini par la différence entre un individu i exposé et un autre non exposé j , soit :

$$ICE = Y_{i,1} - Y_{j,0}$$

De la même façon, nous pourrions prendre en compte la variabilité interindividuelle si nous soumettions plusieurs individus à la même exposition.

Il serait possible de déduire également un effet causal moyen $E = E(X1) - E(X0)$ (Shinozaki, Euzaki, 2020).

Soit l'exemple suivant proposé par ces auteurs :

A_i = un médicament contre l'hypertension,

$A_i = 1$, le traitement est pris par un groupe « cas » dans une étude de *design* « cas-témoins »

Y = le résultat potentiel, soit l'occurrence de maladies cardiovasculaires

L_i = un ensemble de covariables (l'état de santé antérieur des patients, les conduites à risque, le support social dont ils bénéficient) observées à l'échelle des individus $i = 1, \dots, n$

Y_{i0} = le résultat potentiel inobservé si contrefactuellement $A_i = 0$ (le traitement n'est pas administré dans le groupe « témoin »)

Alors, l'effet causal moyen E d'une administration du médicament A (ou d'une exposition) sur le résultat Y peut être défini comme :

$E[Y1] - E[Y0]$ = qui compare les résultats contrefactuels d'une exposition binaire de $Y = 1$ et $Y = 0$

Le calcul de l'effet causal moyen peut également s'avérer très utile pour comparer une stratification de résultats au sein d'une population, ou bien à différentes échelles spatiales et/ou temporelles.

Enfin, *“The potential outcomes framework does not delimit the bounds of causality; it is concerned with a subset of causal questions that can be defined as a contrast of hypothetical interventions. By focusing on this subset, effect estimates have a relatively well-defined causal interpretation, which is especially important when they are used to guide decisions in the real world”* (VanderWeele et al., 2016). *“Like microarrays or logistic regression, potential outcomes are a tool for scientific reasoning—not the only tool, but an extremely important one for causal thinking. Potential outcomes do not provide the only form of evidence and they do not preclude the use of other forms of evidence”* (VanderWeele et al., 2016).

1.4.2. La théorie de la causalité & les réseaux bayésiens

La formalisation à proprement parler de la théorie de la causalité et des réseaux bayésiens, outil permettant de mieux concevoir cette causalité est issue des travaux de Judea Pearl (chercheur à l'UCLA) et de Finn Jensen (Chercheur à l'université d'Aalborg au Danemark) dans les années 80. La naissance des réseaux bayésiens vient du fait que ces chercheurs souhaitaient à la fois intégrer un peu plus d'incertitude dans les systèmes experts jugés comme étant trop déterministes, et parallèlement approfondir, formaliser, cette notion de causalité.

Leurs travaux vont s'inspirer de ceux du mathématicien britannique Thomas Bayes (1702-1761) qui travailla sur la théorie des probabilités. Le théorème de Bayes cherche à déterminer la distribution *a posteriori* de la probabilité p d'une loi binomiale. Une loi binomiale repose sur ce que l'on appelle une épreuve de Bernouilli, soit une expérience aléatoire qui ne peut prendre que 2 issues (succès de probabilité p ou échec de probabilité $q = 1 - p$). C'est à cette expérience que s'adonna Bayes en essayant de connaître la position d'une balle lancée dans son dos et sur une table par un assistant, auquel il demandait à chaque lancer où la balle avait atterri par rapport à la précédente (à droite, à gauche, devant, derrière). A chaque itération (nouveau lancer de balle) sa connaissance de la position de la première balle s'affinait. Ainsi, le nombre de succès obtenus à l'issue des n épreuves correspond à la variable aléatoire X qui suit une loi binomiale, soit une loi de probabilité discrète ($p = 1 =$ succès, $q = 0 =$ échec), de formule :

$$P(k) = P(X = k) = \binom{n}{k} p^k q^{n-k}$$

Où :

$P(k)$: Probabilité de k succès dans une répétition de n expériences

$\binom{n}{k}$: Coefficient binomial qui correspond au nombre de combinaisons, et se calcule à partir d'une fonction factorielle de type $\binom{n}{k} = \frac{Ckn}{k!} = \frac{n!}{k! (n - k)!}$

p^k : Probabilité de réussite

q^{n-k} : Probabilité d'échec

Bayes a présenté cela comme une probabilité sur le paramètre P . Nous pouvons ainsi déterminer **la probabilité d'une hypothèse** (position de la 1^{ère} balle) **connaissant certaines observations** issues d'une expérience comme nous venons de le voir, ainsi que des paramètres relatifs à ces probabilités.

Pearl (1995a, 2000a) établit ainsi une Théorie générale de la causalité qu'il nomme *Structural Causal Model* (SCM). Il souligne que la causalité n'est pas dans la relation en elle-même (*association statistic*), mais plutôt dans **le sens de la relation**. Ainsi, Pearl se demande comment on peut mathématiquement formuler le fait que ce ne sont pas les symptômes qui sont la cause de la pathologie, mais que c'est bien la pathologie qui est la cause des symptômes. C'est le travail du généticien Sewal Wright (1921) qui va lui apporter la réponse. En effet, Wright a été le premier à utiliser une combinaison d'équations et de graphes pour communiquer des relations causales.

Par exemple, si x correspond à une pathologie, et Y à un symptôme de x , Wright écrit une équation linéaire de type :

$$y = \beta x + uY$$

Où x correspond au degré, ou bien à la sévérité de la maladie, y au degré, ou bien à la sévérité du symptôme, et uY à tous les autres facteurs (co-facteurs) qui peuvent possiblement jouer sur le symptôme quand x demeure inchangée (constante). Dans cette équation β est un paramètre qui est appelé « *path coefficient* ». Ce paramètre permet de quantifier l'effet direct et causal de x sur y , c'est-à-dire que connaissant les valeurs de β et de uY , une augmentation de x engendrera une augmentation de y via β , indépendamment du fait des valeurs d'autres variables dans le système, et indépendamment de l'origine des variations de x (influences externes, ou bien internes).

Pearl, propose le graphe ci-après (Figure n°2) pour exprimer les relations causales que nous venons de voir. A noter que l'absence de flèche de Y vers X signifie que le symptôme Y ne figure pas parmi les facteurs UX qui influencent la maladie X . Ainsi, dans cet exemple le modèle complet entre un symptôme et une maladie s'écrit grâce au schéma qui retranscrit la possible existence d'une influence causale directe de X sur Y , et l'absence d'influence causale de Y sur X , tandis que les équations écrivent les relations quantitatives entre les variables, déterminées à partir de données.

Les variables U_x et U_y sont appelées exogènes, c'est à dire que leurs influences peuvent s'étendre ou pas à d'autres variables, mais elles ne sont pas influencées par les variables endogènes du système. Elles sont en quelque sorte au bord du système.

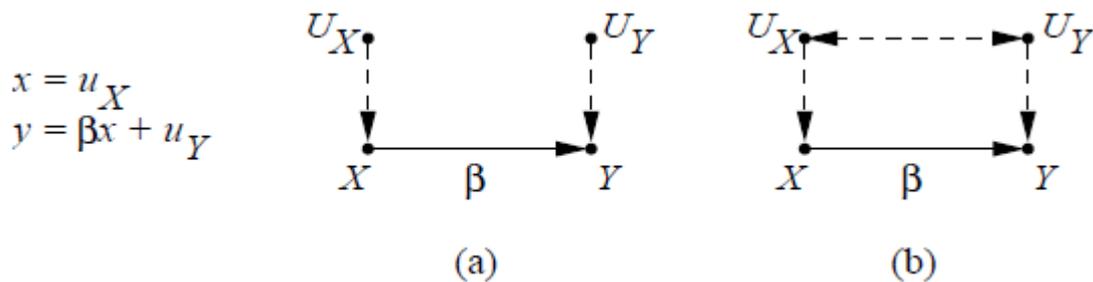


Figure 2. Un modèle d'équation structurelle simple, et ses diagrammes associés. Les variables exogènes non observées sont reliées par des flèches en pointillés.

Source : *Causal inferences in statistics* (Pearl, 2000a)

Le modèle causal structurel développé par Pearl (*Structural Causal Model*, SCM) (Pearl, 1995a, 2000a) combine les caractéristiques des modèles d'équations structurelles (Goldberger, 1973 ; Duncan, 1975), le travail de Neyman et de Rubin sur les résultats potentiels (Neyman, 1923 ; Rubin, 1974), et les modèles graphiques développés pour le raisonnement probabiliste et les analyses causales (Pearl, 1988 ; Lauritzen, 1996 ; Spirtes et al., 2000 ; Pearl, 2000a).

1.4.3. Éléments de compréhension de l'expression de la causalité dans un DAG

Un graphe causal peut être défini comme suit :

Soit, un ensemble de variables endogènes (en lien présumé avec une problématique donnée, faisant en quelque sorte partie du « système » que nous souhaitons étudier), réunies au sein d'un graphe direct acyclique (Direct Acyclic Graph, DAG). Les variables sont représentées par des nœuds, les relations entre ces variables exprimant la causalité par des flèches, et l'absence de flèche, nous venons de le voir, suppose l'absence de causalité. Deux variables sont dites adjacentes si elles sont directement connectées par un arc. Tandis que le positionnement (ascendant ou descendant) d'un nœud par rapport à un autre va prendre les adjectifs « parent » ou bien « enfant ».

Soit, 3 nœuds, N1, N2, N3 représentant 3 variables. Le nœud N1 est le nœud ascendant ou bien parent des nœuds N2 et N3, et les nœuds N2 et N3 sont dits les nœuds enfants de N1, nous pouvons aussi dire que N2 et N3 ont comme parent commun N1 :

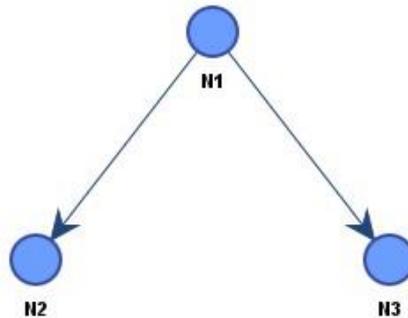


Figure 3. La généalogie des nœuds

Par conséquent, N1 est la cause commune à N2 et N3. N1, est la seule variable indépendante du réseau puisqu'elle n'est la conséquence a priori d'aucune autre, on dit qu'elle a une **probabilité marginale**. Si N1 est vrai, alors N2 et N3 le sont également, autrement-dit, si la cause est vraie, l'effet l'est lui aussi. Ainsi, si nous modifions N1, les distributions de N2 et N3 changent en fonction.

Dans la structure ci-dessous (inverse de la précédente, dite structure en V), les nœuds 1 et 2, ont comme enfant commun N3 qui est appelé aussi dans ce cas le nœud *collider*.

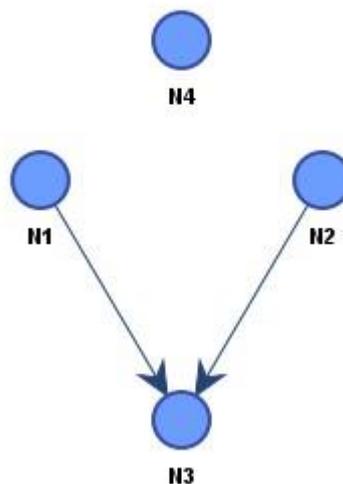


Figure 4. La structure en V formée par les nœuds

En termes de causalité, N3 est sous l'influence commune des nœuds N1 et N2 qui sont des variables indépendantes marginalement dans U , soit $N1 \perp N2 | U$ (l'information ne circule pas entre elles), à moins qu'elles ne soient influencées par une cause exogène non mesurée qui nécessiterait un ajustement sur un nœud N4 (*voir supra*). Cette cause exogène peut correspondre à ce que certains appellent un élément intermédiaire, un médiateur, qui a donné lieu aux modèles de type médiation.

1.4.4. Les modèles de type Médiation

Situé entre une variable telle que l'exposition et l'effet de celle-ci, le médiateur peut avoir un effet direct, indirect, entrer en interaction ou pas, avec la variable d'intérêt (effet de l'exposition, pathologie) (Vanderweele, 2014, Shi, 2021). Ainsi, la causalité entre 2 variables peut être modulée par une troisième (médiateur) qui est le long du chemin causal.



Figure 5. L'élément médiateur dans le réseau

L'effet indirect peut être le fait par exemple de la socio-économie, de déterminants de santé individuels qui vont se surajouter à l'exposition environnementale, ou bien au contraire qui vont la minorer du fait d'un meilleur accès aux soins, d'une meilleure qualité de vie. Ces médiateurs peuvent correspondre par conséquent à des éléments de confusion (*confounder factor*). Par leur prise en compte, ils mettent selon les auteurs mieux en évidence le chemin causal, et peuvent favoriser une meilleure compréhension de ce dernier par les décideurs. Ils s'avèrent très utiles dès lors que ces facteurs de confusion ne peuvent pas être directement mesurés, et se distinguent en ce sens de la statistique classique (associations). En effet, la causalité peut se cacher dans une variable explicative, latente, qui influencerait simultanément 2 autres variables. Autrement-dit, l'association entre 2 variables pourrait être due à une cause commune, et dès lors « *l'association ne présuppose pas la causation (causalité)* »...

Lange (2016) recense dans son *overview* une vingtaine d'articles qui portent sur l'analyse de médiation publiés en 2016, dont voici un extrait, nous pouvons voir quelles sont les variables médiatrices auxquelles les auteurs font appel pour mieux comprendre les effets entre une exposition et une pathologie (Cf. Tableau ci-dessous). Selon lui, *“the natural indirect effect is the change we would observe on a given outcome if we could change each person’s mediator to the value it would naturally take when the exposure was intervened on, but without actually changing the exposure from reference level”*.

Tableau 1. La *meta review* de Lange sur les modèles de type médiation

TABLE. Overview of Papers Presenting Novel Statistical Mediation Analyses in 2015

Paper	Comments
Jonker et al. ¹² (Epi)	The effect of income level on geographical variations in cardiovascular disease and the mediating role of health-related behavioral risk factors. Data is partly ecological. The mediation analysis is based on difference of coefficients methods, but the implementation is slightly unusual because of ecological data
Jackson et al. ¹³ (Epi)	The mortality effect of first generation versus second-generation antipsychotics and the mediated role of various individual events (stroke, ventricular arrhythmia, AMI, etc). Mediation analysis implemented using SAS or SPSS (not specified) macros ²
Rao et al. ¹⁴ (Epi)	Behavioral factors (smoking and alcohol consumption) as mediators between childhood socioeconomic conditions and oral cancer. Implemented using SAS macros ²
Lu et al. ¹⁵ (Epi)	The effect of BMI on coronary heart disease risk and the mediating role of metabolic risk factors, inflammatory, and prothrombotic biomarkers. Mediation analysis conducted using difference of coefficients approach with a full causal interpretation due to a rare outcome
Zhang et al. ¹⁶ (Epi)	Studies the effect of neighborhood deprivation on BMI mediated by the healthful and unhealthful food vendor density. The mediation analysis is conducted using G-computation implemented in Stata
Mendola et al. ¹⁷ (Epi)	Preterm birth as a mediator for the relation between preeclampsia and neonatal health. Mediation analysis implemented in SAS based on marginal structural models
Sieurin et al. ¹⁸ (EJE)	Presents an analysis of smoking as a mediator between personality traits and Parkinson disease. Implemented using natural effect models as suggested by Lange et al. ¹
Küpers et al. ¹⁹ (IJE)	Studying DNA methylation as a mediator between maternal smoking during pregnancy and the birth weight. Mediation analysis is conducted using the product of coefficients approach and implemented in Stata
Amato et al. ²⁰ (NEJM)	Examines driving pressure as a mediator between lung-protective mechanical ventilation and 60-day survival. Estimation done using the mediation package for R by Imai ²¹
Chalder et al. ²² (Lancet Psychiatry)	Examines fear avoidance (and cognitive behavioral responses) as a mediator between treatment assignment of chronic fatigue syndrome and physical functioning (and fatigue). Estimation uses the product of coefficients method based in Stata
Freeman et al. ²³ (Lancet Psychiatry)	Examines worry as a mediator between cognitive behavior therapy and delusions among patients with psychotic disorders. Estimation is done using structural equation methods in Mplus
Victoria et al. ²⁴ (Lancet Global Health)	Examines IQ at 30 years of age as a mediator between being breastfed and income at 30 years of age. Mediation analysis conducted in Stata, but details are not provided. Interpretation is focused on the mediated proportion
Singla et al. ²⁵ (Lancet Global Health)	Examines parenting practices and mother-care variables (spousal support) as mediators between parenting interventions and cognitive and receptive language development of a child. Estimation uses product of coefficients in SAS and bootstrapping for confidence intervals
Huang et al. ²⁶ (AJE)	Examines mothers' behavior (childhood maltreatment) and health status as mediators between grandmothers' education and child's birth weight. Estimation uses SAS for mediation analysis
Côté et al. ²⁷ (AJE)	Examines how feeling unsafe at school, youth screen time, and physical activity mediate the association between poverty exposure and weight status. Estimation uses structural equation modeling
Ekamper et al. ²⁸ (AJE)	Examines mediating role of health status at 18 years of age as a mediator between pre-/perinatal famine exposure and mortality between age of 18 and 63. Estimation uses the difference of coefficients method based on Cox models
Ivory et al. ²⁹ (AJE)	Examines mediating role of household smoking between neighborhood smoking and individual smoking as well as between neighborhood deprivation and individual smoking. Estimation uses the difference of coefficients

Le temps peut également être un médiateur intéressant, on parle alors de modèle de médiation longitudinale, dont voici un exemple ci-dessous :

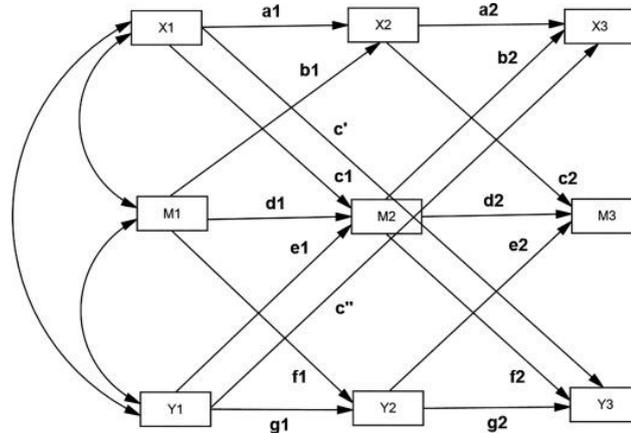


Figure 6. Un modèle de médiation longitudinale à 3 vagues (Asendorpf, 1996)

Ainsi, « *The effect of X at time 1 affects the mediator M at time 2 which, in turn, affects the outcome Y at time 3, and vice versa* » (Asendorpf, 1996). Les réseaux bayésiens dynamiques permettent de modéliser cela, de même que les modèles structuraux marginaux (Marginal structural model, MSM), qui peuvent s'avérer très utiles pour mesurer au sein d'une analyse longitudinale les effets cumulés par exemple de la pollution, ou bien des effets différés. Ainsi, « *Lorsqu'une exposition antérieure affecte les facteurs de confusion des expositions ultérieures, l'estimation des effets des expositions variant dans le temps nécessite des techniques statistiques spéciales, éventuellement avec des modèles structurels (c'est-à-dire contrefactuels) pour les effets ciblés* » (Shinozaki, Suzuki, 2020). Ces modèles font appel à des pondérations de probabilité inverse (Talbot et al., 2015), ils sont relativement récents et font partie du cadre contrefactuel de l'inférence causale.

A noter également que les scores de propension (c'est-à-dire la probabilité d'exposition conditionnelle aux facteurs de confusion mesurés), et le *back door criterion* qui permet de « fermer » l'influence d'un médiateur dans le chemin causal afin de ne conserver que l'effet le plus direct entre 2 variables, sont autant de procédés qui peuvent s'avérer bien utiles dans nos recherches.

Le réseau de la figure ci-dessous représente une connexion en série indirecte de N1 sur N3 *via* N2. Autrement-dit, N1 a un potentiel effet causal sur N3 par l'intermédiaire de N2. Les nœuds N1 et N3 sont dépendants marginalement, ce qui signifie que même si nous ne connaissons pas l'exacte valeur de N2, N1 nous informe sur N3, et inversement. L'information n'est pas bloquée, elle circule librement dans les deux directions (de N1 à N3 et de N3 à N1) (Conrady, Jouffe, 2015). Dans une connexion en série, l'information ne circule de N1 à N3 que si N2 est inconnu.

Si N1 est vrai cela entraîne que N2 l'est également, mais si N1 est faux, N2 peut être vrai ou faux. Il s'agit ainsi de relations causales, orientées, mais qui peuvent être réversibles de l'effet vers la cause. Toute information que je connais sur N1 peut être modifiée par ma connaissance de N2, et inversement de N2 vers N1.

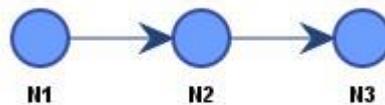


Figure 7. Connexion en série de type indirect

Dans une connexion cette fois convergente, N1 et N3 causent N2. L'information ne circule de N1 à N3 que si N2 est connu.

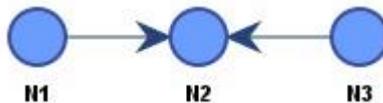


Figure 8. Connexion convergente

Et dans son contraire, la connexion divergente, N2 est la cause à la fois de N1 et de N3. L'information peut circuler entre N1 et N3, seulement si la cause commune N2 est inconnue.

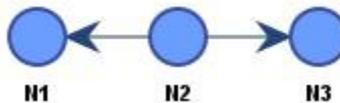


Figure 9. Connexion divergente

Nous voyons bien que **c'est le graphe qui donne la structure causale entre les variables**. A noter que le sens de la relation n'est pas exprimé, dans la figure n°7, N1 est une cause de N3 sans que nous sachions pour l'instant à ce niveau si cette cause est positive ou négative. La longueur des arcs est informative, plus les arcs sont courts et plus la relation entre les nœuds (variables) est forte. La causalité est donc exprimée par le graphe ou réseau, tandis que les valeurs de chaque nœud sont calculées généralement à partir des données sur la base de probabilités et selon le théorème de Bayes :

$$P(H/E) = (P(E/H) * P(H)) / P(E)$$

Où :

$P(H/E)$ = Probabilité que l'hypothèse H soit vraie suivant un évènement donné E

$P(E/H)$ = Probabilité que l'évènement ait lieu sachant que l'hypothèse est vraie

* $P(H)$ = Probabilité antérieure (*prior probability*) avant l'évènement

$P(E)$ = Probabilité totale de l'évènement

Le théorème de Bayes donne donc la probabilité qu'une hypothèse soit vraie sachant qu'un évènement s'est produit. Les valeurs des variables peuvent être continues, le plus souvent discrètes, binaires (présence, absence), ou bien instanciées (occurrences d'appartenance à une classe). En résumé, le graphe est une représentation qualitative de la distribution de probabilités jointes (Conrady, Jouffe, 2015). Nous pouvons par conséquent dire que les réseaux bayésiens sont des **modèles probabilistes de causalité**.

Les réseaux bayésiens peuvent s'apparenter à un graphe dont les nœuds définissent les variables du système, et dont les arcs définissent l'existence de relations entre ces variables. Ils décrivent de manière qualitative (*via* le graphe) des relations de dépendances ou d'indépendances conditionnelles entre les variables et quantifient ces mêmes relations par le biais des probabilités conditionnelles.

Les relations causales ne sont pas absolues, mais associées à une probabilité indiquant le degré de croyance que l'on a dans l'événement. Autrement-dit, ces réseaux permettent de calculer des probabilités conditionnelles d'événements reliés les uns aux autres par des relations de cause à effet. Cette utilisation s'appelle **l'inférence**. L'inférence est le raisonnement qui part de présupposés (appelés *prémises*) pour en tirer en toute logique des conclusions (*assertions*). La mise en place du réseau se fait en associant au formalisme graphique (structure du réseau), un contenu probabiliste (paramètre du réseau). **La correspondance existante entre la structure graphique et la structure probabiliste associée permet de ramener l'ensemble des problèmes de l'inférence à des problèmes de théorie des graphes plus simples à comprendre.**

Nous voyons là l'humanisme, la générosité de Judea Pearl qui rend accessible des éléments compliqués : *“Trained to accept the human mind as a reliable transducer of experience, and human experience as a faithful mirror of reality, Bayesian statisticians are beginning to accept the language chosen by the mind to communicate experience – the language of cause and effect.”* (Pearl, 2000a). Nous pouvons dire la même chose de Lionel Jouffe qui est à l'origine du logiciel dédié aux réseaux bayésiens : BayésiaLab (<https://www.bayesia.com/>)⁹, et dont les qualités pédagogiques rendent abordable au plus grand nombre la statistique bayésienne, qui peut être ardue. D'ailleurs, lors d'une visite à Judea Pearl pour lui présenter le logiciel, ce dernier l'a complimenté d'un : « *Good stuff!* ».



Photographie. Judea Pearl, le père des réseaux bayésiens à gauche, et Lionel Jouffe cofondateur de BayésiaLab à droite, le 12/11/2010, UCLA

Les réseaux bayésiens permettent d'établir des relations de dépendances ou d'indépendances entre différentes variables aussi bien quantitatives que qualitatives (recodées), d'analyser leurs interrelations, leurs combinaisons, mais surtout de quantifier ces relations à l'aide de paramètres ou de probabilités. Le contexte est celui d'un formalisme mathématique de représentation des connaissances incertaines (probabilistes) particulièrement bien adaptées à la problématique du risque sanitaire, **car les relations entre les variables en matière de santé ne sont pas toujours de type déterministe, mais plutôt de type indirect (probabiliste).**

⁹ J'ai suivi 2 formations aux réseaux bayésiens : en mars et juin 2014, à Paris, délivrées par BAYESIA S.A.S

De plus, l'outil de type « réseau bayésien » a l'avantage de pouvoir rassembler et de fusionner des connaissances à partir de données de diverses natures, de types expert, de les faire cohabiter dans un même modèle, mais également d'introduire le retour d'expériences (données historiques ou empiriques), voire de créer de la connaissance manquante en utilisant le *data mining*. En effet, un réseau bayésien peut être construit à partir de données, soit par apprentissage (apprentissage incrémental, inférence), soit à partir d'une modélisation explicite du domaine de type expert, soit il permet de coupler les connaissances des experts et les données disponibles. Les données peuvent être exprimées sous forme de règles logiques, de type équations statistiques, ou bien encore de probabilités subjectives, voire d'observations.

Une des principales difficultés à laquelle se heurte tout éco-épidémiologiste est l'hétérogénéité des mesures, des données, qui toutes ont un rapport de près ou de loin avec l'état de santé d'une personne. Or, grâce à la fusion de données, le réseau bayésien a la capacité de prendre en compte des données incomplètes, ou incertaines et de vérifier ensuite ces informations. En effet, connaissant les valeurs d'une variable mais méconnaissant les valeurs d'autres, il est possible grâce au calcul des probabilités de retrouver les données manquantes ou lacunaires (principe de l'inférence). En outre, par le biais de procédures de type *Data mining* les meilleures variables à sélectionner sont proposées, de même que des informations *a priori* inconnues et à valeur prédictives peuvent être « automatiquement » extraites de la base de données. Ainsi, les réseaux bayésiens permettent à la fois de modéliser des connaissances et de **produire des connaissances (fonction hautement heuristique) à travers la mise en évidence de relations causales jusqu'à présent cachées ou de variables latentes, et ce, dans le cadre d'une analyse non supervisée (Causal knowledge discovery)**. En effet, les algorithmes non supervisés permettent de découvrir de nouvelles assertions au sens de Greenland (1999), notamment au niveau des hypothèses faites sur les relations entre les variables. Ils permettent de découvrir de nouvelles relations jusque-là insoupçonnées, car **ils ne sont pas fondés sur nos présupposés**, sur nos connaissances limitées, mais sur la connaissance contenue dans les données. C'est pour cette raison qu'il vaut mieux ne pas se restreindre dans le choix des variables (*no causes in, no causes out*).

Les réseaux bayésiens apportent donc de la **valeur ajoutée** en termes de connaissances, ils représentent un formidable outil pour la modélisation quantitative de systèmes complexes en domaine incertain de type sanitaire ou autres. Il s'agit d'une modélisation évolutive, car le réseau bayésien s'adapte aux changements en apprenant au fur et à mesure les nouvelles informations¹⁰, et robuste, fondée sur les théories des graphes et des probabilités. Par rapport à d'autres outils de type Système Expert, les réseaux bayésiens permettent de mieux conditionner les risques, et donc de mieux évaluer les pertes ou dommages encourus, ou bien d'identifier les leviers de réduction de ces risques (politiques sanitaires).

¹⁰ D'autant plus utile que l'environnement et nos relations avec ce dernier sont en perpétuel changement

Ainsi, pour Philippe Weber (2015) : *“The robustness and flexibility of this class of modeling formalism are demonstrated by the wide spectrum of real-world problems in which they have been successfully applied. In particular, they have gathered a lot of attention and have provided a set of relevant solutions in industrial fault detection, identification and recovery and in dependability analysis.”*

Toute application mettant en œuvre des connaissances peut donc relever de l’utilisation des réseaux bayésiens, qu’il s’agisse :

- d’inférer une association entre plusieurs variables,
- d’estimer la probabilité d’évènements passés et futurs,
- de mettre à jour l’estimation de la probabilité de certains évènements lorsque de nouvelles mesures deviennent possibles,
- de formaliser la connaissance d’experts (élicitation),
- d’extraire de la connaissance contenue dans une base de données,
- de combiner la connaissance experte, et celle contenue dans la base de données.

« *Les réseaux bayésiens, initiés par Judea Pearl dans les années 1980 se sont révélés des outils très pratiques pour la représentation de connaissances incertaines et le raisonnement à partir d’informations incomplètes, dans de nombreux domaines comme la bio-informatique, la gestion du risque, la sécurité informatique, le transport etc..* » (Leray, 2008). Ils ont acquis leurs lettres de noblesse dans le domaine du diagnostic médical, informatique et industriel et leurs champs d’application ne cessent de s’étendre : risques opérationnels, gestion de projet, marketing, optimisation de processus, sûreté.

Par rapport à l’analyse statistique standard, l’analyse causale va en effet plus loin ; son but est d’inférer non seulement des croyances, ou des probabilités dans des conditions statiques, mais aussi **la dynamique de ces croyances dans des conditions changeantes**, par exemple, des changements induits par des traitements thérapeutiques ou des interventions externes (politiques publiques). Cette distinction implique que les concepts causaux et associatifs sont différents. Pour Pearl le slogan *"la corrélation n’implique pas la causalité"* peut être traduit en un principe plus utile : on ne peut pas étayer d’affirmations causales à partir des seules associations, même à un niveau populationnel - derrière chaque conclusion causale doit se trouver une hypothèse causale qui peut être établie sur la base de (pré)-connaissances elles-mêmes issues de régularités, ou de fréquences repérées. Tout ce que nous pouvons dire, c’est que deux événements sont dépendants, c’est-à-dire que si nous en trouvons un, nous pouvons nous attendre à rencontrer l’autre, mais nous ne pouvons pas distinguer la dépendance statistique, quantifiée par la probabilité conditionnelle $P(\text{maladie}|\text{symptôme})$, de la dépendance causale, pour laquelle nous n’avons aucune expression dans le calcul de probabilité standard.

Judea Pearl, dans son ouvrage (2000a) rapporte ainsi les propos de Nancy Cartwright (1989) "no causes in, no causes out", ce qui signifie que nous ne pouvons pas convertir la connaissance statistique en une connaissance causale. Pour cela, nous devons disposer d'outils *ad hoc* comme les réseaux bayésiens. En effet, les algorithmes d'apprentissage automatique traditionnels ne peuvent pas récupérer les structures causales contenues dans les données, l'expression causale étant donnée par le graphe (réseau) bayésien.

De plus, pour Pearl (2000a) si les suppositions fondées sur les associations (statistique standard) sont par principe testables, et ce, d'autant plus que l'échantillon est grand, et les mesures suffisamment fines, la sensibilité à l'hypothèse de base tend à augmenter avec la taille de l'échantillon. *A contrario*, les suppositions causales ne peuvent pas être testées à moins d'une expérience scientifique, cependant, la sensibilité à l'hypothèse de base tend à diminuer au fur et à mesure que l'échantillon grandit, c'est-à-dire que les causes, une fois trouvées, restent inchangées. A l'heure du *Big data* cela peut s'avérer important (Calude et al., 2017).

Si nous regardons d'ailleurs l'évolution du terme « *Causality* » dans les publications PubMed entre 1945 et 2021, celui-ci ne cesse d'augmenter au fil du temps atteignant même un pic en 2020, exprimant le réel besoin de raisonner en ce sens. Contrairement, à ce que l'on pourrait penser, il ne s'agit pas d'une modélisation dépassée, au contraire, plus que jamais elle a tout son sens, et il est dommage qu'elle soit si méconnue.

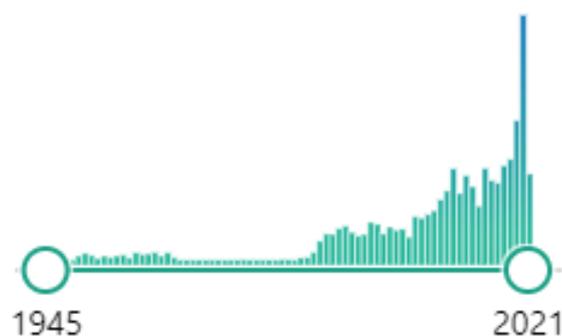


Figure 10. Emploi du terme « *Causality* » dans les publications PubMed entre 1945 et 2021

Pour bien comprendre la différence entre la statistique classique et le paradigme causal, en statistique classique (associations) on chercherait à répondre à la question suivante : « *What is association between pollution exposure and health?* », et sous un paradigme causal à celle-ci : « *What is the causal effect of differential exposure to pollution on health ?* » (Pearl, 2000a). En ce sens, ces théories sont particulièrement bien adaptées aux problématiques de santé environnementale, où nous essayons de coller au plus près de la réelle exposition des habitants, et surtout de comprendre l'effet différentiel d'une exposition délétère dans l'espace géographique sur leur santé.

1.4.5 La création d'un réseau bayésien

Nous allons voir à présent comment créer un réseau bayésien. Un réseau bayésien est composé de deux éléments : **une carte conceptuelle** et une **base de données**. La carte conceptuelle définit la **structure du réseau**, il s'agit d'un outil pour l'organisation et la représentation des connaissances. Cette carte est soit construite à partir des données issues de la base de données, soit à partir de la connaissance de l'expert sur le sujet d'étude. La structure du réseau traduit par des arcs orientés l'ensemble des relations causales entre les variables. Les probabilités de cause à effet entre les noeuds constituent les paramètres du réseau, elles sont calculées soit à partir de la base de données, soit entrées par l'expert (élicitation). Il s'agit généralement de variables discrètes, si elles sont continues, elles doivent être au préalable discrétisées. Il convient de souligner que les réseaux bayésiens ne peuvent pas contenir de boucles de rétroactions entre les variables. Nous ne pouvons donc pas développer de réseaux bayésiens à partir de modèles contenant de la causalité circulaire parmi les variables.

La construction des réseaux bayésiens s'effectue de la manière suivante. Une fois le sujet d'étude identifié, vient l'étape d'acquisition de connaissances, il est alors nécessaire de se documenter sur le sujet de manière **exhaustive**, ou bien de consulter les experts qui traitent de cette thématique afin de n'oublier aucune variable latente susceptible de jouer un rôle (*no causes in, no causes out*). Par la suite, les variables et l'espace d'étude (délimitations, échelle spatiale) sont définis.

Lors de la construction d'un réseau bayésien, il est possible de créer soit la structure et les paramètres du réseau, soit uniquement la structure, mais les paramètres proviennent alors de la base de données, soit de créer entièrement les deux à partir de la base de données.

Dans le premier cas, le réseau est entièrement créé par l'expert qui a une grande connaissance du sujet, notamment des relations de cause à effet entre les variables. Dans une première étape de formalisation de la connaissance, l'expert élabore manuellement la carte conceptuelle, et ajoute à l'aide des tables de probabilités les informations connues sur certaines variables, leur valeur réelle, (par exemple la prévalence d'une pathologie). Il peut donc inférer des causalités à partir de données non observationnelles, ce qui lui permet, toujours par inférence bayésienne d'obtenir des probabilités sur d'autres variables.

Dans la seconde situation (création de la structure uniquement du réseau) l'expert élabore comme précédemment la carte conceptuelle, mais en revanche, il utilise des algorithmes d'apprentissage pour connaître les paramètres à partir de la base de données, c'est-à-dire les différentes probabilités conditionnelles entre les variables.

Enfin, dans le dernier cas, des algorithmes non supervisés permettent d'apprendre à la fois la structure et les paramètres du réseau. L'expert n'intervient que dans le choix des variables. Il s'agit alors de découvrir de la connaissance causale (*Causal Knowledge Discovery*) parmi celles-ci. De la base de données vont alors spontanément émerger la structure et les paramètres du réseau.

Nous possédons à présent un graphe causal où chaque relation entre une paire de variables est valable indépendamment. Ainsi, le graphe causal représente des relations autonomes entre les nœuds parents et les nœuds enfants. C'est comme si chaque nœud « écoutait les instructions » de ses parents, et uniquement de ses parents : les valeurs du nœud enfant sont uniquement déterminées par la valeur de ses parents, pas par d'autres nœuds du système, et ces relations restent invariantes quelles que soient les valeurs que les autres nœuds prennent (Conrady & Jouffe, 2015).

Un des grands intérêts des réseaux bayésiens est la prise en compte de toutes les variables d'exposition, et les effets de celles-ci, au sein d'un même modèle, et de manière **simultanée**, ce qui peut nous donner une idée de l'**exposome** dans lequel vivent des populations. Car il n'y a pas bien souvent, une seule cause qui prévaudrait à l'émergence d'une pathologie, mais plusieurs d'origine environnementale, sociale, génétique, ou bien faisant partie de ce que nous appelons les déterminants individuels. Ces différents facteurs vont alors interagir, se combiner (effet synergétique), se potentialiser (effet supra-additif), ou bien au contraire s'atténuer (infra-additif), nous le verrons plus loin.

Ainsi, *“Within that context it is often not appropriate to seek the discrete cause or causes of a disease, but rather to identify a complex of interrelated and often interacting factors that influence the risk of disease”* (Lucas, McMikael, 2005). Ce qui fait dire à Bach (2019) en parlant de la multiplicité des facteurs qui peuvent expliquer une maladie *“In many cases, there remains a considerable amount of uncertainty, which is particularly regrettable because it is always preferable to treat causes than symptoms »*.

De cette complexité émerge l'incertitude et le goût pour la recherche causale. Et cette recherche causale ne peut pas mieux s'exprimer selon moi que par les réseaux bayésiens.

En effet, le critère de causalité en épidémiologie est le plus souvent établi selon la présence/absence, ou bien selon la sur/sous-représentation, dans les études de type cas/témoin. La théorie de la causalité nous permet d'aller plus loin.

Ainsi, dans le cas où une variable cause l'autre, nous pouvons dire que la relation est causale symétrique, lorsque les 2 variables s'influencent mutuellement, la relation est causale réciproque. Deux variables doivent reproduire le même effet indépendamment d'une autre variable. Deux variables peuvent être influencées par une autre variable antécédente qui leur est alors causale. Si l'effet n'est pas systématiquement produit, ou bien produit dans la même proportion que la cause, cela signifie que d'autres variables doivent intervenir. Nous avons vu qu'une relation temporelle lie les variables entre elles, car l'effet ne peut pas précéder la cause. Il existe également une relation spatiale, par exemple, en santé environnementale, l'effet se fait sentir par rapport à la distance du lieu d'émission, selon un gradient, mais qui n'est pas toujours absolu, puisque selon les vents dominants une direction peut être privilégiée.

Ce qui nous importe c'est de comprendre les processus, les mécanismes à l'œuvre en santé environnementale, et pour cela nous avons besoin de la causalité. Ainsi, d'un ensemble de causalités interdépendantes émerge ou pas une pathologie, mais selon quelles causalités ou plutôt quels enchaînements ou chemins de causalité, et avec quelle proportion (probabilités) ? Et comment estimer la causalité d'une exposition passée dans les résultats sanitaires d'aujourd'hui ? Si un lien est trouvé entre 2 variables, comment la force de ce lien évolue dans l'espace ? dans le temps ? Voilà, des questions auxquelles nous souhaitons apporter des éléments de réponse.

L'avantage des réseaux bayésiens, est par conséquent, de permettre une formalisation complète d'un domaine de connaissances. Le fait que nous puissions enrichir la connaissance de l'expert par la connaissance contenue dans les données est un véritable atout. De plus, leur fonction heuristique est évidente. Selon Naïm et al., (1999) ils diffèrent des réseaux de neurones en ce sens que ces derniers apprennent des données, ils reposent sur une approche fondée sur la connaissance « auto-apprise » on dirait aujourd'hui auto-supervisée, et sur l'emploi de l'induction, ils permettent d'apprendre, mais non de comprendre le raisonnement, ils savent, mais ne savent pas le pourquoi. Ainsi, les réseaux neuronaux sont capables de fournir des prédictions, mais pas d'explications, permettant d'acquérir uniquement des données, mais pas d'expertises. De plus, ils ne fonctionnent pas en milieu incertain. Par contre, ils sont très efficaces pour les simulations dynamiques de données (prédictifs), ou bien encore pour la reconnaissance de formes, d'images et de signaux, raisons pour lesquelles ils sont très utilisés à l'heure actuelle en IA pour de multiples APIs. Il n'est pas question ici d'opposer les différentes méthodes d'IA entre elles, cela n'a aucun intérêt. Elles sont toutes utiles et appropriées selon l'objectif de recherche mené. En fait, comme bien souvent, il apparaît plus judicieux de les coupler, comme nous le verrons dans la dernière partie de ce manuscrit.

Il est temps, désormais de présenter à travers les résultats d'une étude de type Air Santé, menée récemment sur des communes limitrophes de l'étang de Berre tout l'intérêt de travailler en santé environnementale avec des réseaux bayésiens afin d'analyser à fine échelle les liens existants entre exposition à des polluants et leurs effets sur la santé des riverains, et ce, à travers une démarche discursive de type top down.

2. La preuve de concept, les résultats de l'étude Air Santé en Pays de Martigues

Cette étude vise à **analyser les liens potentiels entre pollution de l'air, et les pathologies qui peuvent en découler, au sein du territoire du Pays de Martigues**, où la population est soumise à un cocktail de polluants très fourni, en rapport avec la diversité des sources émettrices (transport routier, maritime, aérien, industries de la pétrochimie, de la sidérurgie, du raffinage, de la production d'énergie, des cimenteries).

2.1 Etat de l'art des études menées sur ce territoire

L'agglomération de Martigues est une zone urbano-industrielle située dans le sud de la France, ouverte sur la mer Méditerranée, au bord de la plus grande zone industrielle du sud de l'Europe, à savoir l'étang de Berre. L'étang de Berre est entouré de grands complexes industriels et de raffineries de pétrole, il connaît des niveaux élevés de polluants atmosphériques ayant un impact sur la santé et le bien-être des citoyens (Allen et al., 2017 ; Goix et al., 2018 ; Mantey et al., 2019).

Tardivement, à partir de 2013, des études épidémiologiques et scientifiques ont été menées pour étudier les effets sanitaires de la pollution de l'air sur la population locale. Tout d'abord, l'étude REVELA13 (Rein, VEssie et Leucémie Aigüe dans le Département des Bouches du Rhône) menée par Santé Publique France entre 2013 et 2016, portant sur les cancers du rein et de la vessie et les leucémies aiguës myéloïdes, a montré que les taux d'incidence des cancers de la vessie, chez les > 20 ans, sont plus élevés que ceux observés au niveau national pour les deux sexes (Mantey et al., 2019).

Ensuite, l'étude Fos-EPSEAL (Étude Participative en Santé Environnement Ancrée Localement) lancée en 2015, dans deux agglomérations, Fos-sur-Mer et Port-Saint-Louis-du-Rhône, a permis de collecter des données de santé sur un échantillon aléatoire d'habitants par une approche participative (Allen et al., 2017 ; Fortin et al., 2017 ; Jeanjean et al., 2021). L'étude a révélé que les maladies chroniques, et les symptômes aigus étaient une expérience de santé partagée dans les deux villes. La prévalence chez l'adulte de l'asthme cumulatif (débutant la plupart du temps à l'âge adulte), des cancers (notamment chez la femme), et du diabète (de type 1) est plus élevée à Fos-sur-Mer et Port-Saint-Louis-du-Rhône qu'au niveau national (Allen et al., 2017 ; Jeanjean et al., 2021). Les affections respiratoires (hors rhume des foies) concernent près d'un adulte sur deux, et un enfant sur quatre (Jeanjean et al., 2021).

L'étude INDEX (Etude d'imprégnation de la population aux polluants atmosphériques de la zone industrialo-portuaire de Fos-sur-Mer) a été conduite en 2016 à Fos-sur-Mer par l'Institut Ecocitoyen pour la Connaissance des Populations à partir d'échantillons de sang et d'urine de 138 habitants préalablement sélectionnés. Cette étude a révélé une sur-imprégnation de la population générale par inhalation de plomb et de deux heptachlores furanes (1,2,3,4,6,7,8-HpCDF, 1,2,3,4,7,8,9-HpCDF), caractéristiques des émissions industrielles, et au benzène pour les plus âgés (Goix et al., 2018). Dans cette zone, le jardinage était associé à une imprégnation plus élevée en polychlorobiphényles (PCBs) par rapport à la zone témoin (Goix et al., 2018). La consommation de légumes du jardin impliquait une plus forte imprégnation en cadmium dans la zone exposée. De même, la consommation de produits de la mer locaux (poissons, crustacés et fruits de mer) était liée à un niveau d'imprégnation plus élevé en PCBs, dioxines/furanes, mercure et chrome (Goix et al., 2018).

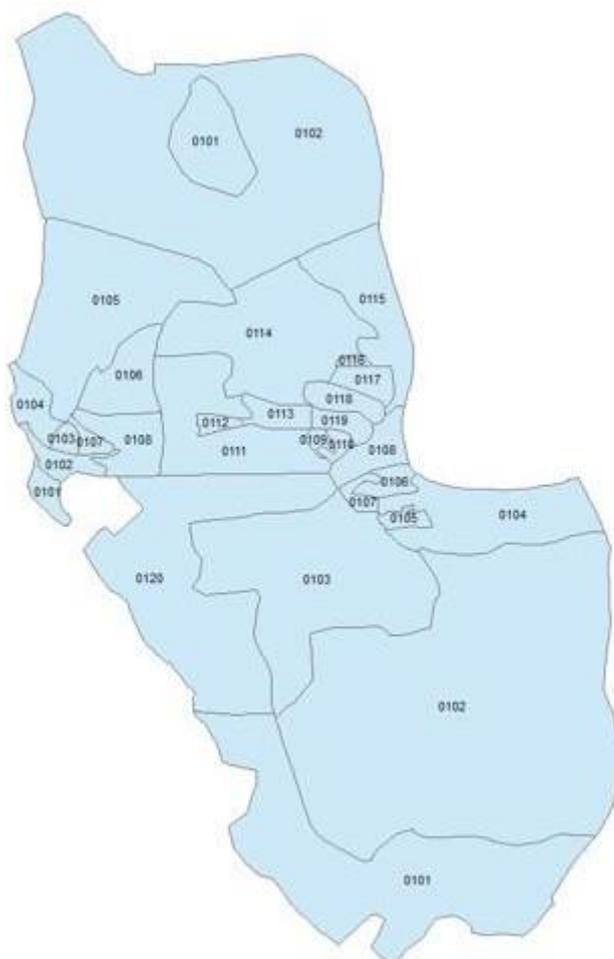
Enfin, les études SCENARII (Simulation de scénarii de pollution atmosphérique pour une Evaluation des Risques Sanitaires - Zone Etang de Berre), et POLIS (Plan d'Observation des Polluants d'Intérêt Sanitaire) ont concerné, en 2010, 66 communes autour de l'Etang de Berre, et ont permis de calculer un excès de risque dans différentes zones surexposées en prenant en compte 39 substances (AtmoSud, 2015, 2016). Pour 7 des substances étudiées, les valeurs limites ont été dépassées : dioxyde d'azote, particules PM10, particules PM2,5, particules de diesel, benzène, 1,3-Butadiène et 1,2-Dichloroéthane. Certaines zones présentant un niveau de risque plus élevé (Golfe de Fos, Martigues, Berre, Marignane, et aux abords des axes routiers).

Nous allons présenter les résultats d'une étude rétrospective de géographie de la santé faisant appel à des données croisées sur la pollution, l'impact de la pollution sur l'environnement (lichens), la précarité des résidents et les séjours hospitaliers par tranche d'âge (soit 171 variables au total) réalisée entre 2015 et 2019 dans trois communes proches de l'étang de Berre à l'échelle infra-communale, à la demande du Pays de Martigues. La plupart des recherches en santé environnementale s'arrêtent à l'échelle de la commune. Or, une grande variabilité spatiale des données environnementales, comme socio-économiques, s'observent à l'échelle infra-communale des IRIS. Le pays de Martigues (Carte n°1), composé des communes de Martigues, Port-de-Bouc et Saint-Mitre-les-Remparts est découpé en 30 IRIS : 20 à Martigues, 8 à Port-de-Bouc, et 2 à Saint-Mitre-les-Remparts (Carte n°2)¹¹.

¹¹ Nous sommes conscients que nous sommes juste à la limite statistique concernant le nombre d'unités minimum pour lancer des analyses statistiques multivariées poussées (30 IRIS)



Carte 1. Les 3 communes du Pays de Martigues



Carte 2. L'espace d'étude à l'échelle des IRIS

Au total **171** variables ont été mobilisées. 18 variables de nature sanitaire qui correspondent à des pathologies d'intérêt selon les 3 tranches d'âge classique (<15 ans, 15- 65 ans, > 65 ans) ayant suscité un séjour hospitalier soit au Centre Hospitalier de Martigues (CHM), soit à la Clinique privée de Martigues, soit dans les 5 hôpitaux de l'AP-HM¹² ou bien à l'Institut Paoli Calmettes (IPC), en 2015. Les données correspondent donc à des nombres de séjours hospitaliers et non à des nombres de patients. Il était, au moment où cette recherche a commencé (fin 2015), malheureusement impossible d'obtenir des « nombres patients »¹³.

¹² Hôpital de la Timone : adultes et enfants, Hôpital de la Conception, Hôpital Nord, et les 2 Hôpitaux Sud : Sainte Marguerite et Salvator

¹³ Les doublons ont été enlevés par les DIMs avant que nous récupérions les données

Ces nombres de séjours/tranches d'âge varient bien entendu suivant le nombre, par exemple, de personnes âgées dans chaque IRIS (Cf. Tableau en Annexe). Afin de ne pas complexifier encore plus les choses et donner ainsi au lecteur la possibilité de faire le lien entre les probabilités calculées par les réseaux bayésiens et les nombres de séjour, nous n'avons pas à ce niveau « contextualisé » le nombre de séjours selon la part de personnes âgées dans chaque IRIS, mais cela est évidemment pris en compte dans les résultats finaux, notamment au niveau de la Classification Ascendante Hiérarchique qui regroupe les IRIS suivant les probabilités de maladies rencontrées en leur sein. Le genre, en plus de la segmentation par âge n'a pu être retenu, car l'association « âge » et « genre » est trop discriminante et pourrait conduire à une potentielle identification des patients.

2.2 Les données sanitaires

L'objectif de cette étude était d'étudier finement les liens potentiels entre pathologies avérées dans la population et la pollution atmosphérique avec laquelle cette population est exposée, dans un contexte industriel. Nous avons donc retenu, après avis du Docteur Laurence Pascal de Santé Publique France qui a participé à l'étude de faisabilité, les pathologies reconnues dans la littérature comme étant en lien avec ce type de pollution, à savoir, les pathologies correspondantes aux **codes CIM-10**¹⁴ suivants :

- . **I00 à I199** toutes les maladies de l'**appareil circulatoire** dans lesquelles on trouve par exemple les cardiopathies ischémiques, les maladies coronariennes,
- . **J00 à J199** toutes les maladies de l'**appareil respiratoire** : laryngites, pharyngites, sinusites, trachéites, bronchopneumopathies, bronchites, bronchiolites, asthme, etc...
- . **E10 à E14** tous les types de **diabète** (type 1 insulino-dépendant et type 2 non insulino-dépendant),
- . **C34** tous les types de **cancer des bronches et du poumon**,
- . **C67** qui correspond au **cancer de la vessie**,
- . **C64** pour le **cancer du rein**,

¹⁴ Classification Internationale des Maladies CIM-10 organisée en 21 chapitres, couvrant l'éventail complet des états morbides classés par organe ou appareil fonctionnel : http://www.departement-information-medicale.com/wp-content/uploads/2017/06/CIM10_2017_sts_20170009_0001_p000.pdf

Le fait de s'intéresser à une échelle territoriale fine impose d'avoir suffisamment de cas par IRIS et donc de travailler plutôt par lots de codes que par codes séparés. Cependant, les DIMs ne sont pas en mesure de nous fournir les données sanitaires au niveau de l'échelle territoriale qui nous intéresse (échelle infra-communale de l'IRIS). Par conséquent, une table de correspondance entre les adresses des patients et les IRIS, doit leur être fournie, et ce n'est qu'après que nous pouvons obtenir les données demandées, à savoir le nombre de cas en 2015/pathologies/tranches d'âge/IRIS. Les DIMs extraient donc les adresses (sans les numéros de rues) des patients qui ont séjourné en 2015 dans leur établissement et qui habitent dans les 3 communes étudiées (codes postaux : 13500, 13110, 13920, et 13117 - Lavéra). À cette étape de l'étude, nous recevons un lot d'adresses sans numéro de rues et sans les codages CIM10. Pour chaque adresse nous avons dû retrouver l'IRIS d'appartenance.

Au total, ce sont plus de **11 683 adresses** qui ont été ainsi recodées en codes IRIS. Cela représente un travail considérable, particulièrement chronophage, car il ne peut malheureusement être fait de manière systématique, en utilisant par exemple, les bases de données TOPO ou Adresses de l'IGN, seulement 34% des adresses ont pu être trouvées correctement en les croisant avec les bases de données existantes. Cela s'explique par le fait que le PMSI (Programme de Médicalisation des Systèmes d'Information), dont sont issues les données sanitaires, est à la base un outil de métronomie, conçu pour évaluer l'activité hospitalière et attribuer proportionnellement les ressources, et non dans une optique épidémiologique. Par conséquent, les adresses des patients hospitalisés ne sont pas standardisées, leur imprécision, leur inexactitude au niveau par exemple de l'orthographe (l'adresse du patient est encore entrée rapidement par les secrétaires lors de son admission) rend difficile toute procédure de recodage automatique.

2.3 Les données de pollution atmosphérique

Ces variables sanitaires sont mises en rapport avec des données de pollution atmosphérique fournies par AtmoSud, partenaire de cette étude, et datant de 2015. AtmoSud a développé ces dernières années, dans le cadre d'un Plan Régional, Santé-Environnement (PRSE) intitulé SCENARII, un outil informatique qui produit entre autres, des cartographies des **niveaux annuels de polluants dans l'air ambiant**. Nous disposons ainsi de mesures relatives à **44** polluants, pour la moitié de nature gazeuse, et l'autre moitié de nature particulaire, leurs niveaux ont été estimés à la fois sur des modélisations (inventaires d'émissions atmosphériques + dispersion), et des mesures (longs historiques ou campagnes ponctuelles menées ces dernières années). Nous avons le choix entre disposer de **mesures de pollution** à l'échelle des IRIS, ou bien à l'échelle des **zones urbaines** (zones bâties auxquelles un *buffer* de 100 m a été appliqué sur leur périmètre) (Figures n°11 & 12). Nous avons préféré cette seconde option, qui nous permet d'avoir une **mesure plus précise de la pollution dans les lieux de vie de la population** (Carte n°3), et ce, d'autant plus que certains IRIS sont vastes (13560101, 13560102, 13560103 et 13560114 à Martigues, 130980102 à Saint-Mitre-les-Remparts, et 130770105 à Port-de-Bouc, avec une population parfois ramassée dans une seule zone, comme nous pouvons le voir sur la carte n°3).

Tableau 2. Les polluants de nature gazeuse fournis par AtmoSud

	Polluant	Formule	Modélisation/mesure	Niveau de fond local (µg/m3)
Gaz	Dioxyde de soufre	SO2	M+m *	6,00E-01
Gaz	Dioxyde d'azote	NO2	M+m *	8,00E+00
Gaz	Monoxyde de carbone	CO	M+m	2,63E+02
Gaz	Acide chlorhydrique	HCl	M seulement *	8,23E-04
Gaz	Acide fluorhydrique	HF	M seulement *	5,29E-04
Gaz	Benzène	C6H6	M+m	7,00E-01
Gaz	Benzo[a]pyrène	BaP	M seulement *	2,42E-06
Gaz	Hydrocarbures aromatiques polycycliques	HAP	M seulement *	2,12E-03
Gaz	Dioxines et furanes	PCDD et PCDF	M seulement *	2,31E-10
Gaz	Mercure	Hg	M+m	1,06E-03
Gaz	Benzo[a]anthracène	BaA	M seulement *	3,42E-06
Gaz	Benzo[b]fluoranthène	BbF	M seulement *	2,68E-06
Gaz	Benzo[k]fluoranthène	BkF	M seulement *	1,88E-06
Gaz	Indéno[1,2,3-c,d]pyrène	IcdP	M seulement *	1,52E-06
Gaz	Dibenzo[a,h]anthracène	DahA	M seulement *	3,12E-07
Gaz	Fluoranthène	Fluoran	M seulement *	2,24E-05
Gaz	Polychlorobiphényles	PCB	M seulement *	5,66E-07
Gaz	1,3-butadiène	C4H6	M+m	1,30E-01
Gaz	Hydrogène sulfuré	H2S	M seulement *	5,72E-06
Gaz	1,2-dichloroéthane	DCE	M+m *	3,00E-02
Particule	Poussières de moins de 10 µm de diamètre	PM10	M+m	1,60E+01
Particule	Poussières de moins de 2,5 µm de diamètre	PM2,5	ratio PM2,5/PM10	1,00E+01

Légende : M : modélisation, m : mesure, * le fond est pris au niveau d'un point éloigné de toutes sources (zone étang de Rolland). Les données sont toutes exprimées en µg/m³ sauf pour le Mercure (HG) exprimé en ng/m³

Tableau 3. Les polluants de nature particulaire fournis par AtmoSud

PARTICULES		
Polluant	Formule	Modélisation/mesure
Benzo[a]pyrène	BaP	M+m
Hydrocarbures aromatiques polycycliques	HAP	M seulement*
Dioxines et furanes	PCDD et PCDF	M+m
Plomb	Pb	M+m
Arsenic	As	M+m
Cadmium	Cd	M+m
Nickel	Ni	M seulement*
Mercure	Hg	M+m
Chrome	Cr	M+m
Chrome VI	Cr VI	ratio Cr VI / Cr total
Cuivre	Cu	M+m
Sélénium	Se	M+m
Vanadium	V	M+m
Zinc	Zn	M+m
Benzo[a]anthracène	BaA	M+m
Benzo[b]fluoranthène	BbF	M+m
Benzo[k]fluoranthène	BkF	M+m
Indéno[1,2,3-c,d]pyrène	IcdP	M+m
Dibenzo[a,h]anthracène	DahA	M+m
Fluoranthène	Fluoran	M seulement*
Polychlorobiphényles	PCB	M seulement*
Particules diesel	PM10_diesel	Ratio diesel sur PM10_corr

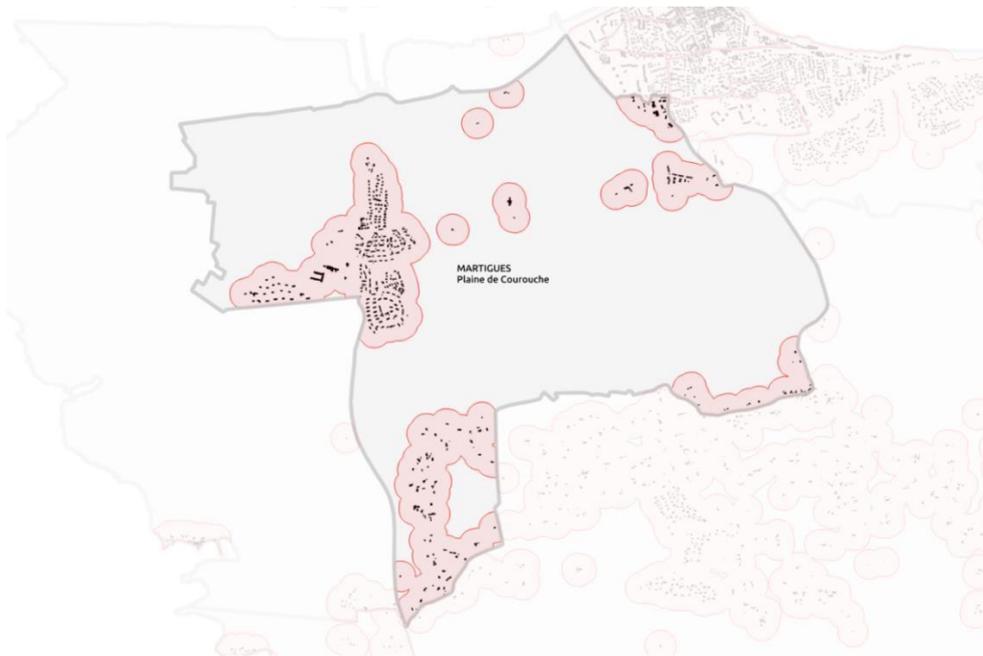


Figure 11. Buffer de 100 m appliqué sur le périmètre des zones bâties, exemple de l'IRIS de la plaine de Courouche à Martigues. Source AtmoSud

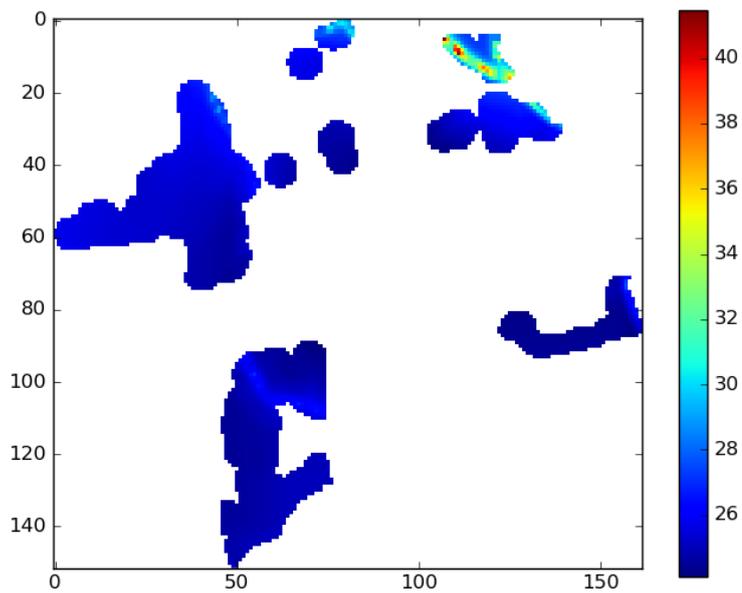
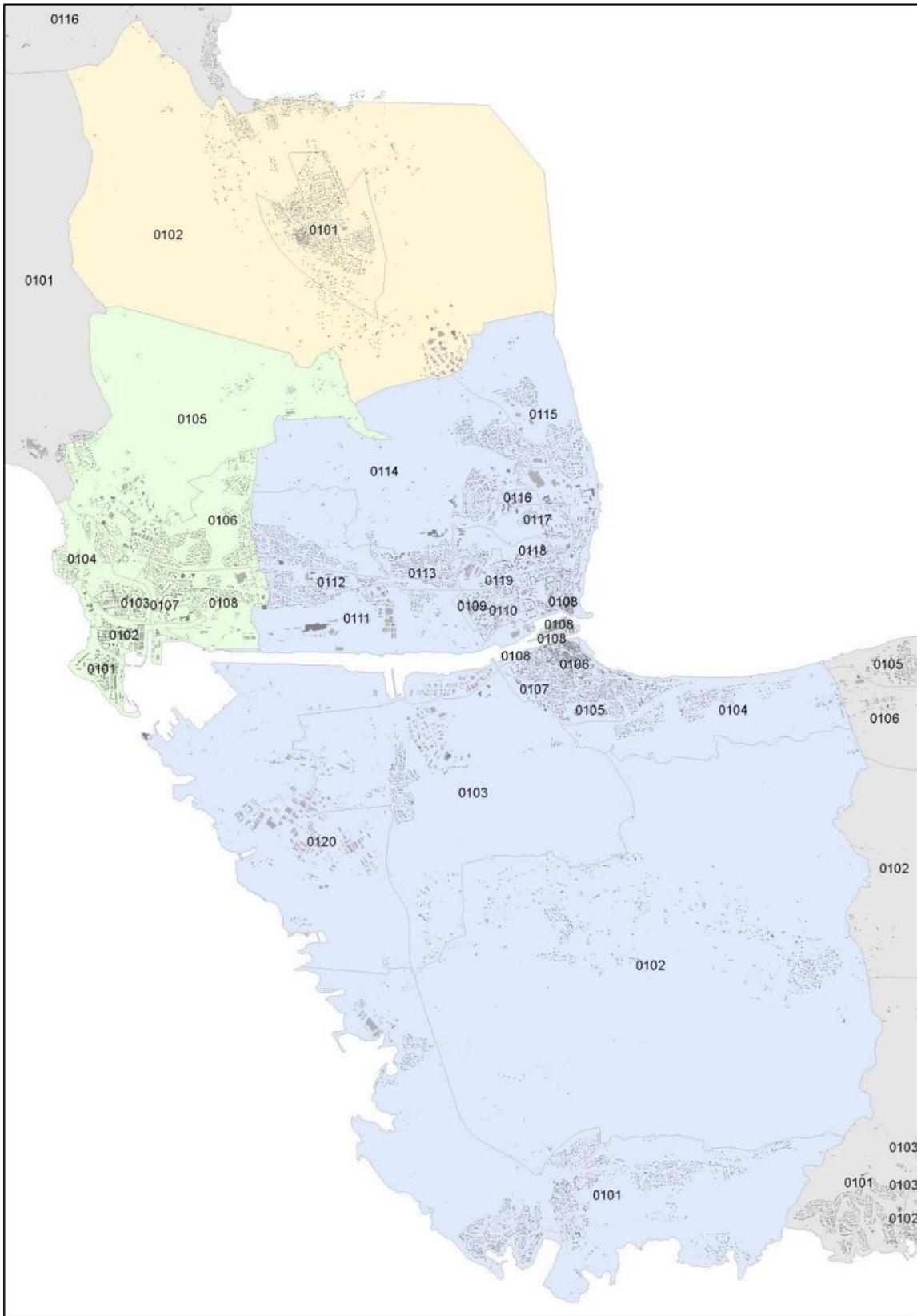


Figure 12. Intersection des zones urbaines de l'IRIS de la plaine de Courouche avec les données de pollution atmosphérique. Source AtmoSud

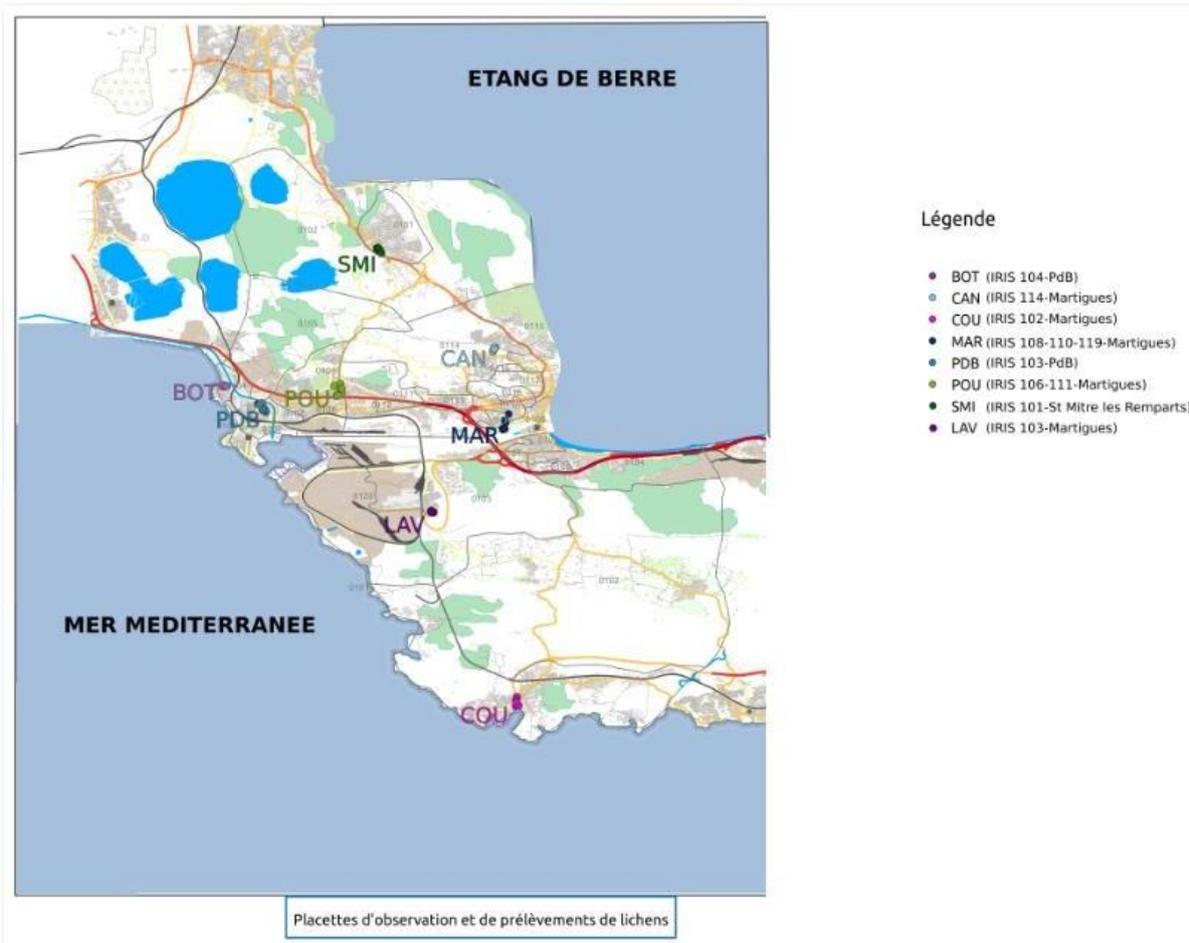


Carte 3. Les zones urbaines dans les IRIS. Source AtmoSud

À noter qu'en ce qui concerne l'ozone, ce polluant est relativement présent au sein de ce territoire en période estivale, où des niveaux dépassant les seuils réglementaires pour la santé humaine y sont fréquemment observés. L'origine de ces phénomènes de pollution à l'ozone réside dans la présence des polluants précurseurs à sa création que sont les oxydes d'azote et les composés organiques volatils. De plus, les conditions météorologiques rencontrées en période estivale (fort ensoleillement) sont très favorables aux réactions chimiques qui vont permettre la création de ce composé. Les sources émettrices industrielles et celles liées aux transports routiers et non routiers sont responsables de la survenue de ces phénomènes de pollution photochimique à l'ozone. Néanmoins, les phénomènes de pollution à l'ozone couvrent de vastes territoires de plusieurs dizaines de kilomètres carrés, et le domaine géographique de l'étude (Pays de Martigues) rencontre dans son ensemble sensiblement les mêmes concentrations en ozone, jour après jour. Ce composé est par conséquent peu discriminant à l'échelle de ce territoire (Martigues, Port-de-Bouc, Saint-Mitre-les-Remparts), lorsqu'il y a une pollution à l'ozone elle affecte l'ensemble de la zone dans des niveaux comparables. De plus, cette pollution n'est qu'estivale (une partie de l'année), alors que les autres composés sont présents durant toute l'année. Pour ces raisons, nous n'avons pas retenu ce polluant dans cette étude.

2.4 Les relevés lichéniques : bio-indication et bio-imprégnation

À ces 44 variables de pollution atmosphérique, nous disposons de **109 autres** (Cf. Liste pour la bio-imprégnation en Annexe) qui correspondent à des mesures effectuées sur des prélèvements lichéniques, réalisés en septembre 2017, par l'Institut Ecocitoyen pour la Connaissance des Pollutions (IECP) sur des placettes spécifiques (Cf. Carte n°4). Ces placettes ont été sélectionnées en concertation avec les différents acteurs de ce projet, et adaptées selon la présence ou non d'arbres *ad hoc* sur le terrain, permettant effectivement la réalisation des relevés. Concernant la validité spatiale des placettes, un rayon de 500 m autour du point GPS retenu pour les placettes de bio-imprégnation peut être considéré. En effet, selon une étude réalisée par l'IECP (Goix et al., 2018), une variabilité de 35 % maximum pour les métaux, et de 30 % pour les différents congénères d'HAP, a été mesurée sur une placette dans un rayon de 500 m. Les résultats d'imprégnation obtenus représentent 6 mois d'intégration. Ces mesures sur des organismes vivants tels que des lichens permettent de **limiter un facteur de confusion** qui pourrait être plus lié au mode de vie des habitants qu'aux effets de la pollution atmosphérique (tabagisme), et ce, d'autant plus que la socio-économie est faible (lien paradoxal connu dans la littérature entre la précarité et le tabagisme).



Carte 4. Les placettes des prélèvements lichéniques. Source IECP

Tableau 4. Les informations concernant les prélèvements

INFOS - MESURES DE BIOACCUMULATION DANS LES LICHENS				
PRELEVEMENTS (IECP JD/CED/GR/AA)				
humidification eau milliQ et prélèvement couteau				
Lieu	code site	date	Lat-Nord	Lon-Est
Port-de-Bouc - Botaï	BOT	20/09/2017	43° 24' 58.64"N	4°58'27.28E
Martigues - La Couronne	COU	18/09/2017	43° 20' 5.01"N	5°2'59.596E
Martigues - Lavéra	LAV	18/09/2017	43°23'2.31"N	5° 1'40.21"E
Martigues - Canto	MAC	18/09/2017	43° 27' 5.357"N	5°0'52.012E
Martigues - Pouane	MAP	20/09/2017	43° 24' 54.119"N	5°0'25.65E
Martigues - Mairie	MAR	18/09/2107	43°24'27.78"N	5° 2'47.09"E
Port-de-Bouc - Centre	PDB	20/09/2017	43°24'44.56"N	4°59'14.28"E
Saint-Mitre - Ville	SMI	19/09/2017	43° 27' 4.874"N	5° 0' 52.585"E
PREPARATION DES ECHANTILLONS (IECP JD/CED/GR/AA)				
Conservation +4°C pendant 5 jours maxi				
Tri: retrait des résidus d'écorces, poussières...				
Congélation -45°C				
Lyophilisation à -55°C / 0.05 mbar				
Broyage à billes (matériaux en Zr)				
Congélation -45°C				
ANALYSES				
HAP et PCB : Laboratoire d'Analyse Chimique de la Timone (Univ Aix Marseille / Fac de Pharma) - Emmanuel Wafo				
Métaux: Laboratoire de la Timone (Univ Aix Marseille / Fac de Pharma) - Florence Chaspoul				
PCDDF: La Drôme Laboratoires (Valence, France)				

2.5 Les données socio-économiques

Enfin, nous avons fait appel à 7 variables de nature socio-économique, car les effets de la pollution ne vont pas concerner tout le monde de la même manière. **Les personnes défavorisées risquent d'y être plus exposées, mais également d'avoir moins de ressources pour y faire face (report des consultations médicales faute de moyens¹⁵, asthme et diabète moins bien contrôlés).** Ces variables sont issues des bases de données infra-communales de l'INSEE en 2014, elles correspondent à :

1. la population des ménages qui a emménagé dans l'IRIS en 2014 depuis plus de 10 ans, cette variable est très utile pour approcher l'exposition des personnes P14_PMEN_ANEM10P
2. le nombre de personnes non scolarisées de 15 ans ou plus, titulaires d'aucun diplôme P14_NSCOL15P_DIPLO
3. le nombre de personnes des ménages dont la famille principale est une famille monoparentale C14_PMEN_MENFAMMONO
4. le nombre de personnes en résidences principales occupées par des locataires P14_NPER_RP_LOC
5. le nombre de personnes immigrées P14_POP_IMM¹⁶
6. le nombre de chômeurs de 15 à 64 ans P14_CHOM1564¹⁷
7. les revenus médians DEC_MED14

Ces variables permettent d'approcher le **niveau de précarité ou de défaveur sociale** des IRIS. **Nous aurions pu utiliser un indice composite**, il en existe plusieurs, le plus connu étant celui de Townsend (1987), ou bien encore celui calculé en local, à l'ARS PACA (ARS, 2012) disponible lui aussi à l'échelle des IRIS, qui prend en compte les revenus médians, la part des résidences principales en location, le taux de chômage, la part des non diplômés pour les plus de 15 ans, et la part des familles monoparentales. Il est donc très proche des variables que nous avons sélectionnées, mais **il nous paraît important de conserver ces variables désagrégées, afin de voir si des différences entre les IRIS apparaissent à ce niveau.**

¹⁵ Bien que cela soit interdit, la CMU est parfois refusée par certains praticiens, ce « refus » non explicite se manifeste par des délais de rendez-vous exagérément longs, ce qui peut décourager certaines personnes, qui peuvent alors décider soit de différer leur consultation, soit de ne pas recourir aux soins, soit de s'orienter vers les services des urgences pour une prise en charge plus rapide, mais avec alors le problème du suivi

¹⁶ Variable rapportée ensuite à la population totale de l'IRIS pour avoir un % plus parlant

¹⁷ Variable rapportée ensuite à la population active de 15 à 64 ans de l'IRIS pour avoir un % plus parlant

2.6 Méthodologie : la modélisation bayésienne

L'ensemble de toutes ces variables (171) * 30 IRIS, correspond à 5 130 cellules dans notre base de données. Ces données sont intégrées au sein d'une même **modélisation**, de type **bayésien**, dont le **formalisme probabiliste** est nous l'avons vu dans la première partie particulièrement adapté aux questions de santé environnementale.

Nous avons choisi d'extraire (pour plus d'impartialité) la connaissance contenue dans notre base de données¹⁸.

Rappelons que La **structure du réseau** (ou carte conceptuelle) est visuellement représentée par des **arcs** orientés figurant l'ensemble des relations causales **reliant les variables**, représentées par des **nœuds**. Tandis que les relations causales entre les variables sont calculées à partir de notre base de données. Les probabilités étant calculées à partir de variables discrètes, si des variables sont continues nous devons au préalable les discrétiser, c'est ce que nous avons fait en utilisant la méthode du K-means qui permet d'obtenir des intervalles optimaux par la recherche de valeurs moyennes en minimisant la variance intra-classe et en maximisant la variance interclasses¹⁹. Un rapport d'importation rend compte des résultats obtenus pour chaque variable (découpages en différentes classes).

Les **variables** sont ensuite représentées sous la forme de **nœuds** :

¹⁸ Tous les traitements ont été réalisés avec le logiciel spécifique dédié aux réseaux bayésiens : BayesiaLab 7.0.1 (version disponible à ce moment-là)

¹⁹ Lorsque la méthode du K-means n'est pas adaptée à une variable, le logiciel nous propose une autre méthode de discrétisation plus appropriée à la distribution de notre série



Figure 13. L'ensemble des 171 variables représentées sous la forme de nœuds

Par la suite, **la carte conceptuelle correspondant à nos 171 variables émerge de manière non supervisée (IA)**, elle représente le réseau bayésien à proprement parler (Figure n°15). Au préalable, nous avons changé les orientations de certains arcs, car si le logiciel repère les relations entre les variables en traçant nous venons de le voir un lien entre celles qui sont en rapport, il ne sait pas à priori quel est le sens de cette relation, il détecte par exemple que les variables A et B sont liées par une certaine probabilité, mais il ne sait pas à ce stade si c'est la variable A qui influe sur B, ou bien si c'est l'inverse, c'est à l'expert de le lui indiquer. De même, à ce niveau nous ne savons pas si la corrélation est positive ou bien négative, il faut pour cela consulter un tableau qui correspond à l'analyse des relations. Afin de mieux distinguer la nature des variables sur le réseau bayésien, nous l'avons différenciée par des couleurs : **orange** pour les **pathologies d'intérêt**, **vert** pour les variables issues des **relevés lichéniques**, **violet** pour les **polluants atmosphériques**, et **bleu** pour les variables de nature **socio-économique**.

Une **cartographie 3D** sur notre réseau permet de représenter graphiquement la force des nœuds²⁰, ainsi que l'intensité et le sens des arcs qui les relient (plus l'arc est épais et plus la corrélation est forte, donc la probabilité d'avoir les 2 variables liées est grande, lorsque l'arc est de couleur rouge la corrélation est négative, bleu : elle est positive, Figure n°16).

L'analyse des relations va donc nous indiquer **le sens et l'intensité des relations entre toutes nos variables, prises 2 à 2**, sur la base du calcul du coefficient de corrélation de Pearson.

Nous observons dans un tableau qui correspond à l'analyse des relations, plusieurs cas de figure :

- des corrélations attendues, entre par exemple les phases gazeuses et particulières de chaque polluant, de même qu'entre les polluants dont on sait qu'ils sont potentiellement présents ensemble (ex. Cr_p et Cr6_p), ou bien encore les données socio-économiques dont on sait aussi qu'elles sont liées (par exemple, entre le nombre de locataires et le nombre de familles monoparentales),
- des relations qui paraissent intéressantes entre différents polluants, nous avons des polluants qui sont corrélés, c'est-à-dire que la présence de l'un explique la présence de l'autre, car ils sont émis ensemble depuis la même source, ils se retrouvent alors dans ce que l'on appelle communément un panache,
- des relations entre les données relatives aux lichens (polluants atmosphériques retrouvés dans les lichens),
- des relations entre pathologies et polluants, entre données issues de la socio-économie et polluants, ou bien entre pathologies, voire entre socio-économie et des variables relatives aux lichens.

²⁰ C'est-à-dire leur importance relative par rapport à l'ensemble du réseau

- **Attardons-nous sur ces dernières relations, en orange, puisqu'elles sont l'objet même de cette étude²¹. La première qui apparaît concerne le nombre de ménages habitant dans leur IRIS depuis plus de 10 ans (P14_PMEN_ANEM10P) et l'abondance relative des espèces de lichen²² (cette relation est positive à hauteur de 70%). Elle pourrait être interprétée comme suit : plus l'abondance relative des espèces de lichen est importante et plus les personnes ont tendance à rester là où elles habitent (bien-être). Une corrélation négative existe même entre cette variable et CB194²³ (-0,70), cependant, une légère corrélation (+0.6) la lie avec CO (Monoxyde de carbone), et avec les maladies cardiovasculaires chez les 15-65 ans (I00I1991565ans, +0.65).**

Une seconde corrélation s'observe entre les patients de plus de 65 ans atteints de maladies cardiovasculaires (I00I19965ans) et respiratoires (J00J19965ans). Il y a là une mise en évidence d'une comorbidité bien connue des médecins, et qui apparaît déjà chez les patients plus jeunes (dont l'âge est compris entre 15 et 65 ans : I00I1991565ans et J00J1991565ans), elle s'explique par la chronicité de ces pathologies (I00I1991565ans - I00I19965ans, +0.67), tout comme pour le diabète (E10E1465ans - E10E141565ans, +0.73).

Notons également, une relation assez forte (+0.78) **entre maladies respiratoires chez les plus de 65 ans et un PCB dioxin-like reconnu pour sa toxicité : CB169²⁴.**

Un lien « intéressant » est relevé (+0.63) entre **maladies cardiovasculaires chez les plus de 65 ans (I00I19965ans)** et une variable de nature socio-économique : **sans diplôme P14_NSCOL15P_DIPLMIN**, de même qu'entre le **revenu (DEC_MED14)** et **l'exposition à un polluant HCl** (acide chlorhydrique, relation négative de l'ordre de 0.65) qui signifie que plus le revenu des ménages augmente et moins ils sont soumis à cette pollution.

Au niveau des **cancers des bronches et du poumon chez les plus de 65 ans (C3465ans)**, nous constatons plusieurs relations positives avec des polluants, notamment avec **CB187 (+0.80), TEQ-2005²⁵ (+0.83), 2,3,7,8 TeCDF (+0.86), 2,3,4,7,8 PeCDF (+0.84), As (+0.78) et SO2 (+0.62).**

²¹ Nous n'avons souligné que les corrélations significatives (> 0.65), à l'exception toutefois de celles impliquant une pathologie qui sont systématiquement repérées

²² « L'abondance correspond à la somme des fréquences de chaque espèce sur une placette. Le relevé étant réalisé sur 10 arbres par placette, sur les 4 faces de l'arbre et selon une grille d'observations à 5 mailles, la fréquence de chaque espèce sur une placette est un rapport du nombre d'observations de l'espèce sur la placette divisée par 200 (10 arbres * 4 faces * 5 mailles). La valeur d'abondance est la somme des fréquences de chaque espèce relevées sur la placette », Source IECF

²³ Famille des PCBs, voir explications page suivante

²⁴ **Facteur de toxicité (TEF) CB169 : 0.03**

²⁵ Le TEQ 2005 correspond à la somme des PCBs et dioxines/furanes exprimés en équivalents toxiques. L'équivalent toxique de chaque congénère est exprimé en multipliant sa concentration par son facteur de toxicité (TEF), qui permet de pondérer la concentration de chaque congénère par son "efficacité toxique". Les TEF utilisés sont ceux définis par l'Organisation Mondiale de la Santé (OMS) en 2005 (Van den Berg, 2006)

Pour le **cancer de la vessie** des relations apparaissent là encore pour les plus de 65 ans (**C6765ans**) avec **VDL (+0.67)**, **CB20 (+0.64)**, **CB114 (+0.85)** et dans une **moindre mesure CB157 (+0.52)**, ces 2 derniers PCB (114 et 157) étant également des PCBs dioxin-like²⁶. Les éléments que nous venons d'énoncer²⁷ correspondent à des PCBs (pour ceux commençant par CB), et des furanes (de la famille des dioxines/furanes) pour 2,3,4,7,8 PeCDF et 2,3,7,8 TeCDF. Les polychlorobiphényles (PCBs) sont des composés aromatiques organochlorés dérivés du biphényle, industriellement synthétisés. Ils forment une famille de composés ayant la même structure chimique de base, mais dont le nombre et la position des atomes de chlore sont variables. Il existe ainsi 209 congénères qui correspondent aux 209 possibilités de configuration des atomes de chlore. Du fait de leur toxicité, les PCBs ont été interdits en France pour leur production et utilisation en 1987, mais ils ont été très largement utilisés depuis 1930, notamment dans les matériels électriques, (comme isolants dans les transformateurs et gros condensateurs), dans les échangeurs thermiques et hydrauliques (fluides caloporteurs, fluides hydrauliques), dans les matières plastiques (ignifugeant), dans les peintures, laques, vernis, colles, encres, câbles, textiles synthétiques, joints d'isolation et mastics, linoléum, PVC, caoutchouc, papier (comme agents plastifiants et adhésifs, anti-oxydants), dans les huiles de l'industrie mécanique (comme additif lubrifiant) et dans des produits phytosanitaires (bactéricides, insecticides...).

Les sources de PCBs sont strictement anthropiques. Il s'agit de sources historiques dues à l'ancienne utilisation et production des PCBs avant 1987 (sols et sédiments pollués qui contaminent l'environnement lors de l'érosion des sols ou de crues), et de sources actuelles dues à l'utilisation de matériaux contenant encore du PCB (comme les transformateurs) ou à **l'incinération ou le stockage de déchets contenant du PCB**.

Certains PCBs ont un mode d'action similaire aux dioxines, ce sont les PCBs-dioxin like (PCB-DL). Les autres PCBs, qui ont un mode d'action spécifique, sont appelés PCBs-non dioxin like (PCB-NDL). Les PCBs ont la particularité d'être lipophiles et de se bioaccumuler le long de la chaîne alimentaire (INERIS, 2012). Une étude nationale sur l'imprégnation en PCBs de pêcheurs avait montré des concentrations plus élevées chez ceux consommant fréquemment des poissons provenant de cours d'eaux contaminés (ANSES, 2011). Les PCBs absorbés s'accumulent principalement dans les tissus adipeux, mais aussi dans le foie, les muscles, le sang, la peau, les poumons, le cerveau et le système nerveux. Les PCBs peuvent-être transmis durant la grossesse et l'allaitement au fœtus, et à l'enfant, car ils sont stockés dans le lait maternel et passent la barrière placentaire.

²⁶ Facteur de toxicité (TEF) PCB 114 & PCB 157 : 0.00003

²⁷ CB169, CB114, CB149, CB157, CB187, CB20, CB28, CB194

Le terme « dioxine » regroupe quant-à lui deux familles de composés aromatiques tricycliques chlorés : les « polychlorodibenzodioxines » (PCDD) qui comportent 75 congénères, et les « polychlorodibenzofuranes » (PCDF) qui comportent 135 congénères. Les PCB-dioxin like sont parfois également désignés sous le terme « dioxines ». Les dioxines ont la particularité d'être lipophiles et de se bioaccumuler le long de la chaîne alimentaire. **Les dioxines sont des résidus qui se forment essentiellement lors de combustions, qu'elles soient naturelles ou industrielles.** Les principales émissions résultent des activités industrielles (**incinération de déchets**, production de chaleur, **métallurgie et sidérurgie**, réactivation du charbon actif, **industrie du chlore et de ses dérivés**, industrie de la pâte à papier et du textile). Les échappements des moteurs à explosion émettent également des dioxines, de même que les combustions de bois et de charbons, qu'elles soient industrielles ou résidentielles. Leurs facteurs de toxicité correspondent pour 2,3,4,7,8 PeCDF à 0.3 (très toxique), et à 0.1 (très toxique) pour 2,3,7,8 TeCDF.

Analyse des nœuds (taille) : Force des nœuds - Analyse des arcs: Corrélation de Pearson

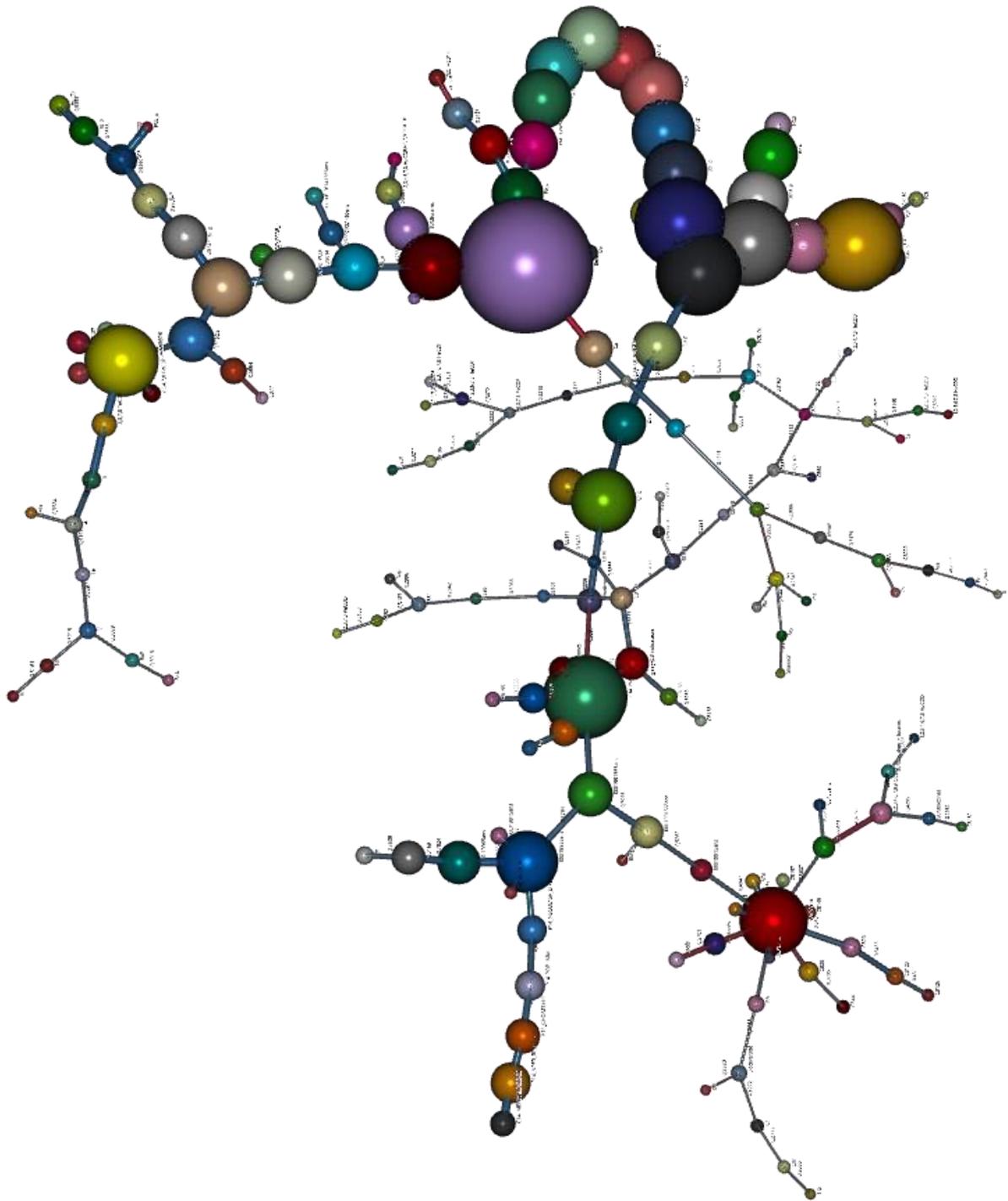


Figure 16. Cartographie 3D, analyse des nœuds & des arcs

Au point où nous sommes parvenus de l'analyse, il est temps de regarder d'un peu plus près les relations que nous venons de trouver entre les pathologies, les polluants, et la socio-économie. Pour cela, nous allons monitorer ces relations afin d'observer comment elles interagissent entre elles.

2.7 Les interactions entre les variables : le monitoring

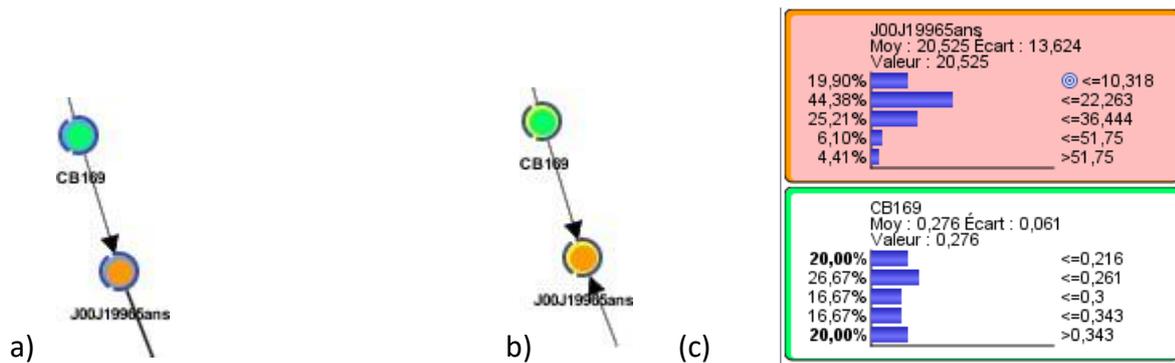
Nous avons fait le choix d'étudier toutes les relations qui concernaient les pathologies, ou bien des variables socio-économiques, même si ces relations étaient faibles (coefficient de corrélation : $R < 0.5$), voire négatives.

L'analyse va donc porter sur les relations suivantes :

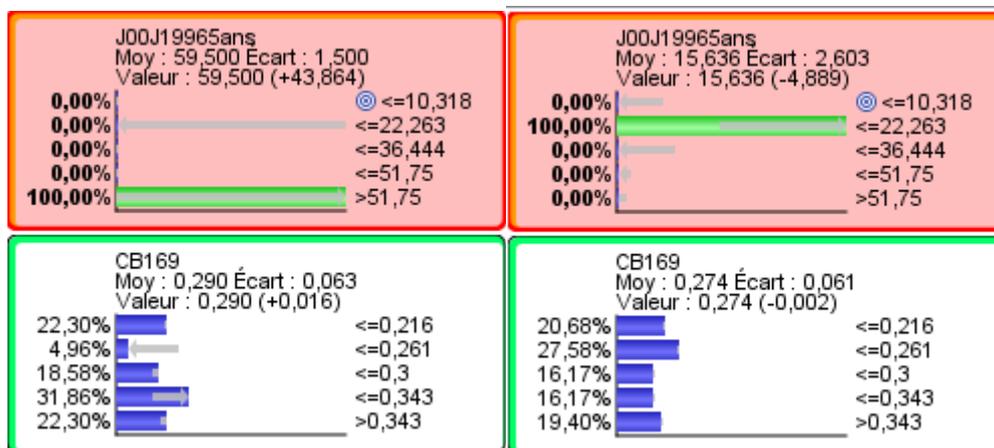
. Interactions entre J00J19965ans & CB169 (R+0,78)	. Interactions entre E10E1465 et Cd_p (R +0.52)
. Interactions entre J00J1991565ans & BkF_p (R +0,35)	. Interactions entre I00I19965ans & plusieurs variables
. Interactions entre C3465ans & plusieurs variables	. Interactions entre DEC_MED14 et HCl (R -0,65)
. Interactions entre C341565ans & HF (R +0.46)	. Interactions entre P14_PMEN_ANEM10P & plusieurs variables
. Interactions entre C6765ans & plusieurs variables	

2.7.1. Les interactions entre les maladies respiratoires chez les plus de 65 ans (J00J19965ans) & CB169 (R + 0,78)

Nous avons vu précédemment dans l'analyse des relations **que les pathologies respiratoires chez les plus de 65 ans** sont en lien avec CB169, et que cette relation est globalement positive à hauteur de 78%. **Mais qu'en est-il plus précisément au niveau des différentes classes (modalités) de ces 2 variables, et comment ces variables interagissent entre elles ?** Pour répondre à cette question nous allons les **monitorer**. La relation est tout d'abord repérée dans le réseau bayésien (a), puis nous monitorons les variables (b), et nous regardons ce qu'il en est au niveau de certaines de leurs modalités (c). Le nombre de modalités correspond au nombre de classes demandées (K-means 5 classes).



Ce qui est particulièrement intéressant à présent c'est de **forcer une de ces classes**, par exemple la modalité qui correspond au nombre de séjours les plus importants (ici : nombre de séjours > 51,75), qui devient alors verte, et d'observer comment les probabilités associées de CB169 sont alors instantanément mises à jour. Ainsi, nous pouvons constater que si nous nous focalisons sur les nombres de séjours les plus importants, les valeurs de CB169 partent à la hausse, surtout pour les 2 dernières modalités (évolution par exemple de 16,67 dans le monitoring initial à 31,86 en d.1). Ce qui signifie que lorsque nous considérons les nombres de séjours hospitaliers les plus importants, enregistrés pour des pathologies respiratoires chez les plus de 65 ans, au sein de ce territoire, cela est en relation avec les valeurs les plus élevées de CB169 retrouvées dans les relevés lichéniques²⁸. La modalité la plus fréquente (44,38%) a également été forcée sans que nous voyions dans ce cas de modifications notables sur le monitoring.



²⁸ Il convient toutefois d'être prudent dans nos interprétations, car les relevés lichéniques ne sont pas exhaustifs sur l'ensemble du territoire d'étude (faute de temps et de moyens), néanmoins les placettes de relevés rendent compte de l'impact de l'environnement en certains points qui n'ont pas été choisis par hasard, en conséquence les relations observées méritent ultérieurement des investigations plus larges (plus de placettes de relevés)

Nous allons voir à présent que chez les 15-65 ans ces mêmes pathologies respiratoires semblent plus en relation avec le Benzo[k]fluoranthène (notamment dans sa phase particulaire : BkF_p), qui est un HAP, qu'avec CB 169. Il semblerait donc que selon l'âge la susceptibilité aux polluants diffère.

2.7.2 Les interactions entre les maladies respiratoires chez les 15-65 ans (J00J1991565ans) & BkF_p (R +0,35)²⁹

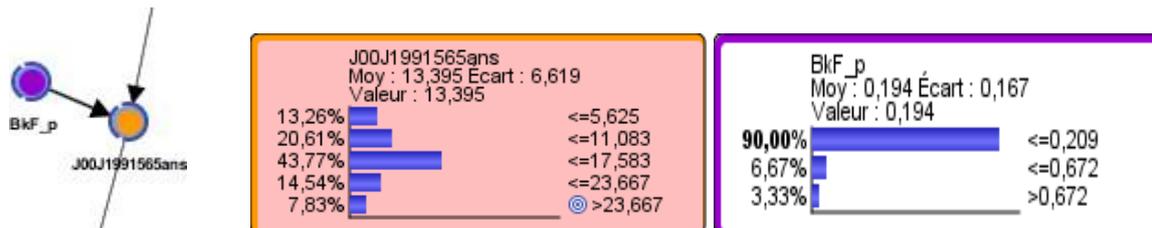
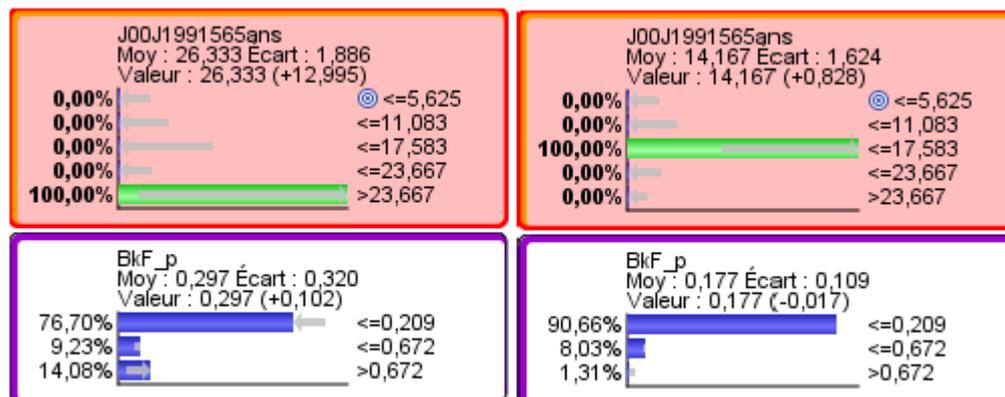


Figure 18. Monitorings initiaux entre J00J1991565 & BkF_p



Modalité J00J1991565ans la plus élevée . Modalité J00J1991565ans la plus fréquente

Ce monitoring initial nous indique que dans 90% des cas, les valeurs de BkF_p sont inférieures ou égales à 0.209 $\mu\text{g}/\text{m}^3$. Si nous nous focalisons sur la modalité la plus élevée, qui correspond aux 7.83% d'IRIS pour lesquels le nombre de séjours pour pathologies respiratoires chez les 15-65 ans est supérieur à 23,667, les valeurs du Benzo[k]fluoranthène passent de 3.33% à 14.08% des plus forts taux de ce polluant : 0.672 $\mu\text{g}/\text{m}^3$.

²⁹ Même si la relation n'est pas significative (R + 0.35), nous avons fait le choix de laisser cette relation, car elle peut évoluer dans le temps

Réitérons cette méthodologie sur les autres pathologies. Il est coutume de dire que les cancers sont des pathologies multifactorielles, qu'en est-il au sein de ce territoire du Pays de Martigues pour les cancers du poumon et de la vessie³⁰ ? Quelles variables interagissent avec ces pathologies, et à quelles hauteurs, sur ce territoire ?

2.7.3 Les interactions entre les cancers des bronches et du poumon chez les plus de 65 ans (C3465ans) & plusieurs variables (TEQ-2005, CB187, [2.3.4.7.8 PeCDF], As, SO₂, [2,3,7,8, TeCDF]

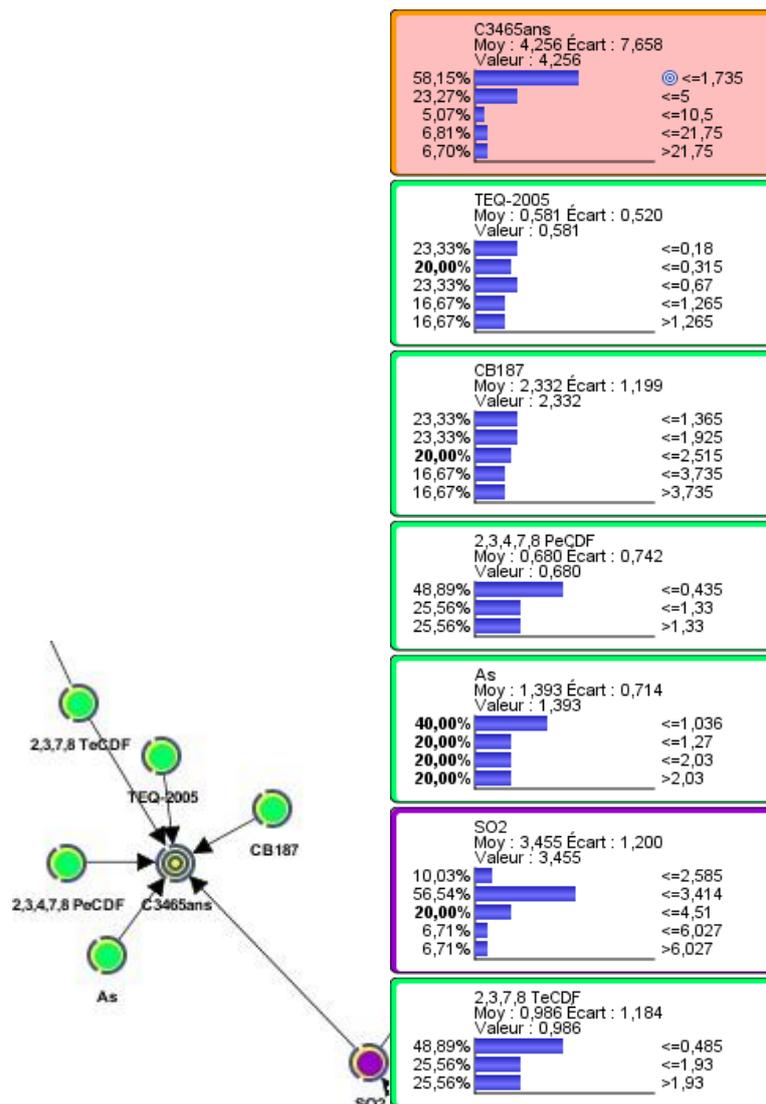
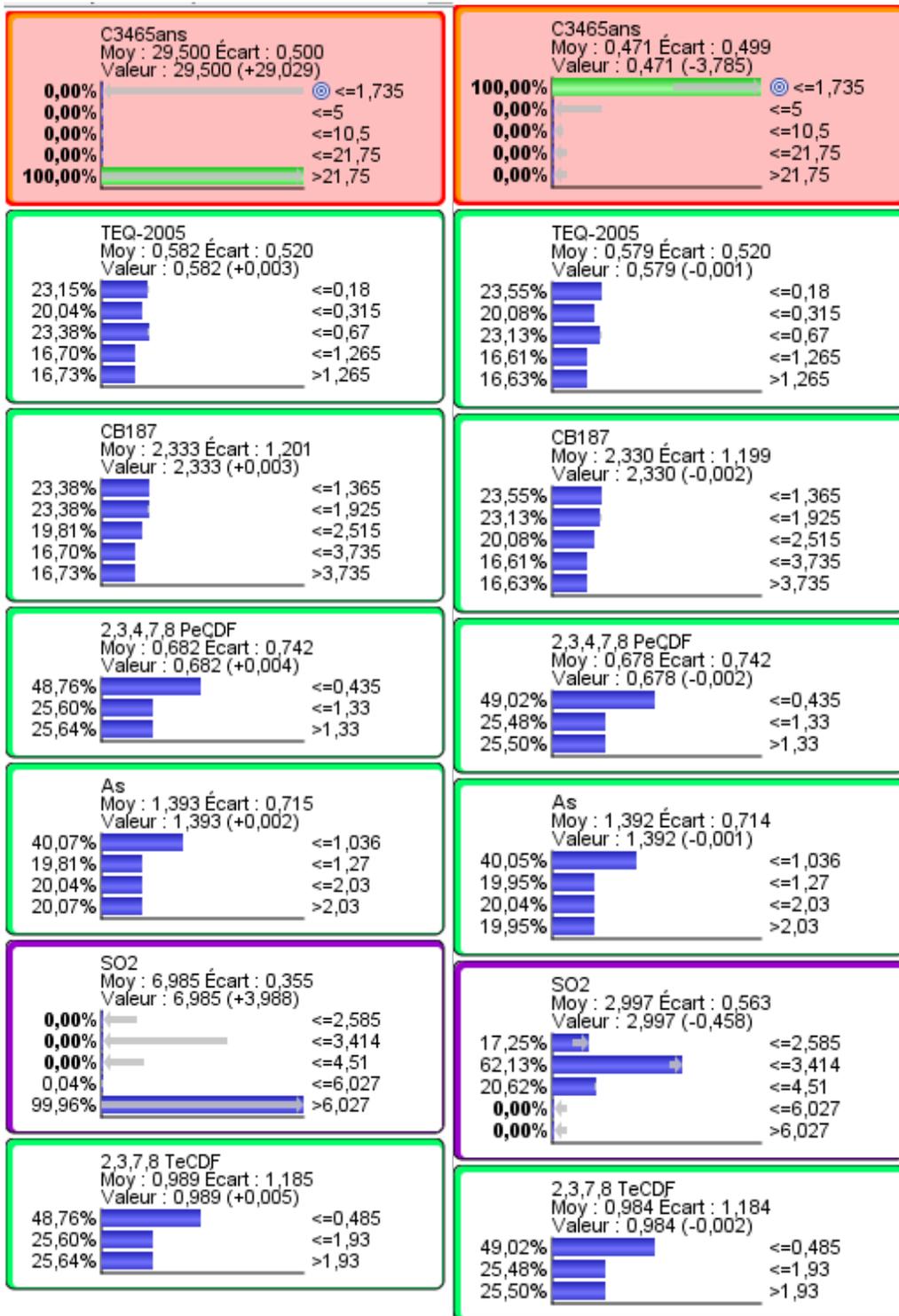


Figure 19. Monitorings initiaux entre C3465ans & plusieurs variables

³⁰ Le cancer du rein (C64) n'apparaît pas dans ces traitements, car le nombre de séjours est globalement faible, seul l'IRIS Jonquières Boudème Font Sarade (130560105) à Martigues se distingue des autres par une valeur maximale de 7 séjours pour cette pathologie pour les plus de 65 ans, en 2015



Modalité C3465ans la plus élevée . Modalité C3465ans la plus fréquente

Au vu des monitorings, les **cancers des bronches et du poumon chez les plus de 65 ans** sont en **lien simultanément avec plusieurs variables**, notamment avec **6 variables issues des relevés lichéniques (TEQ-2005, CB187, [2.3.4.7.8 PeCDF], As, SO₂, [2,3,7,8, TeCDF])**.

Néanmoins, il semble que ce soit le **SO₂** qui soit **le plus impactant**, surtout **quand il dépasse le seuil de 6,027 µg/m³**, ce qui est une faible dose.

L'OMS est consciente du fait que le SO₂ agirait dès de plus faibles doses, sans savoir très bien encore comment cela est possible³¹. Nous pensons que ce polluant interagit avec d'autres (effet cocktail), ce qui le rendrait « actif » dès une faible concentration.

Nous voyons là tout l'intérêt de réaliser ces traitements par modalités, car sinon, nous risquerions de passer à côté d'une relation intéressante, puisqu'au niveau global, la relation entre C3465ans et le SO₂ peut paraître faible (R+0,62), par rapport aux autres variables, où la relation est plus forte de l'ordre de +0,8. **Ainsi, si nous forçons le modèle sur les plus forts nombres de séjours, nous constatons qu'ils sont en lien avec cette valeur de SO₂ (6,027 µg/m³) qui est pourtant une valeur basse.**

C34 est aussi en relation, chez cette fois les 15-65 ans avec un autre polluant : HF, (l'acide fluorhydrique).

2.7.4 Les interactions entre les cancers des bronches et du poumon chez les 15-65 ans (C341565ans) & l'acide fluorhydrique (HF) (R+0.46³²)

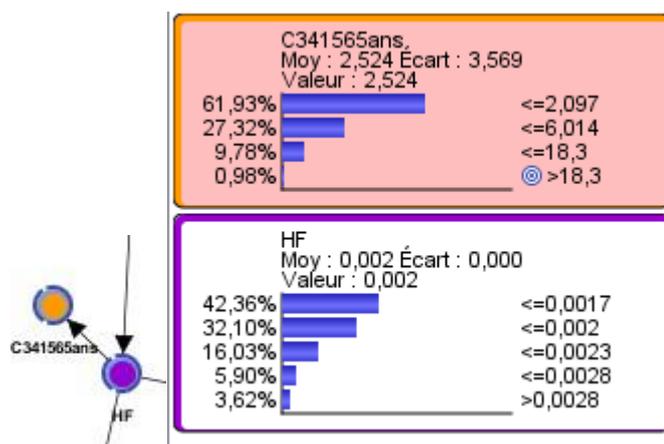
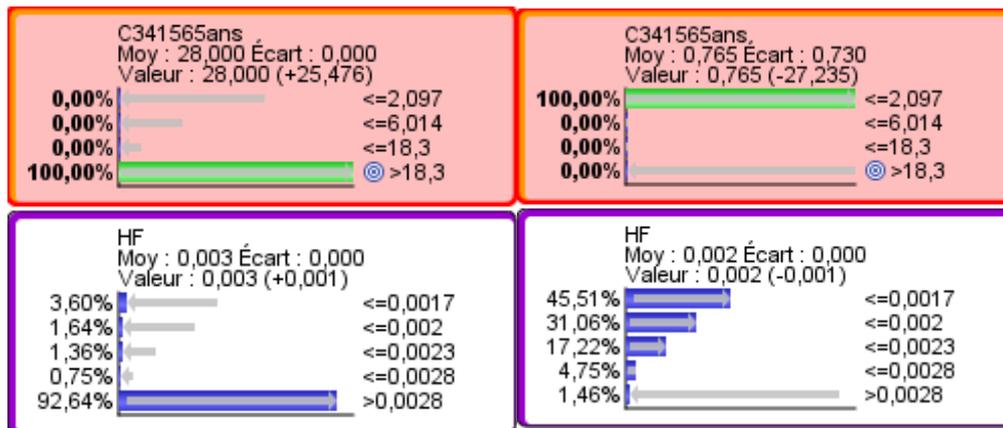


Figure 20. Monitorings initiaux entre C341565ans & HF

³¹ « On sait maintenant que le SO₂ a des effets sur la santé à des concentrations bien plus faibles qu'on ne le soupçonnait auparavant », OMS, 2018 [http://www.who.int/fr/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](http://www.who.int/fr/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)

³² Limite de significativité et faible nombre de séjours



. Modalité C341565ans la plus élevée . Modalité C341565ans la plus fréquente

L'acide fluorhydrique est très agressif pour les muqueuses, c'est la raison pour laquelle son impact apparaîtrait chez les patients plus jeunes, dès que le seuil de $0,0028\mu\text{g}/\text{m}^3$ est dépassé.

Observons maintenant ce qu'il en est des cancers de la vessie, qui sont des cancers qui apparaissent tardivement, et qui peuvent être dus à une exposition prolongée à certains xénobiotiques.

2.7.5 Les interactions entre les cancers de la vessie chez les plus de 65 ans(C6765ans) & plusieurs variables

Tout comme les cancers des bronches et du poumon, les cancers de la vessie chez les plus de 65 ans sont multifactoriels, et en lien avec des variables là encore retrouvées dans les lichens : CB114 (R +0.85), CB20 (R +0.64), CB157 (R +0.52). Des corrélations cette fois-ci négatives ont cependant été observées avec CB28 (R -0.50), crustacés (R -0.59), et CB149 (R -0.65). C'est-à-dire que plus nous avons des nombres de séjours élevés pour C67 et moins nous avons de fortes valeurs pour ces variables. Si cela peut se comprendre facilement pour les lichens de type crustacés, cela pose question pour CB28 et CB149, mais peut être que **la relation n'est pas linéaire entre ces PCBs et leurs effets sur la santé, mais plutôt de type curvilinéaire ?**

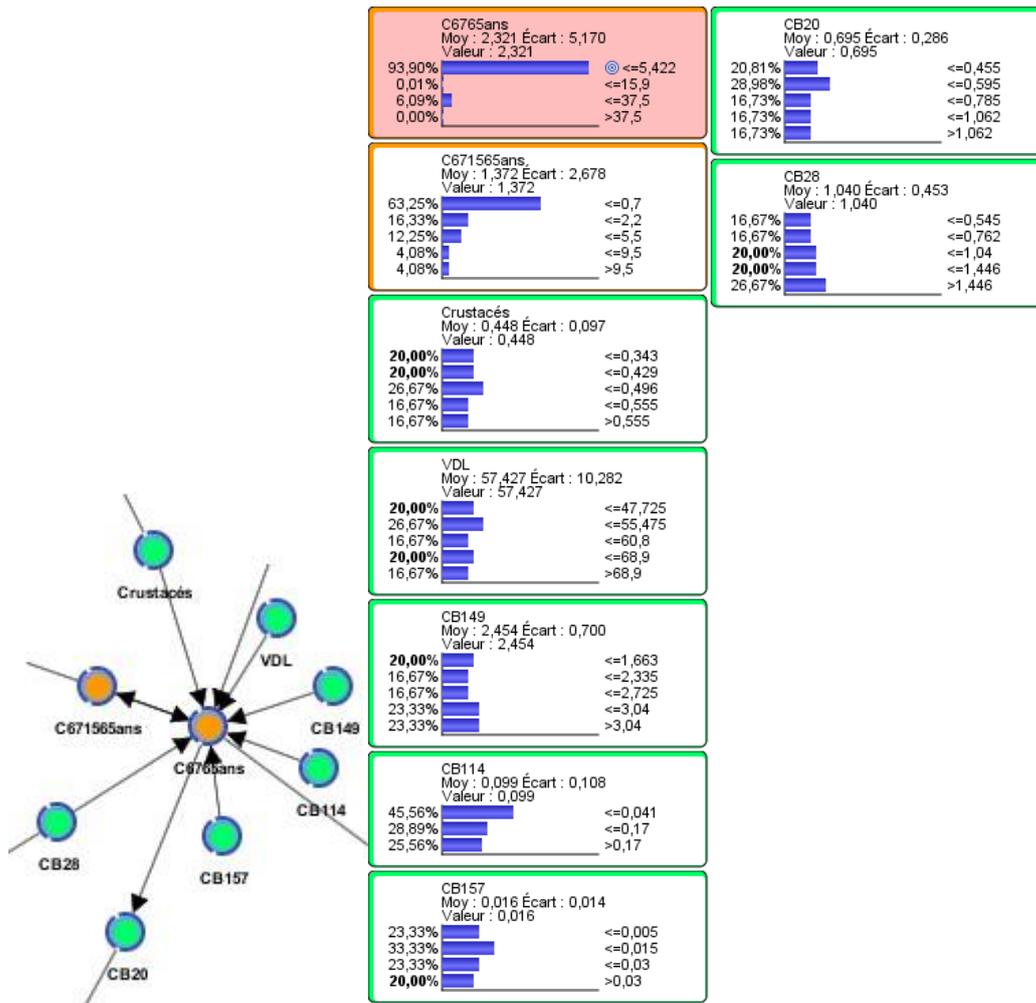


Figure 21. Monitorings initiaux entre C6765ans & plusieurs variables

La majeure partie des séjours appartenant à la 1ère classe (93.90 ; 5.422), nous n’allons pas représenter la classe la plus fréquente puisque nous l’avons là dans le monitoring initial, ni la classe la plus élevée qui concerne très peu de séjours. Il y a en fait peu de cancers de ce type dans le territoire d’étude, ce qui est plutôt une bonne nouvelle.

Seulement deux IRIS, semblent plus concernés que les autres par cette pathologie, à savoir l’IRIS Saint Macaire PlanFossan - 13050114 à Martigues (53 séjours), et l’IRIS 130770102 Centre à Port-de-Bouc (22 séjours). Les monitorings nous indiquent que le nombre de lichens de type crustacés n’est pas très élevé lorsque cette pathologie est observée (corrélation négative, la majeure partie des lichens de type crustacés sont concentrés dans les plus faibles valeurs, 40% < 0.429), tout comme CB114 et CB20, cependant, la corrélation étant cette fois positive cela signifie qu’ils agissent dès ces plus basses valeurs sur C6765 (seuil <0.17µg/m³ pour CB114 ; seuil <0.595 µg/m³ pour CB20, et seuil <0.015 µg/m³ pour CB157).

2.7.6. Les interactions entre le diabète chez les plus de 65 ans (E10E1465) & Cd_p (R +0.52)³³

Une relation est apparue entre le **diabète chez les plus de 65 ans** et le **Cadmium** dans sa **phase particulaire**. Le cadmium est absorbé par l'alimentation (*via* des dépôts, transfert de pollution air-sol), mais apparemment le fait d'inhaler des particules présentes dans l'air ambiant serait capable de perturber le métabolisme de l'insuline au niveau du pancréas. **Cette relation est très forte pour la modalité la plus élevée de patients diabétiques (> 45.1 nombre de séjours) et concerne une valeur de Cd_p comprise entre 0.214 et 0.250 µg/m³. Pour la modalité la plus fréquente, la relation apparaît dès les plus basses valeurs de Cd_p (38,18% inférieur ou égal à 0.181 µg/m³).**

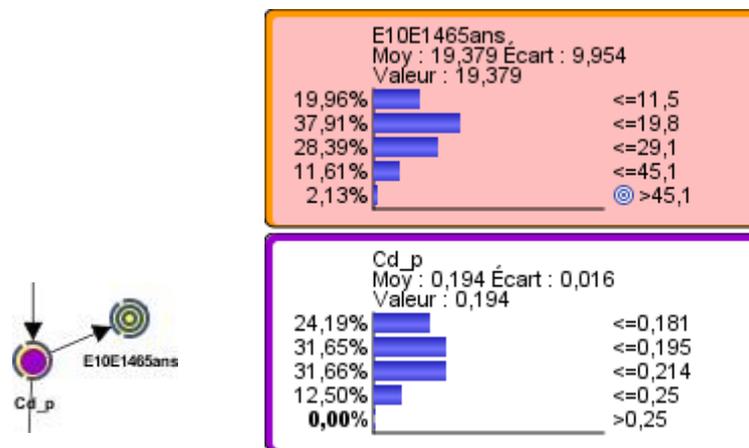
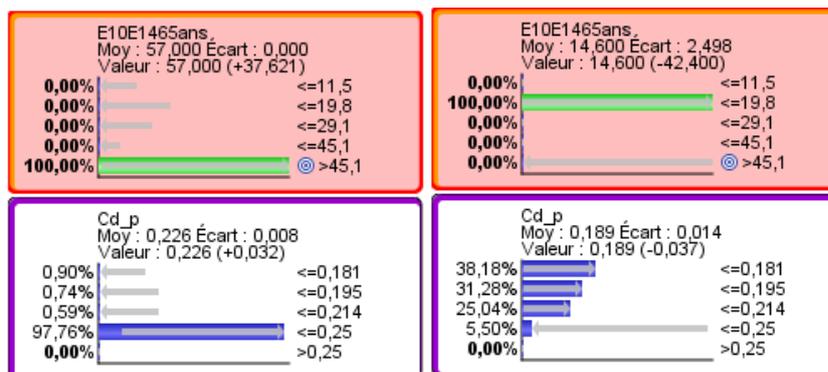


Figure 22. Monitorings initiaux entre E10E1465ans & plusieurs variables



. Modalité E10E1465ans la plus élevée . Modalité E10E1465ans la plus fréquente

³³ <http://www.agence-nationale-recherche.fr/Projet-ANR-13-CESA-0008>

2.7.7 Les interactions entre les maladies cardiovasculaires chez les plus de 65 ans (I00I19965ans) & plusieurs variables

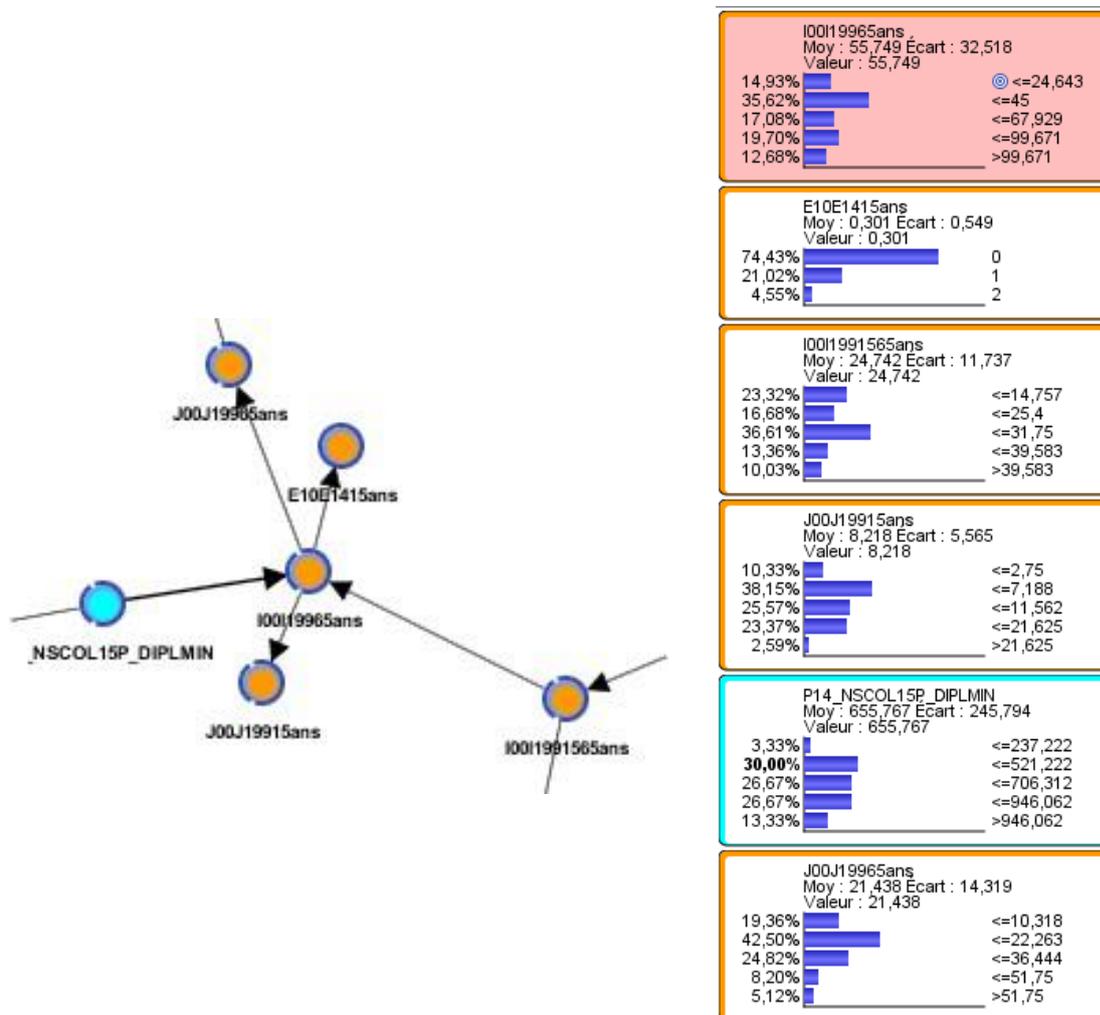
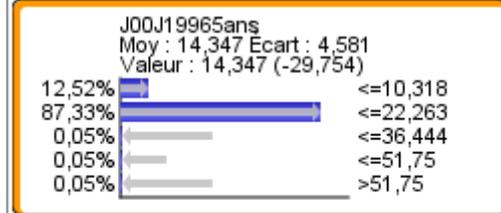
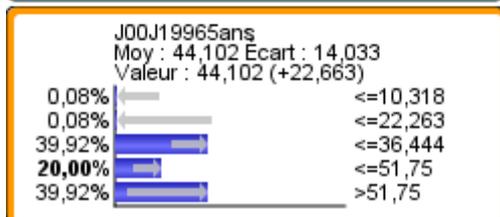
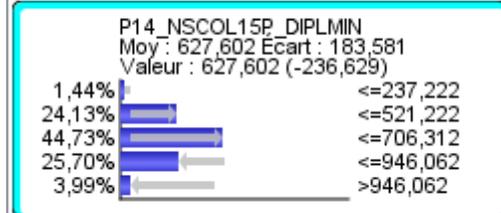
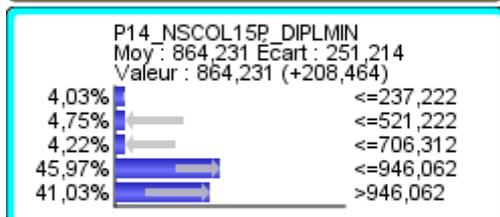
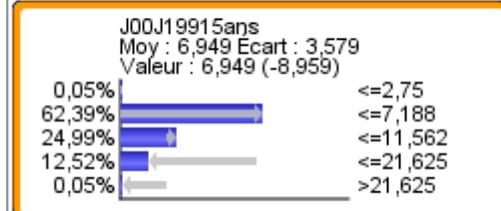
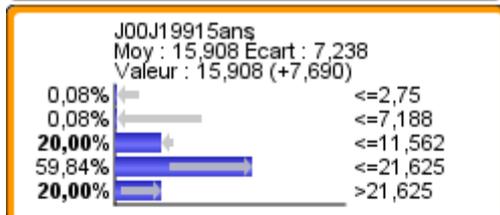
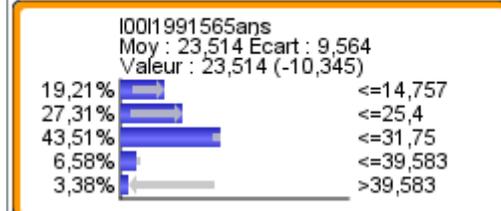
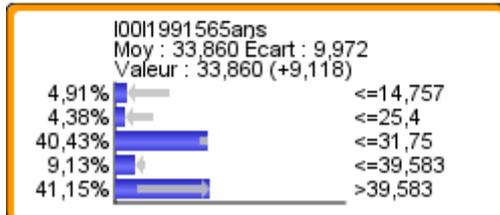
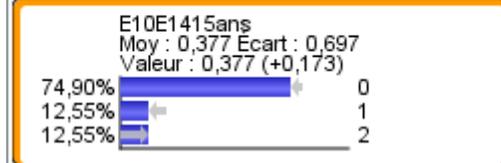
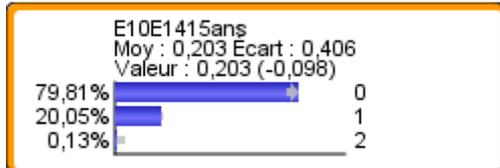
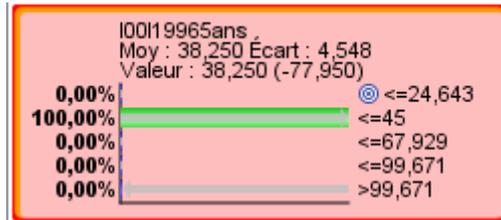
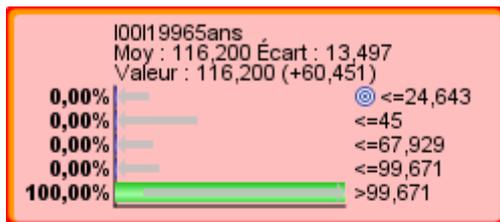


Figure 23. Monitorings initiaux entre I00I19965ans & plusieurs variables

Les **maladies cardiovasculaires chez les plus de 65 ans** sont souvent corrélées avec d'autres pathologies (**comorbidités**), c'est ce que nous retrouvons ici, où elles sont liées avec les **maladies respiratoires** à la fois chez les patients du même âge (J00J19965ans, R +0.86), mais aussi chez les patients les plus jeunes (<15-ans) (J00J19915ans, R +0.62). Les pathologies cardiovasculaires chez les plus de 65 ans sont également en lien avec une variable de nature socio-économique à savoir les personnes sans diplômes (P14_NSCOL15P_DIPLMIN), dont les conditions de travail et/ou de vie peuvent à la longue influencer sur leur santé cardiaque (R +0.63). Ainsi, **la modalité la plus élevée des maladies cardiovasculaires (> 99.671) est en lien avec le nombre élevé de personnes sans diplômes à 15 ans** (évolution de +27.7% : Monitoring initial 13.33 => modalité la plus élevée 41.03%).



. Modalité I00J19965ans la plus élevée

. Modalité I00J19965ans la plus fréquente

2.7.8 Les interactions entre les revenus (DEC_MED14) et l'exposition à l'acide chlorhydrique (HCl)

Une relation négative ($R = -0,65$) est particulièrement intéressante à observer entre les revenus des ménages (DEC_MED14) et l'exposition à un polluant, à savoir l'acide chlorhydrique (HCl). Ainsi, **plus les revenus sont élevés et moins les ménages semblent exposés**, puisque pour les revenus les plus élevés (> à 24 603 euros et au-delà), l'exposition à HCl est pratiquement nulle, tandis que **les personnes dont le revenu annuel avoisine les 18 000 euros (situation la plus fréquente au sein de ce territoire) sont beaucoup plus concernées par cette exposition**.

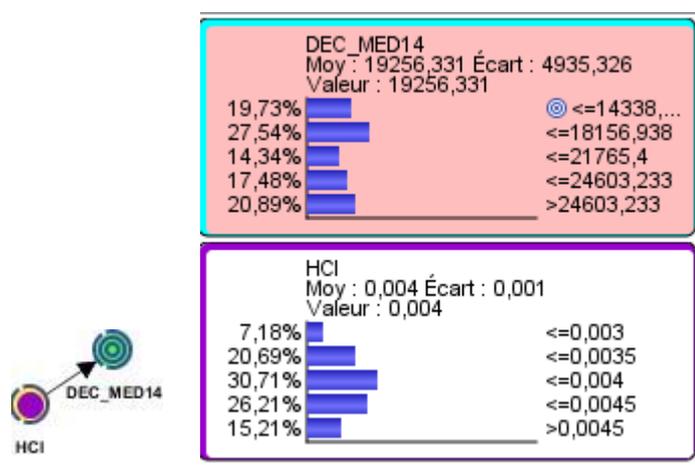
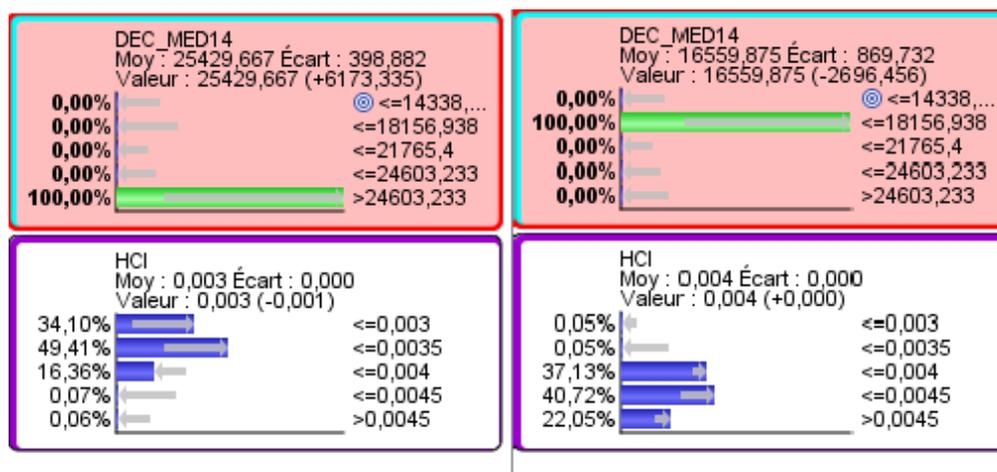


Figure 24. Monitorings initiaux entre DEC_MED14 & HCl



Modalité DEC_MED14 la plus élevée . Modalité DEC_MED14 la plus fréquente

Enfin, nous terminerons par l'étude des interactions entre la population qui habite dans son IRIS d'appartenance depuis plus de 10 ans, avec plusieurs variables.

2.7.9 Les interactions entre le fait d’habiter dans un IRIS depuis plus de 10 ans (P14_PMEN_ANEM10P) & plusieurs variables³⁴

Des corrélations sont positives avec les maladies cardiovasculaires chez les 15-65 ans (I00I1991565) (R +0.65) et CO (R +0.60). **P14_PMEN_ANEM10P est une variable importante**, car elle rend compte de ce à quoi sont exposés les habitants depuis plusieurs années, dans ce cas : l’exposition au CO (monoxyde de carbone). La corrélation négative avec le butadiène (-0,53) peut s’expliquer par son odeur qui ressemble à celle issue des pots d’échappement des automobiles, et qui peut incommoder à la longue les personnes, préférant dans ce cas déménager. Si on regarde le monitoring sur la modalité la plus élevée, 99,5% des ménages qui habitent dans leur IRIS depuis plus de 10 ans, sont exposés à <0.264µg/m³ de butadiène en moyenne annuelle, ce qui est en deçà de la valeur de référence à ne pas dépasser (2 µg/m³) selon l’Ambient Air Quality Criteria (AAQCs) de la Province de l’Ontario au Canada.

La corrélation négative avec CB194 et PCB indicateurs peut s’expliquer par le fait qu’il n’y a pas de variations statistiques entre P14_PMEN_ANEM10P et ces variables, en effet, ces dernières donnent l’impression qu’elles sont uniformément réparties, que ce soit pour la modalité la plus élevée, ou bien la modalité la plus fréquente, donc le nombre de ménages habitant dans un IRIS depuis plus de 10 ans augmente ou diminue indépendamment de CB194 et PCB indicateurs, ce qui entrainerait la corrélation négative. En revanche, une **relation positive semble exister entre le fait d’habiter dans la zone d’étude depuis plus de 10 ans et les maladies cardiovasculaires chez les 15-65 ans (I00I1991565), entre autres dues à une exposition au CO** (Lucas et al., 2009).

³⁴ NB : Les corrélations sont négatives entre P14_PMEN_ANEM10P et CB194 (R -0.70), P14_PMEN_ANEM10P et PCB indicateurs (R -0.60), et avec le butadiène (R -0.53)

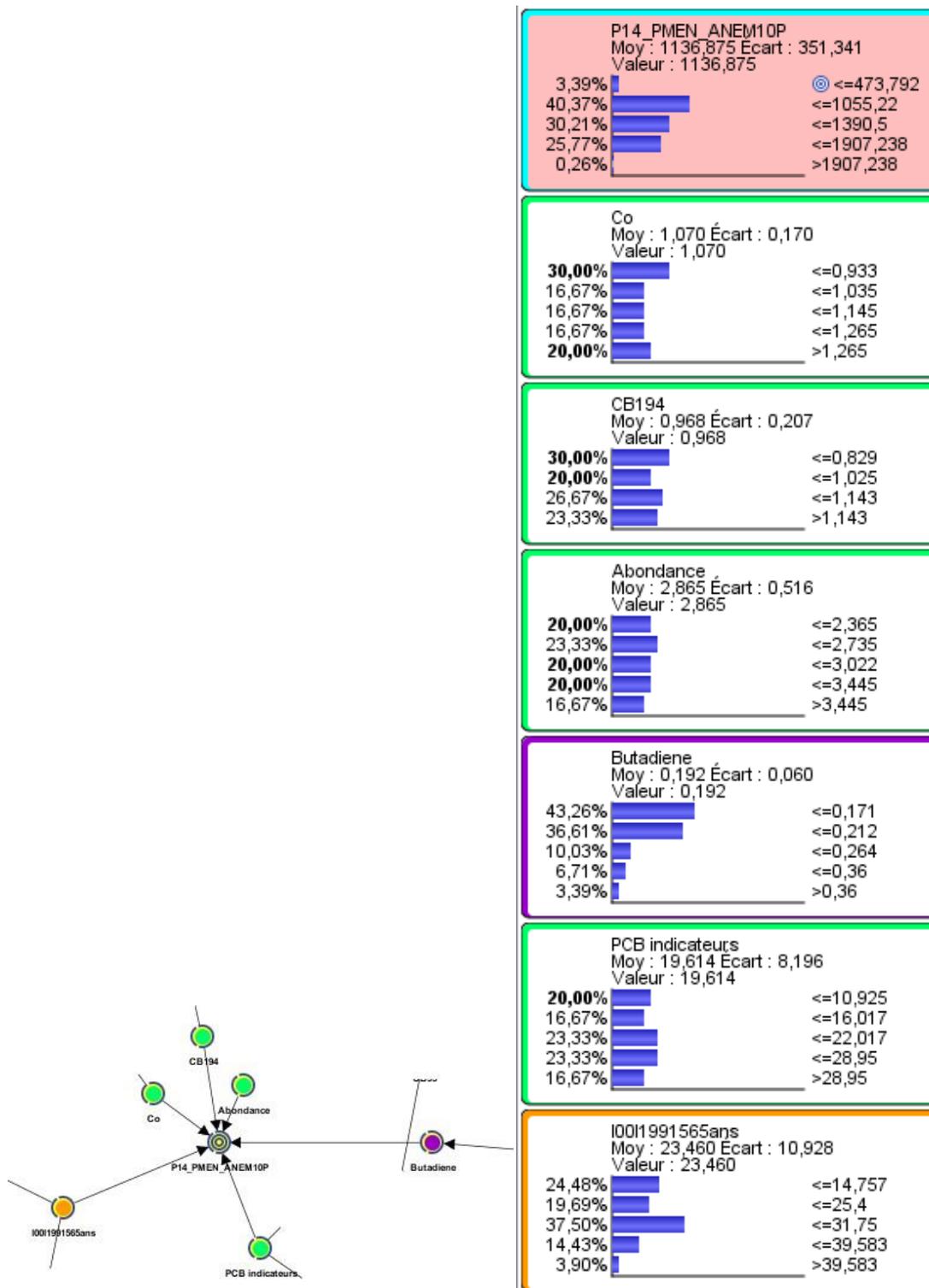
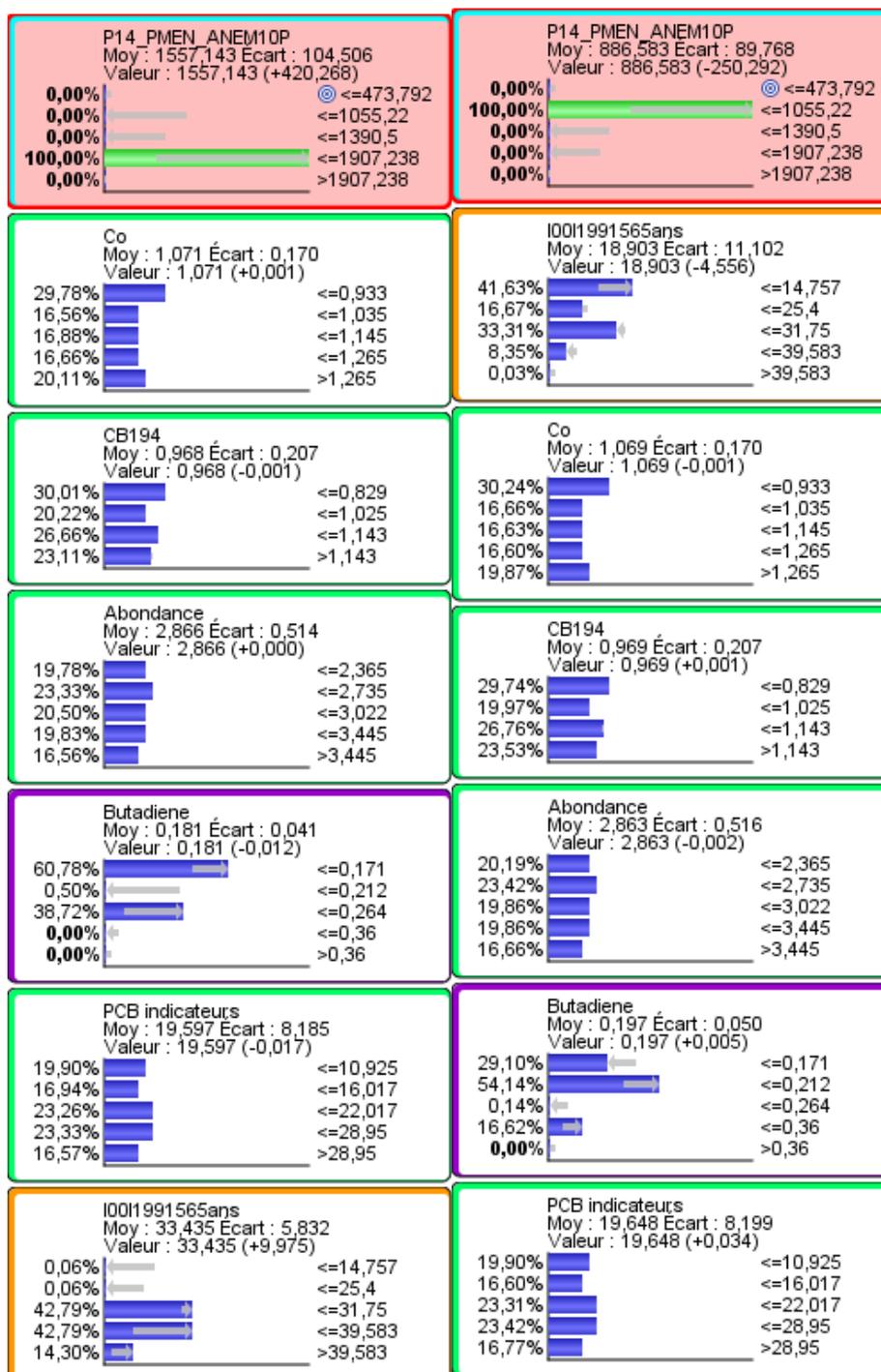


Figure 25 . Monitorings initiaux entre P14_PMEN_ANEM10P & plusieurs variables



. Modalité P14_PMEN_ANEM10P la plus élevée. Modalité la plus fréquente

Jusqu'ici les analyses menées étaient globales, sur l'ensemble des IRIS du territoire du Pays de Martigues. Descendons à une échelle plus fine, afin de voir comment ces relations varient selon les IRIS, et si certains sont plus exposés que d'autres. Cela grâce à un nouveau traitement qui va produire des analyses de sensibilité pour chacun des 30 IRIS de notre espace d'étude.

2.8 Les analyses de sensibilité par IRIS

La **variable IRIS** va donc devenir à présent la variable la plus importante, elle devient une **variable cible**. Nous allons mesurer dans chaque IRIS l'**importance probabilisée de chaque variable**. Cependant, dans un souci de lisibilité, nous allons le faire par grands domaines de variables (santé, polluants atmosphériques, socio-économie, lichens).

Voici le résultat, ci-dessous, de l'algorithme de **traitement supervisé** de type « naïve augmentée » lancé sur la variable cible IRIS, **pour les pathologies**.

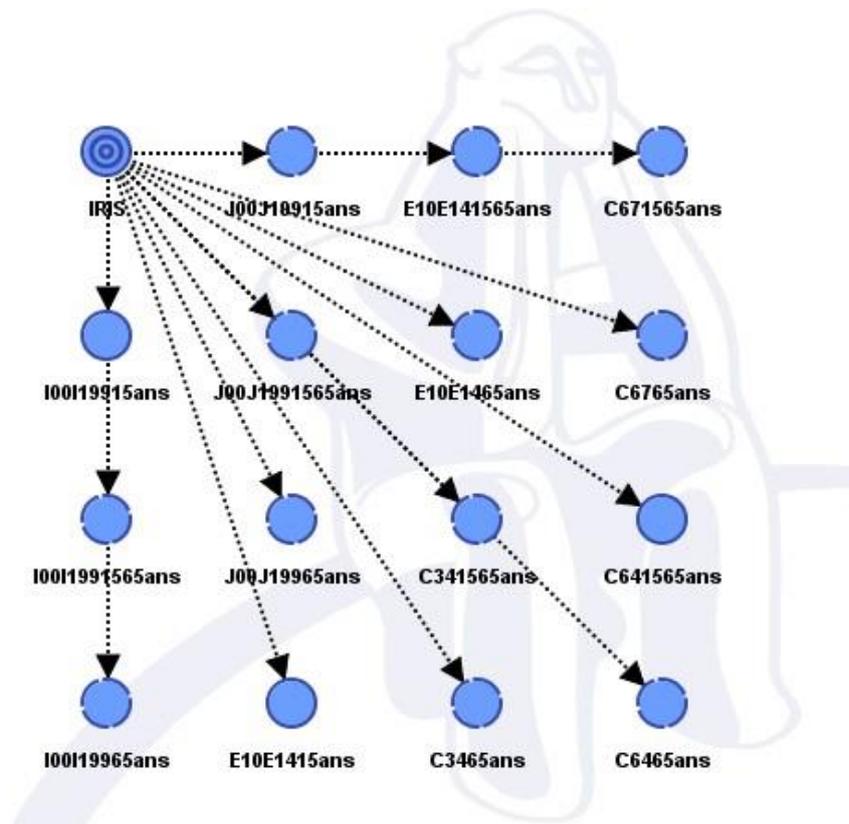


Figure 26. Traitement supervisé sur la variable cible IRIS pour les pathologies

C'est à partir de cet algorithme que les analyses de sensibilité sont établies. La figure ci-dessous représente toutes les analyses de sensibilité par pathologies, elle nous permet d'observer **l'hétérogénéité des IRIS** pris ensemble, et nous conforte dans la nécessité de disposer d'analyses de sensibilité IRIS par IRIS.

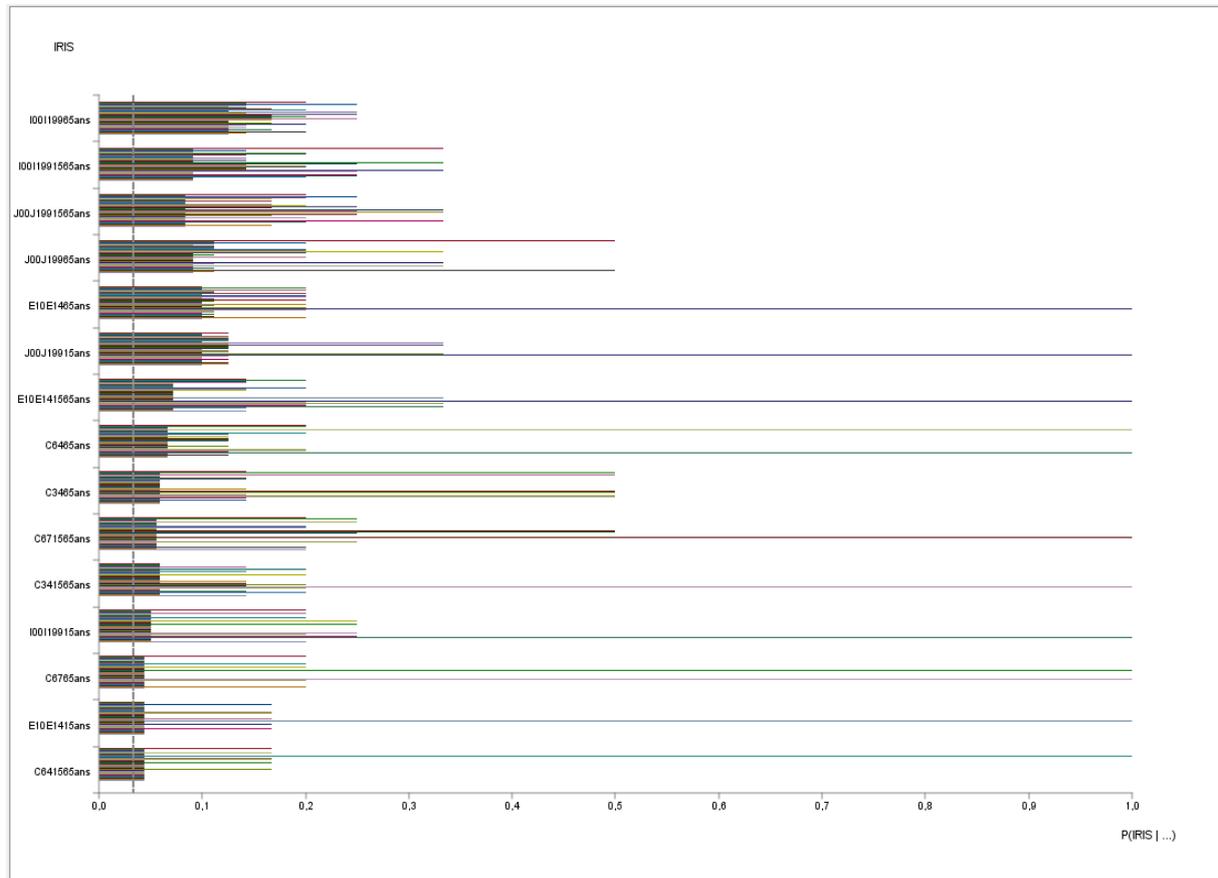


Figure 27. L'ensemble des analyses de sensibilité par pathologies et tranches d'âge pour tous les IRIS

Ces analyses de sensibilité sont trop nombreuses pour être placées ici ($30 \times 3 + 10^{35}$). Voici, ci-dessous, et à titre d'exemple, l'analyse de sensibilité sur les pathologies, menée sur l'IRIS 130560101 (Côte bleue), à Martigues, sur laquelle nous pouvons voir que les maladies respiratoires pour les plus de 65 ans (J00J19965 ans) ont une probabilité élevée d'être retrouvées dans cet IRIS, de l'ordre de 50%, suivies des maladies cardiovasculaires pour les 15-65 ans pour 35%, et ainsi de suite.

³⁵ 30 IRIS pour les pathologies, 30 pour les polluants atmosphériques et la socio-économie et 10 IRIS pour les lichens (là où il a été possible d'en prélever)

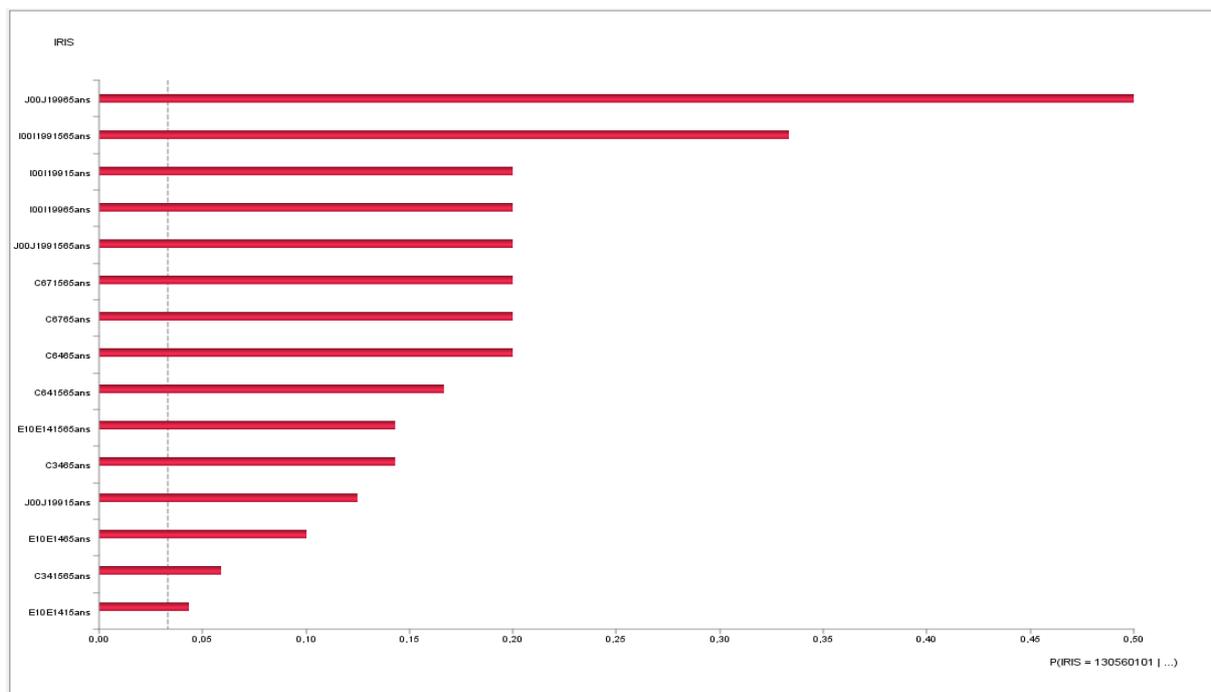


Figure 28. L'analyse de sensibilité/pathologies pour l'IRIS Côte bleue (13560101) à Martigues

Par la suite, nous avons récupéré ces valeurs de probabilité³⁶ pour chacun des IRIS et uniquement pour le domaine des pathologies, qui nous intéresse le plus, et nous les avons reportées dans le tableau n°5, ci-après.

³⁶ Par exemple, en ce qui concerne l'IRIS (130560101) de la figure n°28, la probabilité pour les maladies respiratoires chez les plus de 65 ans est de 0.5

Puis, ces **valeurs de probabilité** ont été **divisées par le % de personnes dans chaque tranche d'âge** afin de les **contextualiser** et d'**éliminer les éventuels effets de taille**, dans le cas où nous aurions dans un IRIS, proportionnellement, plus de personnes âgées qu'ailleurs (Cf. Tableau en Annexe).

Les résultats sont représentés dans le tableau n°6, **une valeur élevée est donc le signe d'un nombre de séjours importants pour la pathologie et l'IRIS considérés, et ce, indépendamment du % de personnes âgées dans cet IRIS.**

Tableau 6 . Les valeurs de probabilité contextualisées sur le % des différentes tranches d'âge

IRIS	I00I19315ans	I00I19315ans	I00I19315ans	J00I19315ans	J00I19315ans	E10E145ans	E10E145ans	E10E145ans	E10E145ans	C34I5ans	C34I5ans	C34I5ans	C34I5ans	C67I5ans	C67I5ans	C67I5ans	C67I5ans	C67I5ans	C64I5ans	C64I5ans	C64I5ans	C64I5ans	C64I5ans	C64I5ans
13056001	0,82	0,59	0,96	0,51	0,35	2,40	0,00	0,25	0,18	0,00	0,10	0,69	0,00	0,35	0,96	0,00	0,30	0,96	0,00	0,00	0,30	0,96	0,00	0,96
13056002	0,00	0,16	0,66	0,44	0,15	0,52	0,00	0,36	0,36	0,00	0,11	2,33	0,00	0,45	0,20	0,00	0,00	0,45	0,20	0,00	0,00	0,00	0,00	0,93
13056003	0,00	0,23	1,20	0,49	0,41	0,96	0,81	0,23	0,16	0,00	0,10	0,28	0,00	0,09	0,21	0,00	0,10	0,09	0,21	0,00	0,10	0,00	0,10	0,32
13056004	0,38	0,15	0,51	0,61	0,33	0,40	0,00	0,24	0,33	0,00	0,24	1,80	0,00	0,09	0,16	0,00	0,00	0,09	0,16	0,00	0,00	0,00	0,00	0,24
13056005	0,00	0,38	0,65	0,46	0,31	0,47	0,00	0,13	0,21	0,00	0,11	0,30	0,00	0,47	0,23	0,00	0,00	0,47	0,23	0,00	0,00	0,00	0,00	5,19
13056006	0,00	0,34	0,71	0,58	0,14	0,55	0,00	0,12	0,19	0,00	0,34	0,29	0,00	0,09	0,22	0,00	0,00	0,09	0,22	0,00	0,00	0,00	0,00	0,33
13056007	0,00	0,25	0,79	0,59	0,29	0,52	0,00	0,12	0,35	0,00	0,10	0,67	0,00	0,10	0,20	0,00	0,00	0,10	0,20	0,00	0,00	0,00	0,00	0,31
13056008	1,05	0,15	0,96	0,65	0,14	0,53	0,00	0,12	0,17	0,00	0,24	0,69	0,00	0,09	0,96	0,00	0,00	0,09	0,96	0,00	0,00	1,65	0,96	
13056009	0,00	0,15	0,82	0,38	0,14	1,31	0,00	0,34	0,34	0,00	0,10	0,39	0,00	0,34	0,29	0,00	0,00	0,34	0,29	0,00	0,00	0,00	0,00	0,82
13056010	0,00	0,26	1,05	1,47	0,30	0,84	0,74	0,26	0,36	0,00	0,11	0,25	0,00	0,36	0,18	0,00	0,00	0,36	0,18	0,00	0,00	0,00	0,00	0,28
13056011	1,21	0,17	0,59	0,61	0,36	1,37	0,81	0,26	0,18	0,00	0,36	0,24	0,00	0,10	0,82	0,00	0,00	0,10	0,82	0,00	0,00	0,30	0,51	
13056012	0,00	0,27	1,67	1,03	0,48	1,33	0,00	0,14	0,21	0,00	0,11	0,39	0,00	0,11	0,29	0,00	0,00	0,11	0,29	0,00	0,00	0,00	0,00	0,44
13056013	0,00	0,16	0,71	0,69	0,29	0,39	0,00	0,12	0,34	0,00	0,10	0,25	0,00	0,86	0,18	0,00	0,00	0,86	0,18	0,00	0,00	0,00	0,00	0,53
13056014	1,07	0,49	1,07	0,53	0,12	0,60	0,00	0,10	0,16	0,00	0,09	0,32	0,00	0,73	5,37	0,00	0,00	0,73	5,37	0,00	0,24	0,67	0,67	
13056015	0,00	0,46	0,82	0,50	0,61	0,45	0,00	0,13	0,18	0,00	0,11	0,29	0,00	0,46	0,21	0,00	0,00	0,46	0,21	0,00	0,00	0,00	0,00	0,61
13056016	0,00	0,25	3,08	0,29	0,44	2,46	0,48	0,12	0,17	0,00	0,10	0,72	0,00	0,10	0,53	0,00	0,00	0,10	0,53	0,00	0,00	0,00	0,00	0,82
13056017	0,00	0,34	1,39	0,44	0,56	0,76	0,00	0,12	0,34	0,00	0,24	1,19	0,00	0,09	0,36	0,00	0,00	0,09	0,36	0,00	0,00	0,07	0,55	
13056018	0,00	0,35	0,59	0,48	0,14	0,43	3,82	0,58	0,19	0,00	0,10	0,28	0,00	0,10	0,20	0,00	0,00	0,10	0,20	0,00	0,00	0,00	0,00	0,31
13056019	0,00	0,28	0,45	0,49	0,48	0,33	0,00	0,14	0,19	0,00	0,28	1,80	0,00	1,93	0,16	0,00	0,00	1,93	0,16	0,00	0,00	0,00	0,00	0,24
13056020	0,00	0,27	0,56	1,74	0,32	0,31	0,00	0,14	0,38	0,00	0,38	1,73	0,00	0,11	0,15	0,00	0,00	0,11	0,15	0,00	0,00	0,32	0,43	
13077001	0,00	0,67	0,33	3,28	0,17	1,56	0,55	2,00	2,00	0,00	0,29	0,27	0,00	0,11	0,20	0,00	0,00	0,11	0,20	0,00	0,00	0,00	0,00	0,31
13077002	1,12	0,45	0,64	0,56	0,36	0,49	0,00	0,36	0,36	0,00	1,80	0,64	0,00	0,10	4,45	0,00	0,00	0,10	4,45	0,00	0,00	0,00	0,00	0,30
13077003	0,82	0,18	0,52	0,41	0,16	0,38	0,00	0,65	0,22	0,00	0,39	2,07	0,00	0,49	0,83	0,00	0,00	0,49	0,83	0,00	0,00	0,00	0,00	0,83
13077004	0,00	0,17	0,66	0,38	0,16	1,54	0,00	0,38	0,21	0,00	0,11	2,31	0,00	0,11	0,20	0,00	0,00	0,11	0,20	0,00	0,00	0,00	0,00	0,92
13077005	0,89	0,43	0,88	0,45	0,57	0,64	0,60	0,34	0,17	0,00	0,10	1,01	0,00	0,10	0,31	0,00	0,00	0,10	0,31	0,00	0,00	0,00	0,00	0,88
13077006	3,88	0,44	0,98	0,39	0,35	0,66	0,00	0,58	0,19	0,00	0,25	0,35	0,00	0,10	0,26	0,00	0,00	0,10	0,26	0,00	0,00	0,00	0,00	5,90
13077007	0,00	0,35	0,45	0,37	0,15	0,33	0,00	0,25	0,17	0,00	0,35	0,52	0,00	0,10	0,16	0,00	0,00	0,10	0,16	0,00	0,00	0,00	0,00	0,24
13077008	0,00	0,16	0,64	0,63	0,15	1,59	0,00	0,12	0,19	0,00	0,10	0,19	0,00	0,35	0,14	0,00	0,00	0,35	0,14	0,00	0,00	0,00	0,00	0,40
13098001	0,00	0,17	0,58	0,60	0,31	0,45	0,00	0,13	0,37	0,00	0,11	0,24	0,00	0,10	0,82	0,00	0,00	0,10	0,82	0,00	0,00	0,00	0,00	0,27
13098002	0,94	0,15	0,77	0,47	0,13	0,56	0,00	0,23	0,16	0,00	0,23	0,36	0,00	0,32	0,27	0,00	0,00	0,32	0,27	0,00	0,00	0,00	0,00	0,41

Les valeurs du tableau n°6 sont ensuite intégrées dans une **Classification Ascendante Hiérarchique** (ci-dessous) afin de réunir au sein de groupes, des **IRIS** au **profil sanitaire similaire**. Au total, **5 groupes** d'IRIS se distinguent nettement.

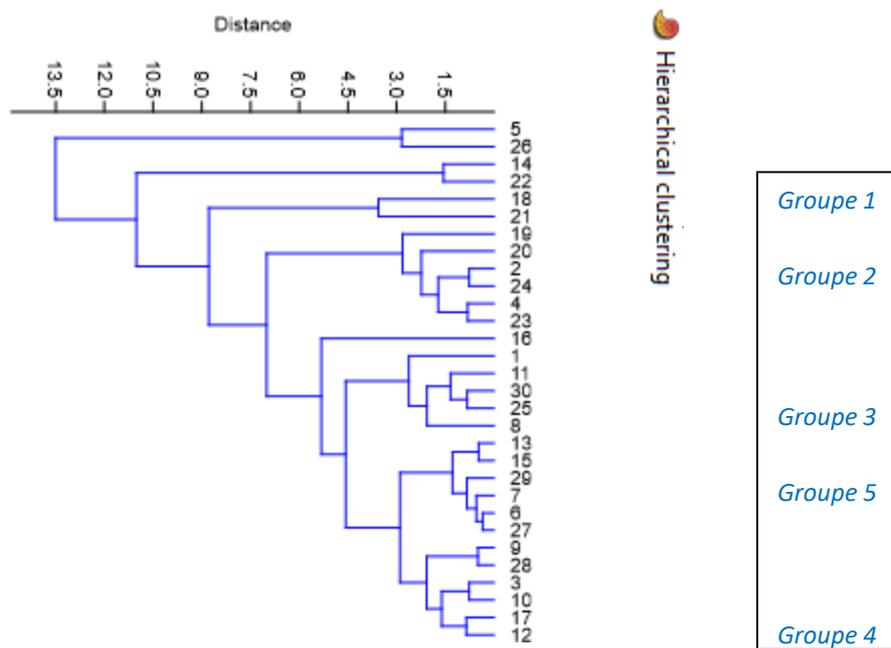


Figure 29. Les résultats de la CAH (Algorithme méthode Ward, distance euclidienne)

Nom IRIS	Numéro IRIS INSEE	Numéro IRIS CAH
Cote Bleue	130560101	1
Plaines Saint-Pierre et Saint-Julien	130560102	2
Plaine de Courouche	130560103	3
Jonquieres Est	130560104	4
Jonquieres Boudeme Font Sarade	130560105	5
Jonquieres Centre	130560106	6
Jonquieres Les Foulettes	130560107	7
Ferrieres et Ile	130560108	8
Les Moulins	130560109	9
Paradis Saint-Roch	130560110	10
Croix-Sainte Saint-Jean La Gafette	130560111	11
Croix-Sainte Mas de Pouane	130560112	12
Croix-Sainte Coudouliere	130560113	13
Saint-Macaire Plan Fossan	130560114	14
Figuerolles Touret de Vallier	130560115	15
La Colline	130560116	16
Canto-Perdrix	130560117	17
Les Capucins Rayettes	130560118	18
Notre Dame Paradis	130560119	19
Lavera	130560120	20
La Leque	130770101	21
Centre	130770102	22
Tassy Est	130770103	23
Tassyouest	130770104	24
Ecartis	130770105	25
Saint-Jean Bergerie	130770106	26
Les Comtes Ouest	130770107	27
Les Comtes Est	130770108	28
Centre	130980101	29
Peripherie	130980102	30

Groupe 1 CAH																				
IRIS	I00119915ans	D011991565ans	I00119965ans	J00119915ans	D011991565ans	I00119965ans	E10E1415ans	I0E1415ans	I0E141565ans	OE1465ans	C3415ans	341565ans	C3465ans	C6715ans	E71565ans	C6765ans	C6415ans	E641565ans	C6465ans	Moyenne
130560105	0.00	0.38	0.65	0.46	0.31	0.47	0.00	0.13	0.21	0.00	0.11	0.30	0.00	0.47	0.23	0.00	0.31	0.519	0.31	0.51
130770106	3.88	0.44	0.88	0.39	0.35	0.66	0.00	0.58	0.19	0.00	0.25	0.35	0.00	0.10	0.26	0.00	0.00	5.90	0.00	0.80
130560114	1.07	0.49	1.07	0.53	0.12	0.60	0.00	0.10	0.16	0.00	0.09	0.32	0.00	0.73	5.37	0.00	0.24	0.67	0.00	0.64
130770102	1.12	0.45	0.64	0.56	0.36	0.49	0.00	0.36	0.00	0.00	1.80	0.64	0.00	0.10	4.45	0.00	0.00	0.30	0.00	0.65
130560118	0.00	0.35	0.59	0.48	0.14	0.43	3.82	0.58	0.19	0.00	0.10	0.28	0.00	0.11	0.20	0.00	0.00	0.31	0.00	0.42
130770101	0.00	0.67	0.93	3.28	0.17	1.56	0.00	2.00	2.00	0.00	0.29	0.27	0.00	0.10	0.20	0.00	0.00	0.31	0.00	0.68
Moyenne	1.01	0.46	0.81	0.95	0.24	0.70	0.73	0.63	0.62	0.00	0.44	0.36	0.00	0.27	1.78	0.00	0.09	2.11	0.00	0.62
Groupe 2 CAH																				
IRIS	I00119915ans	D011991565ans	I00119965ans	J00119915ans	D011991565ans	I00119965ans	E10E1415ans	I0E1415ans	I0E141565ans	OE1465ans	C3415ans	341565ans	C3465ans	C6715ans	E71565ans	C6765ans	C6415ans	E641565ans	C6465ans	Moyenne
130560119	0.00	0.28	0.45	0.49	0.48	0.33	0.00	0.14	0.19	0.00	0.00	0.28	1.80	0.00	1.93	0.00	0.00	0.00	0.24	0.38
130560120	0.00	0.27	0.58	1.74	0.32	0.31	0.00	0.14	0.38	0.00	0.38	1.73	0.00	0.11	0.15	0.00	0.32	0.43	0.00	0.38
130560102	0.00	0.16	0.66	0.44	0.15	0.52	0.00	0.36	0.36	0.00	0.11	2.33	0.00	0.45	0.20	0.00	0.00	0.93	0.00	0.37
130770104	0.00	0.17	0.66	0.38	0.16	1.54	0.00	0.38	0.21	0.00	0.11	2.31	0.00	0.11	0.20	0.00	0.00	0.92	0.00	0.40
130560104	0.98	0.15	0.51	0.61	0.33	0.40	0.00	0.24	0.33	0.00	0.24	1.80	0.00	0.09	0.16	0.00	0.00	0.24	0.00	0.34
130770103	0.82	0.18	0.52	0.41	0.16	0.38	0.00	0.65	0.22	0.00	0.39	2.07	0.00	0.49	0.83	0.00	0.00	0.83	0.00	0.44
Moyenne	0.30	0.20	0.56	0.68	0.27	0.58	0.00	0.32	0.28	0.00	0.25	2.01	0.00	0.53	0.83	0.00	0.00	0.83	0.00	0.48
Groupe 3 CAH																				
IRIS	I00119915ans	D011991565ans	I00119965ans	J00119915ans	D011991565ans	I00119965ans	E10E1415ans	I0E1415ans	I0E141565ans	OE1465ans	C3415ans	341565ans	C3465ans	C6715ans	E71565ans	C6765ans	C6415ans	E641565ans	C6465ans	Moyenne
130560116	0.00	0.25	3.08	0.29	0.44	2.46	0.48	0.12	0.17	0.00	0.10	0.72	0.00	0.10	0.53	0.00	0.00	0.82	0.00	0.53
130560101	0.82	0.59	0.96	0.51	0.35	2.40	0.00	0.25	0.18	0.00	0.10	0.69	0.00	0.35	0.96	0.00	0.30	0.96	0.00	0.52
130560111	1.21	0.17	0.59	0.61	0.36	1.37	0.81	0.26	0.18	0.00	0.36	0.24	0.00	0.10	0.82	0.00	0.30	0.51	0.00	0.44
130980102	0.94	0.15	0.77	0.47	0.13	0.56	0.00	0.23	0.16	0.00	0.23	0.36	0.00	0.32	0.27	0.00	0.00	0.41	0.00	0.28
130770105	0.89	0.43	0.88	0.45	0.57	0.64	0.60	0.34	0.17	0.00	0.10	1.01	0.00	0.10	0.31	0.00	0.00	0.88	0.00	0.41
130560108	1.05	0.15	0.96	0.65	0.14	0.53	0.00	0.12	0.17	0.00	0.24	0.69	0.00	0.09	0.96	0.00	1.65	0.96	0.00	0.46
Moyenne	0.82	0.29	1.21	0.50	0.33	1.33	0.31	0.22	0.17	0.00	0.19	0.62	0.00	0.18	0.84	0.00	0.38	0.76	0.00	0.44
Groupe 4 CAH																				
IRIS	I00119915ans	D011991565ans	I00119965ans	J00119915ans	D011991565ans	I00119965ans	E10E1415ans	I0E1415ans	I0E141565ans	OE1465ans	C3415ans	341565ans	C3465ans	C6715ans	E71565ans	C6765ans	C6415ans	E641565ans	C6465ans	Moyenne
130560109	0.00	0.15	0.82	0.38	0.14	1.31	0.00	0.34	0.34	0.00	0.10	0.39	0.00	0.34	0.29	0.00	0.00	0.82	0.00	0.30
130770108	0.00	0.16	0.64	0.63	0.15	1.59	0.00	0.12	0.19	0.00	0.10	0.19	0.00	0.35	0.14	0.00	0.00	0.40	0.00	0.26
130560103	0.00	0.23	1.20	0.49	0.41	0.96	0.81	0.23	0.16	0.00	0.10	0.28	0.00	0.09	0.14	0.00	0.10	0.32	0.00	0.31
130560110	0.00	0.26	1.05	1.47	0.30	0.84	0.74	0.26	0.36	0.00	0.11	0.25	0.00	0.36	0.18	0.00	0.30	0.28	0.00	0.38
130560117	0.00	0.34	1.39	0.44	0.56	0.76	0.00	0.12	0.34	0.00	0.24	1.19	0.00	0.09	0.36	0.00	0.07	0.55	0.00	0.36
130560112	0.00	0.27	1.67	1.03	0.48	1.33	0.00	0.14	0.21	0.00	0.11	0.39	0.00	0.11	0.29	0.00	0.00	0.44	0.00	0.36
Moyenne	0.00	0.24	1.13	0.74	0.34	1.13	0.26	0.20	0.27	0.00	0.13	0.45	0.00	0.22	0.24	0.00	0.00	0.47	0.00	0.33
Groupe 5 CAH																				
IRIS	I00119915ans	D011991565ans	I00119965ans	J00119915ans	D011991565ans	I00119965ans	E10E1415ans	I0E1415ans	I0E141565ans	OE1465ans	C3415ans	341565ans	C3465ans	C6715ans	E71565ans	C6765ans	C6415ans	E641565ans	C6465ans	Moyenne
130560113	0.00	0.16	0.71	0.69	0.29	0.39	0.00	0.12	0.34	0.00	0.10	0.25	0.00	0.86	0.18	0.00	0.00	0.53	0.00	0.26
130560115	0.00	0.46	0.82	0.50	0.61	0.45	0.00	0.13	0.18	0.00	0.11	0.29	0.00	0.46	0.21	0.00	0.00	0.61	0.00	0.27
130980101	0.00	0.17	0.58	0.60	0.31	0.45	0.00	0.13	0.37	0.00	0.11	0.24	0.00	0.10	0.82	0.00	0.00	0.27	0.00	0.23
130560107	0.00	0.25	0.79	0.59	0.29	0.52	0.00	0.12	0.35	0.00	0.10	0.67	0.00	0.10	0.20	0.00	0.00	0.31	0.00	0.24
130560106	0.00	0.34	0.71	0.58	0.14	0.55	0.00	0.12	0.19	0.00	0.34	0.29	0.00	0.09	0.22	0.00	0.00	0.33	0.00	0.22
130770107	0.00	0.35	0.45	0.37	0.15	0.33	0.00	0.25	0.17	0.00	0.35	0.52	0.00	0.10	0.16	0.00	0.00	0.24	0.00	0.19
Moyenne	0.00	0.29	0.68	0.55	0.30	0.45	0.00	0.15	0.27	0.00	0.18	0.38	0.00	0.28	0.30	0.00	0.00	0.38	0.00	0.23

Tableau 7. Les 5 profils sanitaires issus de la CAH

2.9 Les profils sanitaires des IRIS

Le groupe n° 1 rassemble des IRIS atypiques du point de vue de leur profil sanitaire, il se distingue nettement des autres, puisque nous trouvons là les plus forts taux, pour globalement toutes les pathologies. Certaines valeurs méritent attention :

- **les maladies cardiovasculaires pour les moins de 15 ans (I00I19915ans) pour l'IRIS Saint Jean Bergerie - 130770106 (3,88),**
- **les maladies respiratoires pour les moins de 15 ans (J00J19915ans) pour l'IRIS la Lègue - 130770101 (3,28),**
- **le diabète pour les moins de 15 ans (E10E1415ans) pour l'IRIS Les Capucins-Rayettes 130560118 (3,82),**
- **les cancers des bronches et du poumon chez les 15-65 ans pour l'IRIS Port de Bouc Centre - 130770102 (1,80),**
- **les cancers de la vessie pour les plus de 65 ans (C6765ans) pour les IRIS Macaire-Plande Fossan- 130560114 (5,37) et Port de Bouc Centre 130770102 (4,45),**
- **les cancers du rein pour les plus de 65 ans (C6465ans) pour les IRIS Jonquières Boudème Font Sarade - 130560105 (5,19) et Saint Jean Bergerie - 130770106 (5,90). Notons l'IRIS la Lègue - 130770101 qui détient de fort taux, dans pratiquement, toutes les pathologies.**

. Le second groupe diffère des autres par moins de pathologies de type cardiovasculaires, respiratoires et de diabète³⁷, mais plus de taux de **cancers du poumon chez les plus de 65 ans (C3465ans) (2,01 en moyenne)**. Deux IRIS ressortent particulièrement à ce niveau : l'IRIS **Plaines Saint Julien et Saint Pierre - 130560102 (2,33)** et l'IRIS **Tassy Ouest - 130770104 (2,31)**. En ce qui concerne le **cancer de la vessie (C671565ans)** l'IRIS **Notre Dame Paradis 130560119** détient la valeur la plus élevée pour ce type de cancer (**1,93 pour la tranche des 15-65 ans**). Certaines valeurs se démarquent des autres : le taux de **maladies cardiovasculaires chez les moins de 15 ans** pour l'IRIS **Jonquières Est 130560104 (0,98)**, celui des **maladies respiratoires toujours pour les moins de 15 ans (IRIS Lavéra 130560120 ; 1.74)**.

. Ce qui distingue le groupe n°3 ce sont les forts taux de **maladies cardiovasculaires chez les moins de 15 ans** et surtout les **maladies cardiovasculaires chez les plus de 65 ans (1,21 de moyenne), soit plus que dans le premier groupe (0,81)**.

Il en est de même pour **les maladies respiratoires à partir de 15 ans** dont les taux sont presque deux fois plus élevés que ceux du groupe 1 (**1,33** contre 0,70). **Le diabète chez les moins de 15 ans individualise également ces IRIS. Les cancers des bronches et des poumons pour les plus de 65 ans**, sans atteindre les valeurs du second groupe sont également notables (**0,62 en moyenne**), **tout comme les cancers de la vessie pour la même tranche d'âge** sans dépasser toutefois le groupe 1 (**0,64 en moyenne**).

. Le groupe 4 est en quelque sorte l'inverse du groupe 2, dans le sens où nous observons plus de **pathologies de type cardiovasculaires chez les plus de 65 ans (1.13 en moyenne)**, **respiratoires à la fois chez les plus de 65 ans, mais aussi chez les moins de 15 ans, et moins de taux de cancers**.

. Enfin, le dernier groupe réunit des IRIS où les **taux de pathologies** sont **plus faibles** comparativement aux autres. Il va nous servir de **groupe témoin** pour notre seconde vérification du modèle. En effet, nous pouvons être tentés de prendre un groupe témoin « hors zone » pour lequel on estime qu'il subit moins de pollution. En général, un espace à la « campagne » est sélectionné. Or, les perturbateurs endocriniens auxquels la population peut être exposée *via* les pesticides peuvent induire les mêmes pathologies (diabète), cela peut donc brouiller les comparaisons « cas-témoins » puisque nous risquons de nous retrouver avec des 2 côtés des taux de diabète élevés. Comme le groupe numéro 5 ne ressort pas particulièrement il ne sera pas utilisé pour valider au niveau des IRIS les relations trouvées par le Réseau Bayésien (première vérification « cas », mais il nous servira pour la seconde vérification « témoin »).

Après avoir donc établi le profil sanitaire de chacun des IRIS de notre zone d'étude (différents groupes de la CAH), il est temps à présent de procéder aux étapes de validation des résultats trouvés par le réseau bayésien. Celles-ci vont être effectuées à deux niveaux.

2.10 La validation de la modélisation bayésienne

Le premier niveau va examiner si nous retrouvons bien les relations qui ont émergé à l'échelle globale avec le réseau bayésien (p.77) au sein des IRIS concernés par ces relations (échelle locale). Nous allons nous concentrer que sur les relations avec des pathologies, et nous ne retenons que les plus significatives où $R >$ (ou autour) de +0.50. Par conséquent, les relations entre J00J1991565ans & BkF_p ($R +0,35$), entre DEC_med14 et HCI ($R -0.65$), ainsi qu'entre P14_PMEN_ANEM10P & plusieurs variables ne seront pas prises en compte dans cette étape de validation du modèle.

Nous avons vu précédemment que dans le profil sanitaire n°1 de la CAH, des IRIS se différenciaient par un nombre plus élevé de cancers des bronches et des poumons chez les 15-65 ans (IRIS PdB Centre - 130770102), et que pour cette pathologie et cette tranche d'âge une relation avec HF (acide fluorhydrique) avait été trouvée, tandis que d'autres IRIS se démarquaient par des cancers de la vessie chez les plus de 65 ans pour lesquels un lien avec plusieurs PCBs avait cette fois été identifié (IRIS Saint Macaire Plan Fossan - 130560114). Afin de valider le modèle bayésien nous allons voir si ces relations se retrouvent au niveau des analyses de sensibilité (par ex. un niveau important dans l'IRIS concerné, du polluant incriminé). Par exemple pour l'IRIS 130770102 (Port de Bouc centre), regardons si le niveau d'acide fluorhydrique (HF) est important dans l'analyse de sensibilité de cet IRIS (Figure n°30, ci-après) :

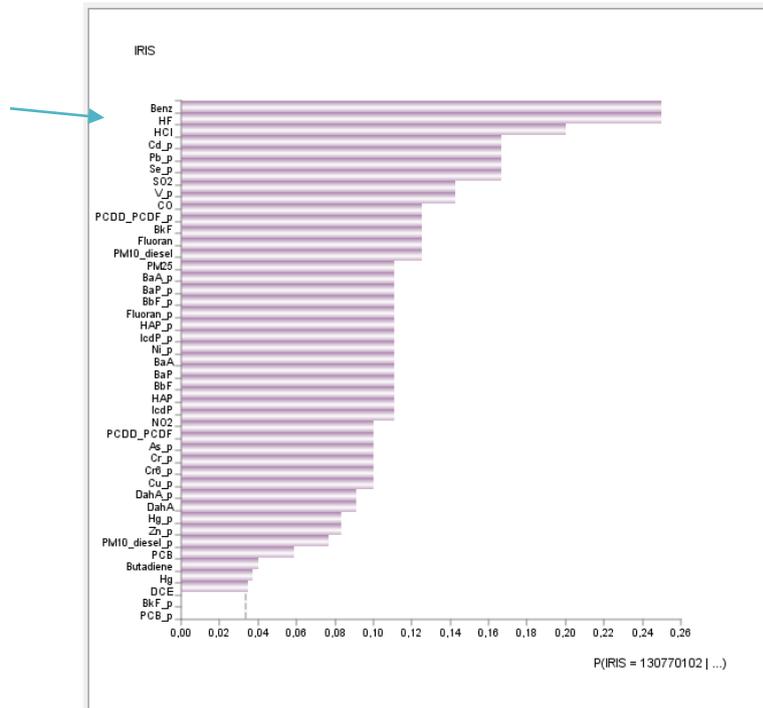


Figure 30. Analyse de sensibilité de l'IRIS 130770102

Avec une probabilité de 0.26, HF est un des principaux polluants qui caractérise cet IRIS, la relation est donc « validée »³⁸.

Regardons maintenant ce qu'il en est pour l'IRIS 130560114 (Plan Fossan) et sa relation avec plusieurs PCBs. Les PCBs incriminés (CB114, CB20 et CB157) ressortent parmi les principaux xénobiotiques présents au sein de cet IRIS.

³⁸ Même si nous sommes conscients qu'il nous faudrait plus d'IRIS

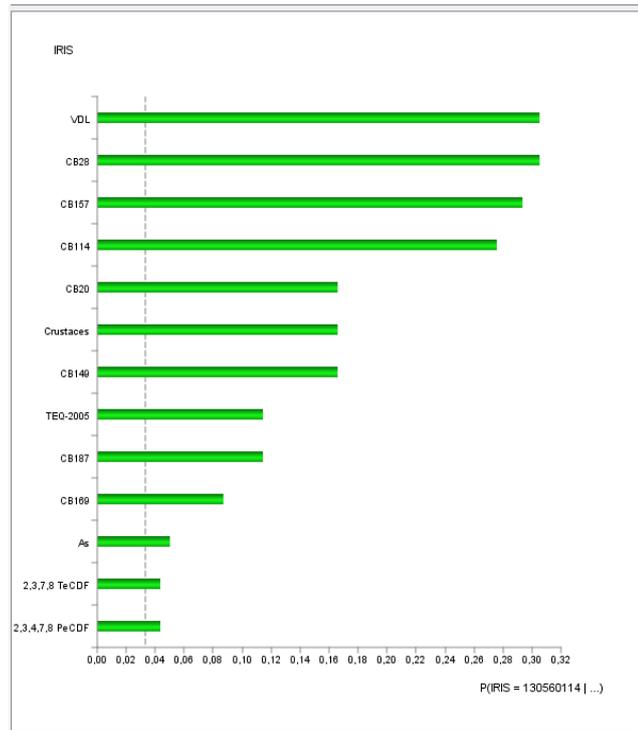


Figure 31. Analyse de sensibilité de l'IRIS 130560114

Le tableau n°8 synthétise toutes les vérifications faites entre les relations retrouvées au sein des IRIS³⁹. Il ressort qu'hormis celle qui lie, dans le profil sanitaire 2 de la CAH, HF et l'IRIS 130560102, où HF ne ressort qu'au niveau du 4^{ème} groupe de polluants⁴⁰ dans l'analyse de sensibilité correspondante, toutes les relations calculées par le réseau bayésien sont validées (en bleu-vert clair).

³⁹ Les relevés lichéniques ne couvrant pas tous les IRIS l'étude des relations entre C3465ans ne peut se faire que pour le SO2 et non avec CB187, TEQ2005, 2.3.7.8 TeCDF et 2.3.4.7.8 PeCDF et As

⁴⁰ Notons toutefois que la corrélation est à la limite de la significativité (R +0.46)

	. relations entre J00J19965ans & CB169 (R +0,78)	. relations entre E10E1465 et Cd_p (R +0,52)	
	. relations entre J00J1991565ans & BKF_p (R +0,35)	. relations entre I00I19965ans & P14_NSCOL15P_DIPLOM (R +0,63)	
	. relations entre C3465ans & plusieurs variables	. relations entre DEC_MED14 et HCl (R -0,55)	
	. relations entre C341565ans & HF (R +0,46)	. relations entre P14_PMEN_ANEM10P & plusieurs variables	
	. relations entre C6765ans & plusieurs variables		
Groupe 1 CAH		. relations entre C341565ans & HF (R +0,46)	. relations entre C6765ans & plusieurs variables : (CB114, CB20, CB157)
	IRIS	0,26 (polluant le + fort)	CB114, CB20, et CB157 (second groupe de polluants)
	130770102		
	130560114		
Groupe 2 CAH		. relations entre C3465ans & plusieurs variables* (SO2) (R +0,62)	. relations entre C341565ans & HF (R +0,46)
	IRIS	0,55 (second polluant le plus élevé cf. analyse sensibilité)	0,18 (4ème groupe de polluants)
	130560102	0,50 (1er polluant le plus élevé cf. analyse sensibilité)	0,25 (second polluant après SO2)
	130770104	<i>*bien que C3465 soit reliée à plusieurs variables, nous ne pouvons vérifier le modèle que sur la variable SO2, car nous ne disposons pas de relevés lichéniques pour ces IRIS</i>	
Groupe 3 CAH		. relations entre J00J19965ans & CB169 (R +0,78)	. relations entre I00I19965ans & P14_NSCOL15P_DIPLOM (R +0,63)
	IRIS	pas de relevé lichénique	0,13
	130560116	0,31*	0,25
	130560101	0,16*	0,12
	130560111	0,09*	0,25
	130560108	0,32 mais pour IRIS proche 13090101	0,11
	130980102	pas de relevé lichénique	0,13
	130770105	<i>*attention ce n'est pas parce que la valeur est faible qu'elle est sans effet, de plus ; par exemple 0,16 correspond à 16% + toxicité CB169</i>	
Groupe 5 CAH		. relations entre J00J19965ans & CB169 (R +0,78)	. relations entre I00I19965ans & P14_NSCOL15P_DIPLOM
	IRIS	pas de relevé lichénique	0,11
	130560109	pas de relevé lichénique	0,12
	130070108	pas de relevé lichénique	0,11
	130560103	0,16 (2nd groupe)	0,11
	130560110	0,085 (max polluant 0,12, 2nd groupe)	0,11
	130560117	pas de relevé lichénique	0,11
	130560112	pas de relevé lichénique	0,13

Tableau 8. Les relations modélisées retrouvées dans les analyses de sensibilité des IRIS concernés

Enfin, la seconde vérification correspond à l'étude du groupe « témoin », le n°5 de la CAH, qui se caractérise par de faibles taux de pathologies. **Nous allons donc voir, si ces faibles taux de pathologies s'expliquent, cette fois, par de faibles taux des variables incriminées** (Tableau n°9).

Là encore, le modèle est validé, les valeurs des polluants pour lesquels une relation avec une pathologie a été trouvée sont faibles dans ces IRIS, à l'exception toutefois au niveau de l'IRIS Figuerolles Touret de Vallier (130560115) entre le diabète chez les plus de 65 ans (E101465 ans), et le Cadmium (phase particulaire), où la valeur de Cd_p est plus élevée comparativement aux autres IRIS qui composent ce groupe, sans que cela ne se traduise pour autant par un surplus de diabète dans cette tranche d'âge.

Cela peut s'expliquer par le fait que le diabète chez les plus de 65 ans est une pathologie liée plus au vieillissement qu'à peut-être une exposition à un polluant, ou bien que la socio-économie compense dans cet IRIS, rappelons quand même que R est à la limite de la significativité (+0.52).

En ce qui concerne la relation entre les maladies respiratoires chez les plus de 65 ans (J00J19965ans) et CB169, la relation qui est pourtant forte (R +0.78) ne se vérifie pas sur l'IRIS Saint Mitre Centre - 130980101, la valeur de CB 169 est élevée mais elle ne se traduit pas par un taux de J00J19965ans haut (effet d'une socio-économie favorisée dans cet IRIS qui compenserait l'exposition par notamment un meilleur accès aux soins ?).

Au stade où nous sommes, il est temps d'aller encore plus loin dans la recherche des facteurs causaux des profils sanitaires mis en évidence au sein de ce territoire du Pays de Martigues.

Tableau 9. Des relations plus faibles dans les IRIS du groupe témoin

	130560113	130560115		130980101	130560107	130560106	130770107
Interactions entre I00119965ans & P14_NSCOLI5P_DIPLOM (R+0,63)	0,11	0,11		0,12	0,12	0,12	0,13
Interactions entre J0011991565ans & BkF_p (R + 0,35)	0	0		0	0	0	0
Interactions entre J00119965ans & CB169 (R + 0,78)	No lichens	No lichens		0,30	No lichens	No lichens	No lichens
Interactions entre E10E1465ans et Cd_p (R+0,52)	0,09	0,20		0,17	0,08	0,13	0,10
Interactions entre C341565ans & HF (R +0,46)	0,09	0,17		0,17	0,08	0,09	0,26
Interactions entre C3465ans & plusieurs variables	S02 0,03	S02 0,34		TEQ2005:0,17 CB187:0,12 2,3,4,7,8 PeCDF:0,03 As:0,04 S02:0,22 2,3,7,8 TeCDF:0,04	S02 : 0,04	S02 : 0,04	S02 : 0,16
Interactions entre C6765ans & plusieurs variables	No lichens	No lichens		Crustacés:0,160 VDL:0,170 CB149:0,318 CB114:0,040 CB157:0,05 CB20:0,075 CB28:0,170	No lichens	No lichens	No lichens

2.11 L'introduction du spatial dans les réseaux bayésiens

Les études menées en santé environnementale souffrent la plupart du temps de critiques à leur encontre au regard de la difficulté à remonter aux facteurs causaux des pathologies. Un autre argument souvent entendu est le caractère multifactoriel qui prévaut à la santé d'un individu. Néanmoins, en prenant en compte le plus de variables possibles en termes de polluants, de socio-économie, qui peuvent potentiellement influencer la santé des habitants, et ce, à une échelle fine (la plus proche de celle de la majorité des individus situés en zones urbaines), sur la base de pathologies avérées, et en utilisant un formalisme mathématique approprié, fondé sur des probabilités conditionnelles et le graphe causal, nous pouvons alors présupposer d'un certain nombre de facteurs « causaux ».

Des analyses de sensibilité ont été à présent menées sur chacun des 30 IRIS en prenant simultanément en compte **tous les polluants atmosphériques** fournis par AtmoSud (44 polluants), les 17 variables issues des relevés de lichens précédemment **repérées comme ayant un rôle sur les pathologies**⁴¹ et les **7 variables de nature socio-économique, soit un total à présent de 68 variables** (Figure n°22). En effet, il peut être intéressant de distinguer les IRIS où les facteurs environnementaux et/ou socio-économiques sont influents. L'analyse de sensibilité globale menée sur tous les IRIS nous indique là encore une forte variabilité entre ces derniers. Nous pouvons observer **3 grandes catégories** : parfois un IRIS est dominé par une famille de polluants tels que des HAP ou des PCB (**Cocktail polluants HAP ; Cocktail polluants PCB**), parfois par des polluants qui n'appartiennent ni à l'une ni à l'autre des 2 grandes familles de polluants que nous venons d'évoquer (**Cocktail polluants Autres**). De plus, à y regarder de plus près en repérant les variables qui sont les plus influentes dans chaque IRIS, **la socio-économie s'invite parfois au milieu de variables environnementales soit en amplifiant leurs effets** dans le cas d'une socio-économie de type défavorisée, **soit au contraire en les compensant** dans le cas d'une meilleure socio-économie.

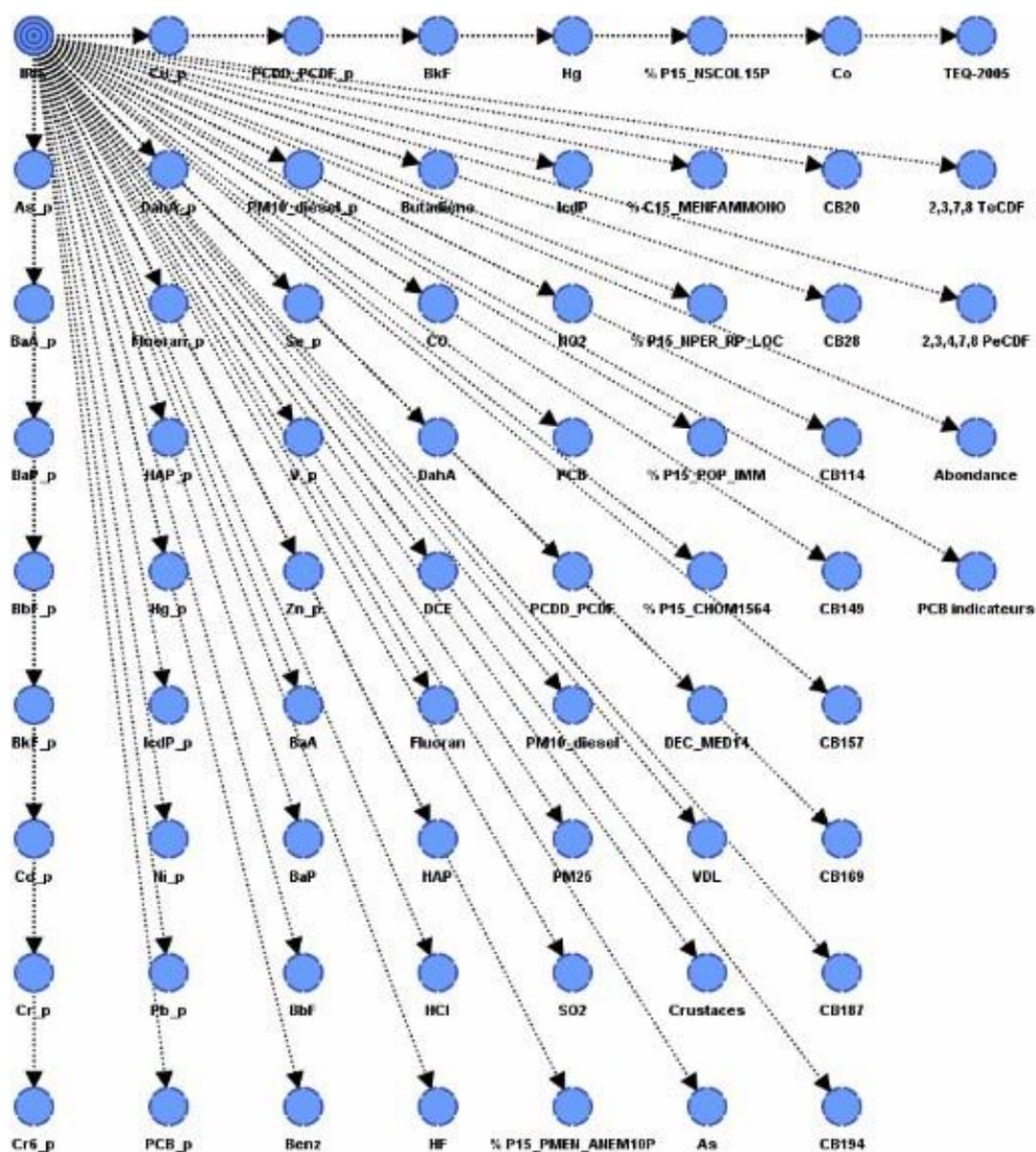


Figure 32. Traitement supervisé sur la variable cible IRIS pour les polluants, la socio-économie et les relevés lichéniques

⁴¹ CB169, TEQ-2005, CB187, 2,3,4,7,8 PeCDF, As, 2,3,7,8 TeCDF, crustacés, VDL, CB149, CB114, CB157, CB20, CB28, Co, CB194, Abondance, PCB indicateurs

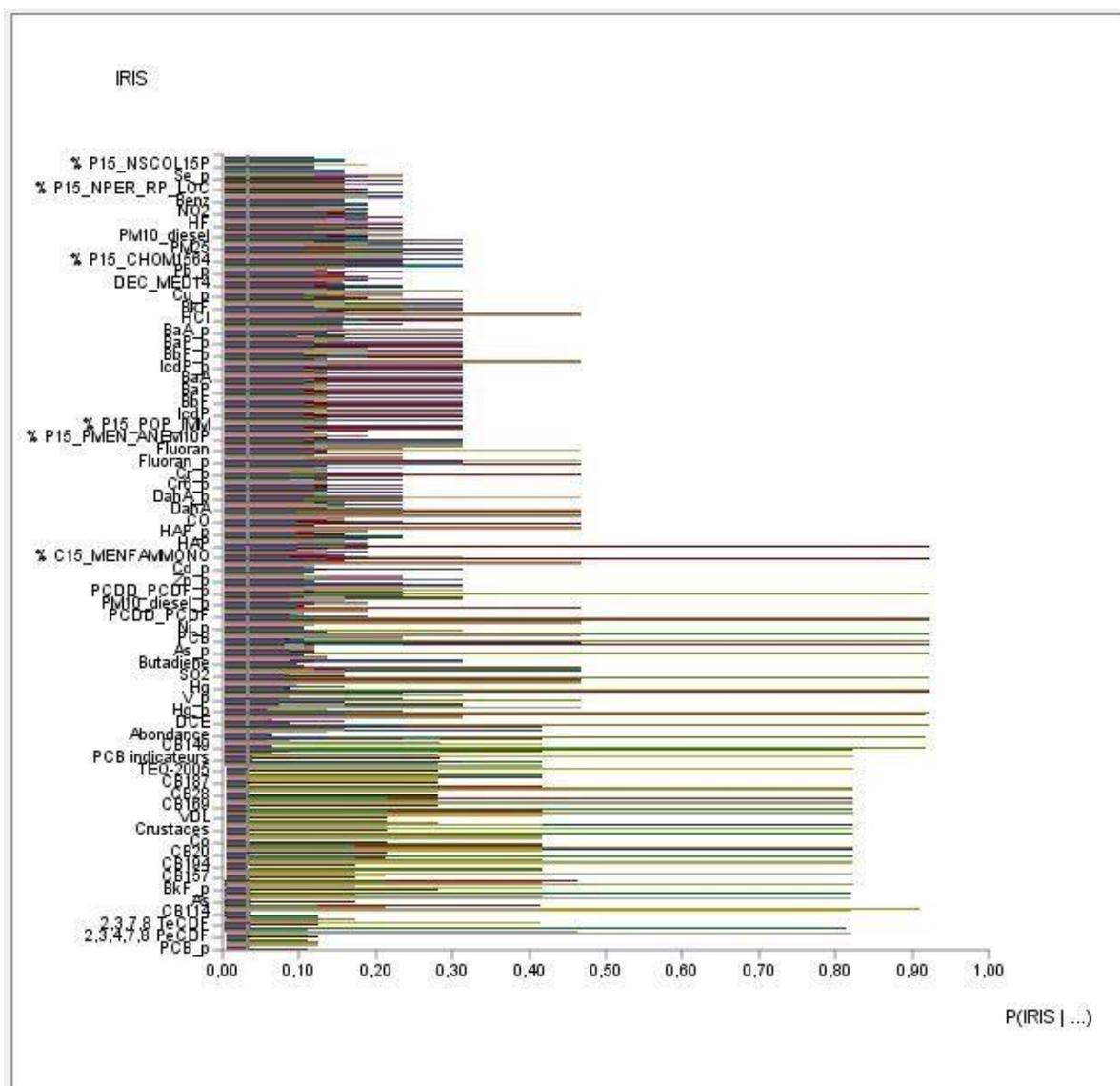


Figure 33. L'ensemble des analyses de sensibilité sur les 68 variables et pour tous les IRIS

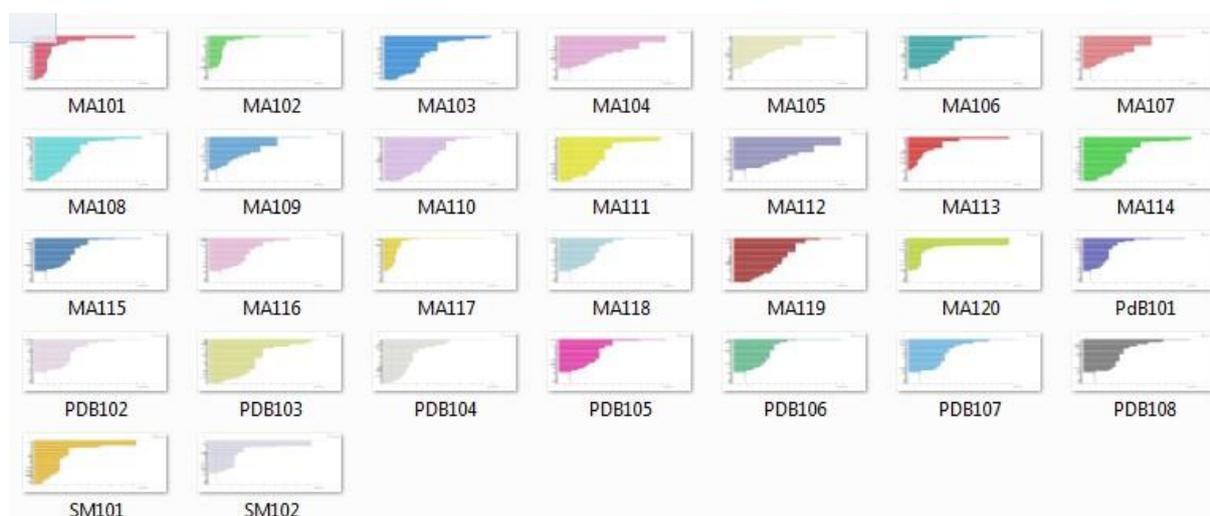


Figure 34. Le profil des analyses de sensibilité sur les 68 variables pour chaque IRIS

Là encore, nous pouvons voir sur les figures précédentes que la situation est très différente dans les IRIS suivant l'importance des variables présentes en leur sein.

Puis, pour chaque analyse de sensibilité, nous avons relevé les variables probabilisées les plus importantes^{42, 43} que nous avons reportées dans le tableau ci-après, tandis qu'un code couleur spécifie leur appartenance au même groupe de la CAH (Tableau n°7). Par exemple, les IRIS en violet correspondent au premier groupe de la CAH, où nous avons un plus grand nombre de séjours dans toutes les pathologies.

Tableau 10. Code couleur et indication des pathologies relevées dans chaque groupe de la CAH

Groupe 1 CAH : Pathologies ++
Groupe 2 CAH : Cancers poumon (> 15 ans) & Cancers vessie (15-65 ans)
Groupe 3 CAH : Maladies cardiovasculaires (<15 ans, et >65 ans) & Maladies respiratoires (<15 ans) & Diabète (<15 ans)
Groupe 4 CAH : Maladies cardiovasculaires (>65 ans) & Maladies respiratoires (<15 ans, et >65 ans)
Groupe 5 CAH : Groupe témoin Pathologies

⁴² Sans toutefois dépasser le seuil de 0.20 de probabilité, sauf exception lorsque, par exemple, un polluant est connu pour être impactant (de type PCB ou autres à une valeur de 0.18)

⁴³ Les variables barrées ne jouent pas

Tableau 11 . Les probabilités de prédominance des facteurs environnementaux et socio-économiques dans chaque IRIS

IRIS	Probabilités de prédominance des facteurs environnementaux et/ou socio-économiques	10 ans	Nscol	Monol	Loc	Imm	Chom	Rev
130560105	Daha_p_0.46, Daha, Fluoran_p, Fluoran, MONO, Zn_p, BaA_p, BbF_p, IcdP_p	53	43	23	69	15	15	13980
130560104	Butadiène 0.31, BaA, BaP, BaA_p, BaP_p, BbF_p, IcdP_p, BkF, PCB, Zn_p, Daha_p, Daha_p, Cu_p, Fluoran_p	50	31	15	35	9	10	25576
130560109	Butadiène 0.47, Cu_p, BaA_p, BaP_p, BbF_p, IcdP_p, BaA, BaP, BbF, IcdP, %P15_CHOM1564, BkF, PM10_diesel, PM25	56	40	36	95	12	25	15004
130560107	Butadiène 0.46, %P15_PMEN_ANEM10P, Cu_p, BaA_p, BaP_p, BbF_p, BaA, BaP, BbF, IcdP_p, IcdP	36	32	18	47	8	14	21014
130560113	PM10_diesel_p_0.95, HAP_p, Daha_p, Daha_p, Fluoran_p, Fluoran, Zn_p, BaA_p	55	26	13	21	5	16	23684
130560117	BkF_p_0.91, % P15_POP_IMM, % P15_PMEN_ANEM10P, DEC_MED14	45	20	16	36	7	14	19500
130560114	Co_0.29, VDL, CB28, Abondance, CB157, CB114	60	33	12	39	9	10	23824
130560102	PCB_0.92, Hg, SO2, PM10_diesel_p, DCE, Hg_p, V_p, CO	60	23	10	20	3	9	26060
130980102	PCB_0.46, Zn_p, Co, Cr_p, Cr6_p, PCDD_PCDF_p, HCl, Pb_p, BkF_p, SO2, Cd_p	57	28	9	17	4	11	25544
130980101	PCB_0.46, Co, Zn_p, PCDD_PCDF_p, Cr_p, Cr6_p, HCl, Pb_p, SO2, CB_169, CB149	59	28	13	26	5	11	25544
130560106	%P15_PMEN_ANEM10P_0.31, PCB, DEC_MED14, NO2	35	34	19	64	13	22	19244
130560108	%P15_PMEN_ANEM10P_0.31, % C15_MENFAMMONO, %P15_CHOM1564, PCB, PM10_diesel, PM2,5	32	39	33	76	14	24	15951
130560103	Butadiène 0.31, DCE, Hg_p, CB20, Crustacés, V_p, Hg, Benz	66	39	11	42	3	17	23710
130770103	Butadiène 0.31, % P15_CHOM1564, CB20, CB187, PCB indicateurs, TEQ-2005	64	50	37	86	14	26	13030
130560101	V_p_0.92, As_p, Ni_p, DCE, Hg, SO2, PCDD_PCDF, BkF_p, PM10_diesel_p, Hg_p, Co, CB169, As, CO, HF	49	29	12	26	3	12	24230
130560120	V_p_0.94, Hg, Butadiène, As_p, PCDD_PCDF, Ni_p, Cd_p, PCDD_PCDF_p, DCE, Hg_p, PCB_p, SO2, PM10_diesel_p, Cr_p, Cr6_p	51	21	7	13	4	11	ND
130770104	SO2_0.46, Co, Crustacés, CB187, TEQ-2005, As, CB194, 2.3,7,8 TeCDF, 2.3,4,7,8 PeCDF, HF	67	49	20	49	12	29	17380
130560111	Zn_p_0.31, CB20, VDL, CB28, CB149, PCB indicateurs, As	60	34	8	13	11	17	24974
130770108	DCE_0.31, V_p, Hg, HF, DEC_MED14	61	45	18	34	14	14	19258
130560115	SO2_0.31, CO	62	23	11	34	4	11	24880
130770106	%P15_POP_IMM_0.31, DEC_MED14, HF, %P15_CHOM1564,	53	24	9	9	6	14	23436
130770102	% C15_MENFAMMONO_0.31, % P15_PMEN_ANEM10, HF, Benz	42	42	25	74	10	21	16688
130560119	%P15_PMEN_ANEM10P_0.23, NO2, % P15_NSCOL15P	43	42	17	68	11	19	17574
130560118	% C15_MENFAMMONO_0.31, %P15_PMEN_ANEM10P, NO2	49	38	22	62	10	20	15686
130560110	% P15_NPER_RP_LOC_0.23, % P15_PMEN_ANEM10P, NO2	48	36	20	55	11	17	17414
130560116	% P15_POP_IMM_0.31, % P15_PMEN_ANEM10P, % C15_MENFAMMONO, % P15_CHOM1564, Cd_p, % P15_NSCOL15P	46	44	41	94	18	32	11642
130770105	% P15_POP_IMM_0.31, % P15_NPER_RP_LOC, Hg_0.17, CO_0.17, Cu_p_0.17, HF_0.17	61	36	16	47	19	18	16782
130770107	% P15_POP_IMM_0.31, Hg, % C15_MENFAMMONO, HF, HCl, Benz, % P15_NSCOL15P	60	54	38	89	18	22	13036
130560112	MONO_0.31, Cu_p, BaA_p, BaP_p, BbF_p, IcdP_p, BaA, BaP, BbF, IcdP, BkF	54	51	31	93	16	30	9682
130770101	PCDD_PCDF_0.47, % C15_MENFAMMONO, HF, % P15_CHOM1564, Benz	58	54	35	80	13	33	11328

Ainsi, à titre d'exemple, pour l'IRIS 130560105 (Boudème) nous avons relevé 10 polluants qui appartiennent à la famille des HAP, et dont la probabilité est comprise entre 1 (très important dans l'IRIS) à 0.2 (seuil d'arrêt des relevés, car la variable joue un moins grand rôle). La première variable qui sort pour cet IRIS est le DahA_p, il est suivi d'un chiffre en l'occurrence ici **0.46** qui indique son niveau de probabilité dans cet IRIS, soit 46% (Tableau n°12, ci-après). Ces relevés ont permis de dégager 3 grandes familles d'IRIS : des IRIS où les polluants étaient nombreux, et formaient ce que l'on appelle communément un **cocktail de polluants** qui pouvait contenir soit une majorité de **HAP**, soit une majorité de **PCB**, soit des polluants qui n'entraient ni dans la famille des HAP, ni dans celle des PCB et que nous avons qualifiés d'« **Autres** », mais dont la potentielle action sur la santé mérite néanmoins que l'on s'y attarde.

Au total nous avons donc 3 profils d'IRIS :

1. Cocktail polluants HAP (7 IRIS)
2. Cocktail polluants PCB (8 IRIS)
3. Cocktail polluants « Autres » (15 IRIS)

Recensés dans le tableau ci-dessous :

Tableau 12. Le profil attribué à chacun des IRIS

IRIS	Nom	Profil	Variable polluant la + importante	Valeur Variable	Variable socio-eco la + importante	Valeur Variable
130560105	Boudeme	Cocktail polluants HAP	DahA_p	0,46	MONO	0,32
130560104	Jonquières Est	Cocktail polluants HAP	Butadiène	0,31	-	0,00
130560109	Les Moulins	Cocktail polluants HAP	Butadiène	0,47	CHOM	0,31
130560107	Jonquieres Foulettes	Cocktail polluants HAP	Butadiène	0,46	-	0,00
130560113	Coudouliere	Cocktail polluants HAP	HAP_p	0,95	-	0,00
130560117	Canto Perdrix	Cocktail polluants HAP	BkF_p	0,91	IMMIG	0,33
130560114	Plan Fossan	Cocktail polluants PCB	CB28 Tres toxique	0,30	-	0,00
130560102	Plaines St Pierre & St Julien	Cocktail polluants PCB	PCB	0,92	-	0,00
130980102	Saint Mitre périphérie	Cocktail polluants PCB	PCB	0,46	-	0,00
130980101	Saint Mitre centre	Cocktail polluants PCB	PCB	0,46	-	0,00
130560106	Jonquieres Centre	Cocktail polluants PCB	PCB	0,24	-	0,00
130560108	Ferrières et l'Ile	Cocktail polluants PCB	PCB	0,23	MONO	0,32
130560103	Plaine de Courouche	Cocktail polluants PCB	CB20	0,29	-	0,00
130770103	Tassy Est	Cocktail polluants PCB	CB20	0,28	CHOM	0,31
130560101	Cote bleue	Cocktail polluants Autres	V_p	0,92	-	0,00
130560120	Lavera	Cocktail polluants Autres	V_p	0,94	-	0,00
130770104	Tassy Ouest	Cocktail polluants Autres	SO ₂	0,46	-	0,00
130560111	Croix Sainte	Cocktail polluants Autres	Zn_p	0,31	-	0,00
130770108	Les Comtes Est	Cocktail polluants Autres	V_p	0,24	-	0,00
130560115	Figuerolles	Cocktail polluants Autres	SO ₂	0,31	-	0,00
130770106	Saint Jean Bergerie	Cocktail polluants Autres	HF	0,16	-	0,00
130770102	PdB Centre	Cocktail polluants Autres	HF	0,23	MONO	0,31
130560119	Notre Dame Paradis	Cocktail polluants Autres	NO ₂	0,19	NSCOL	0,19
130560118	Rayettes	Cocktail polluants Autres	NO ₂	0,19	MONO	0,31
130560110	Paradis Saint Roch	Cocktail polluants Autres	NO ₂	0,19	LOCATAIRES	0,23
130560116	La Colline	Cocktail polluants Autres	Cd_p	0,19	IMMIG	0,31
130770105	Ecarts	Cocktail polluants Autres	Hg	0,17	IMMIG	0,31
130770107	Les Comtes Ouest	Cocktail polluants Autres	Hg	0,23	IMMIG	0,31
130560112	Mas de Pouane	Cocktail polluants Autres	Cu_p	0,31	MONO	0,31
130770101	La Lecque	Cocktail polluants Autres	PCDD_PCDF	0,47	MONO	0,33

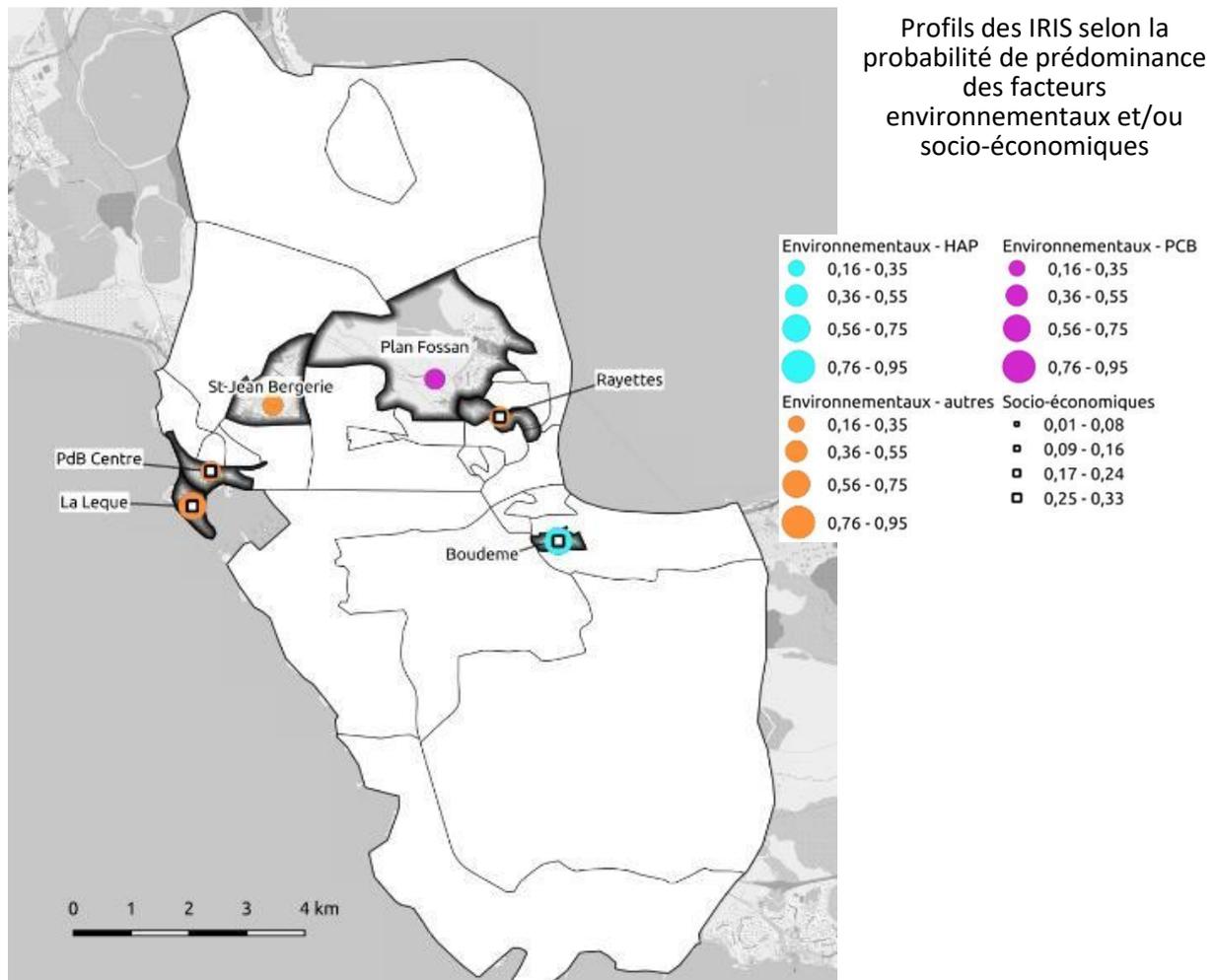
Nous avons vu que pour **la première variable qui sort correspond une valeur de probabilité repérée dans les analyses de sensibilité, elle indique en quelque sorte son niveau d'importance dans l'IRIS. Parfois, cette importance a été nuancée en fonction des autres variables qui l'entouraient.** En effet, si nous prenons l'exemple de l'IRIS de la Plaine de Courouche (130560103), la 1ère variable qui sort dans l'analyse de sensibilité est le Butadiène à 0.31, mais, nous avons préféré retenir CB20 qui est juste derrière, à 0.29, car il s'agit d'un PCB potentiellement plus impactant pour la santé. De même, nous avons indiqué lorsque cela était pertinent la variable socio-économique qui arrivait en premier, et qui était intégrée dans le cocktail de polluants afin d'avoir une idée de ce contexte.

Les HAP regroupent plusieurs polluants qui sont émis en même temps, leur nombre varie ici du simple (6) au double (12), avec une répercussion sur les pathologies. Plus on a de HAP et plus la pathologie est sévère. Cela est dû au caractère cancérigène de chacun d'eux (BaP, BaA, BbF, BkF, DahA, Fluoranthène, IcdP). La socio-économie favorable de l'IRIS Jonquières Foulettes (130560107) semble compenser l'exposition de ses habitants (revenus supérieurs à la moyenne), de même que pour l'IRIS Coudoulière (130560113), mais dans une moindre mesure, car les HAP sont moins nombreux dans cet IRIS.

En ce qui concerne la catégorie Cocktail de polluants de type PCB leur impact sanitaire se traduit essentiellement, et même à de faibles doses (en raison de leur toxicité) par, par exemple, des cancers de la vessie (IRIS Plan de Fossan - 130560114) chez les plus de 65 ans.

Enfin, la dernière catégorie « Autres » correspond à des polluants qui ne peuvent pas être classés dans les catégories précédentes. Nous trouvons là des polluants « traditionnels » comme le NO₂, ou bien encore le SO₂, mais également l'acide fluorhydrique (HF) dont l'action est déjà apparue au niveau du monitoring (p.82), sans oublier d'autres éléments potentiellement impactants comme : V_p, Hg, Zn_p, Cd_p, Cu_p et les PCDD_PCDF.

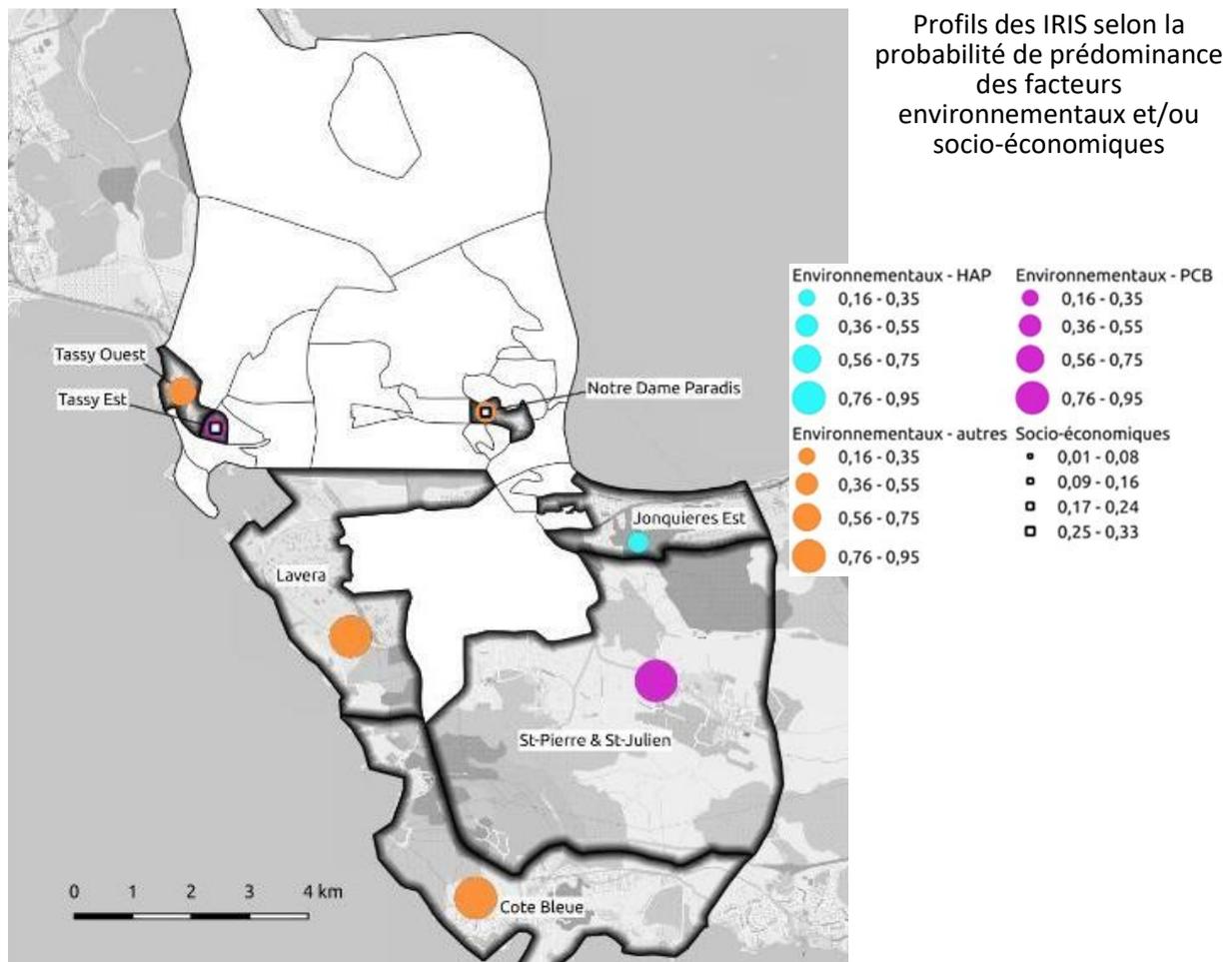
*Regardons à présent, les **pathologies retrouvées dans les IRIS** et mettons-les en relation avec les facteurs les plus importants relevés en leur sein (tableau n°11).*



Carte 5. Groupe 1 CAH, IRIS caractérisés par une prévalence accrue des hospitalisations pour **l'ensemble des pathologies étudiées**

En ce qui concerne le groupe 1 de la CAH, il est exposé à des polluants particulièrement forts de type PCDD_PCDF à La Lègue, **CB28**⁴⁴ à Plan Fossan, HF à PdB Centre, ainsi qu'à St-Jean Bergerie, et au Dibenzo[a, h]anthracène (DahA_p) à Boudème. Dans 4 des 6 IRIS la socio-économie (carrés blancs) se cumule aux polluants et correspond aux familles monoparentales. C'est d'ailleurs la variable prédominante pour l'IRIS des Rayettes (0,31).

⁴⁴ Lorsque les PCBs ont été mesurés dans les lichens ils sont écrits en vert, et en violet lorsqu'ils ont été mesurés dans l'atmosphère

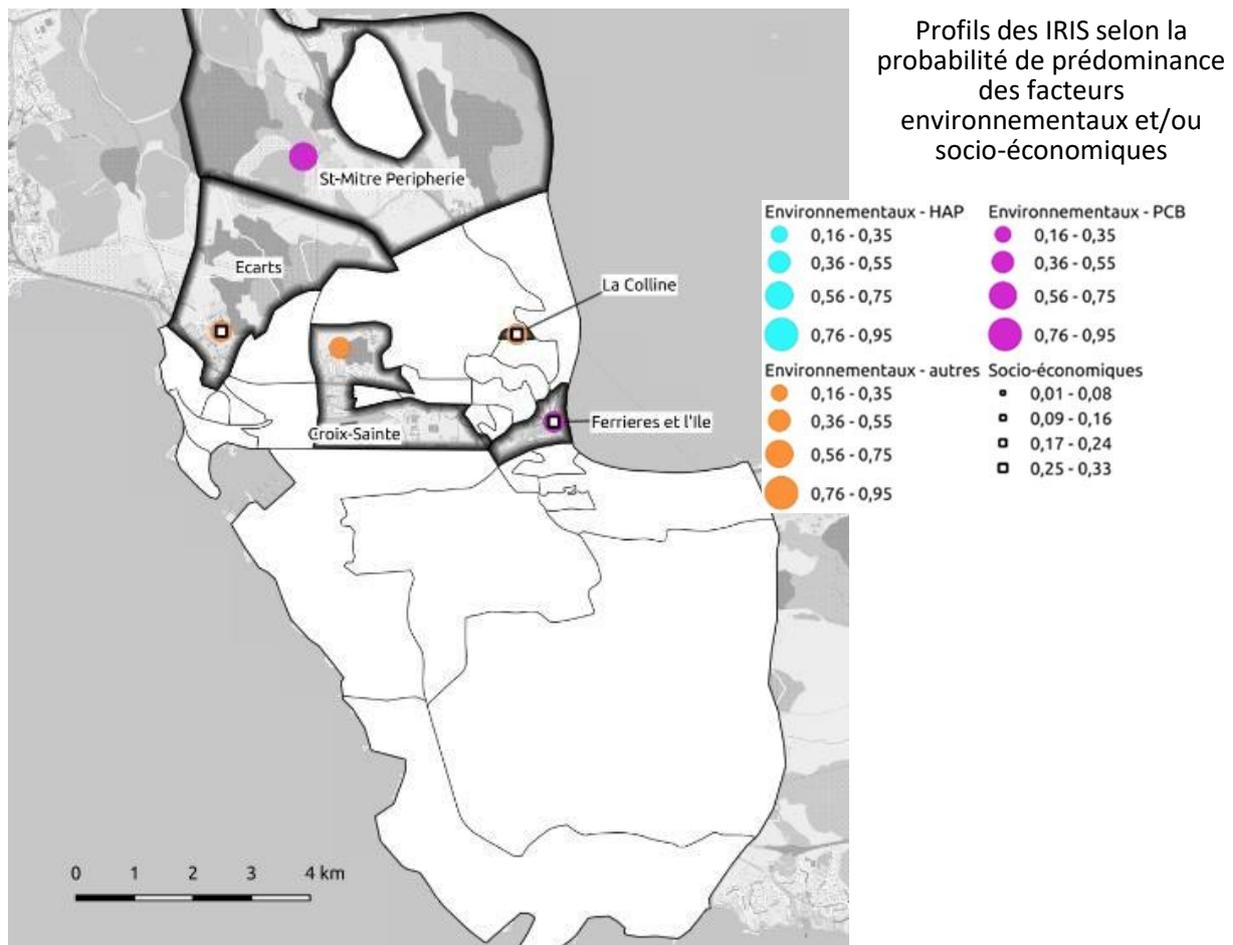


Carte 6. Groupe 2 CAH, IRIS caractérisés par une prévalence accrue des hospitalisations liées aux **cancers du poumon (> 15 ans) et de la vessie (15-65 ans)**

Les IRIS du groupe 2 sont pour leur part exposés aux PCBs (Tassy Est **CB20**, St-Pierre & St-Julien **PCB**), au SO₂ (Tassy Ouest), et au Vanadium⁴⁵ dans sa phase particulaire, et à des niveaux bien plus élevés que dans les autres IRIS, pour les IRIS de Lavéra et de la Côte bleue. Tandis que l'IRIS Jonquières Est, est exposé à plus de 11 HAPs différents, l'IRIS Notre Dame Paradis à du NO₂, mais dans ce dernier cas, comme à Tassy Est, la socio-économie semble également jouer un rôle délétère (non scolarisés à 15 ans et chômage).

La nature des polluants (PCBs, Vanadium), leur multiplicité (11 HAPs différents), peuvent expliquer les pathologies rencontrées dans les IRIS de ce second groupe de la CAH.

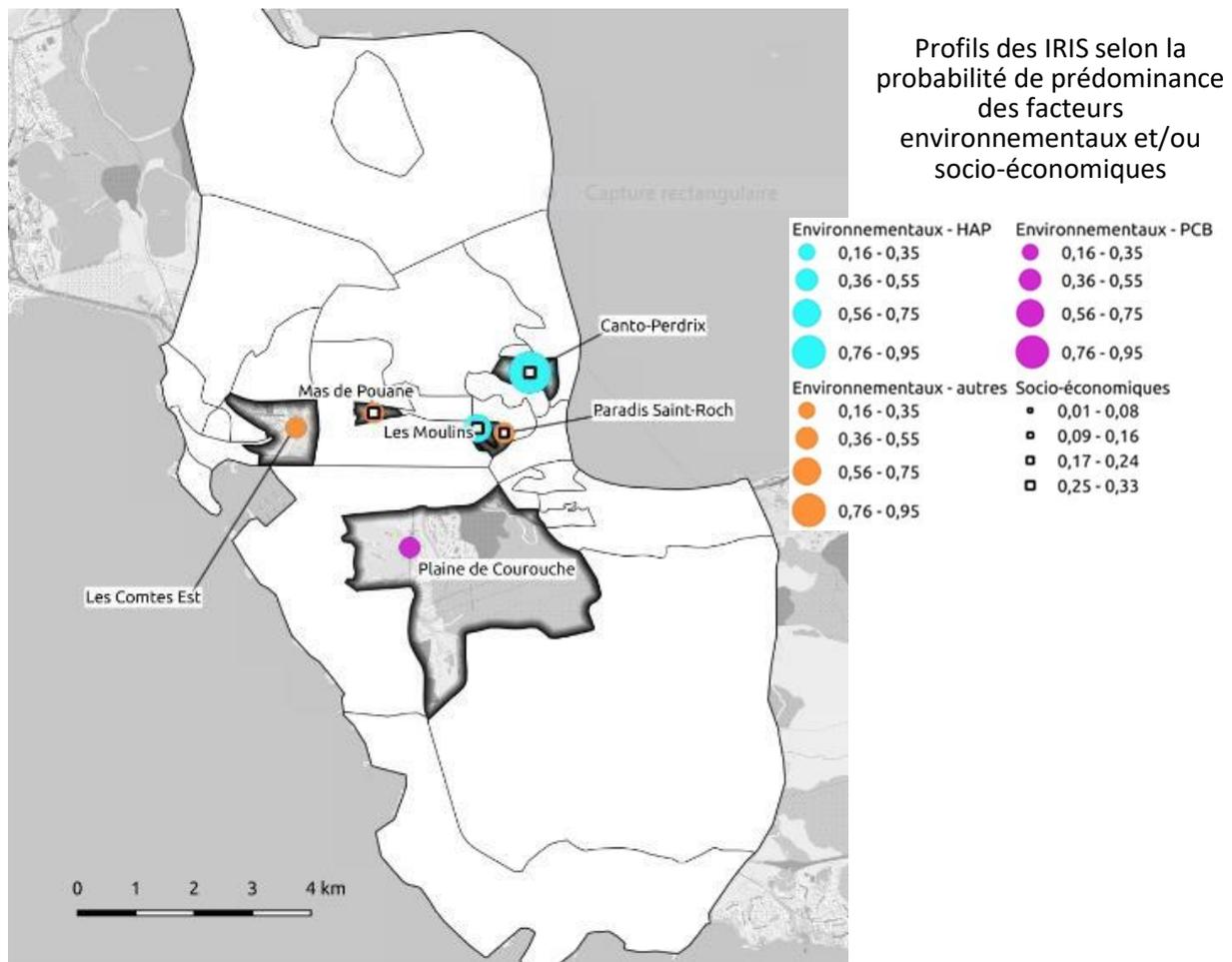
⁴⁵ https://www.atsdr.cdc.gov/toxfaqs/ToxFAQS_Foreign_Language_PDFs/tfacts58_french.pdf



Carte 7. Groupe 3 CAH, IRIS caractérisés par une prévalence accrue des hospitalisations liées aux **maladies cardiovasculaires (< 15 ans et > 65 ans)**, aux **maladies respiratoires (<15 ans)** et au **diabète (< 15 ans)**

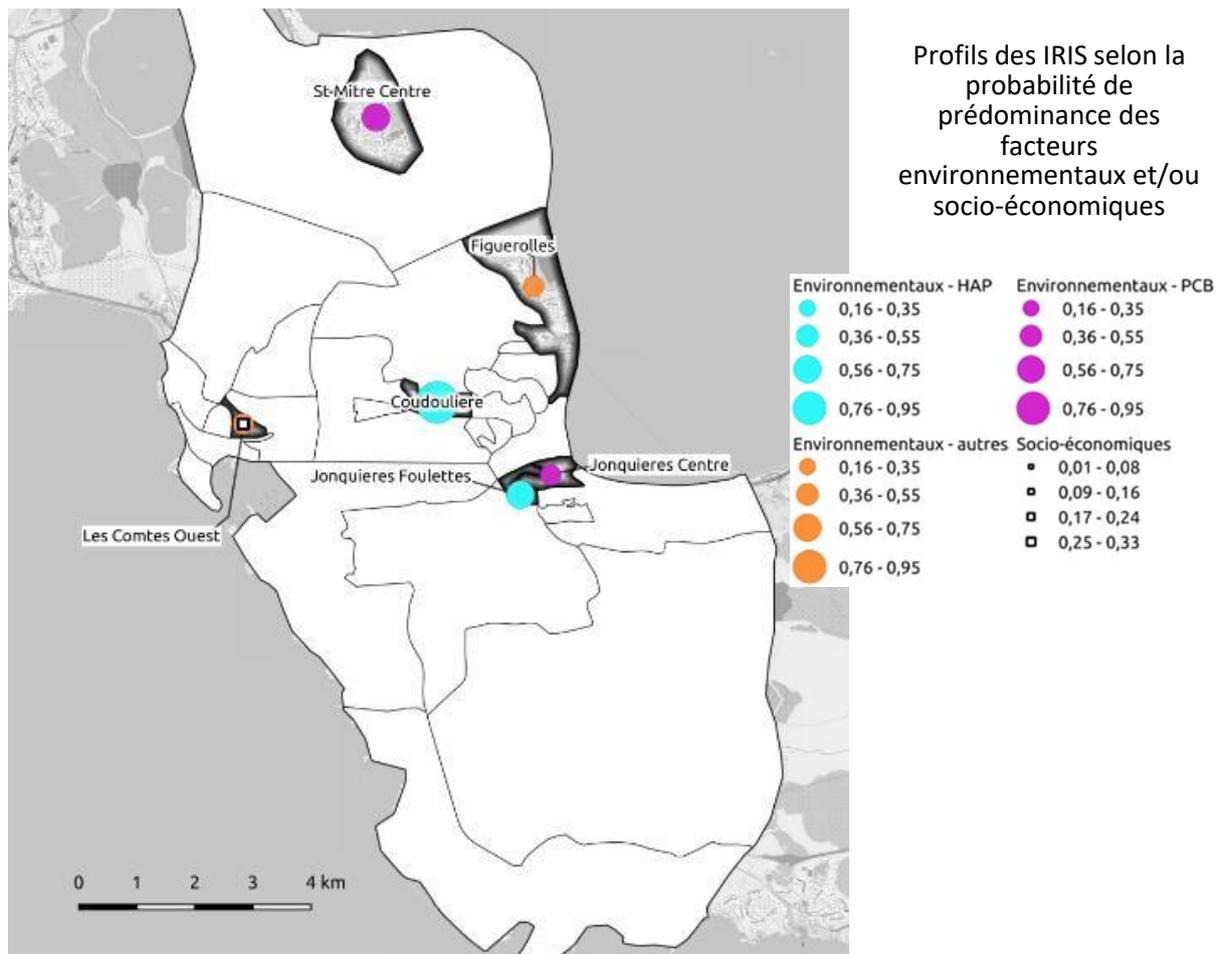
Regardons à présent ce qu'il en est pour les IRIS du groupe 3. L'IRIS Ferrière a un profil de type PCB comme l'IRIS St-Mitre Périphérie, cependant lorsque l'on examine les données pour St-Mitre, les valeurs de PCB ne sont pas très élevées comparativement à d'autres IRIS, cependant, c'est le premier polluant qui ressort sur l'analyse de sensibilité correspondante. En fait, cet IRIS est soumis simultanément à plusieurs polluants (Co, Zn_p, Cr_p, Cr6_p, PCDD_PCDF, HCl, BkF_p), chacun n'étant pas très élevé comparativement aux autres IRIS, à l'exception toutefois de BkF_p dont nous avons pu voir le rôle dans les maladies respiratoires (monitorings).

En ce qui concerne les Ecart, Hg est le premier polluant qui sort, suivi de CO, Cu_p, et de HF, tandis que Croix Sainte se caractérise plutôt par la présence de Zn_p, suivi par plusieurs polluants de type PCB, et As. Enfin, l'IRIS de la Colline semble plus marqué par des variables de nature socio-économique (population immigrée, familles monoparentales, chômage et personnes non scolarisées à 15 ans, dont on a vu dans les monitorings l'impact de cette dernière variable sur les maladies cardiovasculaires), seul Cd_p qui perturbe le métabolisme de l'insuline se démarque, comme polluant, dans cet IRIS.



Carte 8. Groupe 4 CAH, IRIS caractérisés par une prévalence accrue des hospitalisations liées aux **maladies cardiovasculaires (> 65 ans)** et aux **maladies respiratoires (<15 ans et >65 ans)**

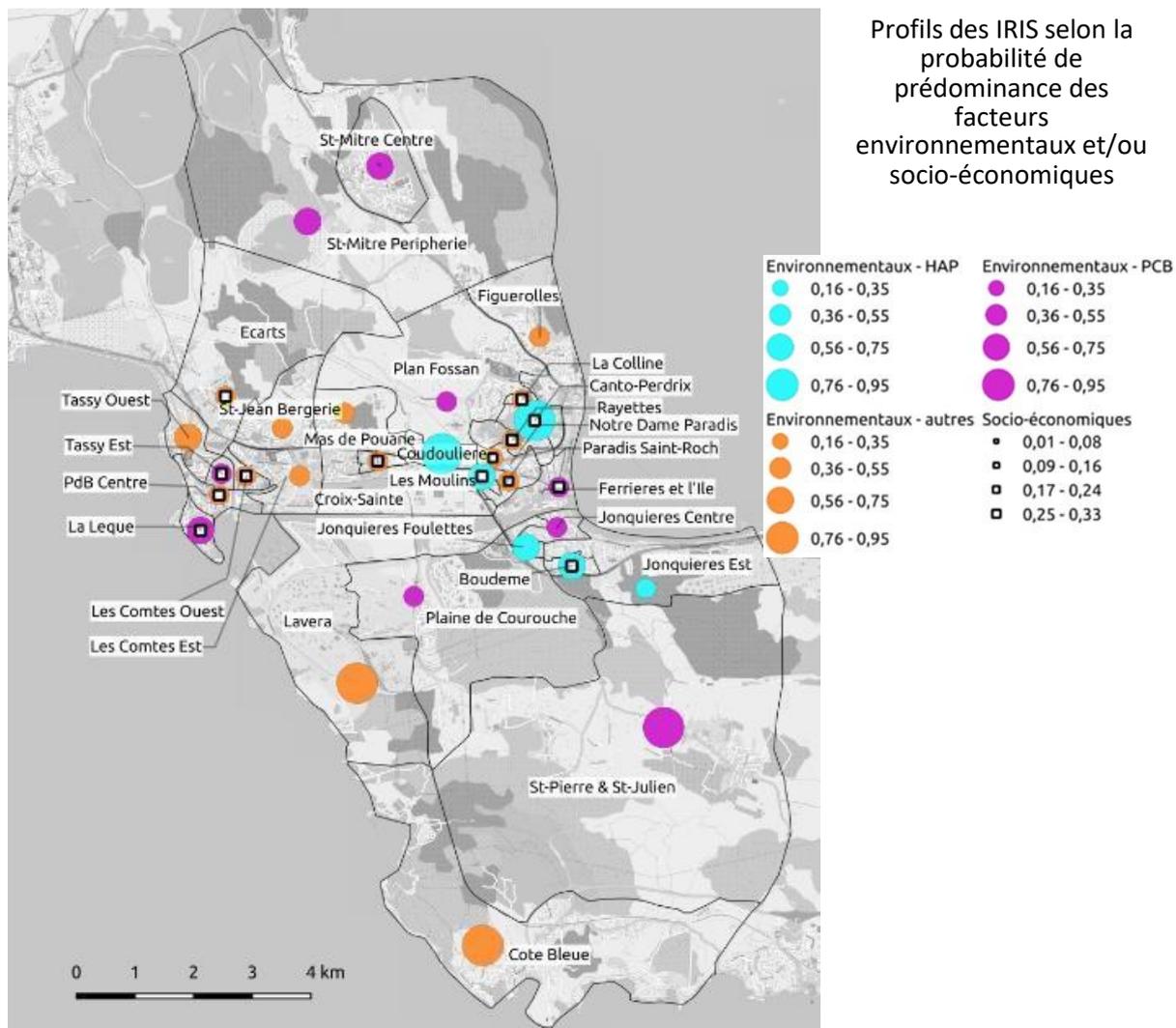
À partir de ce 4^{ème} groupe les nombres de séjour sont relativement moins importants, Canto Perdrix et les Comtes Est s'individualisent au niveau des maladies cardiovasculaires pour les plus de 65 ans, avec une prédominance de BkF_p pour Canto Perdrix et une socio- économie plutôt défavorisée, et un cocktail qui semble délétère pour les Comtes Est (DCE, V_p, Hg et HF), puisque l'effet est observé en plus sur les maladies respiratoires pour les plus de 65 ans.



Carte 9. Groupe 5 CAH, IRIS caractérisés par une prévalence moindre pour les pathologies étudiées « **Groupe témoin** »

Enfin, le groupe témoin se caractérise par globalement moins de pathologies qu'ailleurs, et surtout des polluants moins dangereux et/ou en moindre grande quantité que dans les IRIS des précédents groupes. De plus, la socio-économie étant plus favorable⁴⁶, elle tend à en compenser les effets. Cela est particulièrement vrai pour St-Mitre Centre. Néanmoins, il est « intéressant » de remarquer que la seule présence de SO₂ a une influence sur les maladies respiratoires chez les 15-65 ans dans l'IRIS de Figuerolles.

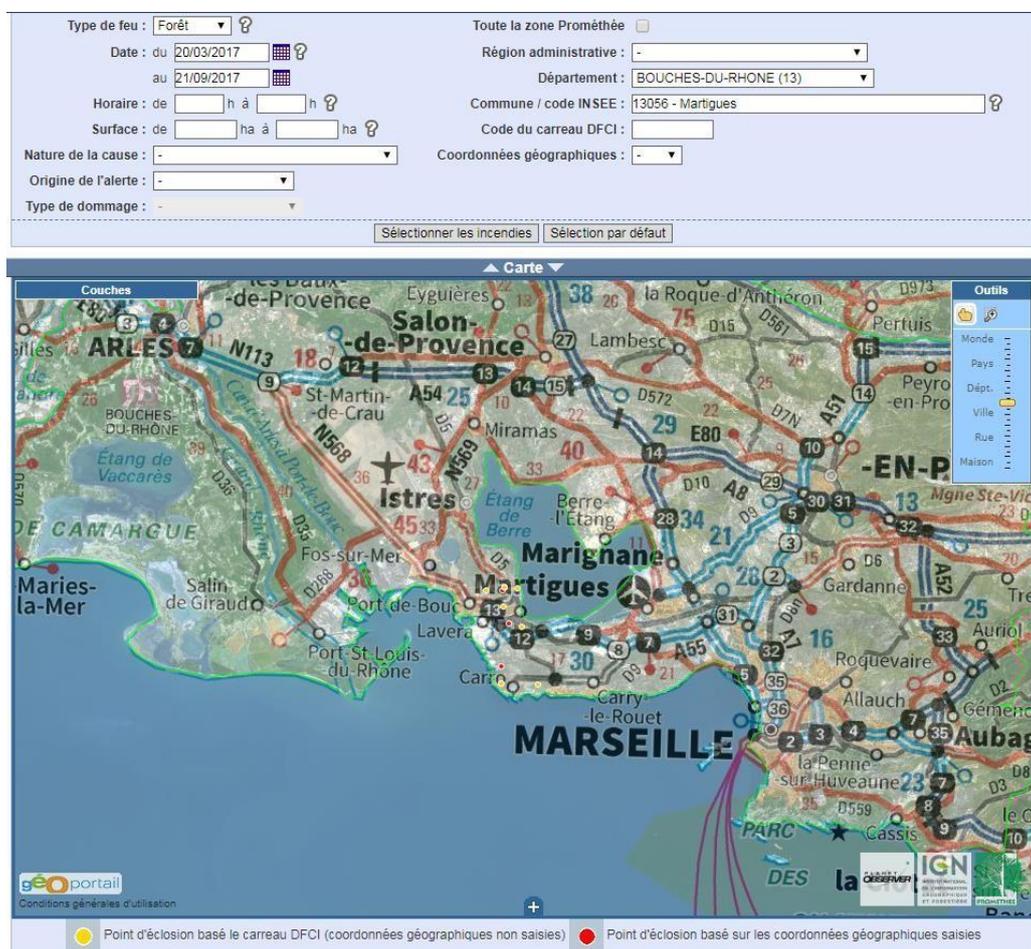
⁴⁶ Hormis aux Comtes Ouest



Carte 10. Carte de synthèse

La carte de synthèse ci-dessus rend compte des émissions de HAPs dans la partie nord-est du Pays de Martigues (Jonquières, Boudème), d'une pollution au Vanadium et au SO₂ dans les quartiers sud de Martigues (Lavéra, Cote bleue), et d'une pollution aux PCB répartie sur le territoire. Ceci alors qu'une socio-économie défavorisée, au centre du Pays de Martigues se cumule aux polluants.

La pollution aux PCBs pose question. Parmi les hypothèses évoquées, celle des incendies, particulièrement récurrents sur le territoire peut être avancée et expliquer ce caractère relativement diffus, que nous observons sur la carte de synthèse. Pour infirmer ou confirmer cette hypothèse nous avons donc fait appel à la base de données Prométhée⁴⁷ et recensés les départs de feux qui ont eu lieu dans les 6 mois précédents les relevés lichéniques (si l'on se réfère à la capacité reconnue des lichens de rendre compte des polluants entre 3 à 6 mois avant un prélèvement). Par conséquent, ici, l'intervalle à retenir s'étendait de fin mars à fin septembre 2017, puisque les prélèvements ont été effectués entre le 18 et le 20 septembre 2017 (Cf. Carte ci-dessous). Cela dit, dans le tableau, la date la plus ancienne est le 16/6, car les feux antérieurs (depuis mars) n'étaient pas suffisamment importants pour être pris en considération. En effet, nous n'avons retenus que les feux qui ont brûlé une surface supérieure à 2 000 m².



Carte 11. Les incendies de forêt sur le territoire martégal entre mars et septembre 2017.
Source. Prométhée.com

⁴⁷ Recense tous les incendies de forêt en région méditerranéenne, en France : <http://www.promethee.com/incendies>

	<u>Num.</u>	<u>Type</u>	<u>Alerte↑</u>	<u>Dpt.</u>	<u>Commune</u>	<u>Surface(ha)</u>	<u>Nature de la cause</u>
	1318	Forêt	18/04/2017	13	Martigues	0,1000	Malveillance
	1851	Forêt	29/05/2017	13	Martigues	0,2000	Chemin De Fer
	2222	Forêt	16/06/2017	13	Martigues	4,1400	Malveillance
	2226	Forêt	16/06/2017	13	Martigues	0,0020	Malveillance
	11109	Forêt	16/06/2017	13	Martigues	0,0100	Malveillance
	2288	Forêt	21/06/2017	13	Martigues	0,1500	-
	2624	Forêt	22/06/2017	13	Martigues	0,0010	-
	2622	Forêt	23/06/2017	13	Martigues	0,0200	-
	2692	Forêt	30/06/2017	13	Martigues	4,7500	-
	3164	Forêt	06/07/2017	13	Martigues	0,0010	-
	3360	Forêt	10/07/2017	13	Martigues	0,2000	Accidentelle
	4387	Forêt	26/07/2017	13	Martigues	158,8000	Jet D'Objets Incandescents
	4392	Forêt	26/07/2017	13	Martigues	4,4600	Jet D'Objets Incandescents
	4797	Forêt	03/08/2017	13	Martigues	0,0500	-
	6304	Forêt	14/08/2017	13	Martigues	0,1500	-
	11251	Forêt	12/09/2017	13	Martigues	0,1200	-

Nombre de lignes : 20 ▾ |< 1 >=> 16 incendies Page : ok

[Télécharger au format CSV] [Documentation]

Critères de sélection :

Type de feu : Forêt, A partir du : 20/03/2017, jusqu'au : 21/09/2017, Département : BOUCHES-DU-RHONE (13), Commune / code INSEE : 13056 - Martigues

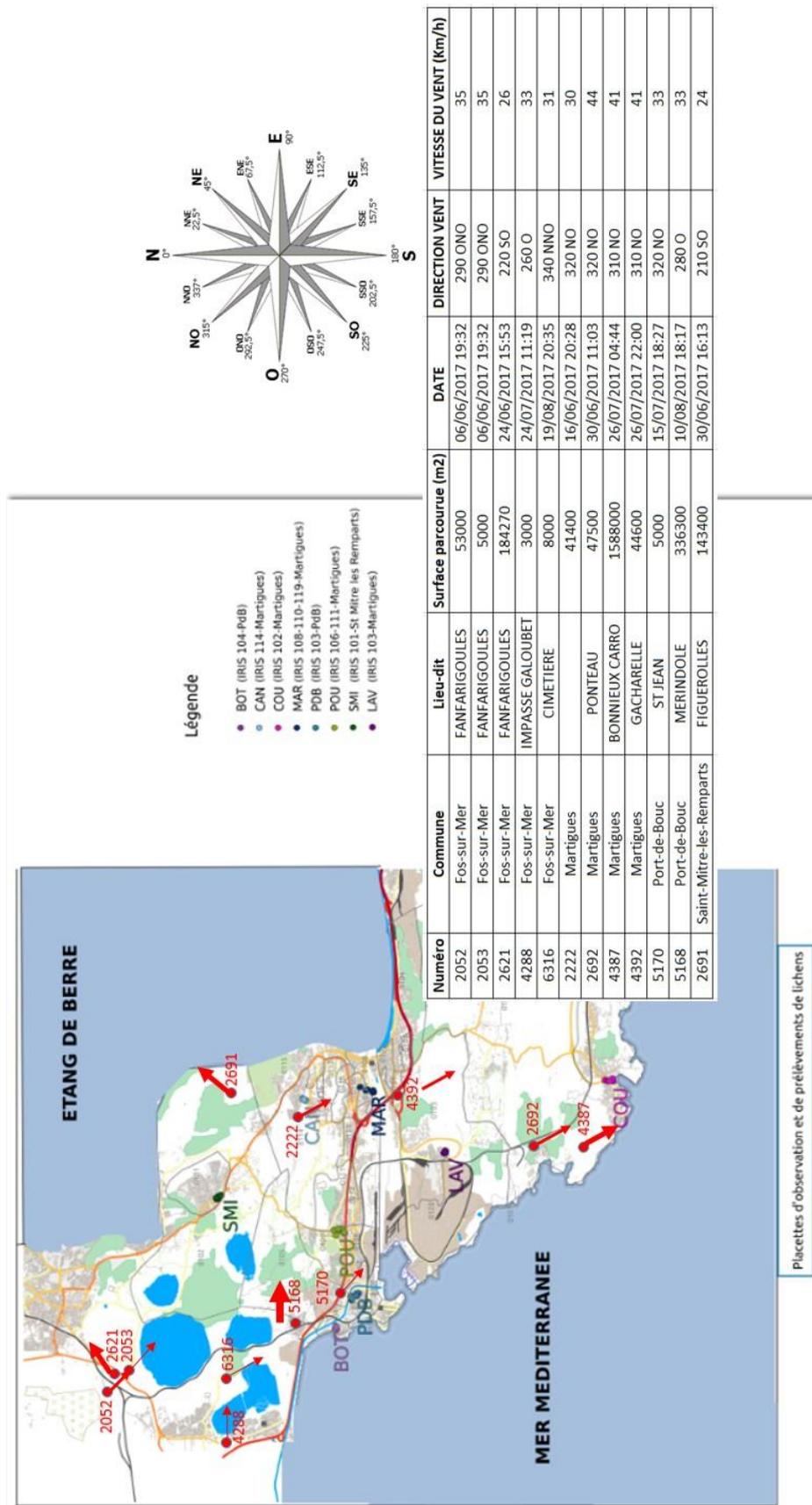
Synthèse du tableau :

- Nombre d'incendies : 16 (dont forêt : 16, AFERPU : 0)
- Surface totale des feux de forêt : 173,15 ha (1,73 km²)

Tableau 13. Extraction de la BDD Prométhée, date d'extraction : 23/04/19
Type de feu : Forêt, à partir du : 20/03/2017, jusqu'au : 21/09/2017, Codes INSEE
communes : 13056 Martigues ; 13039 Fos-sur-Mer

Puis, nous avons reportés sur la carte de localisation des placettes d'observation des lichens, les feux de forêt, qui avaient eu lieu sur nos 3 communes, ainsi que dans la commune limitrophe de Fos-sur-Mer (afin d'éviter un effet de bord). Chaque feu est affecté d'un numéro qui correspond à celui qui lui est attribué dans la base de données Prométhée. Nous avons par la suite recherché le sens du vent au moment de l'incendie et dans les heures qui suivent, en nous référant pour cela aux archives de la station météorologique la plus proche, à savoir Martigues Cap Couronne. Ce sens est indiqué par une flèche, alors que l'épaisseur de la flèche représente l'importance de la surface brûlée.

La mise en perspective de la localisation des lichens et de la localisation des feux de forêt, de leur importance, et du sens du vent à ce moment-là, indique qu'il s'agit de feux relativement peu étendus, de l'ordre de 4 ha (à l'exception de celui de Carro n° 4387 qui a été de 158 ha). La plupart du temps, le vent ne soufflait pas au moment de l'incendie et les quelques heures après vers les placettes de lichens, où lorsque c'était le cas (feux n°2222 vers MAR, et 5170 vers PDB, on peut penser que leur faible taille a eu un impact limité sur les émissions de PCBs).



Carte 12. Direction du vent et importance des feux de forêt à proximité spatiale et temporelle des prélèvements lichéniques

Les incendies ne semblent donc pas être la cause de la pollution observée aux PCBs. Cette pollution peut avoir plusieurs sources différentes, les PCBs seraient alors émis en plusieurs lieux, ce qui expliquerait le caractère diffus de cette pollution, ou bien il pourrait s'agir de dépôts terrestres anciens transportés sous l'effet de forts vents ? Une étude réalisée en 2014, par Air Rhone Alpes démontre qu'en raison de leur caractère particulièrement persistant dans l'environnement, la part de PCBs qui parvient à migrer d'un compartiment à un autre (du sol vers l'atmosphère) est non négligeable⁴⁸.

Dans tous les cas des investigations complémentaires doivent être menées. Des solutions de dépollution de type phytomanagement pourraient être envisagées afin d'extraire ces PCBs du sol, et de limiter d'éventuels transferts dans l'air ambiant.

La méthodologie employée dans cette preuve de concept (Figure, ci-dessous) est de type « top-down », elle pourrait très bien être couplée, à présent, avec d'autres études plus « bottom-up », avec des investigations de terrain dans certains IRIS, ou sur des signaux plus faibles, ou bien en servant d'appui à de futures campagnes de prévention de la population.

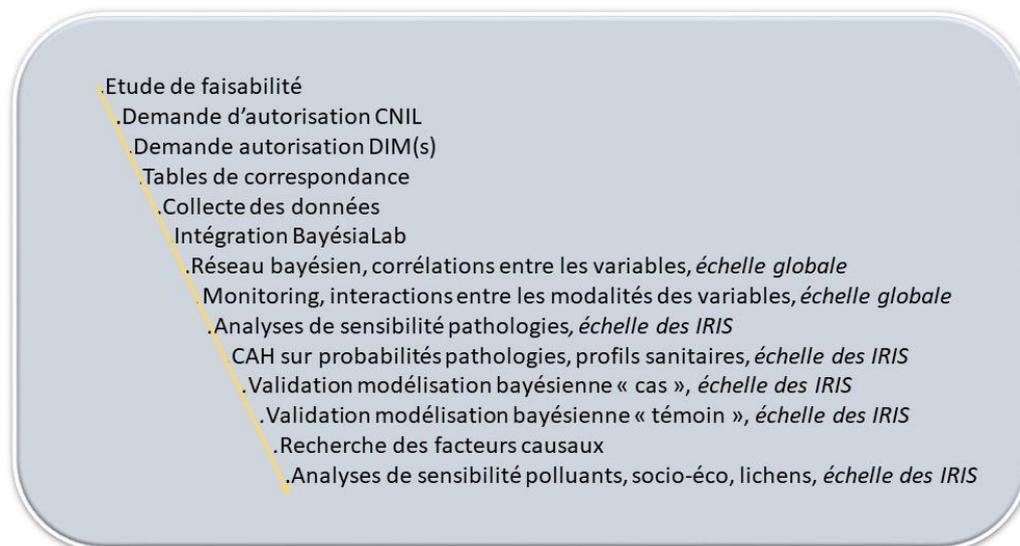


Figure 35. Le *design* de l'étude de type « top down »

⁴⁸https://www.atmo-auvergnerhonealpes.fr/sites/ra/files/atoms/files/rapport_transferts_dioxines_furanes_pcb_sud_grenoblois.pdf

En mobilisant un **grand nombre de variables, désagrégées**, à une **échelle territoriale fine**, et en faisant appel à un **outil approprié** pour rendre compte des liens existants entre ces variables, nous avons pu mettre en évidence plusieurs relations entre : exposition à des xénobiotiques et des **pathologies avérées**, assez importantes pour avoir nécessité une hospitalisation en 2015, et qui suggèrent les préconisations ci-après.

Parmi les 9 relations trouvées, l'effort doit porter sur la **réduction des émissions de SO₂**, dont on a vu qu'il agissait dès un seuil très bas d'exposition, bien en deçà de la norme. Les entreprises doivent poursuivre et intensifier le processus de désulfuration de leurs installations, ou lorsque cela est possible utiliser un combustible moins soufré. Cependant, l'exposition au SO₂ n'est pas que le fait des entreprises présentes au sein de ce territoire, le volume annuel de tankers qui acheminent leur chargement vers la ZIP de Fos-sur-Mer y est aussi pour quelque chose. Là encore, le recours à un carburant moins soufré pourrait contribuer à une meilleure santé des habitants.

Il semble qu'il existe au sein de ce territoire une **pollution diffuse aux PCB**, en lien avec les maladies respiratoires chez les plus de 65 ans, qui doit faire l'objet d'investigations plus précises afin de prendre rapidement des mesures appropriées (dépollution de sites par phytomanagement).

Plusieurs **seuils** sont apparus, à partir desquels certains polluants étaient particulièrement délétères, et en ce sens, la modélisation bayésienne aura permis de les révéler, c'est le cas par exemple de **l'acide fluorhydrique (HF)**, dont l'effet est ressorti dès le seuil de 0.0028 µg/m³. Tandis qu'une exposition au **cadmium dans sa phase particulaire**, pour une exposition comprise entre 0.214 et 0.250 µg/m³ perturbe le métabolisme de l'insuline chez les plus de 65 ans. Enfin, dès que le **Benzo[k]fluoranthène (phase particulaire)** dépasse le seuil de 0.672 µg/m³, cela se traduit par un plus grand nombre de séjours pour maladies respiratoires chez les 15-65 ans.

Le rôle de polluants spécifiques dans la catégorie de profil « Autres » tels que le **Chrome, Chrome VI, et le Vanadium**, dans leur phase particulaire méritent une surveillance *ad hoc*. Il conviendrait donc d'abaisser, lorsque cela est techniquement et financièrement possible, les émissions de ces polluants, en deçà de ces seuils (produits de substitution, *process* industriels moins polluants, filtres plus performants ?).

L'étude a mis en évidence des profils sanitaires différenciés, entre les divers IRIS qui composent le territoire du Pays de Martigues, qui devrait amener au déploiement d'**une offre de soins plus adaptée** aux besoins de la population, et ce, dans une optique de prévention, voire de promotion de consultations chez certains spécialistes. Par exemple, chez un cardiologue, dans les IRIS où le nombre de personnes sans diplômes à 15 ans est important, puisqu'une relation a été trouvée à ce niveau, relation vraisemblablement due à des conditions de vie et/ou de travail plus difficiles pour ces personnes.

De même, des mesures de prévention et d'accompagnement social doivent être menées en direction des familles monoparentales.

Pour finir, les pathologies d'aujourd'hui sont la plupart du temps le résultat d'expositions plus anciennes, où les niveaux de pollution étaient globalement plus hauts. Il serait « intéressant » de disposer sur les polluants incriminés, de mesures plus anciennes, et de relancer les calculs. Tandis que les pathologies avérées mobilisées dans cette étude permettraient de valider les Excès de Risque Individuels (ERI) qui ont été modélisés dans les études SCENARII et POLIS d'AtmoSud.

Il conviendrait maintenant, de réitérer la méthodologie employée à d'autres espaces similaires, voisins ou plus lointains. C'est ce que nous allons aborder dans la dernière partie de cette HDR qui concerne les perspectives de recherche. Mais, nous allons voir que nous allons dépasser le simple stade de la transposition pour cheminer vers une théorisation.

3. Perspectives de recherche

3.1 De la causalité au potentiel

Nous venons de voir avec cette preuve de concept que nous pouvons trouver les éléments causaux des pathologies observées dans cet espace. Cependant, nous ne pouvons toujours pas généraliser ce que nous avons trouvé, car cela est spécifique à ce lieu.

Nous pourrions alors réitérer la méthodologie présentée et l'appliquer à un autre espace industriel de même type (pétrochimique), les exemples ne manquent pas, que ce soit au sein de notre territoire national, ou bien dans d'autres pays. A ce sujet, l'intégration du CS de Santé Publique France qui travaille sur « les études multicentriques autour des bassins industriels et des sols pollués » nous permettrait d'obtenir des données sur les autres sites industriels présents en France, afin de voir si des différences existent en matière de polluants, de leur combinaison, ou de leurs effets. Nous savons par exemple que suivant le pays fournisseur de pétrole brut, une plus ou moins grande quantité de vanadium y est présente.

Nous pourrions dès lors construire un réseau bayésien par site et les comparer entre eux *via* une comparaison de leur structure, ou bien de leur segmentation de variables, ou bien nous pourrions intégrer toutes les données sanitaires, environnementales, socio-économiques, prises en compte dans l'ensemble de ces études, les compléter (ou les inférer) afin d'avoir une matrice carrée, et retrouver l'espace de ces différents sites au niveau de leurs analyses de sensibilité comme nous venons de l'évoquer. C'est ce que nous allons faire, car il est très intéressant de travailler sur les milieux industriels, en raison de la diversité des polluants, de leur dangerosité, et aussi du fait qu'ils retiennent paradoxalement moins l'attention des chercheurs.

De plus, le suivi spatio-temporel de certains indicateurs de santé offrirait une plus grande compréhension des liens entre polluants et santé *via* leur évolution, tandis que le caractère multicentrique lèverait le biais dû aux différences de protocole qui font qu'à l'heure actuelle les résultats obtenus dans les travaux de recherches menés sur ces espaces ne sont pas comparables, car conduits avec des méthodologies différentes.

C'est malheureusement souvent le cas. Ainsi, si nous prenons dans un autre domaine l'exemple de la *meta review* publiée dans le journal *Public Health Nutrition* en 2010, par Charreire et al., qui recense les recherches dans lesquelles l'environnement alimentaire (*food environment*) des habitants est évalué (figure ci-après) nous constatons la diversité des populations cibles, des méthodes utilisées, ainsi que des co-facteurs pris ou pas en compte. Cela rend bien entendu compte, d'une certaine vitalité de la recherche, cependant, les résultats ne sont pas comparables, et les études s'apparentent à des monographies. Certes, elles vont toutes mettre en évidence un lien entre un statut socio-économique défavorable et

un environnement alimentaire qui l'est tout autant, mais nous ne savons pas comment, ni pourquoi ce lien varie entre divers espaces, car tout est différent.

Ainsi, le domaine de la santé environnementale manque cruellement de théorisation. Cela s'explique par la relative jeunesse de ce secteur, par le fait que les études prennent du temps (autorisations CNIL) et qu'elles sont complexes. De plus, ces études ne sont pas récurrentes, ni uniformes, les méthodologies employées varient grandement, de même que les populations cibles. Elles sont menées sur des espaces et des échelles géographiques différentes, apparaissent souvent comme contradictoires, et les caractéristiques intrinsèques des espaces sur lesquels elles se déploient ne sont pas prises en compte.

Tableau 14. Descriptif synthétique des études prises en compte dans la meta review de Charreire et al., (2010)

Author(s) (year)	Location Scale/population	Measures food availability	GIS method	Other variables	Findings
Apparicio <i>et al.</i> (2007) ⁽³⁹⁾	Montreal (Canada) Census tract block	Supermarket: <ul style="list-style-type: none"> Proximity Diversity In the immediate surrounding 	<ul style="list-style-type: none"> Euclidean distance to the closest supermarket Number of supermarkets at a distance less than 1000 m Mean distance to the three closest supermarkets to different companies 	Deprivation index <ul style="list-style-type: none"> Income Lone-parent families Unemployment Education Recent immigrants 	Food deserts do not represent a major problem in Montreal
Austin <i>et al.</i> (2005) ⁽⁴⁴⁾	Chicago (Illinois) Census tract School addresses	Fast food: <ul style="list-style-type: none"> Density Proximity to school 	<ul style="list-style-type: none"> Buffer (400/800 m around school) Euclidean distance to school from the closest fast food 	Contextual data <ul style="list-style-type: none"> Level of commercialization (percentage of commercial land) Household income Inside or outside downtown 	Fast foods are concentrated within short distance from schools
Baker <i>et al.</i> (2006) ⁽⁴⁶⁾	St Louis (Missouri) Census tract	Clustering to: <ul style="list-style-type: none"> Supermarket Fast food 	Spatial clustering <ul style="list-style-type: none"> Circular window Number of food outlets 	<ul style="list-style-type: none"> Composite score of availability to supermarket and fast food Contextual data: race, income (level of poverty) 	Income and race seem to be associated not only with the location of food outlet but also with the selection of food available
Block & Kouba (2006) ⁽⁴⁹⁾	Chicago (Illinois), two neighbouring communities (Austin & Oak Park)	Proximity to: <ul style="list-style-type: none"> Supermarket Independent grocery Convenience store Other food outlets 	<ul style="list-style-type: none"> Buffer (0.25/0.5/0.75/1 mile and >1mile around food store) 	<ul style="list-style-type: none"> Food survey: cost and quality of produce Contextual data: households having no car 	Type and number of grocery stores differ between Austin and Oak Park
Block <i>et al.</i> (2004) ⁽⁴⁵⁾	New Orleans (Louisiana) Census tract	Fast food: <ul style="list-style-type: none"> Density Distance 	<ul style="list-style-type: none"> Buffer (shopping area) (0.5/1 mile around census tract) 	Contextual data: <ul style="list-style-type: none"> Black residents Low income Adjusted by: <ul style="list-style-type: none"> Alcohol outlet density Highways Median home value 	Fast food outlets are geographically associated with predominantly black and low-income neighbourhoods
Bodor <i>et al.</i> (2008) ⁽²²⁾	New Orleans (Louisiana), older urban area 102 households (16 years and more)	Distance to and density of: <ul style="list-style-type: none"> Supermarket Small food store 	<ul style="list-style-type: none"> Buffer (100m/1 km around home) Density of food store Euclidean distance: to the closest small food store, to the closest supermarket 	<ul style="list-style-type: none"> Individual level fruit and vegetable intake and SES characteristics In-store level availability of: fresh, canned and frozen fruit and vegetables 	<ul style="list-style-type: none"> Access to a small food store within 100 m of home was marginally associated with an increased fruit intake No association was found between intake and access to supermarkets

Author(s) (year)	Location Scale/population	Measures food availability	GIS method	Other variables	Findings
O'Dwyer & Coveney (2006) ⁽³⁵⁾	Adelaide (Australia) LGA	Supermarket: ● Density ● Proximity	● Buffer (2-5 km around area (LGA)) ● Distance by road network to the closest supermarket	SEIFA	Socio-economic differences in access to food and the availability of food outlets Food deserts appear to exist
Pearce <i>et al.</i> (2006) ⁽⁴³⁾	New Zealand Census meshblock (smallest unit of dissemination of census data)	Travel time for: ● Recreational amenities ● Food outlet ● Educational facilities ● Health facilities	Modelling travel time depending on: ● Speed limits ● Type of road surface ● Sinuosity/topography		Strong geographical variations in community resource accessibility between neighbourhoods within an urban area
Pearce <i>et al.</i> (2007) ⁽⁴²⁾	New Zealand School Census meshblock	Travel time and distance to: ● Fast food ● Supermarkets ● Convenience store for each school	Modelling travel time to each ● School ● Census meshblock	● NZDep – 2001 ● Socio-economic characteristics of school ● Urban/rural status	Access to fast food outlets is better in more deprived neighbourhoods and around more socio-economically disadvantaged schools
Pearce <i>et al.</i> (2008) ⁽²⁵⁾	New Zealand Census meshblock 12 529 adults (15 years and more)	Travel time to: ● Supermarket ● Convenience store	Modelling travel time depending on: ● Speed limits ● Type of road surface ● Sinuosity/topography	● Consumption of the recommended daily intake of fruit and vegetable ● Individual variables ● Meshblock variables: NZDep – 2001	Little evidence that poor locational access to food retail provision is associated with lower fruit and vegetable consumption
Sharkey & Horel (2008) ⁽⁵¹⁾	Six counties rural region of Texas Census block group	Distance to: ● Supermarket ● Grocery store ● Convenience store ● Discount store	Distance by road to the closest food store	● Deprivation Index ● Minority composition ● Population density	Better spatial access to food store for neighbourhood with high socio-economic deprivation
Smoyer-Tomic <i>et al.</i> (2008) ⁽⁴⁰⁾	City of Edmonton (Alberta, Canada) Census block	Proximity to: ● Supermarket ● Fast food	● Buffer (500, 800 m around geometric centre of each census block) ● Distance to the nearest outlet by street network	Area level ● Race/ethnicity, SES ● Age, family status ● Housing tenure, urbanisation	Fast food outlet exposure was higher in low-income neighbourhood
Timperio <i>et al.</i> (2008) ⁽²⁴⁾	Greater Melbourne (Geelong area) (Australia) 801 children (5-6/10-12 years)	Distance to: ● Supermarket ● Greengrocer ● Convenience store ● Fast food, take-away ● Restaurant, cafes to the closest outlet	● Buffer (800 m around home) ● Existence of one or more outlet ● Distance by road	● Children: frequency of consumption of fruit or vegetables ● Parents: sociodemographic variables	● The availability of convenience stores and fast food outlets close to home may have a detrimental effect on children's fruit and vegetable intake ● The likelihood to consuming vegetables was greater the further children lived from a supermarket or a fast food outlet
Author(s) (year)	Location Scale/population	Measures food availability	GIS method	Other variables	Findings
Wang <i>et al.</i> (2007) ⁽³⁰⁾	Agricultural regions of California Census tract block 7595 adults (25-74 years)	For different type of retail food stores and fast foods: ● Density ● Proximity	● Buffer (0-5 mile around home) ● Shortest Euclidean distance	● Index of the neighbourhood socio-economic environment ● Individual-level variables (sex, age, smoking, physical activity...) ● BMI	● Proximity to various type of food stores was associated with neighbourhood SES ● Higher BMI among residents of low socio-economic neighbourhoods ● Living in an environment where healthy food is not readily available is associated with increased obesity risk
Winkler (2006) ⁽³⁶⁾	Brisbane (Australia) Census collection districts	Distance to the nearest types of shops selling fruits and vegetables (supermarket, greengrocers...)	● Buffer (2-5 km around area) ● Shortest Euclidean distance	● IRSD ● The opening hours of nearby fruit and vegetable outlets	It is unlikely that living in a socio-economically disadvantaged area presents fewer opportunities to purchase fruits and vegetables, at least in an urban area
Zenk <i>et al.</i> (2005) ⁽⁶³⁾	Detroit (Michigan) Census tract block	Supermarket: ● Proximity	● Manhattan distance to the closest supermarket	● Population density ● Non-Hispanic African American residents ● Residents below the poverty line ● Spatial autocorrelation (Moran's I)	The most impoverished neighbourhoods in which African Americans resided were further from the nearest supermarket than were the most impoverished white neighbourhoods
Zenk & Powell (2008) ⁽⁴⁶⁾	Fifty states and twenty largest cities in USA School Census tract	Density to: ● Fast food ● Convenience store	● Buffer (0-5 mile around school)	Independent data ● Racial/ethnic ● Median income ● School level ● Population density ● Level of urbanisation ● Student body size	Within 0-5 miles (walking distance): ● Fast food and convenience stores are more available in lowest-income neighbourhood (except in African American neighbourhood) ● African American neighbourhoods have fewer food outlets than white neighbourhoods ● Urban high schools are exposed to 61% more fast food than urban middle schools

GIS, geographic information systems; LGA, local government area; SES, socio-economic status; SEIFA, Socio-Economic Index For Areas; TV, television; MESA, Multi-Ethnic Study of Atherosclerosis; AHEI, Alternate Healthy Eating Index; NZDep, New Zealand deprivation index; IRSD; Index of Relative Socio-economic Disadvantage.

Il en est de même pour le domaine de recherche qui nous intéresse, à savoir la pollution de l'air, peut-être même encore plus, tant ce domaine a mobilisé l'énergie de beaucoup de chercheurs ces deux dernières décennies. En effet, si nous effectuons une recherche sur *Science direct* pour l'année 2022 en tapant les termes suivants : *air pollution effect ; systematic review*, plus de 5 838 **méta-analyses** apparaissent.

Si nous prenons celle de Boogaard et al., (2022) au sujet des effets sur la santé d'une exposition à long terme à la pollution liée au trafic automobile (*traffic-related air pollution*, TRAP) qui prend appui sur le travail conduit par les experts du *Health Effects Institute* (HEI), sur la base de 353 études sélectionnées dans leur *meta review*, là encore c'est l'association entre la dose et les effets de l'exposition qui est recherchée, entre plusieurs pays, mais sans qu'à aucun moment la question de l'espace d'étude de ces recherches ne soit investigué.

Il existe un cadre pour pouvoir comparer des études entre elles⁴⁹, ce cadre c'est celui des questions PECO (Population, Exposure, Comparator, Outcomes) (Morgan et al., 2022). De sorte que les critères de comparaison entre les études reposent sur 4 piliers : la population (quelle population ?), son exposition (comment est-elle évaluée ?), un élément qui permet la comparaison, et les résultats de l'association (quels effets ?). Comme nous pouvons le voir dans l'exemple ci-dessous, sur l'exposition des nouveaux-nés au bruit, et à travers 5 scénarios les plus utilisés par les chercheurs pour évaluer des expositions :

Tableau 15. Cinq approches paradigmatiques pour identifier l'exposition (Morgan et al., 2022)

Examen systématique potentiel ou contexte de recherche	Approcher	Exemple PECO
1. Calculer l'effet sur la santé d'une exposition ; décrivant la relation dose-effet entre une exposition et un résultat pour la caractérisation des risques.	Explorer la forme et la distribution de la relation entre l'exposition et le résultat dans l'examen systématique.	Chez les nouveau-nés, quel est l'effet supplémentaire d'une augmentation de 10 dB pendant la gestation sur la déficience auditive postnatale ?
2. Évaluer l'effet d'un seuil d'exposition sur les résultats pour la santé, lorsque le seuil peut être éclairé de manière itérative par les résultats de l'examen systématique.	Utilisez des seuils définis en fonction de la distribution dans les études identifiées dans la revue systématique.	Chez les nouveau-nés, quel est l'effet de l'exposition en dB la plus élevée par rapport à l'exposition en dB la plus faible (par exemple, tertiles, quartiles ou quintiles identifiés) pendant la grossesse sur la déficience auditive postnatale ?
3. Évaluer l'association entre un seuil d'exposition et un seuil de comparaison, lorsque les seuils peuvent être identifiés ou sont connus d'autres populations.	Utiliser les seuils moyens des populations externes ou autres (peuvent provenir d'autres recherches).	Parmi les pilotes commerciaux, quel est l'effet du bruit correspondant à l'exposition professionnelle par rapport à l'exposition au bruit subie dans d'autres professions sur la déficience auditive ?
4. Identifier un seuil d'exposition qui améliore les effets sur les résultats de santé.	Utiliser les seuils d'exposition existants associés à des résultats d'intérêt connus pour la santé.	Parmi les travailleurs industriels, quel est l'effet d'une exposition à < 80 dB par rapport à ≥ 80 dB sur la déficience auditive ?
5. Évaluer l'effet potentiel d'un seuil qui peut être atteint grâce à une intervention visant à améliorer les effets de l'exposition sur les résultats de santé.	Sélectionnez le comparateur en fonction des seuils d'exposition pouvant être atteints grâce à une intervention.	Dans la population générale, quel est l'effet d'une intervention qui réduit les niveaux de bruit de 20 dB par rapport à l'absence d'intervention sur la déficience auditive ?

* Seuils est un terme général faisant référence à des seuils, des niveaux, des durées, des plages, des moyennes, des médianes ou des plages d'exposition. dB : décibel ; PECO : population, exposition, comparateur, résultat(s).

⁴⁹ Il y en a d'autres, notons par exemple le standard PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)

Mais, là encore, comme nous pouvons le voir sur la figure ci-après, très souvent les études sont comparées par rapport à l'effet constaté des polluants sur la santé, afin de voir si ceux-ci se trouvent dans le même intervalle de dose-réponse, ce qui tendrait dès lors à valider les études menées.

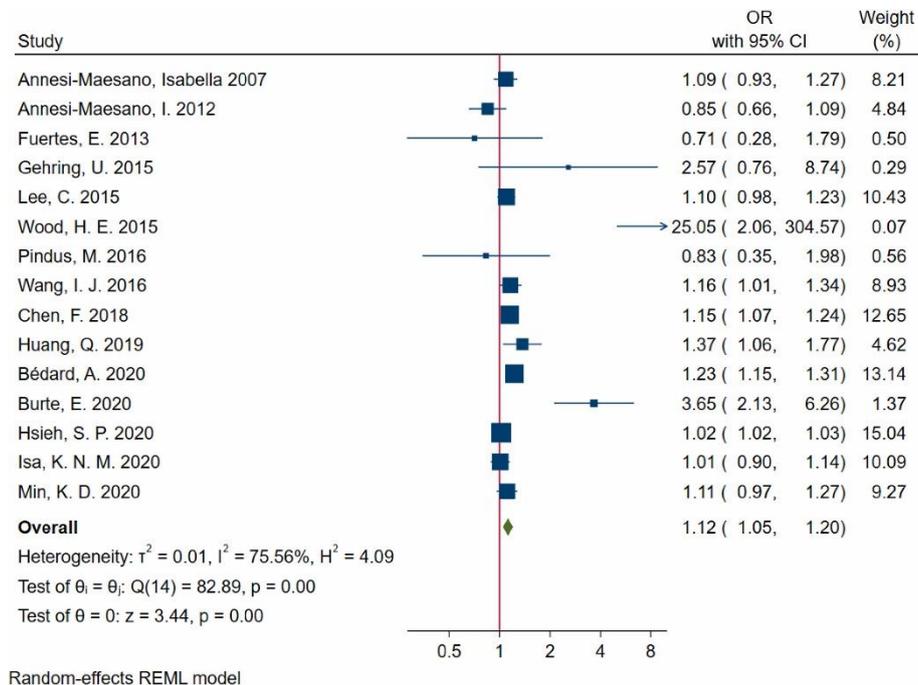


Figure 36. Méta-analyse de la rhinite allergique et de l'exposition aux PM 2,5 pour des paliers de 10 $\mu\text{g}/\text{m}^3$ IC : intervalle de confiance. Le carré bleu représente le poids de chaque étude, le centre du carré représente l'odds ratio (OR) et la longueur de chaque ligne autour du point représente son intervalle de confiance à 95 % (IC à 95 %). L'estimation groupée récapitulative globale (résultat de la méta-analyse) est représentée par un losange vert à la fin du graphique. La ligne verticale continue rouge représente l'estimation nulle (avec une valeur de 1 pour OR). Les lignes horizontales qui traversent la ligne verticale nulle (OR = 1) représentent des études non significatives

*Mais, ce qui va nous importer ce n'est pas tant de comparer, de valider les études réalisées par nos collègues, plutôt que d'**extraire les données sanitaires contenues dans ces études.***

3.2 Le concept d'eXpace

En effet, il me semble également intéressant pour la géographe que je suis, de réitérer la méthodologie présentée dans la preuve de concept, mais cette fois, dans plusieurs contextes (ville, périurbain, rural, naturel, mixtes), et espaces géographiques différents (europe, asie, moyen-orient, afrique, Amérique latine), afin de dégager une catégorisation d'eXpaces, ce qui nous permettra, nous l'espérons, de cheminer à terme vers une généralisation, et donc une théorie.

Je ne viens pas de faire une faute de frappe, un X à la place du s, je propose un nouveau concept, celui d'**eXpace** issu de la contraction des termes « exposome » et « espace ». En effet, nous avons besoin de connaissances à ce sujet, sujet qui n'est pas, il me semble, traité par les épidémiologistes qui s'occupent de l'exposome.

Nous souhaitons revaloriser le rôle du géographe dans les études menées en santé environnementale, qui sont la plupart du temps le fait d'épidémiologistes. Ces derniers partent des cas et vont vers l'espace (espace support), les géographes dans une démarche inverse, à haute valeur ajoutée, devraient partir de l'espace et aller vers les cas (espace acteur). De loin, nous pouvons penser que c'est la même chose, mais c'est loin d'être le cas. En fait, nous avons besoin dans un souci de prévention et de santé publique de posséder la connaissance qui nous permettrait selon l'espace étudié d'évaluer instantanément les cas potentiels. Nous pouvons parler ici d'**intelligence « géosanitaire »**.

En effet, généralement nous mesurons le niveau de pollution d'un espace donné (moyennes annuelles, polluants d'intérêt, c'est-à-dire spécifiques à l'espace d'étude), si possible à fine échelle, et nous le mettons en relation avec la santé par exemple respiratoire des individus qui y habitent. Avec cette notion de potentiel, nous partons cette fois de l'espace d'étude et nous analysons tous les facteurs qui vont moduler la pollution et ses effets.

Mais comment faire ? Nous devons dans un premier temps avoir une idée des causes, sans causes pas de potentiel, pour reprendre l'expression de Nancy Cartwright « *no causes in, no potential out* ». Autrement-dit, c'est parce que nous aurons une bonne connaissance des effets produits par les causes que nous serons capables en observant ces effets de remonter *potentiellement* aux causes sous-jacentes.

Il me semble que les causes ne peuvent pas se limiter à de simples doses-réponses sans prendre en compte l'espace géographique qui va influencer par ses caractéristiques intrinsèques celles-ci.

Un des éléments limitants dans les études de santé environnementale, qui empêche une généralisation, c'est le difficile accès à des données de santé pertinentes, adaptées, exhaustives, et à fine échelle. Par difficulté il faut également entendre le temps, notamment

d'instruction des demandes par la CNIL ou les CPPs⁵⁰ pour l'accès à des données françaises. Si nous prenons le cas d'autres pays cet accès peut être facilité (Royaume Uni), ou bien au contraire interdit à des personnes étrangères au pays.

Ce sont les raisons pour lesquelles nous proposons de **pooler** des données sanitaires collectées à partir de *meta reviews* ou d'articles dans des revues spécialisées comme (*Environmental Health, Environmental Science and Pollution Research, Environmental Health Perspectives, Environmental Research*, liste non exhaustive) Nous nous focaliserons sur les pathologies respiratoires et cardiovasculaires qui sont les plus étudiées afin de disposer d'un nombre de cas suffisants dans diverses configurations spatiales, à fine échelle.

Une fois les données sanitaires collectées nous les mettrons en rapport comme nous l'avons fait dans la preuve de concept avec les données environnementales et socio-économiques de leur espace d'étude. En raison de la variabilité connue des polluants nous souhaitons toujours privilégier une échelle de travail fine (infra-communale).

Le problème étant de pouvoir obtenir des données à la même échelle et pour différents ensembles géographiques (eXpaces pris en Asie, Afrique, Amérique latine, Europe...)⁵¹. En effet, s'il existe bien au niveau de la pollution de l'air des bases de données européennes (Airbase de l'European Environment Agency, EEA), ou mondiales comme celle de l'OMS (AirQ+), elles s'arrêtent aux échelles supérieures, Nuts3 pour l'Europe, et au niveau des grandes villes, pour celle de l'OMS, elles ne descendent pas à l'échelle infra-communale qui nous intéresse.

Nous pourrions faire appel aux données issues de la Télédétection qui sont à une résolution de 30m pour la bande 1 de Landsat 8 Oli (*Operational Land Imager*) qui a été conçue pour suivre les aérosols, mais il me semble que le volume de certains polluants pris à une certaine altitude ne correspond pas forcément à ce qui est réellement respiré à 1.5 m du sol par les habitants, cela discriminerait moins, aussi, les différences de pollution entre les quartiers. De plus, les aérosols mesurés au-dessus d'un point peuvent provenir d'autres lieux. Néanmoins, cela pourrait être utile dans le cas où, pour certains espaces, nous ne parvenons pas à obtenir des données satisfaisantes au sol.

Mais, le mieux sera de faire encore appel aux Associations Agréés pour la Qualité de l'Air (ASQAAs) comme nous l'avons fait pour le pays de Martigues, car il s'agit d'un réseau mondial qui peut fournir au minimum les données de 2 à 3 stations fixes par ville, et qui rendent bien compte de la diversité de la pollution, car leur localisation est très réfléchie. Les données pourront être récupérées *via* le site aqicn.org (les données sont affichées en temps réel, mais il est possible d'obtenir un historique sur 12 mois, voire sur plusieurs années, avec de véritables données autres que des indices de qualité de l'air synthétiques).

⁵⁰ Comité de Protection des Personnes

⁵¹ Dans le cas où cela ne serait pas mentionné dans les articles

L'accès aux données socio-économiques à l'échelle mondiale et à un niveau infra-communal pose quant-à lui moins de problème.

Il est difficile d'énoncer déjà le nombre d'espaces par ensembles géographiques (Europe, Asie, etc) nécessaires pour avoir une puissance statistique suffisante, car cela dépendra des articles que nous trouverons, néanmoins, dans l'idéal, la limite de 30 par type d'espace dans chaque ensemble nous paraît un nombre raisonnable pour pouvoir dégager des caractéristiques suffisamment récurrentes en vue de la théorisation.

Ce découpage par grands ensembles géographiques est bien entendu artificiel, étant entendu que par exemple les grandes villes post-modernes tendent à se ressembler quel que soit le lieu, il est utilisé ici juste pour signaler que nous ne nous limiterons pas à un ensemble géographique précis, afin de voir si par exemple des variations liées au niveau de développement, à la culture, ou bien encore à l'urbanisme et à l'architecture peuvent jouer. De plus, le type de pollution peut être différent d'un espace à l'autre suivant le type d'énergie utilisée, le transport sur de longues distances de poussières désertiques etc.

Aux données environnementales et socio-économiques nous ajouterons cette fois les facteurs géographiques qui à notre connaissance vont amplifier l'effet de cette pollution, c'est à dire par exemple (liste non exhaustive) :

- . la topographie (mesures des pentes),
- . la densité de population,
- . le nombre de véhicules qui transitent dans cet espace, estimable grâce à la densité des routes (à considérer si des niveaux de pollution à fine échelle ne peuvent être obtenus), la densité routière pondérée peut être utilisée comme indicateur indirect de l'exposition à la pollution liée au trafic, à utiliser dans des circonstances où les mesures directes des niveaux de polluants ne sont pas réalisables (Rose et al., 2009),
- . la vitesse du vent qui va diluer avec une plus ou moins grande efficacité les polluants dans l'atmosphère, leur direction dominante (depuis les émissions vers les habitations),
- . la température dans la couche d'air (le nombre de cas d'inversion thermique/an),
- . la densité de végétation (NDVI),
- . la configuration du *built environment* (bâtiments et routes), c'est à dire par exemple la configuration des artères entre elles, par rapport aux vents dominants, ou bien à d'éventuelles brises, le gabarit des rues (mesurable grâce à l'indice de Landsberg) qui va piéger (rues canyon), ou au contraire faciliter la dispersion des polluants (Huang, 2022, Maignant, 2007).

En effet, si le flux de polluants émanant des pots d'échappement est homogène, sa diffusion spatiale se déploie en fonction des paramètres météorologiques (essentiellement de la vitesse

et de la direction du vent) et de la morphologie urbaine qui peut varier énormément au sein d'une ville entre les différents quartiers qui la composent. En effet, à émissions polluantes identiques, la disposition des bâtiments par rapport aux écoulements (champs de vents) et l'écartement entre ceux-ci vont amplifier ou amoindrir ces émissions, augmenter l'inconfort respiratoire, et engendrer des difficultés pour toutes les personnes atteintes de maladies chroniques notamment en cas de fortes chaleurs. Ainsi, le professeur Besancenot a montré une surmortalité plus importante dans les quartiers au bâti resserré durant des vagues de chaleur, et dès que le seuil de 200 000 habitants était dépassé (Besancenot, 2002).

. la forme de l'espace d'étude (lien connu entre par exemple la forme d'une ville et son niveau de pollution) (Ahn et al., 2022, Bereitschaft et al., 2013).

. le feuillage des arbres et l'espacement entre eux qui peuvent avoir le même effet que la morphologie des artères (Selmi et al., 2013) (s'il est possible de le déceler avec des images à haute résolution).

. l'occupation du sol (types de terres agricoles pour tenter d'évaluer la quantité de pesticides épanchée, types d'usines, voisinage, distance vis-à-vis du bâti, etc.).

Nous ajouterons également des paramètres météorologiques afin d'avoir par exemple le nombre de jours de précipitations (qui viennent plaquer la pollution au sol), le taux d'humidité relative, la température moyenne annuelle, estivale, le nombre de vagues de chaleur afin d'avoir une idée de l'inconfort respiratoire associé à la pollution, et dans le but de modéliser à l'aide du monitoring bayésien l'impact de ces paramètres météorologiques sur les pathologies d'intérêt dans un contexte de changement climatique (vagues de chaleur).

Certains polluants seront particulièrement surveillés telles que les particules fines et ultrafines, qui sont déjà très impactantes sur la santé respiratoire des populations, mais qui le seront encore plus avec le transport de poussières durant des phases de sécheresse. Ainsi, « *Particulate matter (PM) is both a major driver of climate change and a source of toxicity for health* » (Mukherjee et al., 2017).

A noter que nous ne sommes pas dans une démarche de modélisation de type LUR (Land Use Regression) (Habermann et al., 2015) puisque nous allons nous appuyer sur des données de pollution issues des ASQAAs, et que les données sanitaires seront connues (extraites des articles, par conséquent avérées).

Il ne faudrait pas penser que de simples extractions d'images satellite retravaillées suffiront à l'objectif visé, l'expertise du chercheur est cruciale, elle ne se décrète pas, car elle est fondée sur de nombreuses lectures, des analogies, voire des intuitions.

Ainsi, c'est la combinaison des facteurs, leur interprétation qui font sens, et non leur prise en compte de manière séparée. Par exemple, la vitesse des vents en un lieu va permettre ou pas la dilution dans l'atmosphère d'éventuels polluants, mais également leur diffusion vers d'autres zones (selon la direction des vents dominants), ou bien l'alternance de brises de terre

ou de mer va diffuser le soir vers l'intérieur des terres (brise de mer) la pollution accumulée le jour sur le littoral...littoral qui a organisé, contraint les flux selon des voies de circulation parallèles au rivage (anisotropie). Tandis qu'un autre littoral, en raison cette fois de son avancée en mer et de sa position géostratégique va être choisi pour abriter des industries pétrochimiques avec toutes les conséquences environnementales et sanitaires qui peuvent en découler.

Dans un premier temps, l'ensemble de ces variables qui peuvent toutes être quantifiées, (si nous ajoutons des variables qualitatives, elles peuvent dans tous les cas faire l'objet d'un codage qui les rendent quantitatives) seront intégrées au sein d'une même modélisation bayésienne, et selon la méthode développée dans la partie relative à la preuve de concept de cette HDR.

Un réseau bayésien sera réalisé par domaine de recherche, par exemple un sur la pollution atmosphérique en contexte urbain, un autre relatif à la pollution industrielle, ou bien rurale, concernant des espaces mixtes, mais dans des espaces géographiques différents (europe, asie, afrique etc.). Ils contiendront bien entendu des variables spécifiques, les polluants incriminés en milieu urbain n'étant pas les mêmes que ceux dont on recherche les effets en contexte industriel ou ruraux, d'où la spécificité du réseau bayésien pour chaque domaine. Egalement, si nous avons évoqué les maladies respiratoires ou bien cardiovasculaires comme pathologies d'intérêt, en contexte rural nous nous focaliserons sur d'autres pathologies, comme par exemple les agénésies transverses des membres qui semblent peu étudiées, et dont le rôle cette fois ci de l'eau est suspecté *via* des résidus de pesticides qui peuvent s'accumuler en fin de canalisations.

Nous devrions aboutir là encore à des résultats inédits concernant l'effet propre, mais surtout combinés des différents facteurs précédemment évoqués sur les pathologies respiratoires, cardiovasculaires, ou autres, et leurs variations pour les différents espaces d'étude.

Puis, dans un second temps, à partir des analyses de sensibilité, et en fonction de l'ensemble de ces facteurs nous attribuerons un **potentiel pathogène** à chacun de ces espaces dans le but de les **catégoriser**.

Attardons-nous un moment sur cette notion intéressante de potentiel. Il y a derrière cette notion quelque chose qui est en puissance, en action, qui est fonction de certaines conditions. Ce sont les physiciens qui ont beaucoup travaillé sur ce sujet avec la question du potentiel thermodynamique, électrique, ou bien encore électromagnétique. Les philosophes l'ont abordé quant à eux plutôt sous l'angle de l'inné et de l'acquis (est ce que le potentiel serait quelque chose d'inné ?), tout comme les psychologues (selon l'angle de l'acquis avec l'éducation). Cette notion concerne également les artistes qui peuvent voir à travers un matériau, comme les veines d'une dalle de marbre un formidable potentiel en vue de la création de sculptures.

Au sens littéral, le potentiel correspond à quelque chose qui est en puissance, ou bien qui a la *capacité à* (mais qui ne s'est pas encore réalisé). Le potentiel est donc lié aux notions d'incertitude, (ce qui est en puissance se réalisera t'il un jour : oui, non, degré de vraisemblance), de possibilisme (est ce qu'il est possible que cela se réalise un jour, oui, non quand), et bien entendu de probabilité (avec quelle probabilité [0 -> 1] cela a t'il une chance de se réaliser).

Il y a également l'idée d'une énergie en mouvement, une tension, une charge, une capacité de charge, une montée en puissance vers le potentiel, on entend aussi des variations de potentiel, ou *a contrario* des lignes équipotentielles.

Le potentiel va de pair avec la notion d'émergence, il lui est antérieur. Du potentiel émerge un certain nombre de caractéristiques capables d'influer sur l'état de santé des populations. Ainsi, c'est quand toutes les causes sont réunies, nous pourrions dire lorsque les conditions sont (dé)favorables que le potentiel d'un espace va émerger sur le plan sanitaire. Mais, c'est peut-être plus complexe que cela, et nous reviendrons plus loin sur cette notion d'émergence.

Pour nous, le potentiel pathogène d'un espace équivaut à la somme des variables influentes sur la santé (polluants, défaveur sociale, etc...) dans le sens de l'« exposome » **eX** qui pèsent d'une certaine charge sur la santé des individus, en fonction de leurs interactions et si certaines conditions (de dépassement de seuil, de vulnérabilité) sont réunies. Le potentiel est donc fonction de propriétés intrinsèques (niveau, dangerosité, interactions avec d'autres variables), mais aussi de sa position dans l'espace (proximité de sources polluantes, contexte socio-économique).

Connaissant en quelque sorte la charge nous pourrions en déduire les pathologies, cela permettrait d'être moins dépendants de données sanitaires coûteuses en termes de temps (autorisations CNIL en France), et ce serait une aide précieuse en matière d'aménagement du territoire, et de politiques sanitaires.

La notion de potentiel est issue des Sciences Physiques. L'énergie potentielle électrique est très proche de l'énergie gravitationnelle, et donc des modèles gravitaires que nous connaissons bien en Géographie.

Examinons à présent comment cette charge pourrait être évaluée.

Soit *a* possède certains caractéristiques (environnement, socio-économie, *built environment*) qui lui donnent une certaine « charge », un certain potentiel *U*, la force s'exerçant sur *a* ne dépend que des propriétés intrinsèques de *a* et de sa position dans l'espace (proximité aux sources polluantes, sous le vent des émissions) ce qui va induire un potentiel spatialisé que nous allons appeler un **géopotential**⁵² qui va produire des différences de potentiel dans l'espace.

⁵² Concept utilisé en météorologie pour désigner un potentiel de pesanteur

Soit le Potentiel $U = \sum$ variables qui influent sur la santé⁵³ en 1 point donné de l'espace a , ce potentiel (s'il est actif) éloigne la population qui vit dans cet espace d'un « bon » état de santé.

L'énergie potentielle (EP) ou Energie Potentielle Gravitationnelle (EPG) s'écrit alors sous la forme :

$$Epg = mgh$$

Où :

m = la masse

Cette masse peut représenter par exemple la dangerosité, la toxicité d'un polluant.

g = la force de gravitation au niveau du sol qui pourrait correspondre à un état en bonne santé d'un individu, ou de plusieurs individus au sein d'un quartier,

h = la hauteur, soit l'ensemble des facteurs qui vont amplifier la dangerosité du polluant en question en a .

Lorsque le potentiel est en quelque sorte libéré (lorsque les conditions sont réunies), le potentiel pathogène se manifeste par des maladies.

Dès lors, nous pouvons observer le potentiel, *via* des différences de potentiel (car l'énergie potentielle n'existe pas en elle-même, elle ne peut pas s'observer directement), suivant différentes conditions de h pour un même polluant :

$$U(a, b) = V(a) - V(b)$$

U = différence de potentiel entre a et b

V = potentiel de pathologies en a et b

⁵³ Derrière cette somme il y a bien entendu des pondérations, de la standardisation, etc.

Le géopotential définit donc la capacité qu'aurait un espace donné à « entrer » dans le corps, soit parce que cet espace possède une charge environnementale particulièrement délétère pour la santé (pollution), soit parce qu'il est habité par des populations sensibles (dites à risques) tels que les enfants, les personnes âgées, ou en situation de précarité. Inversement, la charge d'un espace peut être positive s'il permet de maintenir (**énergie conservative**) un bon état de santé (*walkability*). La charge environnementale peut être très variée aussi bien en quantité, dangerosité, que suivant les compartiments au sein desquels elle est émise (air, eau, sol), et concerner des milieux très divers : urbain (pollution de l'air due au trafic de véhicules), périurbain (pollution de l'air due à la proximité d'industries de type ICPE), et rural (pesticides).

Le potentiel peut être là en puissance, mais il n'a pas encore émergé, il n'est pas en quelque sorte activé, car des forces contraires (favorables) maintiennent la population dans un relatif « bon » état de santé.

Connaissant maintenant ce potentiel de pollution fondé sur la catégorisation d'eXpaces, nous pouvons le comparer avec l'état de santé par exemple respiratoire de la population. Si le nombre de personnes ayant un problème respiratoire est plus élevé que le géopotential ne le voudrait cela peut signifier que le mode de vie (tabac, alimentation), la qualité de l'habitat (air intérieur) sont à investiguer, ou bien la potentialisation d'éléments entre eux (dans le sens cette fois d'une amplification), ou alors que nous avons oublié d'intégrer des facteurs qui influent sur les pathologies. C'est aussi l'intérêt de cette comparaison : faire émerger des facteurs de second ordre qui semblent jouer un rôle plus important que ce que l'on avait préalablement imaginé.

Inversement, si l'état de santé est meilleur que ne le suggère le géopotential, il y aurait dans cet espace des éléments qui joueraient en faveur de l'atténuation de la pollution, cela peut être dû à une socio-économie protectrice (un meilleur accès aux soins), à des politiques publiques sanitaires qui portent leurs fruits etc..

A l'heure actuelle, nous ne sommes pas en capacité de caractériser rapidement ce qui se passe dans un espace. Demain, nous pourrons raisonner en termes de potentiel d'un espace à impacter la santé, et cela changera la donne. Généralement, nous mesurons le niveau de pollution d'un espace donné et nous le mettons en relation (ou pas) avec la santé, par exemple respiratoire, des individus qui y habitent. **Demain, avec cette notion de potentiel, nous partirons cette fois de l'espace d'étude, et nous en déduirons rapidement un certain potentiel pathogène.**

Bien entendu, le géopotential est dynamique, nous pouvons le suivre dans le temps et voir comment il évolue (maintien, accentuation, diminution) à la faveur de changements dans les causes.

Nous pouvons donc diagnostiquer l'espace dans ses multiples dimensions (environnementales, socio-économiques, *built environment*..) au regard de la santé des individus qui y habitent et en déduire des eXpaces.

Cela nous permettrait de répondre à des questions de ce type : Connaissant les pathologies au sein d'une population donnée, quelle est la probabilité pour cette population d'avoir été exposée à des polluants, avec quelle intensité elle a été exposée, et sur combien de temps ? Ou inversement, disposant cette fois de mesures précises de la qualité de l'air, quelle est la proportion d'individus risquant de développer une pathologie particulière ? Pourquoi une exposition *a priori* similaire ne produit pas les mêmes effets ? Quels sont les autres facteurs déterminants qui semblent alors intervenir ? Est-ce la multiplicité de ces déterminants ou bien leurs combinaisons qui présentent le plus de danger pour la santé des individus ? Observons-nous au sein de la ville, des discontinuités spatiales de prévalence ou bien d'incidence qui tendraient à suggérer qu'il y a dans ces espaces des éléments pathogènes que nous devons découvrir ? Que peuvent avoir en commun des patients atteints de la même pathologie ? Quels attributs dans l'espace sont les plus importants relativement à une pathologie donnée ? Quels changements dans les paysages, l'urbanisme, ou dans le climat induisent l'émergence ou la re-émergence de conditions favorables à des pathologies ?

Enfin, **une fois que l'on aura obtenu grâce à la modélisation bayésienne les eXpaces ils pourront être injectés cette fois dans un réseau neuronal convolutif** (Le Cun, 2015) afin que par sa capacité auto-apprenante (apprentissage auto-supervisé), il soit en mesure, **en « visualisant » un type d'eXpace via une image satellite, d'indiquer de manière automatique un potentiel de pathologies au sein de cet espace.**

La connaissance obtenue sur les eXpaces devrait aider à conduire des études omiques ciblées qui permettront de valider en même temps les eXpaces. Ce faisant cela fera sauter le verrou scientifique qui existe dans les études écologiques (*ecological fallacy*), à savoir que nous aurons à la fois des éléments provenant de l'espace de vie (écologie), mais également des mesures au niveau des individus vivant dans cet espace. Cela permettra aussi de faire ressortir le poids des déterminants de santé individuels. **Grâce à la métabolomique les conditions d'expérimentation de l'espace géographique sont alors réunies.**

Une fois le potentiel évalué nous devons porter attention à ses premiers signes d'émergence.

3.3 Du potentiel à l'émergence

Du potentiel, plus exactement des causalités contenues dans le potentiel, peuvent émerger, dans l'espace, des pathologies.

L'émergence est une notion issue cette fois de la psycho-biologie (émergence des émotions, de la conscience...). **Un phénomène est dit émergent lorsqu'on ne peut pas prédire son apparition à partir de la seule connaissance du système au sein duquel il émerge**, pour autant, des causalités sous-jacentes expliqueraient en partie l'émergence du phénomène, rien n'apparaît *ex nihilo*.

L'émergence peut être qualifiée de faible si les nouvelles propriétés apparues dans le système sont le résultat d'interactions à un niveau élémentaire, et au contraire, de forte si les nouvelles propriétés ne sont pas réductibles, c'est-à-dire si **la totalité est supérieure à la somme de ses éléments**. Dans ce cas, il est difficile de rattacher l'émergence à des causes spécifiques comme l'agrégation des potentialités existantes à un niveau micro, ce qui fait dire à Mark Bedeau (1997) « *Although strong emergence is logically possible, it is uncomfortably like magic* ». Avec l'émergence le tout n'est parfois pas réductible à ses parties, d'où la difficulté à prédire le comportement de systèmes complexes "As long as there exist possible configurations of our universe that have not yet occurred, we must be realistic about our ability to predict future dynamics on the basis of formal systems, when the dynamics may be influenced by emergence". Certains auteurs évoquent alors la capacité auto-organisationnelle de ces systèmes. Ainsi, pour Goldstein (1999) l'émergence peut se définir comme l'apparition de structures nouvelles et cohérentes de modèles et de propriétés durant **le processus d'auto-organisation de systèmes complexes**.

Une des hypothèses majeures concernant le phénomène d'émergence est que l'interaction de l'ensemble des éléments de chaque niveau d'organisation serait responsable de l'émergence de nouvelles propriétés fonctionnelles à un niveau supérieur. Autrement-dit, chaque propriété fonctionnelle émergente à un certain niveau serait la résultante de nombreux facteurs d'émergence, provenant de différents niveaux inférieurs. De plus, au fur et à mesure que les niveaux d'organisation sont gravés, l'augmentation des différents types de structure induit une augmentation exponentielle du nombre de combinatoire possible, et donc du nombre de conditions favorables au phénomène d'émergence, puisque l'émergence serait due entre autres à des **processus de synergie, de captation, d'accumulation, de relations causales, voire de rétroactions** au sein du système.

D'un ensemble de causalités interdépendantes émergeraient ou pas une pathologie, mais selon quelles causalités ou plutôt quels enchaînements de causalité, et avec quelles proportions, c'est ce à quoi le formalisme bayésien répond (probabilités) ? Ainsi, si les causes sont connues et réunies nous aurons une idée de la probabilité d'émergence de la pathologie d'intérêt.

L'émergence peut également apparaître lorsque seulement quelques individus obéissant à des règles très simples vont interagir entre eux et former alors en tant que groupe un système complexe, difficilement prédictible. Cette notion d'émergence fait également appel à la **théorie des seuils**, en effet ce n'est qu'après avoir franchi un certain seuil que l'émergence se dévoile, il y a derrière également la question des variables latentes, celle des faibles doses, des nanomatériaux (Marano, 2021), et bien entendu celle des effets cocktails.

Dès lors, en raison de la complexité des processus physiologiques au sein de nos organismes que l'on peut apparenter à une somme de systèmes complexes, en raison aussi de la variabilité interindividuelle, de la susceptibilité génétique, etc, il n'est pas sûr que du déterminisme causal contenu dans le potentiel découle une cascade de causes-conséquences de manière absolue.

Néanmoins, nous sommes persuadés que la démarche présentée, possède des capacités heuristiques dans le domaine qui nous intéresse, de manière relative. En effet, connaissant mieux les causes, dans une perspective de santé publique, les autorités en charge de la santé, les aménageurs, les urbanistes, les collectivités territoriales pourront peut-être être en mesure de limiter, contenir leurs émergences, afin de favoriser des espaces plus sains (Simos & de Leeuw, 2017).

Tout l'objectif de la santé publique étant de prévenir l'émergence de pathologies.

3.4 La prévention de l'émergence

Même si l'émergence ne s'est pas encore (exposition par exemple sur 20 ans avant de développer un cancer) ou suffisamment manifestée (cas isolés), nous avons vu qu'elle peut être là, en puissance, dans le potentiel. C'est donc le principe de précaution qui doit s'appliquer. Le principe de précaution est intimement mêlé à l'incertitude. Ainsi, alors que la science s'établit sur des faits reconnus, partagés, entre les pairs, le principe de précaution peut être pris même si les faits ne sont pas encore clairement établis, et sans attendre qu'ils le soient "*When an activity raises threats to the environment or human health, precautionary measures should be taken, even if some cause-and-effect relationships are not fully established scientifically*" (Raffensperger et al., 1999).

Si nous prenons la définition qui est donnée du principe de précaution, il repose sur l'**anticipation** dans un contexte de connaissances incertaines d'un ou plusieurs effets produits par un ou plusieurs évènement(s), ou substance(s) qui porterai(en)t atteinte à l'intégrité physique des personnes. Les actions prises dans ce cadre ont pour objectif l'atténuation, voire dans le meilleur des cas la suppression du risque, alors que les enjeux sont forts (nombreuses victimes ou personnes potentiellement vulnérables). « *This principle provides a moral justification for acting even though causation is unclear* » (Ricci et al., 2011), c'est là tout son intérêt, et sa force, même si le résultat est lui-même incertain : « *Precautionary principles are the foundations for policy when it has to deal with weakly understood causes of potential catastrophic or irreversible events, and where protective decisions require certain and costly policy interventions that may not solve the problem that they are designed to correct* » (Ricci et al., 2013).

Ce qui fait dire à Hill (1965) *"I have no wish, nor the skill, to embark upon a philosophical discussion of the meaning of 'causation'. The 'cause' of illness may be immediate and direct, it may be remote and indirect underlying the observed association. But with the aims of occupational, and almost synonymously preventive medicine in mind the decisive question is whether the frequency of the undesirable **event B will be influenced by a change in the environmental feature A**. How such a change exerts that influence may call for a great deal of research. However, before deducing 'causation' and taking action we shall not invariably have to sit around awaiting the results of that research. The whole chain may have to be unravelled or a few links may suffice. It will depend upon circumstances* ».

Les travaux de Michel Setbon (2004) suggèrent que le risque est surestimé lorsqu'il est subi (sentiment de non maîtrise), par rapport à un risque « choisi » (mode de vie). Si le risque est évalué de manière globale, populationnelle, par les organismes (1 chance sur 1 million), la potentialité de devenir soi-même une victime rendrait toute exposition inacceptable, ce qui fait que le principe de précaution est généralement bien accepté par la population.

Mais, parfois, la prévention passe à côté de sa cible, les gens n'ont pas conscience du risque, ou bien ils sont trop préoccupés par des conditions matérielles à court terme qui les en détournent. Ainsi, ce ne sont pas forcément les personnes qui seraient le plus exposées qui en sont le plus conscientes, ou si elles le sont, elles peuvent être tentées par une forme de déni, car elles sont captives de leur espace de vie et/ou de travail.

Nous pouvons espérer avec la méthodologie développée ici (résumée dans la figure n°37), et la théorisation qui à moyen terme en sortira, une prévention éclairée, grâce à une meilleure connaissance des causes révélées par les bayésiens et la catégorisation en eXpaces.

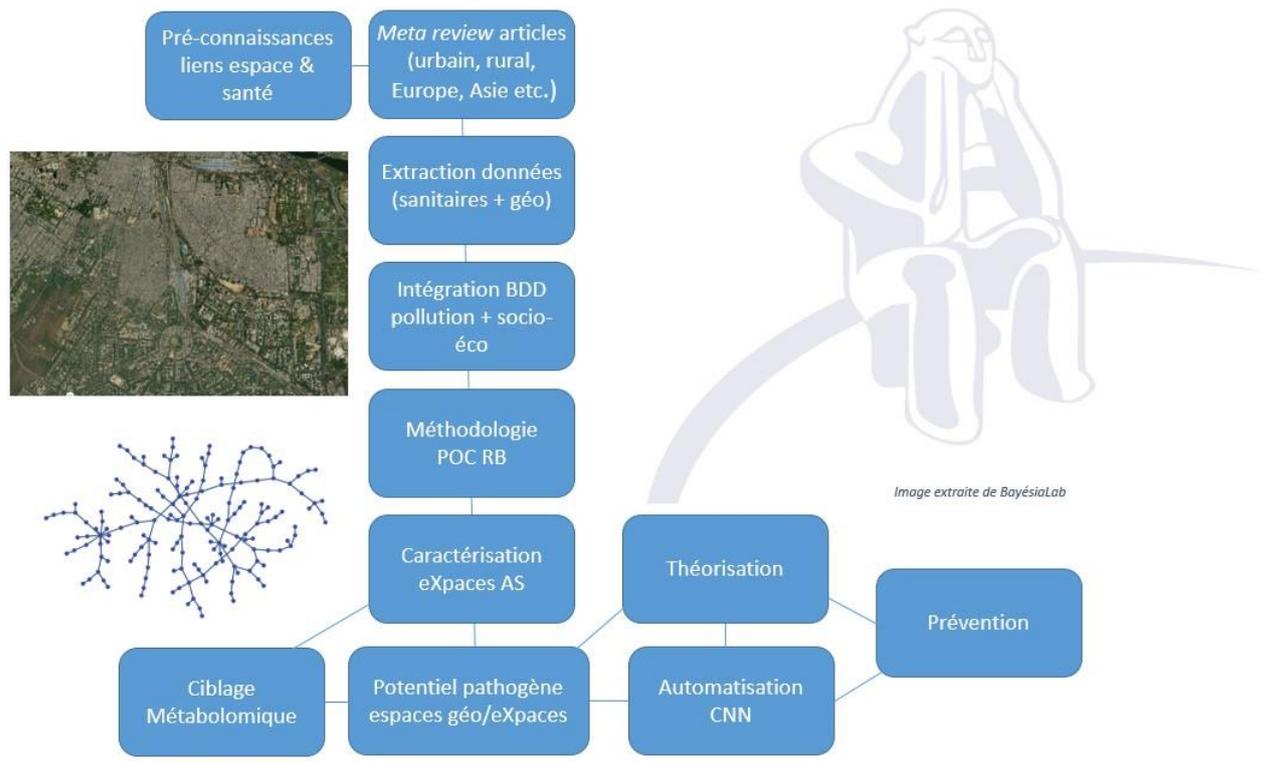


Figure 37. La méthodologie développée pour caractériser des eXpaces

. Conclusion

La complexité de notre monde ne peut qu'amener à la recherche des causes afin de faire face aux nombreux défis qui nous attendent, et prioriser nos efforts dans les lieux les plus problématiques. En espérant que la Géographie saura s'emparer des réseaux bayésiens, les faire siens⁵⁴, alors qu'ils sont déjà très utilisés dans d'autres disciplines, et depuis longtemps, pour tout ce qui touche de près ou de loin au risque (Weber, 2016). Nous avons vu tout leur apport, utilisés seuls, ou bien lorsqu'ils seront combinés à d'autres outils.

À travers l'étude d'un cas concret : la pollution atmosphérique à proximité d'un complexe pétrochimique nous avons démontré l'intérêt de faire appel à la théorie de la causalité et aux réseaux bayésiens. Ils ont permis en mobilisant plus de 171 variables de natures différentes (environnementales, socio-économiques, écologiques, et sanitaires) considérées simultanément de révéler : 9 relations entre polluants et pathologies, des effets de seuils, le rôle de polluants spécifiques, le poids parfois de la socio-économie, notamment sur les maladies cardiovasculaires chez les plus de 65 ans, ainsi qu'une pollution diffuse aux PCBs au sein du territoire du Pays de Martigues.

Nous allons à présent transposer la méthodologie employée dans la preuve de concept non seulement à d'autres espaces similaires (industriels), mais nous l'élargirons à des espaces urbains, ruraux, mixtes, pris dans différents contextes géographiques (Europe, Asie, Afrique, Amérique latine...). Par ce biais, nous espérons nouer de nouvelles coopérations internationales avec mon laboratoire.

Nous proposons également un **nouveau concept**, celui d'**eXpace** afin de catégoriser les espaces géographiques du point de vue de leurs caractéristiques intrinsèques à influencer sur l'état de santé des habitants qui y résident.

L'ensemble de ces caractéristiques conférant un certain potentiel pathogène à ces eXpaces (**géopotential**), qui est nous l'avons vu quantifiable. Comme au temps de Paracelse ce qui nous importe ce sont **les signatures [de l'espace]**.

Nous avons présenté une démarche méthodologique fondée sur le triptyque **causalités/potential/prévention** qui va **ultimement coupler réseaux bayésiens et réseaux de neurones convolutionnels**, afin de systématiser par reconnaissance d'images satellites l'attribution d'un potentiel pathogène à un eXpace, et ce faisant nous plaidons pour **la reconnaissance d'une dose [géographique] réponse en santé environnementale pour une meilleure prévention**.

⁵⁴ A ma connaissance nous ne sommes pas nombreux à les utiliser en Géographie, il faut citer les travaux de Giovanni Fusco et de son équipe sur les liens par exemple entre les caractéristiques urbaines et la valeur immobilière (Venerandi, Fusco et al., 2019), ou bien sur le couplage avec des réseaux de neurones (Fusco, Pérez J, 2019)

La méthodologie présentée doit nous permettre de cheminer à moyen terme vers une **théorie géographique** sur ce sujet, transposable par d'autres chercheurs pour définir des eXpaces qui seront par exemple fortement impactés par le changement climatique, ou bien soumis à des catastrophes naturelles, ou autres. Il serait d'ailleurs intéressant de combiner ces différents aspects entre eux, afin là encore de cibler les espaces sur lesquels l'attention, la prévention, les moyens doivent être mis en priorité. Les connaissances accumulées sur le fonctionnement des systèmes de villes en réseau, et notamment sur leur évolution potentielle à une échelle *meso*, s'avèrera très utile (Rozenblat, 2022). Nous pourrions également étudier la capacité des *smart cities* à générer moins d'eXpaces.

Nous espérons que ce voyage mené à la lecture de cette HDR dans le monde de la géographie de la santé, permettra aux chercheurs de positionner leurs travaux dans ce domaine de recherche en pleine expansion, passionnant, et qui contribue selon nous à un certain renouveau de la géographie, avec notamment toute l'ingéniosité avec laquelle les chercheurs interrogent l'espace, et l'intérêt que ces interrogations suscitent parmi la communauté scientifique, et ce, même si ce lien entre espace et santé est particulièrement difficile à appréhender. Nous avons vu également le rapport étroit qui existe entre les nouvelles technologies au sens large, et une nouvelle façon de considérer l'espace, plus proche des individus. Bien qu'il n'y ait aucun conflit entre les deux, nous avons voulu donner un peu plus de place aux géographes (de la santé) par rapport aux épidémiologistes. Enfin, nous espérons inciter de jeunes géographes à travailler dans ce domaine tant il reste encore à faire.

En arrivant au bout de l'écriture de cette HDR, je me dis qu'un nouveau chantier stimulant s'ouvre, il va orienter mes recherches pour les années à venir, ce n'est pas un aboutissement, au contraire, tout commence !

Je suis consciente qu'il s'agit là d'un programme ambitieux que je ne pourrais certainement pas mener à bien toute seule, mais en collaboration avec d'autres collègues, déjà identifiés, que ce soit en géomatique (télédétection, Sébastien Gadal), en IA (réseaux de neurones convolutionnels, Frédéric Precioso), ou bien encore en métabolomique (Thierry Poucher). Je pourrais également compter sur Didier Josselin, DU de l'UMR ESPACE et son appétence pour la théorisation, comme d'ailleurs beaucoup de mes collègues de cette Unité, eux aussi issus de ce courant de la géographie théorique et quantitative, encore plus que jamais indispensable à l'heure du *Big Data* !

L'aide de doctorants sera également précieuse, notamment celle de Charlotte Sari, ingénieure d'étude à l'EHESP sur les questions d'Urbanisme Favorable à la Santé, d'Anais Squarta sur les liens entre les cas d'agénésies transverses des membres et exposition aux pesticides, et de Clémentine Kerdar, ARC (Attachée de Recherche Clinique) au CHU de Martignes sur de potentiels espaces diabétogènes en contexte industriel.

Elles représentent les premières pierres portées à cet édifice. Nous serons sans doute rejointes en cours de route par, je l'espère, de nombreux doctorants, qui viendront à leur tour apporter leur propre contribution.

. Références

- Académie Nationale de Médecine. L'expert judiciaire face aux éléments médicaux de l'imputabilité et au lien de causalité en responsabilité chirurgicale *Bull. Acad. Natle Méd.*, 2005, 189, no 5, 979-992 Séance du 10 mai 2005.
- Adam M., Schikowski T., Carsin AE., Cai Y., Jacquemin B., Sanchez M., et al. Adult lung function and long-term air pollution exposure. ESCAPE: a multicentre cohort study and meta-analysis. *Eur Respir J* 2015;45:38–50.
- Afrite A., Allonier C., Com-Ruelle L., Le Guen N. L'asthme en France en 2006 : prévalence, contrôle et déterminants, Irdes, Janvier 2011, 122 p.
- Agency for Toxic Substances and Disease Registry (ATSDR). 2012. Toxicological Profile for Vanadium. Atlanta, GA: Department of Health and Human Services, Public Health Service.
- Aken BV., Correa PA., Schnoor JL. Phytoremediation of Polychlorinated Biphenyls: New Trends and Promises. *Environ Sci Technol* 2010;44:2767–76.
- Allard P., Fox D., Picon B. (2008). *Incertitude & environnement*, Société d'Ecologie Humaine, Edisud, 479 p.
- Allen B., Cohen A., Ferrier Y., Lees J. 2017. Fos EPSEAL, Étude participative en santé environnement ancrée localement sur le front industriel de Fos-sur-Mer et Port-Saint-Louis-du-Rhône, Rapport final, Marseille : Centre Norbert Elias.
- Amar E. Malformations et facteurs environnementaux. *Revue de Médecine Périnatale*, 2017, vol 9, p.73-80. <https://doi.org/10.1007/s12611-017-0415-y>
- Amoatey P., Sicard P., De Marco A., Khaniabadi Y.O. 2020. Long-term exposure to ambient PM_{2.5} and impacts on health in Rome, Italy. *Clinical Epidemiology and Global Health* 8: 531-535.
- Amrhein V., Trafimow D., Greenland S. 2018 Inferential statistics as descriptive statistics: there is no replication crisis if we don't expect replication. *PeerJ Preprints* 6e26857v4 <https://doi.org/10.7287/peerj.preprints.26857v4>
- Ahn H., Lee J., Hong A. 2022. Urban form and air pollution: Clustering patterns of urban form factors related to particulate matter in Seoul, Korea, *Sustainable Cities and Society*, 81.
- Asendorpf J.B. (1996) *Modélisation des processus de développement dans Le manuel de la dynamique et des processus de la personnalité*, Springer, 363 p.
- Anses. (2022). Utilisation de micro-capteurs pour le suivi de la qualité de l'air intérieur et extérieur. Rapport d'expertise, 318 p.

- Anses. (2011). AVIS de l'Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail relatif à l'interprétation des résultats de l'étude nationale Anses/InVS d'imprégnation aux PCB des consommateurs de poissons d'eau douce, 8p.
- ARS. (2012) L'indice de désavantage social (IDS) : application aux territoires de Provence Alpes Côte d'azur, Direction de la stratégie et prospective, Département observations & études, ARS Paca, la santé observée, n° 15 août 2012, 6 p.
- Aschan-Leygonie C., Baudet-Michel S., Harpet C., Augendre M., Lavie E., Grésillon E., et Hechinger M. Comment évaluer l'exposition aux pesticides de l'air en population générale ? Enseignements d'une revue bibliographique, *Cybergeog: European Journal of Geography*, Environnement, Nature, Paysage, document 729, mis en ligne le 11 juin 2015, DOI : <https://doi.org/10.4000/cybergeog.27056>
- Aschan-Leygonie C & Baudet-Michel S. Les inégalités de santé respiratoire entre les villes. Dimensions socio-économiques et environnementales, *Métropolitiques*, 9 juin 2014. URL : <https://metropolitiques.eu/Les-inegalites-de-sante.html>
- Atkinson RW., Carey IM., Kent AJ., Van Staa TP., Anderson HR., Cook DG. Long-term exposure to outdoor air pollution and the incidence of chronic obstructive pulmonary disease in a national English cohort. *Occup Environ Med* 2015; 72:42–8.
- AtmoSud, Etude Polis : Plan d'Observation des Polluants d'Intérêt Sanitaire, 2016, <https://www.atmosud.org>
- AtmoSud, Etude SCENARII : Simulation de scénarii de pollution atmosphérique pour une Evaluation des Risques Sanitaires - Zone Etang de Berre, 2015. <https://www.atmosud.org>
- Bach J-F. Causality in medicine, *Comptes Rendus Biologies*, 342 (3–4), 2019, 55–57. <https://doi.org/10.1016/j.crv.2019.03.001>
- Barboza EP., Nieuwenhuijsen M., Ambros A., de Sa TH., Mueller N. The Impact of Urban Environmental Exposures on Health: An Assessment of the Attributable Mortality Burden in Sao Paulo City, Brazil ». *Science of The Total Environment*, vol. 831, juillet 2022, p. 154836. <https://doi.org/10.1016/j.scitotenv.2022.154836>
- Baron RM., Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol*. 1986; 51:1173–1182.
- Bayes T. (1763). *An Essay Towards Solving a Problem in the Doctrine of Chances*.

- Bedeau M.A. Weak Emergence *Philosophical Perspectives : Mind, Causation and World*, vol. 11, 1997, 375:399.
- Benachi A., Babajko S., Barjat T., Beneteau C., Botton J., et al. Premier Rapport sur les Agénésies Transverses des Membres Supérieurs (ATMS) : saisine n° 2018-SA-0242
- Ben Hassine J. La spatialisation des odeurs, Thèse, soutenue le 30/03/21 à UCA, <https://hal.archives-ouvertes.fr/tel-03741104>
- Ben Shlomo Y., Kuh D. A life course approach to chronic disease epidemiology: Conceptual models empirical challenges and interdisciplinary perspectives; *International Journal of Epidemiology*, 2002, 31 (2) 285:293.
- Bereitschaft B., Debbage K. Urban form, air pollution, and CO₂ emissions in large US metropolitan areas *The Professional Geographer*, 65 (4),2013, pp. 612S-635.
- Besancenot J-P. (2002). Vagues de chaleur et mortalité dans les grandes agglomérations urbaines. *Environnement Risques & Santé*, 1, 229-240.
- Boogaard H., et al. Long-Term Exposure to Traffic-Related Air Pollution and Selected Health Outcomes: A Systematic Review and Meta-Analysis. *Environment International*, vol. 164, juin 2022, p. 107262. <https://doi.org/10.1016/j.envint.2022.107262>.
- Bogaert P., Christakos G., Jerrett M., Yu H.L. 2009. Spatio-temporal modelling of ozone distribution in the State of California. *Atmospheric Environment* 43: 2471-2480.
- Boulos MNK., Curtis AJ., Malik PA. Musings on privacy issues in health research involving disaggregate geographic data about individuals. *Int J Health Geogr.* 2009;8:46.
- Boulos MNK. Descriptive review of geographic mapping of severe acute respiratory syndrome (SARS) on the Internet. *Int J Health Geogr.* 2004;3:2.
- Brousmiche D., Occelli F., Genin M., Cuny D., Deram A., Lanier C. 2020b. Spatialized composite indices to evaluate environmental health inequalities: Meeting the challenge of selecting relevant variables. *Ecological Indicators* 111, 106023. <https://doi.org/10.1016/j.ecolind.2019.106023>.
- Buha A., Đukić-Ćosić D., Ćurčić M., Bulat Z., Antonijević B., Moulis J-M., et al. Emerging Links between Cadmium Exposure and Insulin Resistance: Human, Animal, and Cell Study Data. *Toxics* 2020; 8:63.
- Bureau régional de l'OMS pour l'Europe Lignes directrices sur la qualité de l'air - Mise à jour 2021. Copenhague, Danemark.

- Cai Y., Zhang B., Ke W., Feng B., Lin H., Xiao J., et al. Associations of Short-Term and Long-Term Exposure to Ambient Air Pollutants With Hypertension: A Systematic Review and Meta-Analysis. *Hypertension* 2016;68:62–70.
- Calude CS., Longo G. The Deluge of Spurious Correlations in Big Data. *Found Sci* 2017; 22:595–612. <https://doi.org/10.1007/s10699-016-9489-4>.
- Candau, J. L'anthropologie des odeurs : un état des lieux. *Bulletin d'études orientales*, 2016, 64, 43-61. <https://doi.org/10.4000/beo.4642>
- Carlsten, C. Synergistic Environmental Exposures and the Airways Capturing Complexity in Humans. *Chest*, 154, (4), octobre 2018, p. 918-24. <https://doi.org/10.1016/j.chest.2018.06.004>.
- Carter P., Lagan J., Fortune C., Bhatt DL., Vestbo J., Niven R., et al. Association of Cardiovascular Disease with Respiratory Disease. *Journal of the American College of Cardiology* 2019;73:2166–77.
- Cartwright N. (2004). Causality: one word, many things. *Philosophy of sciences*. Cambridge University Press 71 (5): 805–820. doi : 10.1086/426771
- Castano-Vinyals G., Cantor KP., Malats N., Tardon A., Garcia-Closas R., Serra C., et al. Air pollution and risk of urinary bladder cancer in a case-control study in Spain. *Occupational and Environmental Medicine* 2008;65:56–60.
- Charreire H., Casey R., Salze P., Simon C., Chaix B., Banos A., Badariotti D., Weber C., Oppert JM. Measuring the food environment using geographical information systems: a methodological review. *Public Health Nutr.* 2010 Nov;13(11):1773-85. doi: 10.1017/S1368980010000753. Epub 2010 Apr 21. PMID: 20409354.
- Chasles V. (2019). Essai de définition des inégalités environnementales de santé : quelle délimitation ? Quelles matérialités ?, *Revue d'Épidémiologie et de Santé Publique*. 67. S60. [10.1016/j.respe.2018.12.046](https://doi.org/10.1016/j.respe.2018.12.046)
- Chaud P., Mrabt F., Malfait P. Investigation d'un signalement d'agénésies transverses du membre supérieur (ATMS) dans les Bouches-du-Rhône 2019 – 2020. Saint-Maurice : Santé publique France, 2021. 55 p.
- Clougherty J-E., Kinee E-J., Cardet J-C., Mauger D., Bacharier L., et al. Geography, generalisability and susceptibility in clinical trials, *The Lancet Respiratory Medicine*, 9,4, 2021. [doi.org/10.1016/S2213-2600\(21\)00046-1](https://doi.org/10.1016/S2213-2600(21)00046-1)

- Cohen A.J., Brauer M., Burnett R., Anderson H.R., Frostad J., Estep K., et al., 2017. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet* 389: 1907-1918.
- Conrady S., Jouffe L. Bayesian Networks and BayesiaLab: A Practical Introduction for Researchers, 2015.
- Daumalin X. (2003). Du sel au pétrole - *L'industrie chimique de Marseille-Berre au XIXe siècle*, 160p., Tacussel Eds.
- Dauphiné A. (2003). *Les théories de la complexité chez les géographes*. Anthropos, Economica, 248p.
- Daudé. E. Les systèmes pathogènes complexes. Modélisation et Simulation. *Revue francophone sur la santé et les territoires*, Hypotheses.org, 2017, pp.2-6.
- Descartes R. (1637). *Discours de la méthode. Pour bien conduire sa raison, et chercher la vérité dans les sciences, plus la dioptrique, les météores et la géométrie*, 537 p.
- Diebolt W., Helias A., Bidou D., Crepey G. 2005. Les inégalités écologiques en milieu urbain ? Rapport de l'Inspection Générale de l'Environnement, 68 p.
- Doiron D., de Hoogh K., Probst-Hensch N., Fortier I., Cai Y., De Matteis S., et al. Air pollution, lung function and COPD: results from the population-based UK Biobank study. *Eur Respir J* 2019;54:1802140.
- Duncan O.D. *Introduction to Structural Equation Models (Studies in Population)*, Academic Press, 1975, 1st edition 180 p.
- Elliott P., Cuzick J., English D., Stern R. Geographical & environmental Epidemiology, Methods for Small Area Studies, Oxford, 2006, 382 p.
- European Council Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. Official Journal L 152, 11.6.2008 P. 0001-0044.
- European Environment Agency 2019. Air quality in Europe - 2019 report. EEA Technical Report No 10/2019, doi:10.2800/822355.
- European Environment Agency. 2018. Air Quality in Europe - 2018 report. EEA Technical Report No 12/2018, doi: 10.2800/777411.
- European Union. (2016) *Urban Europe Statistics on cities, towns and suburbs*, Eurostat Statistical book doi: 10.2785/91120.

- Evangelou E., Ntritsos G., Chondrogiorgi M., Kavvoura FK., Hernández AF., Ntzani EE., et al. Exposure to pesticides and diabetes: A systematic review and meta-analysis. *Environment International* 2016;91:60–8.
- Ferrier J-P. 1986. Habiter/Penser la Terre : géographie et idée de l'avenir (Nouveau Moyen Age et sortie de «crise»), *Méditerranée*, 1986, 58-3 pp. 3-10.
- Fortin M., Haggerty J., Sanche S., Almirall J. Self-reported versus health administrative data: implications for assessing chronic illness burden in populations. A cross-sectional study. *CMAJ Open* 2017;5:E729–33.
- Flahaut A., Spira A. 2011 Situation et perspectives de développement de l'épidémiologie en France, en 2011, Rapport Académie Nationale de Médecine, 42p.
- Fleuret, S., Gasquet-Blanchard C., & Hoyez, A.-C. (2019). *Abécédaire de la géographie de la santé : Dimension territoriale de la santé*. Éditions matériologiques, 234p.
- Frank LD., Engelke PO., Schmid TL : Santé et conception communautaire : l'impact de l'environnement bâti sur l'activité physique. 2003, Washington, DC : Island Press.
- Fuller R., et al. Pollution and Health: A Progress Update. *The Lancet Planetary Health*, vol. 6, no 6, juin 2022, p. e535-47. [https://doi.org/10.1016/S2542-5196\(22\)00090-0](https://doi.org/10.1016/S2542-5196(22)00090-0).
- Fusco G., Perez J. Bayesian Network Clustering and Self-Organizing Maps under the Test of Indian Districts. A comparison. *Cybergeo : Revue européenne de géographie / European journal of geography*, UMR 8504 Géographie-cités, 2019.
- Gadal S, Ouerghemmi W. Multi-Level Morphometric Characterization of Built-up Areas and Change Detection in Siberian Sub-Arctic Urban Area: Yakutsk. *ISPRS International Journal of Geo-Information*, MDPI, 2019, 8 (3), pp.129. <10.3390/ijgi8030129>
- Gagnère B., Rouget F. 2018. « Études et enquêtes. Investigation d'un signalement d'agénésies de l'avant-bras dans une commune du Morbihan », Santé Publique France, CHU de Rennes : 1-43.
- Galea S. An argument for a consequentialist epidemiology. *Am J Epidemiol*. 2013 Oct 15;178(8):1185-91. doi: 10.1093/aje/kwt172. Epub 2013 Sep 10. PMID: 24022890.
- Gatrell A. Complexity theory and geographies of health: a critical assessment. *Social Science & Medicine*, 2005, 60;12:2661–71
- Gatzweiler F., Coburn J., Moran F-E., Fuders F., Howden-Chapman Ph., Ke X., Rozenblat C., Wang L., Sun W., Zhang Y. 2020. The little book of the health of cities

- Gelman A. The problems with P-values are not just with P-values. *Am Stat.* 2016; 70
https://stat.columbia.edu/~gelman/research/published/asa_pvalues.pdf
- Gnansia E., Michon L., Amar E., Estève J. Evidence for a cluster of rare birth defects in the Ain department (France). *Birth Defects Research.* 2021, p.1–11.
<https://doi.org/10.1002/bdr2.1876>
- Goix S., Periot M., Douib K. Etude INDEX. Etude d'imprégnation de la population aux polluants atmosphériques de la zone industrialo-portuaire de Fos-sur-Mer. Institut Ecocitoyen pour la Connaissance des Pollutions (IECP) 2018, 296p.
- Goldberger A., Duncan O. editors (1973) Structural Equation Models in the Social Sciences. Wiley-Interscience, New York.
- Goldstein J. 1999, Emergence as a Construct : History and Issues, Emergence : Complexity and Organisation *E:CO* Volume1 pp. 49-72
- Gong J, Geng J, Chen Z. Real-time GIS data model and sensor web service platform for environmental data management. *Int J Health Geogr.* 2015;14:2.
- Goodman SN. Introduction to Bayesian methods I: measuring the strength of evidence. *Clin Trials.* 2005;2(4):282-378. doi:10.1191/1740774505cn098oa
- Gotteland C., McFerrin B.M., Zhao X., Gilot-Fromont E., Lélou M. Agricultural landscape and spatial distribution of *Toxoplasma gondii* in rural environment: an agent-based model. *Int J Health Geogr.* 2014;13:45.
- Grandjean P. Paracelsus Revisited: The Dose Concept in a Complex World. *Basic Clin Pharmacol Toxicol* 2016;119:126–32.
- Greenland S., Pearl J., Robins J-M. The sensitivity of a sensitivity analysis. 1997 Proceedings of the Biometrics Section of the American Statistical Association. Alexandria, VA: *American Statistical Association*, 1998; 19-21.
- Greenland S., Pearl J., Robins J-M. Causal Diagrams for Epidemiologic Research, *Epidemiology*, Vol. 1, No. 10, pp. 37-48, January 1999.
- Griffith DA. A comparison of six analytical disease mapping techniques as applied to West Nile Virus in the coterminous United States. *Int J Health Geogr.* 2005;4:18.
- Grimalt J.O., Böse-O'Reilly S., Van den Hazel P. Steps forwars reduction of environmental impact on children's health, *Environmental Research*: 164 pp. 184-185. avril, 2018.

- Guo C., Zhang Z., Lau A.K.H., Lin C.Q., Chuang, Y.C., Chan J., et al. 2018. Effect of long-term exposure to fine particulate matter on lung function decline and risk of chronic obstructive pulmonary disease in Taiwan: a longitudinal, cohort study. *The Lancet Planetary Health* 2, e114–e125.
- Gurgel Helen., et Laques A-E. Dossier thématique sur des exemples d'applications en géographie, santé et environnement. *Confins*, no 42, novembre 2019. <https://doi.org/10.4000/confins.22808>
- Habermann M., Billger M., Eugensson M.H. Land Use Regression as Method to Model Air Pollution. Previous Results for Gothenburg/Sweden. *Procedia Engineering*, vol. 115, 2015, p. 21-28, <https://doi.org/10.1016/j.proeng.2015.07.350>.
- Hansson S.O. The Moral Significance of Indetectable Effects, 10 *RISK* 101 (1999).
- Handschumacher P., Hervouët J.-P. 2004. Des systèmes pathogènes à la santé publique : une nouvelle dimension pour la géographie de la santé tropicale, *Autrepart*, 1;29:47-63.
- Hippocrate. *Traité des airs, des eaux et des lieux*, Tomes 1 et 2.
- Huang C., et al. Effect of Urban Morphology on Air Pollution Distribution in High-Density Urban Blocks Based on Mobile Monitoring and Machine Learning. *Building and Environment*, vol. 219, juillet 2022, p. 109173. <https://doi.org/10.1016/j.buildenv.2022.109173>
- Han D., Rogerson PA., Bonner MR., Nie J., Vena JE., Muti P., Trevisan M., Freudenheim JL. Assessing spatio-temporal variability of risk surfaces using residential history data in a case control study of breast cancer. *Int J Health Geogr.* 2005;4:9.
- Hankey S., Marshall J-D., Brauer M. 2012. Health impacts of the Built Environment: Within-Urban Variability in *Physical Inactivity Environ Health Perspect.* 120(2): 247–253. <https://dx.doi.org/10.1289%2Fehp.1103806>
- HEI (Health Effects Institute), 2022. Systematic Review and Meta-Analysis of Selected Health Effects of Long-Term Exposure to Traffic-Related Air Pollution. Special Report 23. Boston, MA, US: Health Effects Institute.
- Helbich M., Leitner M., Kapusta ND. Geospatial examination of lithium in drinking water and suicide mortality. *Int J Health Geogr.* 2012;11:19.
- Henao G, Anderson J., Precioso F., Staccini P., Riveill M. DiagnoseNET: Automatic Framework to Scale Neural Networks on Heterogeneous Systems Applied to Medical Diagnosis ». IT Convergence and Security, édité par Hyuncheol Kim et Kuinam J. Kim, vol. 712, Springer Singapore, 2021, p. 3-17. https://doi.org/10.1007/978-981-15-9354-3_1

- Hill, A.B. 1965. The environment and disease: Association or causation? *Proc. R. Soc. Med.* 58:295–300.
- Holland PW.1986. Statistics and causal inference. *J Am Stat Assoc.* 1986; 81:945–960.
- Hume D. (1748). *An Enquiry concerning Human Understanding and Concerning The Principles of Morals*. Édité par L.A. Selby-Bigge et Peter Nidditch, en 1777.
- Hwang J., Kwon J., Yi H., Bae H-J., Jang M., Kim N. Association between long-term exposure to air pollutants and cardiopulmonary mortality rates in South Korea. *BMC Public Health* 2020; 20:1402.
- INERIS. 2012. Les Polychlorobiphényles – PCB, Données technico-économiques sur les substances chimiques en France, 89 p.
- InVS. 2008. Programme de surveillance air et santé (PSAS) « Analyse des liens à court terme entre pollution atmosphérique urbaine et mortalité dans neuf villes françaises ». Paris, 2008, 41p.
- Jacquez GM. Current practices in the spatial analysis of cancer: flies in the ointment. *Int J Health Geogr.* 2004;3:22.
- Jeanjean M., Lees J., Allen BL., Cohen A. Interdisciplinary community-based participatory health research across the industrial region of the Étang de Berre: The EPSEAL Fos Crau study. *Revue d'Épidémiologie et de Santé Publique* 2021:S0398762021003254.
- Jeanjean M., Goix S., Périot M., et al. Expositions environnementales et alimentaires à proximité d'un grand port industriel (Fos-sur-Mer, France) identifiée comme une voie importante pour l'accumulation de PCB et de PCDD/F dans le sérum sanguin des résidents. *Expo Santé* 13, 447-464 (2021). <https://doi.org/10.1007/s12403-021-00395-8>
- Johnson S. (2006). *The ghost map: The story of London's most terrifying epidemic--and how it changed science, cities, and the modern world*. Penguin.
- Kawachi I., Berkman L.F. (2003). *Neighborhoods and Health*, Oxford, 352p.
- Keyes K., Galea S. What matters most: quantifying an epidemiology of consequence. *Ann Epidemiol.* 2015 May;25(5):305-11. doi: 10.1016/j.annepidem.2015.01.016. Epub 2015 Feb 7. PMID: 25749559; PMCID: PMC4397182.
- Kihal-Talantikite W, Padilla C M, Zmirou-Navier D, Blanchard O, Le Nir G, et al. A Conceptual Framework for the Assessment of Cumulative Exposure to Air Pollution at a Fine Spatial Scale. *International Journal of Environmental Research and Public Health*, MDPI, 2016, 13 (3), pp.319;

- Khaniabadi Y.O., Daryanoosh M., Sicard P., Takdastan A., Hopke P.K., Esmaeili S., et al. 2018. Chronic obstructive pulmonary diseases related to outdoor PM₁₀, O₃, SO₂, and NO₂ in a heavily polluted megacity of Iran. *Environmental Science and Pollution Research* 25: 17726-17734.
- Khaniabadi Y.O., Sicard P. A 10-year assessment of ambient fine particles and related health endpoints in a large Mediterranean city. *Chemosphere*, 2021, 278: 130502.
- Khaniabadi Y.O., Sicard P., Takdastan A., Hopke P.K., Taiwo A.M., Khaniabadi F.O., et al. 2019. Mortality and morbidity due to ambient air pollution in Iran. *Clinical Epidemiology and Global Health* 7: 222-227.
- Kienberger S., Hagenlocher M. Spatial-explicit modeling of social vulnerability to malaria in East Africa. *Int J Health Geogr.* 2014;13:29.
- Laffont G-H., Martouzet D. Ces lieux qui nous affectent, Colloque de Cerisy, Hermann, 2021, 514p.
- Lai P.C., Low C.T., Wong M., Wong W.C., Chan M.H. Spatial analysis of falls in an urban community of Hong Kong. *Int J Health Geogr.* 2011;8:14.
- Lambin E.F., Tran A., Vanwambeke S.O. et al. Pathogenic landscapes: Interactions between land, people, disease vectors, and their animal hosts. *Int J Health Geogr* 9, 54 (2010). <https://doi.org/10.1186/1476-072X-9-54>
- Lange T; Starkopf, L. Commentary, *Epidemiology*: September 2016 - Volume 27 - Issue 5 - p 677-681 doi: 10.1097/EDE.0000000000000518.
- Langrish J.P., Bosson J., Unosson J., Muala A., Newby D.E., Mills N.L., et al. 2012. Cardiovascular effects of particulate air pollution exposure: time course and underlying mechanisms. *Journal of Internal Medicine* 272: 224-239.
- Larsen K., Gilliland J. Mapping the evolution of “food deserts” in a Canadian city: supermarket accessibility in London, Ontario, 1961–2005. *Int J Health Geogr.* 2008;7:16.
- Latham K.E., Sapienza C., Engel N. The epigenetic loraX: gene–environment interactions in human health. *Epigenomics. Research Gate.* Août 2012. <http://dx.doi.org/10.2217/epi.12.31>
- Le Cun Y., Bengio Y., Hinton G. « Deep learning », *Nature*, vol. 521, no 7553, 2015, p. 436–444 (PMID 26017442, DOI 10.1038/nature14539)
- Lees R. Changes in lung function after exposure to vanadium compounds in fuel oil ash. *Occupational and Environmental Medicine* 1980; 37:253–6.

- Lelieveld J., Evans J.S., Fnais M., Giannadaki D., Pozzer A. The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, 2015, 525: 367-371.
- Le Tourneau, P. (2003). La causalité, La responsabilité civile (pp. 39-54). Paris cedex 14: Presses Universitaires de France.
- Leray Ph. (2008). *Réseaux bayésiens - introduction et apprentissage, modélisation et découverte de connaissances*, Tutoriel, EGC 55p.
- Lucas R-M, McMichael A-J review The environment and disease: association or causation? *Bulletin of the World Health Organization*, October 2005, 83 (10)792:795
- Lucas-Gabrielli V., Mangeney C., Duchaine F., Com-Ruelle L., Gueye A., Denis Raynaud D. Inégalités spatiales d'accessibilité aux médecins spécialistes Proposition de méthodologie pour trois spécialités, *DT n° 87 Mai 2022*, Irdes.
- Lysaniuk B., Ondet A., Prost L. Révéler les traces du toxique. *EchoGéo*, 2020, 10.4000/echogeo.19792
- Maignant G. Dispersion de polluants et morphologie urbaine. *Espace géographique*, 36, 2007, 141-154. <https://doi.org/10.3917/eg.362.0141>
- Mannucci P.M., Harari S., Franchini M. 2019. Novel evidence for a greater burden of ambient air pollution on cardiovascular disease. *Haematologica* 104: 225086.
- Marano F. Des risques avérés aux risques suspectés : particules atmosphériques et nanoparticules manufacturées. *Annales des Mines - Responsabilité et environnement*, 104, 2021, 23-27.
- Mayne D.J., Morgan G.G., Willmore A., Rose N., Jalaludin B., Bambrick H., Bauman A. An objective index of walkability for research and planning in the Sydney Metropolitan Region of New South Wales, Australia: an ecological study. *Int J Health Geogr*. 2013;12:61.
- Mantey K., Pascal L., Cortaredona S. Observatoire des cancers du rein, de la vessie et des leucémies aiguës chez l'adulte dans le département des Bouches-du-Rhône (REVELA13). *Analyses spatiales*, 2013-2016. Saint-Maurice : Santé publique France, 2019. 80 p.
- Mazenq J., Dubus J-C., Gaudart J., Charpin D., Nougairede A., Viudes G, et al. Air pollution and children's asthma-related emergency hospital visits in southeastern France. *Eur J Pediatr* 2017;176:705–11.

- Mendes V., Ribeiro C., Delgado I., Peleteiro B., Aggerbeck M., Distel E., et al. The association between environmental exposures to chlordanes, adiposity and diabetes-related features: a systematic review and meta-analysis. *Sci Rep* 2021;11:14546.
- Moodie E.E.M., Delaney J.A.C., Lefebvre G., Platt R.W. Missing confounding data in marginal structural models : a comparison of inverse probability weighting and multiple imputation. *Int J Biostat*, 2008 ;4(1) :1-23
- Morgan RL., Whaley P., Thayer KA., Schünemann HJ. Identifying the PECO: A framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. *Environ Int.* 2018 Dec;121(Pt 1):1027-1031. doi: 10.1016/j.envint.2018.07.015.
- Mukherjee A., Agrawal, M. World air particulate matter: sources, distribution and health effects. *Environ Chem Lett* 15, 283–309 (2017). <https://doi.org/10.1007/s10311-017-0611-9>
- Münzel T., Sørensen M., Gori T., Schmidt FP., Rao X., Brook FR., et al. Environmental stressors and cardio-metabolic disease: part II—mechanistic insights. *Eur Heart J* 2016: ehw294.
- Naïm P., Wuillemin P-H., Leray P., Pourret O., Becker A. (2007) *Réseaux bayésiens* 386 pp Eyrolles
- Neyman J. 1923. Sur les applications de la théorie des probabilités aux expériences agricoles : Essai des principes, mémoire de master, 1923, réédité en anglais dans la revue *Statistical Science*, Vol. 5, pp. 463-472
- Nkhoma E.T., Hsu C.E., Hunt VI., Harris A. Detecting spatiotemporal clusters of accidental poisoning mortality among Texas counties, US, 1980–2001. *Int J Health Geogr.* 2004;3:25.
- Noack Y., Robert S., Bley D., Lozano P., Gille G., et al. Étude et évaluation multidisciplinaire de la pollution atmosphérique particulaire en milieu périurbain : le projet PACTES-BMP. Méthodologie et premiers résultats. Sud-Ouest Européen, Presses Universitaires du Mirail - CNRS, 2012, Les Observatoires Hommes-Milieus, pp.77-88.
- Novick P. That Noble Dream, The « Objectivity Question » and the American Historical Profession. Cambridge : Cambridge University Press, 1988
- Ocelli F, Lanier C, Cuny D, Deram A, Dumont J, Amouyel P, Montaye M, Dauchet L, Dallongeville J, Genin M, 2020. Exposure to multiple air pollutants and the incidence of coronary heart disease: A fine-scale geographic analysis. *Science of The Total Environment* 714, 136608. <https://doi.org/10.1016/j.scitotenv.2020.136608>

OHAT (Office of Health Assessment and Translation, National Toxicology Program, National Institute of Environmental Health Sciences, U.S. Dept. of Health and Human Services), 2019. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookmarch2019_508.pdf

Ozdenerol E., Williams BL., Kang SY., Magsumbol MS. Comparison of spatial scan statistic and spatial filtering in estimating low birth weight clusters. *Int J Health Geogr.* 2005;4:19.

Paez A., Mercado RG., Farber S., Morency C., Roorda M. Accessibility to health care facilities in Montreal Island: an application of relative accessibility indicators from the perspective of senior and non-senior residents. *Int J Health Geogr.* 2010;9:52.

Paracelsus, *Archidoxis magicae* libri VII, 1526

Park Y.M., and Kwan M.P. 2017. Individual exposure estimates may be erroneous when spatiotemporal variability of air pollution and human mobility are ignored. *Health and Place* 43: 85-94.

Park J., Kim H-J., Lee C-H., Lee CH., Lee HW. Impact of long-term exposure to ambient air pollution on the incidence of chronic obstructive pulmonary disease: A systematic review and meta-analysis. *Environmental Research* 2021; 194: 110703.

Pascal M., Falq G., Wagner V., Chatignoux E., Corso M., Blanchard M., et al. 2014. Short-term impacts of particulate matter (PM₁₀, PM_{10-2.5}, PM_{2.5}) on mortality in nine French cities. *Atmospheric Environment* 95: 175-184.

Passatore L., Rossetti S., Juwarkar AA., Massacci A. Phytoremediation and bioremediation of polychlorinated biphenyls (PCBs): State of knowledge and research perspectives. *Journal of Hazardous Materials* 2014; 278:189–202.

Pavlovsky E.N. *Natural Nidality of Transmissible Diseases, With Special Reference to the Landscape Epidemiology of Zooanthropose.* Urbana, Ill: University of Illinois Press, 1966.

Pászto, V., Jürgens, C., Tominc, P., & Burian, J. (Éds.). (2020). *Spatiality : Spatial exploration of economic data and methods of interdisciplinary analytics.* Springer International Publishing. <https://doi.org/10.1007/978-3-030-26626-4>

Pearl J. Causal diagrams for empirical research. *Biometrika*, 82, 4, 1995, 669:710

Pearl J. (2000a). *Causality: Models, Reasoning, and Inference.* Cambridge, University Press, New York. 2nd edition, 2009. MR1744773

Pearl J. (2009a). Causal inference in statistics: An overview. *Statist Surv*; 3.

- Pearl J. (2009b). *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge: Cambridge University Press.
- Pearl J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, 574 p.
- Pearl J., Mackenzie D. (2018). *The book of why: the new science of cause and effect*. First edition. New York: Basic Books, 418 p.
- Pedrot F. Gouverner une crise sanitaire par le savoir et l'incertitude : le cas des agrégats de « bébés nés sans bras » en France, *Anthropologie & Santé* mis en ligne le 29 mars 2021, <https://doi.org/10.4000/anthropologiesante.9194>
- Pereira Barboza E., Nieuwenhuijsen M., Ambròs A., Herick de Sá T., Mueller N. The impact of urban environmental exposures on health : An assessment of the attributable mortality burden in Sao Paulo city. *Science of The Total Environment*, vol. 831, juillet 2022, p. 154836. doi.org/10.1016/j.scitotenv.2022.154836.
- Pérez S. Modélisation des pathologies respiratoires des écoliers niçois par les Réseaux Bayésiens. *First edition of International Colloquium Health and Space*, Sep 2010, Marseille, France. {hal-01283742}
- Pérez S., Laperrière V., Borderon M., Padilla C., Maignant G., Oliveau S. Evolution of research in health geographics through the International Journal of Health Geographics (2002–2015). *International Journal of Health Geographics*, BioMed Central, 2016, 15 (3), pp.1-9.
- Pérez S. 2016. Etude de faisabilité sur le croisement des informations disponibles sur les zones à risques potentiels et le recensement de pathologies avérées et géo-localisées à l'échelle du territoire du Pays de Martigues, 35p.
- Pérez S. 2019. Étude AIR-SANTE en Pays de Martigues : apports de l'Intelligence Artificielle aux questions de santé environnementale. Université Côte d'Azur, Nice, France, 255p.
- Pérez S., German-Labaume C., Mathiot S., Goix S., Chamaret Ph. Using Bayesian networks for environmental health risk assessment. *Environmental Research*, Elsevier, 2021, pp.112059. {10.1016/j.envres.2021.112059}.
- Pfeiffer D., Robinson T., Stevenson M., Stevens K., Rogers D., Clements A. Spatial Analysis in Epidemiology, Oxford, Biology, 142p.
- Picheral H. La valeur stratégique de l'espace dans les politiques sanitaires, *actualité et dossier en santé publique* n° 19 juin 1997 page XXXVI.
- Picheral H. Le lieu, l'espace et la santé. 1995, *Espace, Populations, Sociétés*, 1 : 19-24.

- Picheral H. Complexes et systèmes pathogènes : approche géographique ?, De l'épidémiologie à la géographie humaine. *Travaux et documents de géographie tropicale*, 1983, n° 48, p. 5-22.
- Picheral H. Géographie médicale, géographie des maladies, géographie de la santé. In: *Espace géographique*, tome 11, n°3, 1982. pp. 161-175; <https://doi.org/10.3406/spgeo.1982.3751>
- Pickle L., Waller LA., Lawson AB. Current practices in cancer spatial data analysis: a call for guidance. *Int J Health Geogr*. 2005;4:3.
- Pourcher T, Suissa L, Guigonis J-M, Liendenthal S et al. Combined Omic Analyzes of Cerebral Thrombi A New Molecular Approach to Identify Cardioembolic Stroke Origin, May 2021 *Stroke* 52(9) DOI: 10.1161/STROKEAHA.120.032129
- Qi F., Du F. Tracking and visualization of space-time activities for a micro-scale flu transmission study. *Int J Health Geogr*. 2013;12:6.
- Quesnel-Barbet A., Soula J., Gamichon M., et al. PoleSat_2018: an optimized, automated, geomatics IT tool based on a gravitational model: strategic decision support in hospital catchment area planning. *SN Appl. Sci.* 2, 992 (2020). <https://doi.org/10.1007/s42452-020-2735-y>
- Raffensperger C., Tickner J. (1999) *Protecting Public Health and the Environment: Implementing the Precautionary Principle*. Island Press, Washington, DC.
- Rafi Z., Greenland S. Semantic and cognitive tools to aid statistical science: replace confidence and significance by compatibility and surprise. *BMC Med Res Methodol*. 2020;20(1):244. Published 2020 Sep 30. doi:10.1186/s12874-020-01105-9
- Rainham D., Krewski D., McDowell I., Sawada M., Liekens B. Development of a wearable global positioning system for place and health research. *Int J Health Geogr*. 2008;7:59.
- Ratier A., Dron J., Revenko G., Austruy A., Dauphin C-E., Chaspoul F., et al. Characterization of atmospheric emission sources in lichen from metal and organic contaminant patterns. *Environ Sci Pollut Res* 2018;25:8364–76.
- Rican S., Rey G., Lucas-Gabrielli V., Bard D., Zeitlin J., et al. Local disadvantages and health: Constructing indices to analyse social and geographic health inequalities in France and their trends. *Environnement, Risques & Santé*, John Libbey Eurotext, 2011, 10 (3), pp.211 - 215. <10.1684/ers.2011.0450>.
- Ricci P., Sheng H-X. 2013, Benefits and Limitations of the Precautionary Principle. 10.1016/B978-0-12-409548-9.01935-7

- Ricci P. 2011, Special Issue on : Causal models in public policy. Dose-Response.
- Richiardi L., Bellocco R., Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int J Epidemiol.* 2013 Oct;42(5):1511-9. doi: 10.1093/ije/dyt127. Epub 2013 Sep 9. PMID: 24019424.
- Robins J.M., Hernán M.A., Brumback, B. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 2000 ;11(5) :550-560.
- Rooney AA., Boyles AL., Wolfe MS., Bucher JR., Thayer KA. 2014. Systematic review and evidence integration for literature-based environmental health science assessments. *Environ Health Perspect* 122:711–718.
- Rose N., Cowie C., Gillett R.; Marks GB. « Weighted Road Density: A Simple Way of Assigning Traffic-Related Air Pollution Exposure ». *Atmospheric Environment*, vol. 43, no 32, octobre 2009, p. 5009-14. <https://doi.org/10.1016/j.atmosenv.2009.06.049>
- Rovira J., Domingo J.L., Schuhmacher M. 2020. Air quality, health impacts and burden of disease due to air pollution (PM₁₀, PM_{2.5}, NO₂ and O₃): Application of AirQ+ model to the Camp de Tarragona County (Catalonia, Spain). *Science of The Total Environment* 703: 135538.
- Rothman KJ., Greenland S. *Modern epidemiology*, 2nd ed. Philadelphia: Lippincott-Raven Publishers; 1998.
- Roux E.; Ignotti E.; Bègue N.; Bencherif H.; Catry T.; Dessay N.; Gracie R.; Gurgel H.; de Sousa Hacon S.; de A. F. M. Magalhães M.; Vieira Monteiro A.M.; Revillion C.; Antunes Maciel Villela D.; Xavier D.; Barcellos C. Toward an Early Warning System for Health Issues Related to Particulate Matter Exposure in Brazil: The Feasibility of Using Global PM_{2.5} Concentration Forecast Products. *Remote Sens.* 2020, 12, 4074. <https://doi.org/10.3390/rs12244074>
- Rozenblat C. 2022. Cities' systems and networks' proximities: toward a multiplex approach in *Handbook on proximity*, Edward Elgar.
- Rozenblat C. 2020. DATABASE: Large Urban Regions of the world - LURs - Version 1.5.2 [10.5281/zenodo.3944181](https://doi.org/10.5281/zenodo.3944181)
- Rubin D. 1974. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *J. Educ. Psychol.* 66 (5): 688–701. doi:10.1037/h0037350
- Salzé P., Banos A., Oppert J-M., Charreire H., Casey R., Simon C., Chaix B., Badariotti D., Weber C. Estimating spatial accessibility to facilities on the regional scale: an extended commuting-based interaction potential model. *Int J Health Geogr.* 2011;10:2.

- Salameh D., Pey J., Bozzetti C., El Haddad I., Detournay A., Sylvestre A, et al. 2018. Sources of PM2.5 at an urban-industrial Mediterranean city, Marseille (France): Application of the ME-2 solver to inorganic and organic markers. *Atmospheric Research* 214: 263-274.
- Sarigiannis DA., Hansen U. Considering the cumulative risk of mixtures of chemicals – A challenge for policy makers. *Environ Health* 2012;11: S18.
- Schwartz S. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. *Am J Public Health* 1994;84:819–24.
- Schwartz G., Il'Yasova D, Ivanova A. Urinary cadmium, impaired fasting glucose, and diabetes in the NHANES III. *Diabetes Care* 2003, 26, 468–470.
- Seddiki K., Saudemont P., Precioso F., Ogrinc N., Wisztorski M., Salzet M., Fournier I., Droit A. 2020. Cumulative learning enables convolutional neural network representations for small mass spectrometry data classification, *Nature Communication*, 2020 Nov 5;11(1):5595. <https://doi.org/10.1038/s41467-020-19354-z>
- Selmi W., Weber Ch., Mehdi L. Multifonctionnalité des espaces végétalisés urbains, *Vertigo*, no Volume 13 Numéro 2, octobre 2013., <https://doi.org/10.4000/vertigo.14133>.
- Setbon M. (2004). *Risques, sécurité sanitaire et processus de décision*, Elsevier, 176p.
- Sheehan M.C., Lam J. Use of Systematic Review and Meta-Analysis in Environmental Health *Epidemiology: a Systematic Review and Comparison with Guidelines. Curr Envir Health Rpt* 2, 272–283 (2015). <https://doi.org/10.1007/s40572-015-0062-z>
- Shekhar S., Evans MR., Kang JM., et Mohan, P. Identifier des modèles dans l'information spatiale : Une enquête sur les méthodes. *WIREs Data Mining Knowl Discov*, 1 , 2011, 193-214. <https://doi.org/10.1002/widm.25>
- Shi Baoyi; Choirat C; Coull Brent A.; VanderWeele Tyler J.; Valeri L. CM. Averse: A Suite of Functions for Reproducible Causal Mediation Analyses, *Epidemiology: September 2021 - Volume 32 - Issue 5 - p e20-e22* doi: 10.1097/EDE.0000000000001378
- Shiba K., Kubzansky LD., Williams DR., VanderWeele TJ., Kim ES. Associations Between Purpose in Life and Mortality by SES. *Am J Prev Med.* 2021;61(2):e53-e61. doi:10.1016/j.amepre.2021.02.011
- Shinozaki T., Suzuki E. Understanding Marginal Structural Models for Time-Varying Exposures: Pitfalls and Tips. *J Epidemiol.* 2020 Sep 5;30(9):377-389.
- Sicard P. 2021, Ground-level ozone over time: an observation-based global overview. *Current Opinion in Environmental Science & Health* 19: 100226.

- Sicard P., Agathokleous E., De Marco A., Paoletti E., Calatayud V. 2021. Urban population exposure to air pollution in Europe over the last decades. *Environmental Sciences Europe* 33: 28.
- Sicard P., Khaniabadi Y.O., Perez S., Gualtieri M., De Marco A. 2019. Effect of O₃, PM₁₀ and PM_{2.5} on cardiovascular and respiratory diseases in cities of France, Iran and Italy. *Environmental Science and Pollution Research* 26: 32645-32665.
- Sicard P., Agathokleous E., Araminiene V., Carrari E., Hoshika Y., De Marco A., Paoletti E. Should we see urban trees as effective solutions to reduce increasing ozone levels in cities? *Environmental Pollution*, 2018, 243: 163-176.
- Sicard P., Talbot C., Lesne O., Mangin A., Alexandre N., Collomp R. 2012, "The Aggregate Risk Index: an intuitive tool providing the health risks of air pollution to health care community and public". *Atmospheric Environment* 46: 11-16.
- Simon D. Quand un concept écologique fait date. L'invention du " complexe pathogène " en géographie. *Revue d'histoire des sciences humaines*, Publications de la Sorbonne, 2016, pp.253-272.
- Simon D. Max Sorre, une écologie humaine, Editions de la Sorbonne, 2021, 319p.
- Simos J., De Leeuw E. *Healthy Cities: The Theory, Policy, and Practice of Value-Based Urban Planning*. New York, NY: Springer; 2017.
- Simos J. Paysage, urbanisme et santé. *Environnement risques & santé* 2015;14(4):272-276.
- Simos J., Cantoreggi N. Vers une nouvelle santé urbaine? Les enseignements à tirer après plus de 25 ans de "Villes-Santé OMS". *Urbia* 2015;18:21-38.
- Slama R., Vrijheid M. Some Challenges of Studies Aiming to Relate the Exposome to Human Health, *Occup Env Med* (Invited Editorial) 72:383-4, 2015.
- Sofia D., Gioiella F., Lotrecchiano N., Giuliano A. 2020. Mitigation strategies for reducing air pollution. *Environmental Science and Pollution Research* 27: 19226-19235.
- Sorre M. L'organisme humain et le milieu biologique naturel, *Bulletin de la Société de Géographie de Lille*, 68, 1926.
- Sorre M. Complexes pathogènes et géographie médicale, *Annales de Géographie*, 1933, vol. 42, n°235, p. 1-18.
- Sorre M. Tome 1 : *Les fondements biologiques. Essai d'une écologie de l'homme*. in *Les fondements de la géographie humaine*. Paris, Armand Colin, 1951.

- Sottas P-E., Lavoué J., Bruzzi R., Vernez D., Charrière N., Droz P-O. An empirical hierarchical Bayesian unification of occupational exposure assessment methods. *Statist Med* 2009; 28:75–93.
- Staccini P. (2014) Sharing data and medical records in *Medical Informatics, e-Health?* pp 315–348 [DOI: 10.1007/978-2-8178-0478-1_13](https://doi.org/10.1007/978-2-8178-0478-1_13)
- Suarez D., Borràs R., Basagaña X. Differences between marginal structural models and conventional models in their exposure effect estimates: a systematic review. *Epidemiology*. 2011;22(4):586-588. doi:10.1097/EDE.0b013e31821d0507
- Suissa L., Fortier M., Lachaud S., Staccini P., Mahagne M-H. Ozone air pollution and ischaemic stroke occurrence: a case-crossover study in Nice, France. *BMJ Open* 2013; 3: e004060.
- Svingen T., Vinggaard AM. The risk of chemical cocktail effects and how to deal with the issue. *J Epidemiol Community Health* 2016; 70:322–3.
- Talbot D., Rossi AM., Bacon SL., Atherton J., Lefebvre G. A graphical perspective of marginal structural models: An application for the estimation of the effect of physical activity on blood pressure. *Stat Methods Med Res*. 2018 Aug;27(8):2428-2436.
- Talbot D., Atherton J., Rossi A.M., Bacon S.L., Lefebvre G. A cautionary note on the use of stabilized weights in marginal structural models. *Stat Med*, 2015 ; 34 (5) : 812-823.
- Tango T., Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. *Int J Health Geogr*. 2005;4:11.
- Tian J., Pearl J. Probabilities of causation, bounds and identification, Uncertainty in Artificial Intelligence, Proceedings 2000, 589:598.
- Townsend P. 1987. Deprivation. *Journal of Social Policy*, 16 (2) : 125-146.
- Tran A., Sudre B., Paz S., Rossi M., Desbrosse A., Chevalier V., Semenza JC. Environmental predictors of West Nile fever risk in Europe. *Int J Health Geogr*. 2014;13:26.
- Vaguet A., Brisset S., Eliot E. 2011. L'insoutenable légèreté de la ville durable : la preuve par l'Inde ? Chapitre 10, Dans : Nicole Mathieu éd., *La ville durable, du politique au scientifique* (pp.145-163). Versailles: Éditions Quæ. <https://doi.org/10.3917/quae.guerm.2011.01.0143>
- Valeri L., VanderWeele T. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods*. 2013;18:137–150.

- Vallée J. Les disparités spatiales de santé en ville : l'exemple de Vientiane (Laos), *Cybergeog*: European Journal of Geography [Online], Space, Society, Territory, document 477, <https://doi.org/10.4000/cybergeog.22775>
- Van den Berg M., Birnbaum LS., Denison M., De Vito M., Farland W., Feeley M., et al. The 2005 World Health Organization Reevaluation of Human and Mammalian Toxic Equivalency Factors for Dioxins and Dioxin-Like Compounds. *Toxicological Sciences* 2006;93:223–41.
- Vandenbroucke JP., Broadbent A., Pearce N. Causality and causal inference in epidemiology: the need for a pluralistic approach. *Int J Epidemiol.* 2016;45(6):1776-1786. doi:10.1093/ije/dyv341
- VanderWeele T.J, Hernán M.A., Tchetgen E.J., Robins J.M. Causality and causal inference in epidemiology: the need for a pluralistic approach, *International Journal of Epidemiology*, Volume 45, Issue 6, December 2016, Pages 2199–2200, <https://doi.org/10.1093/ije/dyw162>
- VanderWeele T.J., Vansteelandt S. Mediation analysis with multiple mediators. *Epidemiol Methods.* 2014;2:95–11.
- Vansteelandt S., Bekaert M., Lange T. Imputation strategies for estimating natural direct and indirect effects. *Epidemiol Methods* 2012 ; 1 : 131-158.
- Venerandi A., Fusco G., Tettamanzi A.G.B., Emsellem D. A Machine Learning Approach to Study the Relationship between Features of the Urban Environment and Street Value. *Urban Science*, MDPI, 2019, 3 (3), pp.100.
- Vernazza-Licht N., Guérenais M-E., Bley D . 2010 Sociétés, environnements santé , 364p, IRD Editions, Objectifs Suds.
- Vineis P., Chadeau-Hyam M., Gmuender H., Gulliver J., Herceg Z., Kleinjans J., Kogevinas M., Kyrtopoulos S., Nieuwenhuijsen M., Phillips D. H., Probst-Hensch N., Scalbert A., Vermeulen R., & Wild C.P. 2017. The exposome in practice : Design of the EXPOsOMICS project. *International Journal of Hygiene and Environmental Health*, 220(2), 142-151. <https://doi.org/10.1016/j.ijheh.2016.08.001>
- Weber P., Simon C. Systems Dependability Assessment: Benefits of Bayesian Networks Models. Wiley-ISTE, 146 pages, August 2016, ISBN: 978-1-84821-992-2 - <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-184821992X.html>
- Weber P., Portinale L. Special Section on “Applications of Probabilistic Graphical Models in Dependability, Diagnosis and Prognosis” *Reliability Engineering & System Safety*, vol. 167, novembre 2017, p. 613-15. <https://doi.org/10.1016/j.res.2017.04.017>

- Weber P., Medina-Oliva G., Simon C., Lung B. Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas - *Engineering Applications of Artificial Intelligence*, 25 (4), 671-682, 2012 – Elsevier.
- Weis Brenda K., et al. Personalized Exposure Assessment: Promising Approaches for Human Environmental Health Research. *Environmental Health Perspectives*, 113 (7), juillet 2005, 840-48. <https://doi.org/10.1289/ehp.7651>
- Wheeler B.W., Lovell R., Higgins S.L., et al. Beyond greenspace: an ecological study of population general health and indicators of natural environment type and quality. *Int J Health Geogr* 14, 17 (2015). <https://doi.org/10.1186/s12942-015-0009-5>
- Wiehe SE., Carroll AE., Liu GC., Haberkorn KL., Hoch SC., Wilson JS., Fortenberry JD. Using GPS-enabled cell phones to track the travel patterns of adolescents. *Int J Health Geogr*. 2008;7:22.
- Wild C-P; Complementing the Genome with an “Exposome”: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiol Biomarkers Prev* 1 August 2005; 14 (8): 1847–1850. <https://doi.org/10.1158/1055-9965.EPI-05-0456>
- World Health Organization. 2016. LWHO, Prévenir la maladie grâce à un environnement sain : une estimation de la charge de morbidité imputable à l'environnement.
- World Health Organization. 2006. Air quality guidelines global update 2005: particulate matter, ozone, nitrogen dioxide and sulfur dioxide. Regional Office for Europe (Copenhagen: WHO Regional Office for Europe, 2006).
- Wright S. Correlation and causation. *Journal of Agricultural Research*, 20: 557–585, 1921.
- Xiao Y., Abrahamowicz M., Moodie E.E.M. Accuracy of conventional and marginal structural Cox model estimators: a simulation study. *Int J Biostat*, 2010 ; 6(2).
- Yang B-Y., Fan S., Thiering E., Seissler J., Nowak D., Dong G-H., et al. Ambient air pollution and diabetes: A systematic review and meta-analysis. *Environmental Research* 2020;180:108817.
- Yang Y., Qi J., Ruan Z., Yin P., Zhang S., Liu J., et al. 2020. Changes in Life Expectancy of Respiratory Diseases from Attaining Daily PM_{2.5} Standard in China: A Nationwide Observational Study. *The Innovation* 1: 100064.
- Zigler CM., Kim C., Choirat C., et al. Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations. *Res Rep Health Eff Inst*. 2016;(187):5-49.

Zmirou-Navier, D. Effets bénéfiques de l'environnement sur la santé en site urbain. *Annales des Mines - Responsabilité et environnement*, 104, 2021, 19-22.
<https://doi.org/10.3917/re1.104.0019>

Zusman Y. 2022. *L'espace aléatoire*, PUF, 262p.

. Annexes

. Liste des polluants recherchés dans les lichens pour la bio-imprégnation

Al	
Ca	
Ti	
V	
Cr	
Mn	
Fe	
Co	Métaux
Ni	
Cu	
Zn	
As	
Mo	
Ag	
Cd	
Sn	
Sb	
Hg	
Tl	
Pb	
Nap	
Acy	
Ace	
Flu	
Phe	
Ant	
FIA	
Pyr	
BaA	HAPs
Chr	
BbF	
BkF	
BaP	
Bpe	
DBA	
Ipy	
CB18	
CB20	
CB31	
CB44	
CB52	
CB126	

CB128
CB138
CB141
CB149
CB151
CB153
CB156
CB157
CB167
CB169
CB170
CB174
CB177
CB180
CB183
CB187
CB189
CB194
CB196

PCBs

1,2,3,7,8 Pentachlorodibenzodioxine
1,2,3,4,7,8 Hexachlorodibenzodioxine
1,2,3,6,7,8 Hexachlorodibenzodioxine
1,2,3,7,8,9 Hexachlorodibenzodioxine
1,2,3,4,6,7,8 Heptachlorodibenzodioxine
1,2,3,4,6,7,8,9 Octachlorodibenzodioxine
2,3,7,8 Tetrachlorodibenzofurane
1,2,3,7,8 Pentachlorodibenzofurane
2,3,4,7,8 Pentachlorodibenzofurane
1,2,3,4,7,8 Hexachlorodibenzofurane
1,2,3,6,7,8 Hexachlorodibenzofurane
2,3,4,6,7,8 Hexachlorodibenzofurane
1,2,3,7,8,9 Hexachlorodibenzofurane
1,2,3,4,6,7,8 Heptachlorodibenzofurane
1,2,3,4,7,8,9 Heptachlorodibenzofurane
1,2,3,4,6,7,8,9 Octachlorodibenzofurane

Dioxines, Furanes

. Le % des différentes tranches d'âges dans les IRIS

IRIS	% 0-19 ans	% 20-64 ans	% >65 ans
130560101	23,92	55,58	20,49
130560102	22,55	55,97	21,48
130560103	20,04	59,53	20,43
130560104	18,80	55,67	25,53
130560105	27,45	53,26	19,28
130560106	21,45	58,45	20,10
130560107	21,15	57,67	21,18
130560108	19,04	60,24	20,71
130560109	26,10	58,67	15,23
130560110	22,05	54,59	23,35
130560111	20,62	55,04	24,34
130560112	32,38	52,62	15,00
130560113	18,22	58,27	23,52
130560114	21,23	61,86	16,91
130560115	24,90	54,69	20,41
130560116	34,50	57,37	8,12
130560117	28,39	59,59	12,02
130560118	24,98	54,82	20,20
130560119	20,42	51,87	27,71
130560120	19,17	51,95	28,87
130770101	29,91	49,08	21,01
130770102	22,24	55,33	22,43
130770103	24,30	51,54	24,15
130770104	26,38	51,96	21,65
130770105	27,91	58,00	14,09
130770106	25,79	57,27	16,94
130770107	24,28	51,16	24,56
130770108	18,38	52,71	28,91
130980101	20,95	54,57	24,48
130980102	21,26	62,41	16,34

. Liste des figures

Figure 1. Les résultats de l'étude menée sur l'asthme chez les écoliers niçois, (Pérez, 2010)	32
Figure 2. Un modèle d'équation structurelle simple, et ses diagrammes associés. Les variables exogènes non observées sont reliées par des flèches en pointillés.....	43
Figure 3. La généalogie des nœuds	44
Figure 4. La structure en V formée par les nœuds.....	44
Figure 5. L'élément médiateur dans le réseau.....	45
Figure 6. Un modèle de médiation longitudinale à 3 vagues (Asendorpf, 1996).....	47
Figure 7. Connexion en série de type indirect.....	48
Figure 8. Connexion convergente.....	48
Figure 9. Connexion divergente	48
Figure 10. Emploi du terme « <i>Causality</i> » dans les publications PubMed entre 1945 et 2021	53
Figure 11. Buffer de 100 m appliqué sur le périmètre des zones bâties, exemple de l'IRIS de la plaine de Courouche à Martigues. Source AtmoSud	63
Figure 12. Intersection des zones urbaines de l'IRIS de la plaine de Courouche avec les données de pollution atmosphérique. Source AtmoSud.....	63
Figure 13. L'ensemble des 171 variables représentées sous la forme de nœuds.....	69
Figure 14. Les liens causaux entre les nœuds	70
Figure 15. La carte conceptuelle des 171 variables, le réseau bayésien.....	75
Figure 16. Cartographie 3D, analyse des nœuds & des arcs	76
Figure 17. Monitorings entre J00J19965ans & CB169	78
Figure 18. Monitorings initiaux entre J00J1991565 & BkF_p.....	79
Figure 19. Monitorings initiaux entre C3465ans & plusieurs variables.....	80
Figure 20. Monitorings initiaux entre C341565ans & HF	82
Figure 21. Monitorings initiaux entre C6765ans & plusieurs variables.....	84
Figure 22. Monitorings initiaux entre E10E1465ans & plusieurs variables.....	85
Figure 23. Monitorings initiaux entre I00I19965ans & plusieurs variables	86
Figure 24. Monitorings initiaux entre DEC_MED14 & HCl	88
Figure 25 . Monitorings initiaux entre P14_PMEN_ANEM10P & plusieurs variables	90
Figure 26. Traitement supervisé sur la variable cible IRIS pour les pathologies	92
Figure 27. L'ensemble des analyses de sensibilité par pathologies et par tranches d'âge pour tous les IRIS.....	93
Figure 28. L'analyse de sensibilité/pathologies pour l'IRIS Côte bleue (13560101) à Martigues	94
Figure 29. Les résultats de la CAH (Algorithme méthode Ward, distance euclidienne)	98
Figure 30. Analyse de sensibilité de l'IRIS 130770102	102
Figure 31. Analyse de sensibilité de l'IRIS 130560114	103
Figure 32. Traitement supervisé sur la variable cible IRIS pour les polluants, la socio- économie et les relevés lichéniques	108
Figure 33. L'ensemble des analyses de sensibilité sur les 68 variables et pour tous les IRIS	109
Figure 34. Le profil des analyses de sensibilité sur les 68 variables pour chaque IRIS.....	109
Figure 35. Le <i>design</i> de l'étude de type « <i>top down</i> »	123
Figure 36. Méta-analyse de la rhinite allergique et de l'exposition aux PM 2,5 pour des paliers de 10 µg/m ³ IC : intervalle de confiance. Le carré bleu représente le poids de chaque étude, le centre du carré représente l'odds ratio (OR) et la longueur de chaque ligne autour du point représente son intervalle de confiance à 95 % (IC à 95 %). L'estimation groupée récapitulative globale (résultat de la méta-analyse) est représentée par un losange vert à la fin du graphique. La ligne verticale continue rouge représente l'estimation nulle (avec une valeur de 1 pour OR). Les lignes horizontales qui traversent la ligne verticale nulle (OR = 1) représentent des études non significatives.....	130
Figure 37. La méthodologie développée pour caractériser des eXpaces.....	143

. Liste des cartes

Carte 1. Les 3 communes du Pays de Martigues.....	59
Carte 2. L'espace d'étude à l'échelle des IRIS	59
Carte 3. Les zones urbaines dans les IRIS. Source AtmoSud	64
Carte 4. Les placettes des prélèvements lichéniques. Source IECF	66
Carte 5. Groupe 1 CAH, IRIS caractérisés par une prévalence accrue des hospitalisations pour l'ensemble des pathologies étudiées	114
Carte 6. Groupe 2 CAH, IRIS caractérisés par une prévalence accrue des hospitalisations liées aux cancers du poumon (> 15 ans) et de la vessie (15-65 ans)	115
Carte 7. Groupe 3 CAH, IRIS caractérisés par une prévalence accrue des hospitalisations liées aux maladies cardiovasculaires (< 15 ans et > 65 ans), aux maladies respiratoires (<15 ans) et au diabète (< 15 ans)	116
Carte 8. Groupe 4 CAH, IRIS caractérisés par une prévalence accrue des hospitalisations liées aux maladies cardiovasculaires (> 65 ans) et aux maladies respiratoires (<15 ans et >65 ans)	117
Carte 13. Groupe 5 CAH, IRIS caractérisés par une prévalence moindre pour les pathologies étudiées « Groupe témoin ».....	118
Carte 10. Carte de synthèse	119
Carte 11. Les incendies de forêt sur le territoire martégal entre mars et septembre 2017. Source. Prométhée.com.....	120
Carte 12. Direction du vent et importance des feux de forêt à proximité spatiale et temporelle des prélèvements lichéniques	122

. Liste des tableaux

Tableau 1. La <i>meta review</i> de Lange sur les modèles de type médiation.....	46
Tableau 2. Les polluants de nature gazeuse fournis par AtmoSud	62
Tableau 3. Les polluants de nature particulaire fournis par AtmoSud.....	62
Tableau 4. Les informations concernant les prélèvements	66
Tableau 5. Les valeurs de probabilité des analyses de sensibilité sur les pathologies/IRIS.....	95
Tableau 6 . Les valeurs de probabilité contextualisées sur le % des différentes tranches d'âge	97
Tableau 7. Les 5 profils sanitaires issus de la CAH	99
Tableau 8. Les relations modélisées retrouvées dans les analyses de sensibilité des IRIS concernés	104
Tableau 9. Des relations plus faibles dans les IRIS du groupe témoin	106
Tableau 10. Code couleur et indication des pathologies relevées dans chaque groupe de la CAH ...	110
Tableau 11 . Les probabilités de prédominance des facteurs environnementaux et socio-économiques dans chaque IRIS.....	111
Tableau 12. Le profil attribué à chacun des IRIS	112
Tableau 13. Extraction de la BDD Prométhée, date d'extraction : 23/04/19	121
Tableau 14. Descriptif synthétique des études prises en compte dans la meta review de Charreire et al., (2010).....	127
Tableau 15. Cinq approches paradigmatiques pour identifier l'exposition (Morgan et al., 2022)	129

. Résumé

Le domaine de la santé environnementale manque cruellement de théorisation. Cela s'explique par la relative jeunesse de ce secteur, par le fait que les études prennent du temps, et qu'elles sont complexes. De plus, ces études ne sont pas récurrentes, ni uniformes, les méthodologies employées varient grandement, de même que les populations cibles. Elles sont menées sur des espaces et des échelles géographiques différentes, apparaissent souvent comme contradictoires, et les caractéristiques intrinsèques des espaces sur lesquels elles se déploient sont rarement prises en compte. Dès lors, leur spatio-dépendance freine toute généralisation nécessaire à une théorisation.

C'est la raison pour laquelle nous proposons ici une chaîne méthodologique fondée sur l'utilisation des réseaux bayésiens pour remonter aux causes des pathologies (étiologie) en santé environnementale. Le concept opérationnel d'eXpace est également présenté. Il est question de mettre l'espace géographique et sa diversité comme indicateur sanitaire *via* un triptyque causalités/potentiel/prévention.

Les résultats pourront servir à de multiples perspectives de recherche, notamment à la métabolomique qui permettra non seulement de valider les eXpaces trouvés, d'introduire l'expérimentation dans les espaces géographiques, mais également d'affiner l'inclusion de patients dans ce type d'analyses.

. Summary

The field of environmental health is severely under-theorized. This is explained by the fact that it is a relatively young sector, and that studies are time-consuming and complex. Moreover, these studies are neither recurrent nor uniform, the methodologies used vary greatly, as well as the target populations. They are conducted on different geographical spaces and scales, often appear to be contradictory, and the intrinsic characteristics of the spaces on which they are deployed are rarely taken into account. From then on, their spatial dependence puts a brake on any generalisation, which is necessary for theorising.

This is the reason why we propose a methodological chain based on the use of Bayesian networks to get to the root causes of environmental health pathologies (aetiology). The operational concept of eXpace is also expounded. The idea is to take the geographical space and its diversity as health index via a causalities/potential/prevention triptych.

The results can be used in numerous research perspectives, notably in metabolomics, making it possible not only to validate the eXpaces found and introduce the experimentation into geographical spaces, but also to fine-tune the inclusion of patients in this type of analysis.